

Context of Evaluation and Data in 21st century Federal Policy: A Text Analysis Approach

May 6, 2025

By Katharyn Loweth



GEORGETOWN UNIVERSITY

Outline

- Motivation
- Research Questions
- Data Source
- Text Analysis
 - Term Frequency
 - Word Embeddings (GloVe)

Motivation

- M.S. Candidate in Data Science for Public Policy
- Professional experience in the research sector and multiple agencies within federal government.
 - NORC at the University of Chicago
 - Department of Health and Human Services
 - U.S. Census Bureau
 - Department of Defense



OFFICE OF COMMUNITY SERVICES

An Office of the Administration for Children & Families



OFFICE OF PEOPLE ANALYTICS
U.S. DEPARTMENT OF DEFENSE

Motivation

- Evaluations and data-informed decision-making help policymakers design and improve programs, policies, and regulations
- My Professional Goal: To utilize data science methods to support policymakers and reduce burden of evaluations on programs and bureaucrats



Motivation

- Context of evaluation and the tools available has changed significantly over the last 25 years
 - New technology/techniques (machine learning)
 - New policies (Evidence Act of 2018)
 - Change in Administration Priorities (Department of Government Efficiency)



Research Questions

1. Since January 20, 2001, how has the federal governments' use of evaluation and data terminology changed?
 - a) Has 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 data and evaluation terminology differently?
3. How do political parties and presidential administrations contextualize evaluation and data differently?



Data Source

Data Source

- Federal Register API
- Used text search to compile **final agency rules** and **presidential documents** from January 20, 2001 through April 19, 2025 based on the following keywords/phrases*:
 - **Keywords:** “evaluation”, “efficiency”, “evidence”, “AI”, “data”**, “analysis”**
 - **Phrases:** “Government efficiency”, “government evaluation”, “data analysis”, “data science”, “data driven”, “program evaluation”, “performance evaluation”, “machine learning”, “artificial intelligence”, “data collection”, “evidence building”, “evaluation plan”, “evaluation policy”, “data management”

Data Source

- Total # of Documents: 36,729
 - 36,048 Agency Final Rules
 - 681 Presidential Documents

Bush Admins.	Obama Admins.	Trump Admins.	Biden Admin.
13,957	12,281	5,587	4,904

- Corpus: Full text versions of documents obtained by using the *body_text_html* feature from the Federal Register API results to web scrape text.

Data Preprocessing

- Added document metadata variables:
 - President's political party
 - President's administration term (first or second term)
- Using corpus, created two objects for analysis:
 - One object that consists of only unigram tokens (“data” and “analysis” are two separate tokens)
 - One object that includes custom bigrams (“data_analysis” is considered one token)

Insight 1: Use of Bigrams Vary Widely Across Corpus

- Frequency of Key Unigrams (Unigram Corpus)

AI	Analysis	Data	Efficiency	Evaluation	Evidence
7,118	327,411	757,524	86,777	117,296	116,752

- Frequency of Key Bigrams (Bigram Corpus)

Artificial Intelligence	Data Analysis	Data Collection	Data Science	Data Driven	Data Management	Evaluation Policy	Evaluation Plan	Evidence Building
493	2,720	22,646	27	114	708	25	363	6

Government Efficiency	Government Evaluation	Machine Learning	Performance Evaluation	Program Evaluation
67	9	181	2,660	611

