

Context of evaluation and data in 21st century Federal  
Policy: A Text Analysis Approach  
PPOL 6801 Text as Data Final Project

Katharyn Loweth

05.02.2025

## **Introduction**

One of the primary goals of the first 100 days of the 2nd Trump administration has been the pursuit of “government efficiency”. The use of “government efficiency” to describe the motivation for President Trump and Elon Musk’s efforts has caused me to reflect on the use of evaluation-related terminology in the federal government.

Although the phrase has now commonly found in federal government news because of the Department of Government Efficiency (DOGE), the desire and goal to evaluate government efforts to improve efficiency has been a focus of multiple administrations. For most of my professional career, I have worked on federal government evaluation efforts, both within and outside of government agencies. Evaluations of programs, policies, and regulations help policymakers be more efficient with their resources, be accountable to Congress and the American people, and better understand how they are currently supporting their populations. The context and tools for federal evaluation has changed significantly in the last 25 years. With President Bush implementing nationwide evaluation efforts like No Child Left Behind to Congress passing the Evidence Act of 2018 which mandates that agency data to be accessible and requires agencies to plan to develop statistical evidence to support policymaking (ASPE, n.d.). However, putting more emphasis on evaluation and data is much easier said than done. One of the key reasons why I decided to return to graduate school is that I saw the potential of implementing data science techniques to support policymakers and reduce burden of evaluations on both constituents and bureaucrats.

The new Trump administration’s ability to quickly change the meaning and context of “government efficiency” highlighted how presidential administration priorities can impact larger discourse. Thus,

for my final project, I explored how the terminology and context of evaluation and data have changed in the federal government in the 21<sup>st</sup> century to gain more knowledge into how the federal government's evolving discourse.

### **Research Questions:**

1. Since 2001, how has the federal governments' use of evaluation and data-focused terminology changed?
  - a. Have the usage of these terms continuously increased?
  - b. When do new concepts (i.e., artificial intelligence, data science) begin to be used in federal documents?
2. Do federal agencies and the Executive Office of the President utilize these terms differently?
3. How do political parties and presidential administrations contextualize evaluation and data differently?

### **Report Overview**

In this report, I will examine the frequency and context of specific evaluation and data-focused terminology in official federal government documents. I will first provide an overview of my data source and the preprocessing steps I completed so that I could compare annual and presidential administrations' usage of specific terms. Then I will detail the text analysis methods I used to further analyze the context of these specific terms and discern any noticeable differences across presidential administrations and parties. Lastly, I will discuss possible future directions of this project.

# Data and Methods

## Data source

The data for this project comes from the Federal Register, which the U.S. Federal Government's official daily publication for Federal agencies and organizations as well as presidential documents like executive orders ("Federal Register, 1936 to Present", 2025). Using the Federal Register API, I compiled search results for final rules of Federal agencies and presidential documents published between January 20, 2001 and April 19, 2025 that included at least one of the following key terms:

- "Government efficiency", "Government evaluation", "Data analysis", "Data science", "Data driven", "Program evaluation", "Performance evaluation", "Machine learning", "Artificial intelligence", "Data collection", "Evaluation", "Efficiency", "AI", "Data" (only for presidential documents), "Analysis" (only for presidential Documents)

After aggregating all search results together and removing duplicates from the dataset, I used the provided body\_text\_html feature to then collect the full text of these documents.

## Variables of Interest

My corpus for analysis is the full text of these federal agency rules and presidential documents ( $n=36,729$ ). My unit of analysis are word tokens within these documents.

## Data wrangling steps

**Cleaning Corpus:** When I initially scraped the text from the html page I removed extra white space characters and html characters like "/n". As part of the tokenization process, I removed

numbers, symbols, punctuation, and word separators. I also removed alphanumeric tokens such as “a1b” or phone numbers that were not removed in the initial removal process. Lastly, I set a minimum term frequency of 5 to reduce dimensionality, but set padding = TRUE to not lose the order of tokens for analysis.

**Creating custom bigrams:** To further examine how the appearance and use of different phrases changes over time, I transformed the following phrases so that in the analysis they were considered to be a single token.

- “Data Collection”, “Artificial Intelligence” “Data Science”, “Machine Learning”, “Data Collection”, “Data Analysis”, “Government Efficiency”, “Government Evaluation”, “Performance Evaluation”, “Program Evaluation”

**Creating new variables:** In addition to the custom bigrams, I also created three new categorical variables to help inform my analysis. Based on the information from the Federal Register search results, I labeled whether the document was either an agency final rule or presidential document. Additionally, using information on the sitting President and publication date, I labeled whether the document was released during a Democrat or Republican presidential administration and during which specific administration term the document was published.