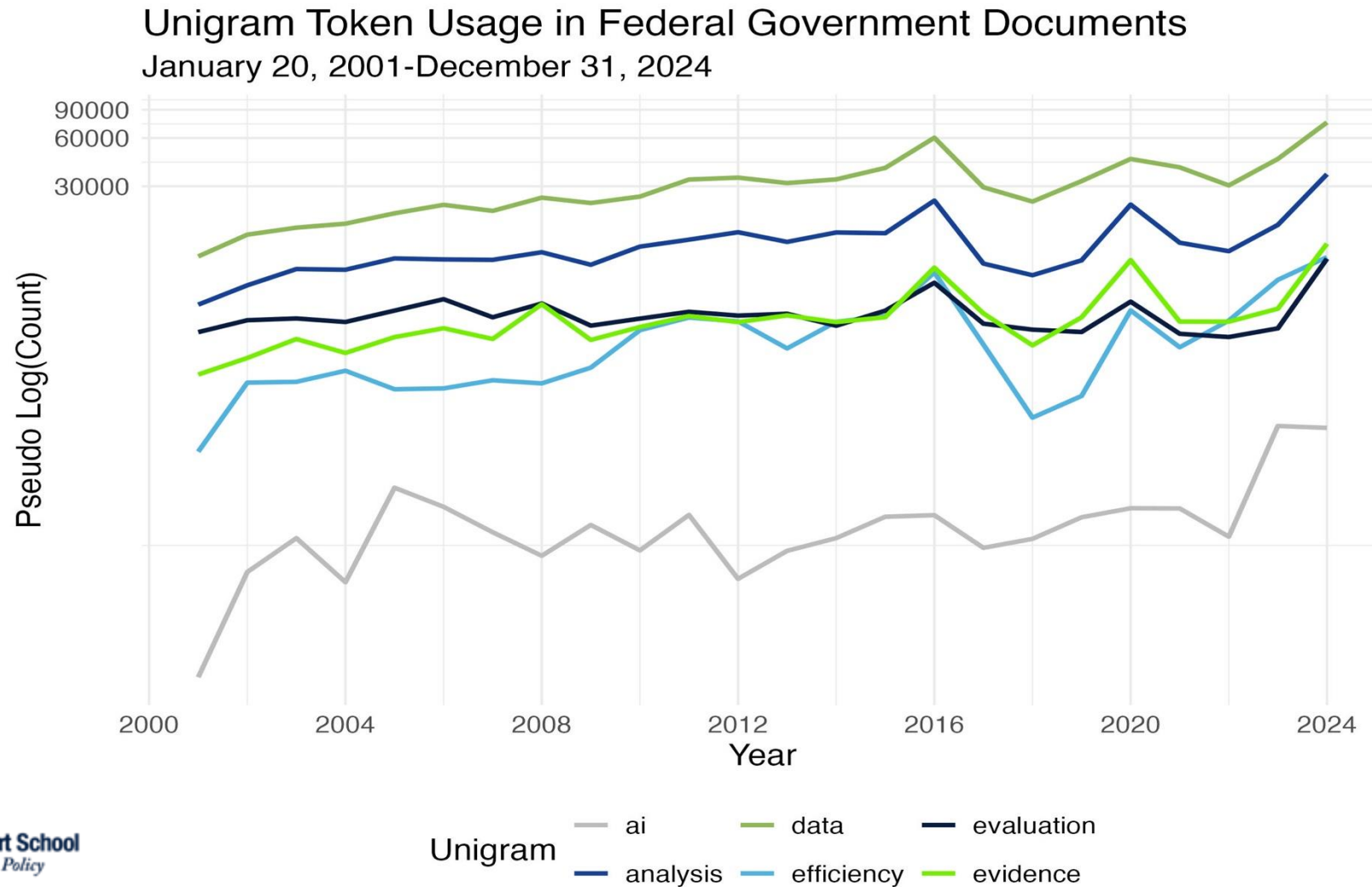


Analysis Part 1: Use of Words Over Time

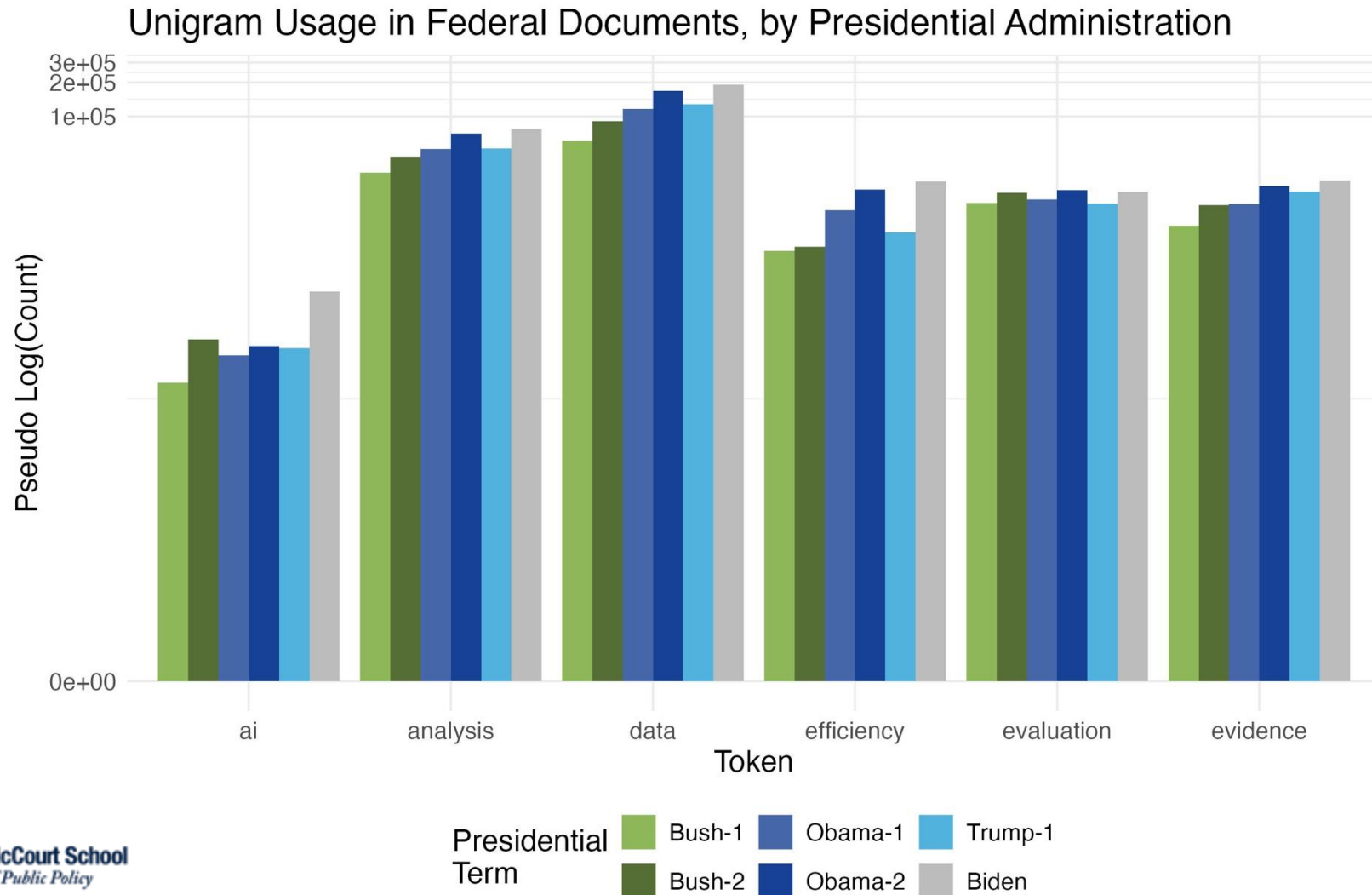
Q1

Since January 20, 2001, how has the federal