The descriptive statistics for these categorical variables are included below

Table 1: Number of Documents, by Document Type

Agency Rules	Presidential Documents
36,048	681

Table 2: Number of Documents, by Presidents' Political Party

Democrat	Republican
17185	19544

Table 3: Number of Documents, by Presidential Administration

Bush-1	Bush-2	Obama-1	Obama-2	Trump-1	Biden	Trump-2
7264	6693	6056	6225	5367	4904	211

## Analysis

### 1. Count Frequency of Specific Terms

With my tokenized corpus, I can identify when a particular token is used across the documents and aggregate counts. Using the corpus variables like publication date of the documents and presidential administration, I also compare how the frequency changes based on subgroups.

### 2. Key Words in Context (KWIC):

The KWIC technique allows for closer examination of the words surrounding a particular target token. It extracts the target token from the corpus and provides the context window (the tokens directly before and after the target) to show how the term is getting used.

### **3. Global Vectors for Word Representation (GloVe) word embedding model**

GloVe is an unsupervised learning algorithm that captures information about tokens through a multi-dimensional vector representation (Pennington et al., 2014). It utilizes feature co-occurrence to measure the linguistic or semantic similarity of words in a corpus (Pennington et al., 2014). It also uses an optimized cost function to help the embedding improve iteratively.

For my word embeddings, I subsetted the corpus by Presidential Administration, with the exception of grouping both Trump administrations together (6 groups), and by political party (2 groups).

For the presidential administration subsets I did not create any custom bigrams due to the corpus' average size, but I did create bigrams within the democrat and republican corpuses.

With these subsets, I created feature co-occurrence matrices using weighted count and a context window of 5. I then fit a GloVe model with 300 dimensions for each of these matrices. I ran each model 12 times to help minimize loss. For the party-level I created two sets of word embeddings, one with bigrams and one without.

## Results

### **Insight 1 (Q1): Most evaluation and data term usage peaks in 2016, 2020, and 2024**

As can be seen in Figures 1-4 below, the usage of the majority of the evaluation and data term usage does not continuously increase over time. Instead, the most frequently used unigrams (data, analysis, evaluation) and bigrams (data collection, data analysis, and performance evaluation) are relatively consistent in how much they appear in federal government documents from 2001 to 2015, and hit an initial peak in 2016 before a period of decline. There is a similar peak in 2020 before another period of decline before all terms reaching their highest annual usage in 2024. Similarly, when we examine term usage by presidential administration, it is clear that while the overall the words are used more frequently there is a dip in usage during the first Trump administration. The results are more varied for bigrams, although most terms do experience a period of decline after 2016 and peak again in 2024.