governments' use of evaluation and data terminology changed?

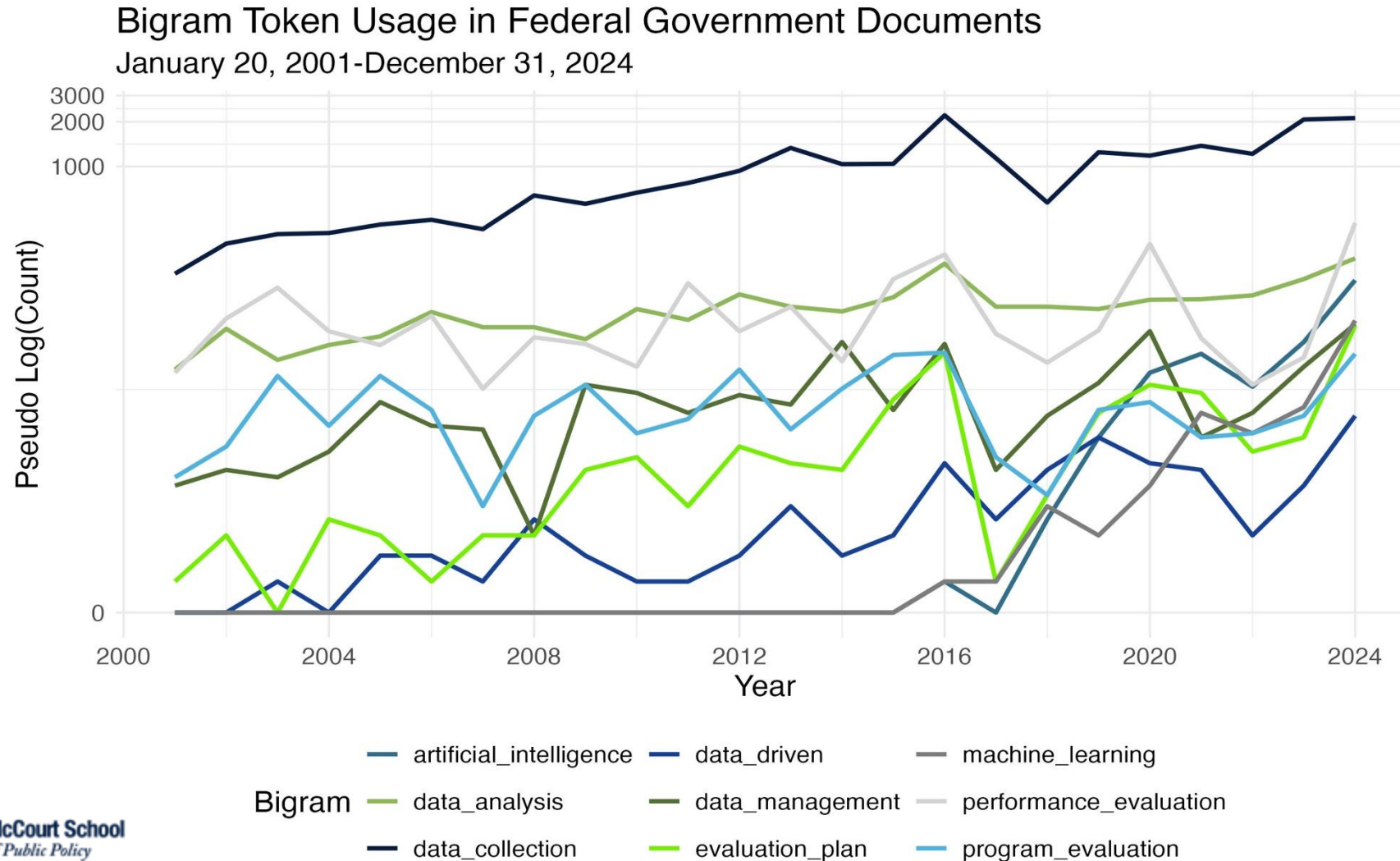
Insight 2: Use of Terminology Peaked in 2016, 2020, and 2024