## Unigram Token Usage in Federal Government Documents

January 20, 2001-April 19, 2025

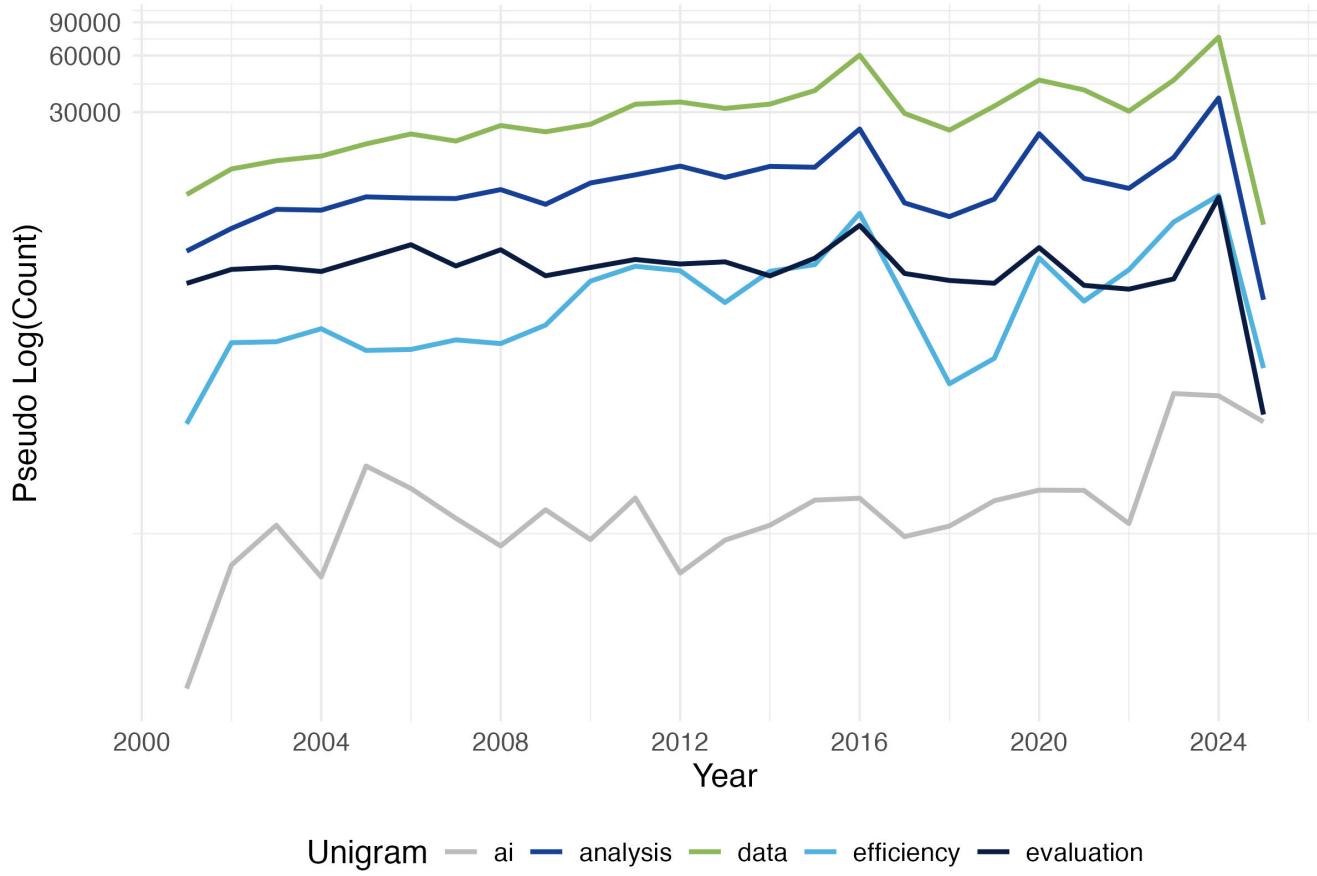


Figure 1: Annual Unigram Term Usage

## Bigram Token Usage in Federal Government Documents

January 20, 2001-April 19, 2025