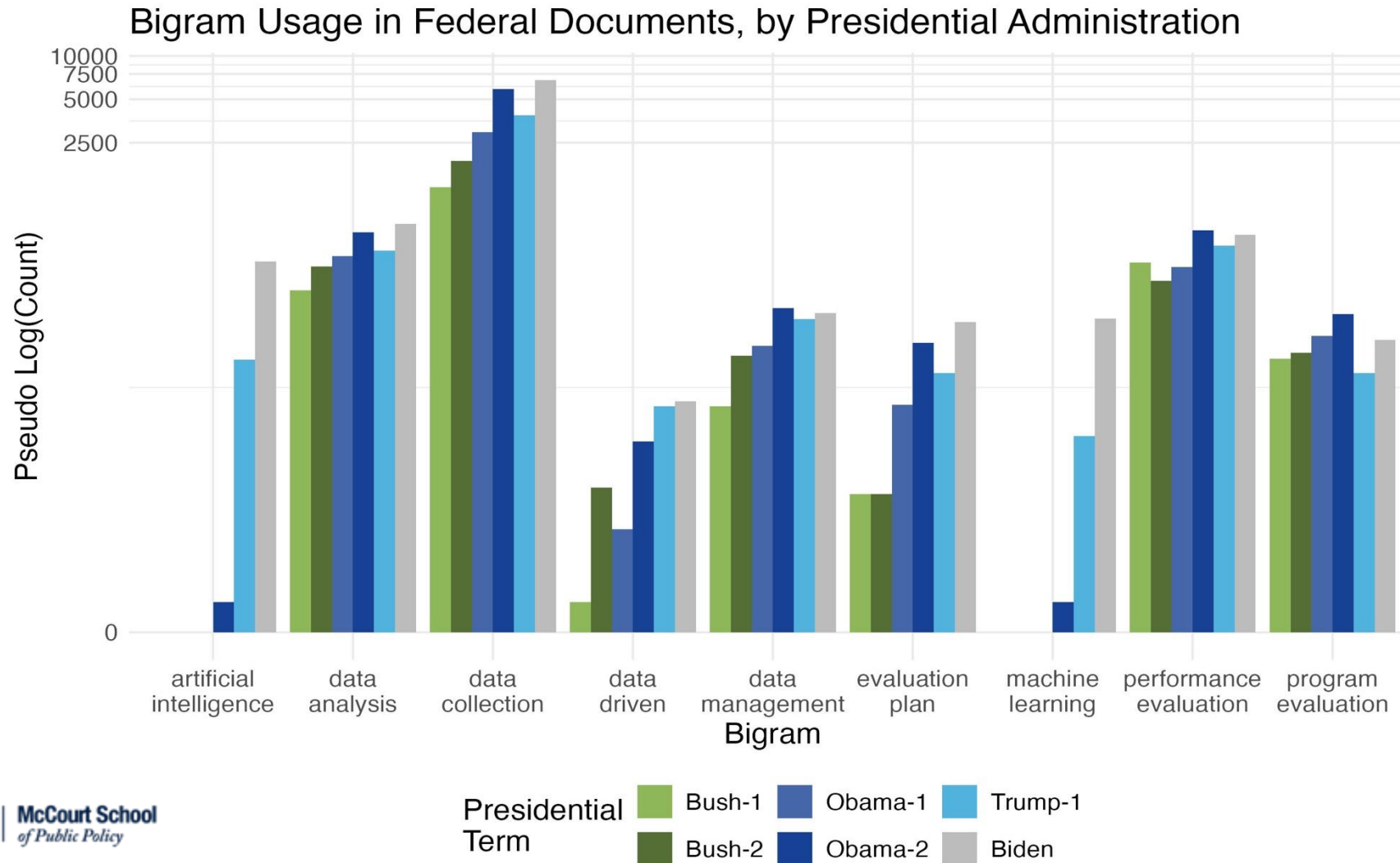
Insight 3: There is a gradual (but not continuous) increase in evaluation-related term usage across administrations



Insight 4: Use of New Bigram Terminology (Artificial Intelligence, Machine Learning) Begins in 2016