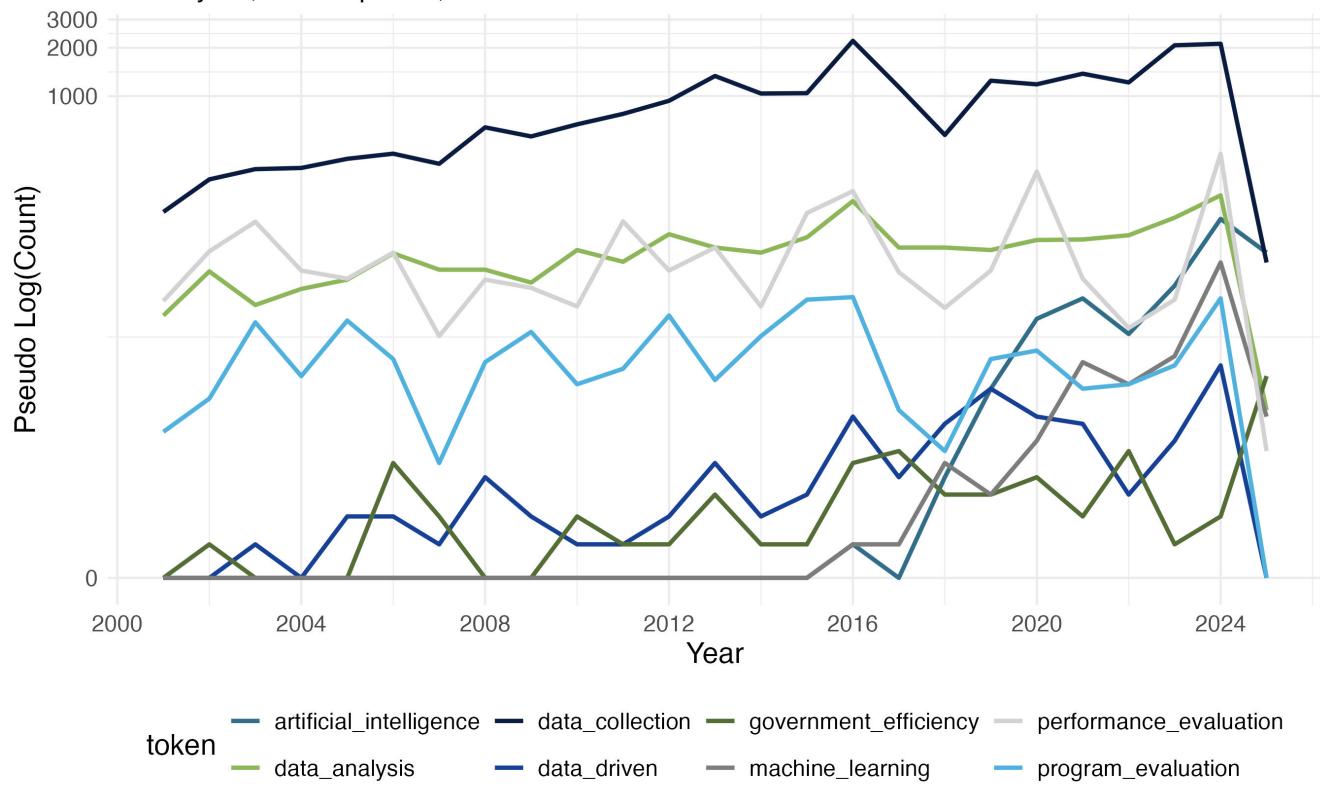


Figure 2: Annual Bigram Term Usage

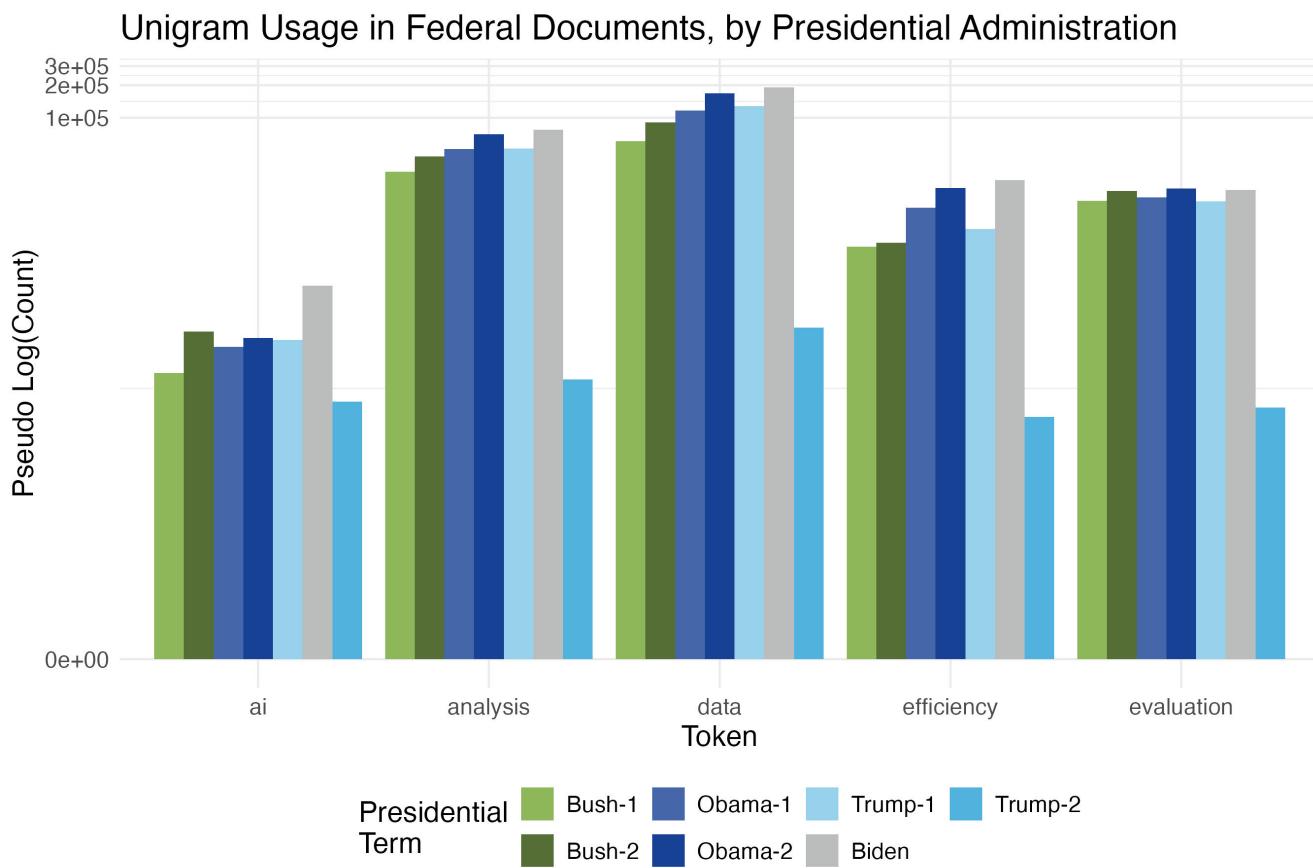


Figure 3: Unigram Usage in Federal Government Documents, by Presidential Administration