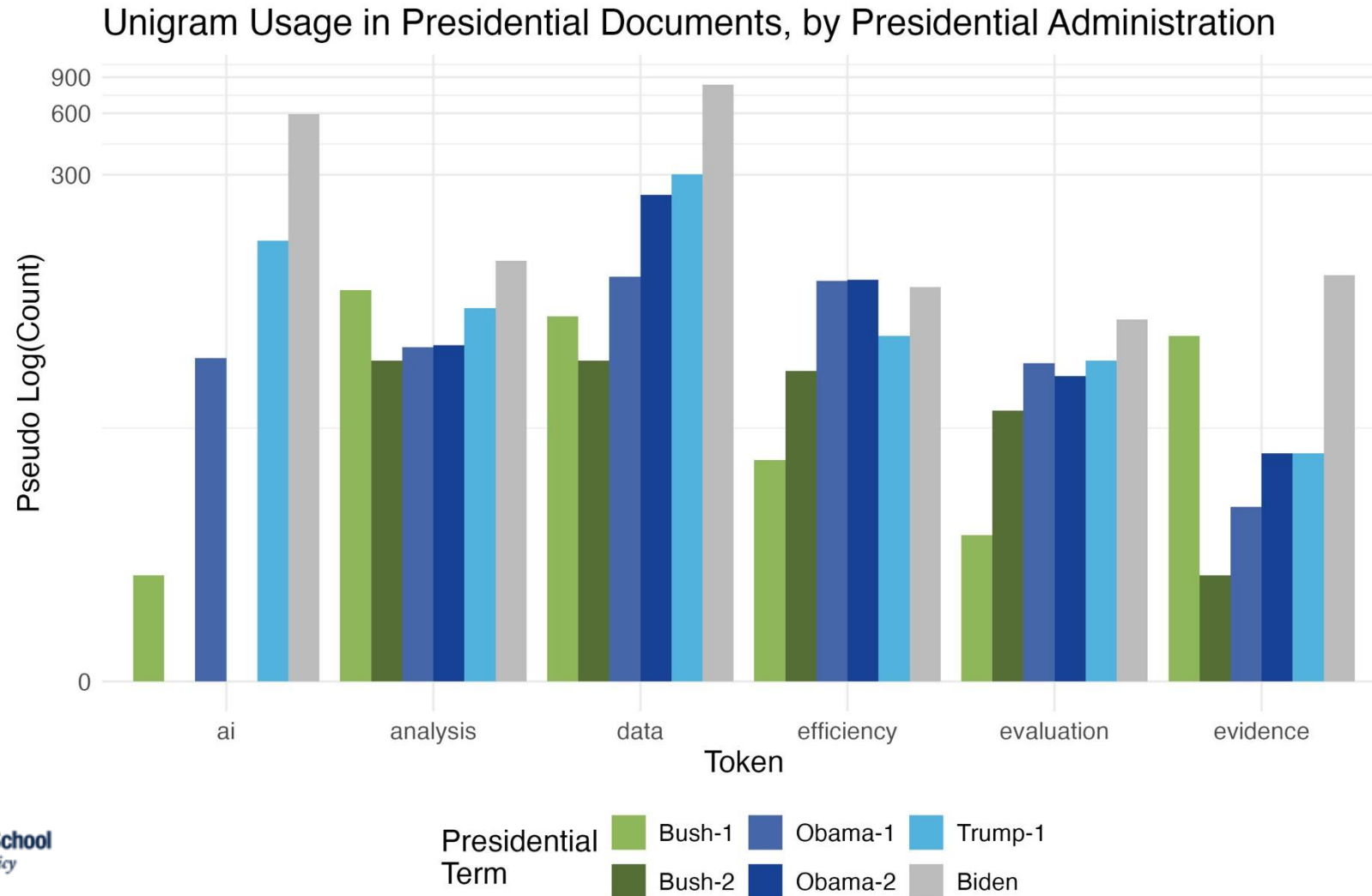
Insight 4: Use of New Bigram Terminology (Artificial Intelligence, Machine Learning) Begins in 2016



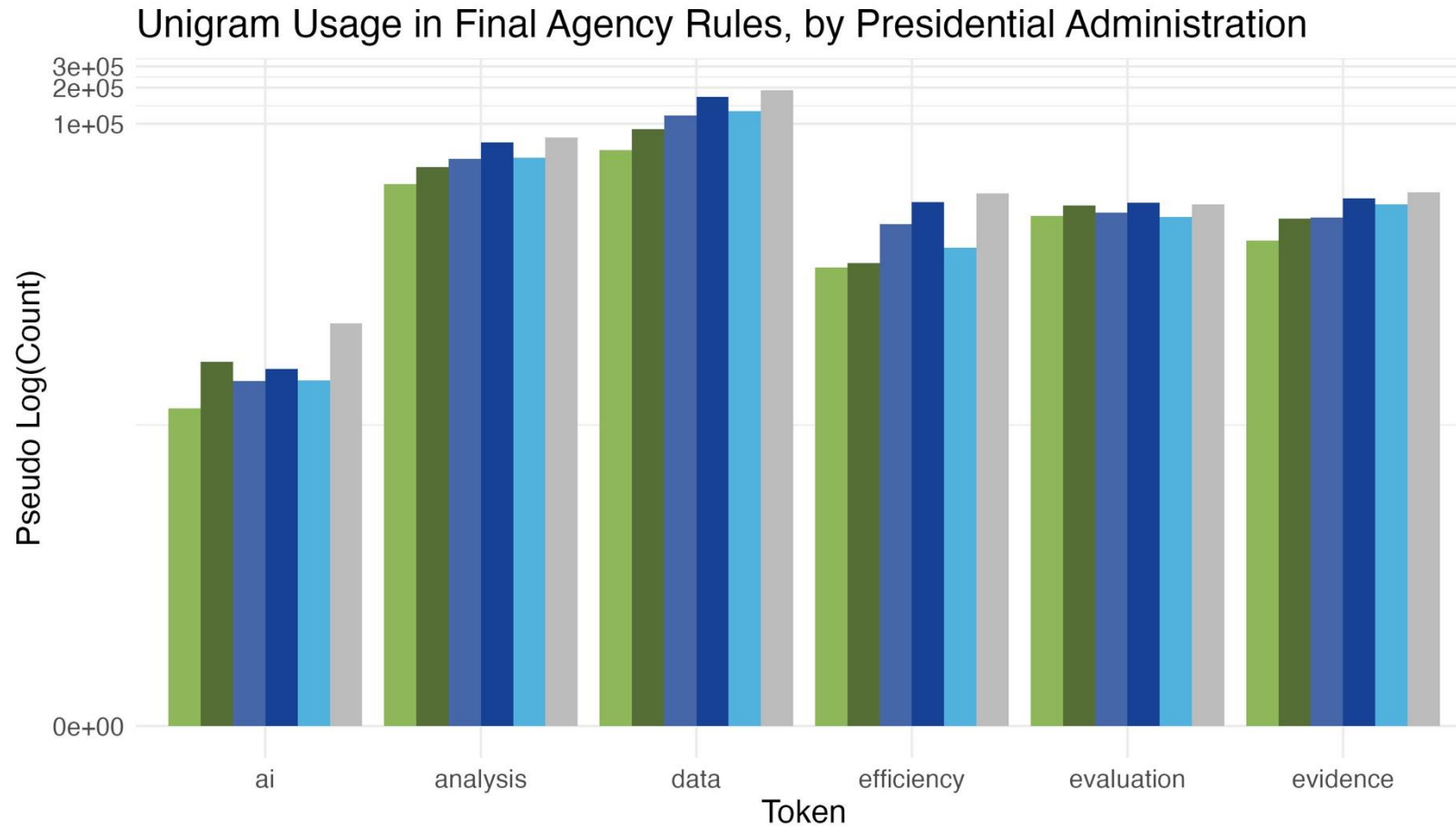
Q2

Do federal agencies and the Executive Office of the President utilize data and evaluation terminology differently?

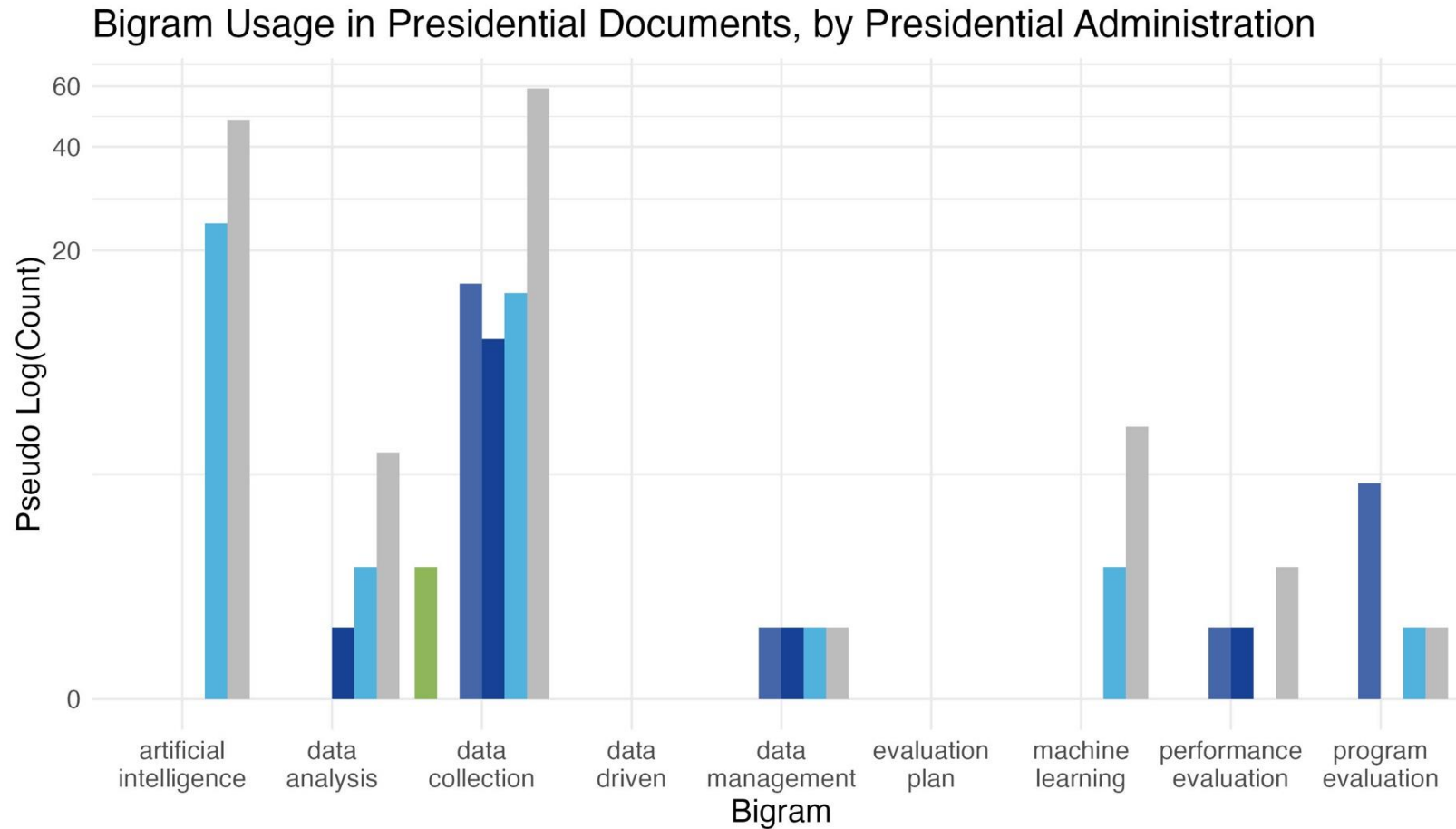
Insight 5: Presidents vary their use of evaluation and data terminology in their documents