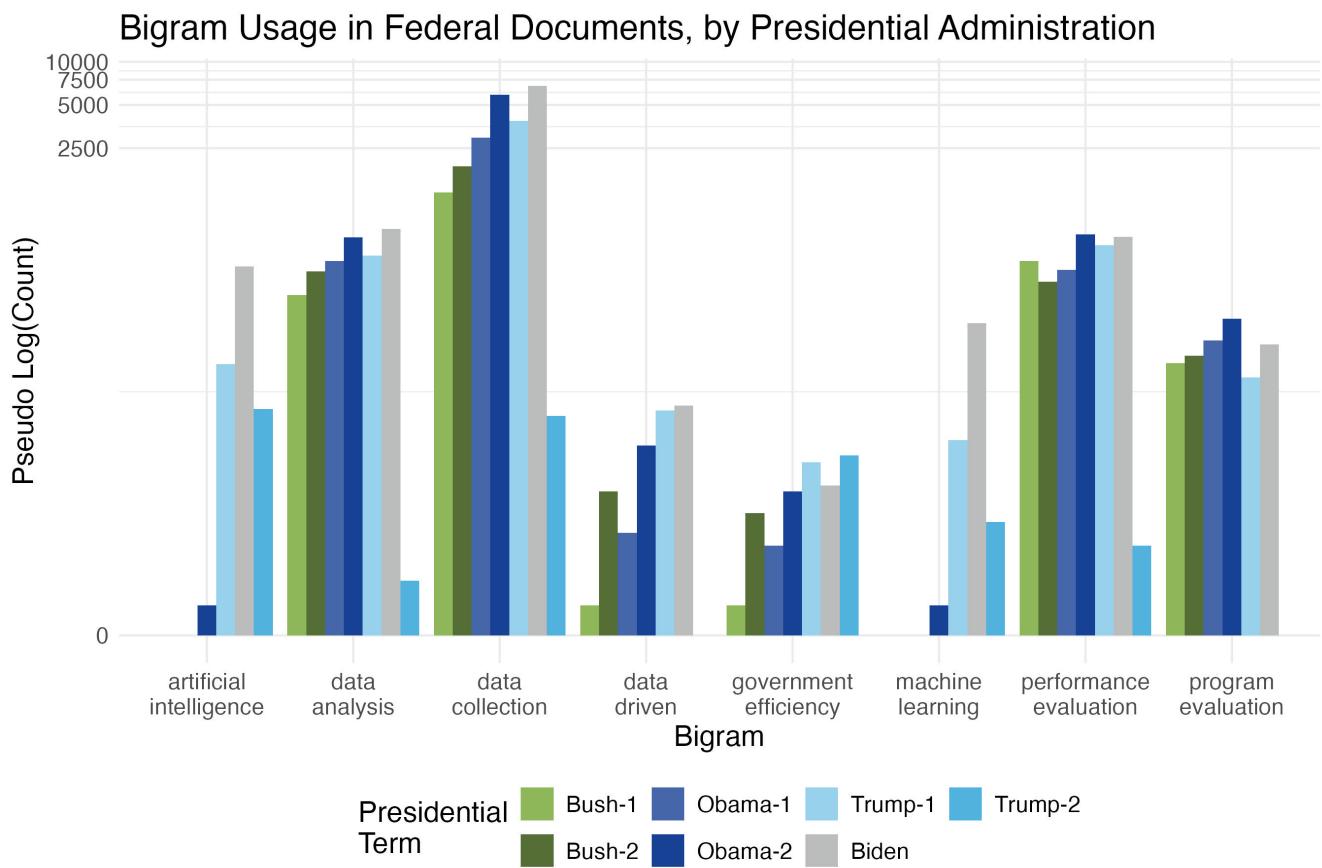


Figure 4: Bigram Usage in Federal Government Documents, by Presidential Administration

## **Insight 2 (Q1): New data-related terminology emerge after 2016**

The key exceptions in this pattern are the emergence of new terminology such as artificial intelligence and machine learning. As can be seen in Figure 2, While other words experience declining or stagnant usage in 2017-2019, these two terms experience noticeable increase in usage. By 2024, they have both become more frequently mentioned than program evaluation in government documents. This is also confirmed by Figure 4, which shows that both artificial intelligence were not mentioned before Obama's second administration, but were used by the Trump administrations and Biden administration.

## **Insight 3 (Q2): Presidential Documents more noticeably change their terminology compared to Final Agency Rules**

In Figures 5-6 that focus on the bigram usage within presidential documents and federal agency rules, respectively, we can see that while five of the seven bigrams were used by federal agencies during the Bush administrations, President Bush only utilized one of the terms – data collection – in his documents. Additionally, while there are fluctuations in term usage in federal agency rules, they are more drastic in presidential documents; there are multiple instances where words terms mentions in one year are not mentioned in the subsequent year. However, these differences may be partially due to the vastly different sample size of the two document types.