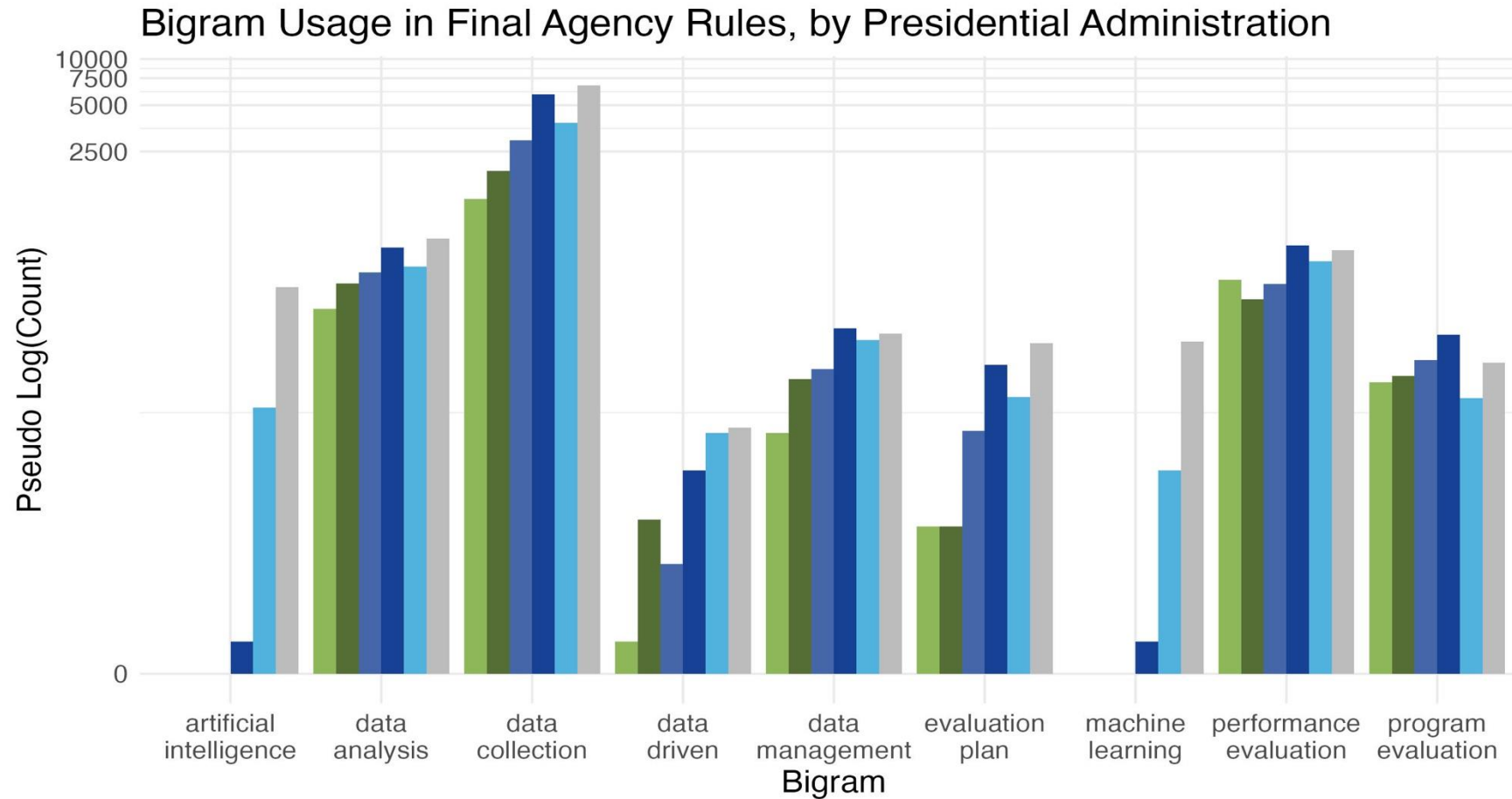
Insight 6: Federal Agencies are more consistent in their use of these terms