## Bigram Usage in Presidential Documents

January 20, 2001-April 19,2025

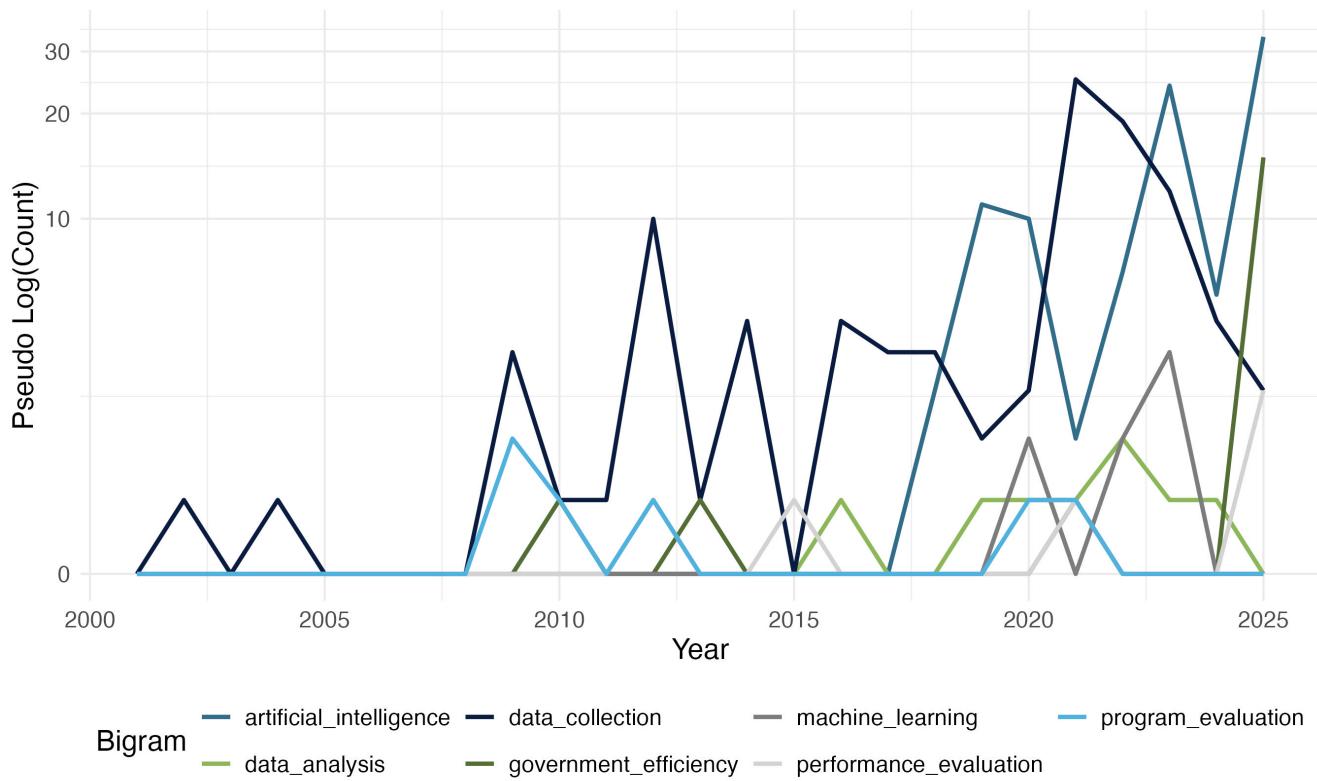


Figure 5: Bigram Term Usage in Presidential Documents

## Bigram Usage in Federal Agency Rules

January 20, 2001-April 19, 2025

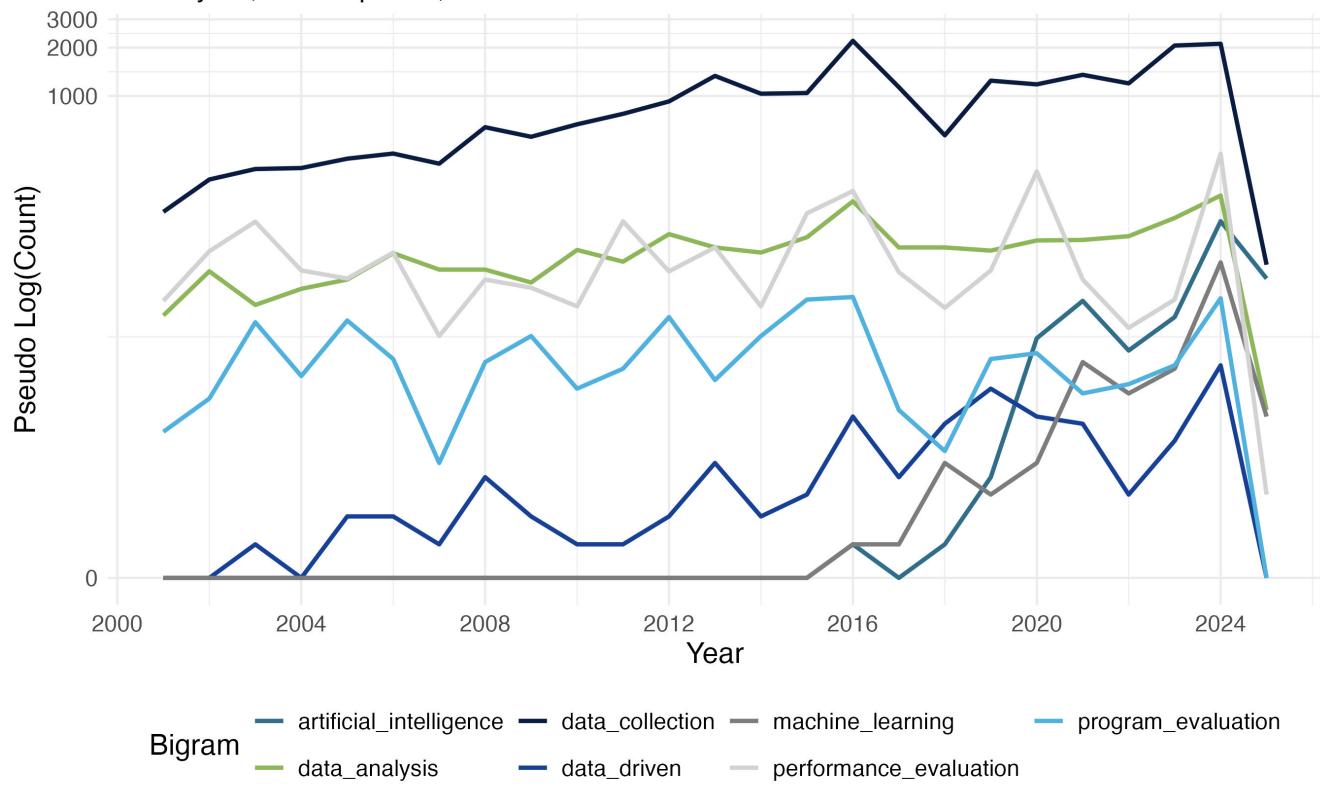


Figure 6: Bigram Term Usage in Agency Rules

**Insight 4 (Q2): Key Words in Context reveal that some terminology are more frequently used within repetitive, boilerplate text than others.**

The KWIC technique allows for the direct examination of immediate context of specific terms. Based on the results captured in Table 4 which include the most common context for 5 bigrams, we see that there are instances where the same exact wording is used dozens or hundreds of times across the corpus.

Table 4: Frequency of Particular Phrases around Specific Terms

Left Context	Key Target/Pattern	Right Context	Frequency
“any other aspect of this...”	data_collection	“...including suggestions for reducing the”	244
“solely for statistical research or...”	program_evaluation	“...purposes and which are not”	24
“office of the managing director...”	performance_evaluation	“...and records management shall send”	21
“recording concurrent computerized real time...”	data_analysis	“...and greater than hours of”	18
“implemented by a evaluator the...”	evaluation_plan	“...must articulate formative and summative”	11

## Insight 5 (Q3): Context of unigram evaluation terms is relatively consistent across administrations

Examining the results of the presidential administration word embedding models demonstrate that the context of unigrams has stayed relatively consistent. In Table 5, while each administration has at least one unique word for its context of “evaluation”, all administrations share a majority of words like “assessment”, “analysis”, “review”, and “results”. In Table 6, the consistency is even more pronounced with only two of the six administrations having unique words.

Table 5: Nearest Neighbors for “Evaluation”, by Presidential Administration

Bush-1	Bush-2	Obama-1	Obama-2	Trump	Biden
analysis	assessment	assessment	assessment	assessment	assessment
assessment	review	analysis	review	analysis	review
review	analysis	review	analyses	review	analysis
results	initial	analyses	consideration	response	consideration
appropriate	results	criteria	process	process	<b>development</b>
final	included	response	response	consideration	results
<b>performance</b>	<b>conducted</b>	initial	include	<b>determination</b>	criteria
page	final	based	<b>evaluations</b>	appropriate	include

Table 6: Nearest Neighbors for “Analysis”, by Presidential Administration

Bush-1	Bush-2	Obama-1	Obama-2	Trump	Biden
assessment	analyses	analyses	analyses	analyses	analyses
results	assessment	results	results	assessment	results
analyses	results	based	assessment	results	estimates
data	impact	impact	based	based	based
economic	estimates	<b>presented</b>	impact	estimates	final
final	economic	assessment	estimates	final	discussed
estimates	based	estimates	discussed	impact	impact
evaluation	final	discussed	evaluation	<b>print</b>	assessment

## **Insight 6 (Q3): Comparing contexts at the party level provides additional nuances in term usage**

While we see minimal variation across administrations term use, this is somewhat expected considering the small size of each sub-corpus (~5,000-7,000 documents). Figure 7 includes the results for the same unigrams in Figures 5-6 but at the larger presidential party subgroup level, where each subcorpus including more than 17,000 documents. Compared to the previous results, the context for “evaluation” at the party level provides more information on how Democrat and Republican administrations use the words differently. Each party has 3 unique terms as nearest neighbors. The results for “analysis” are still consistent, but the unique word for each party demonstrates that Democrat administrations put more emphasis on presenting analysis whereas Republican administrations put more emphasis on the tools used.

Table 7: Nearest Neighbors for “Evaluation” and “Analysis”, by Presidential Party (No Bigrams)

Evaluation (Democrat Corpus)	Evaluation (Republican Corpus)	Analysis (Democrat Corpus)	Analysis (Republican Corpus)
assessment	review	analyses	analyses
review	assessment	results	results
analysis	analysis	estimates	estimate
<b>conducted</b>	<b>testing</b>	based	<b>assessment</b>
results	<b>comprehensive</b>	<b>presented</b>	impact
<b>consideration</b>	results	estimate	based
criteria	criteria	impact	esimate

Evaluation (Democrat Corpus)	Evaluation (Republican Corpus)	Analysis (Democrat Corpus)	Analysis (Republican Corpus)
analyses	<b>initial</b>	final	final

## **Insight 7 (Q3): Comparing the context of bigrams is more revealing in how different political administrations utilize these terms**

While the unigram analysis demonstrates consistent context across groups, the bigram analysis highlights how groups discuss these concepts differently. In Table 8, which provides a comparison of bigrams “data analysis” and “data collection” at the party level, there are multiple difference between the two parties’ most similar words. For “data analysis”, more than half of the words are unique to each party. The democrat corpus results indicate that the Democrat administrations connect it to data collection and showing results, whereas Republican administrations have a more clinical, technical context for data analysis. For “data collection”, each party has three unique words. The nearest neighbor words for data collection for Democrat administrations are more general words whereas the Republican administrations have more of a context of continuous data collection and improvement.

Table 8: Comparing Party results for “Data Analysis” and “Data Collection”

Data Analysis (Democrat Corpus)	Data Analysis (Republican Corpus)	Data Collection (Democrat Corpus)	Data Collection (Republican Corpus)
<b>Data_collection</b>	calculations	reporting	reporting
<b>extension</b>	<b>extensive</b>	data	improve
<b>shows</b>	results	implementation	<b>ongoing</b>
data	<b>performed</b>	<b>collection</b>	<b>facilitate</b>
calculations	<b>tests</b>	<b>monitoring</b>	implementation
<b>detailed</b>	<b>validation</b>	measures	measures

Data Analysis (Democrat Corpus)	Data Analysis (Republican Corpus)	Data Collection (Democrat Corpus)	Data Collection (Republican Corpus)
showed results	data <b>laboratory</b>	believe improve	efforts data

## Discussion

This exploration of data and evaluation terminology helps demonstrate that the context and usage of these terms has varied throughout the 21st century. While overall the usage of these terms has increased since 2001, the growth has not been continuous and dipped during the first Trump administration. Additionally, we see that Presidents use fewer terms than agencies in their documents but have increased their use in the last few years. Lastly, the context of evaluation and data terminology can vary based on factors like administration and political party. While the usage of unigram terms was relatively consistent, the use of bigrams demonstrate notable differences in how Democrat and Republican administrations talk about evaluation.

**Stated Contributions:** While this is not a causal study, it illuminates some of the ways that the federal government has incorporated terminology as new technologies and methodologies are created and how administrations can change the context and interpretation of words in their official documents. Additionally, this study demonstrates some of the capabilities of conducting a text analysis like this to policymakers, who could use this type of study to identify shifts in policy priorities or topics for trainings, detect possible emerging trends, or to help guide resource allocation.

**Future Directions:** As part of the word embedding process, I broke my corpus into subgroups to fit the GloVe models and discern differences in the context across groups. A future direction of this effort would be to utilize a large language model such as Bidirectional Encoder Representation from Transformer (BERT) to fit the entire corpus and develop a contextualized word embedding. This type of model would provide more nuance in how the different administrations and political parties utilize these terminology and help illuminate some of the complex relationships between politics and evaluation.

## Citations:

Office of the Assistant Secretary for Planning and Evaluation (ASPE). (n.d.). Implementing the Foundations for Evidence-Based Policymaking Act at the U.S. Department of Health and Human Services. *U.S. Department of Health and Human Services*. <https://aspe.hhs.gov/topics/data/evidence-act-0>

(2025, April 10). Federal Register, 1936 to Present. *Govinfo*. <https://www.govinfo.gov/help/fr#about>

(n.d.) Federal Register. *National Archives*. <https://www.federalregister.gov/>  
Pennington, J., Socher, R., and Manning, C.D. (2014). *GloVe: Global Vectors for Word Representation*. <https://nlp.stanford.edu/projects/glove/>