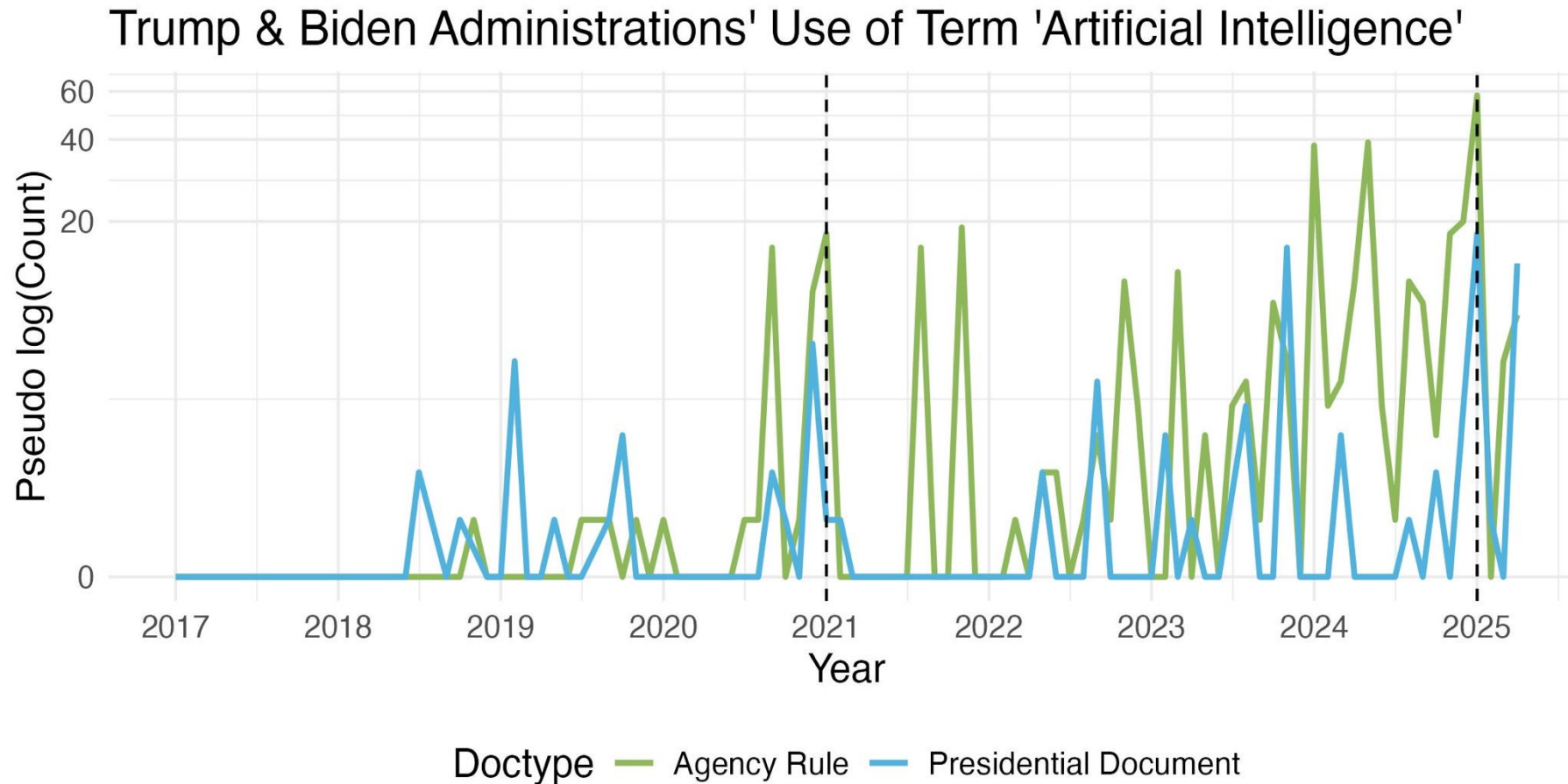
Insight 7: Some bigram terms are only used in Federal Agency Rules



Insight 7: Some bigram terms are only used in Federal Agency Rules



Insight 8: Presidents and Agencies have similar use patterns for emerging technical terms like Artificial Intelligence





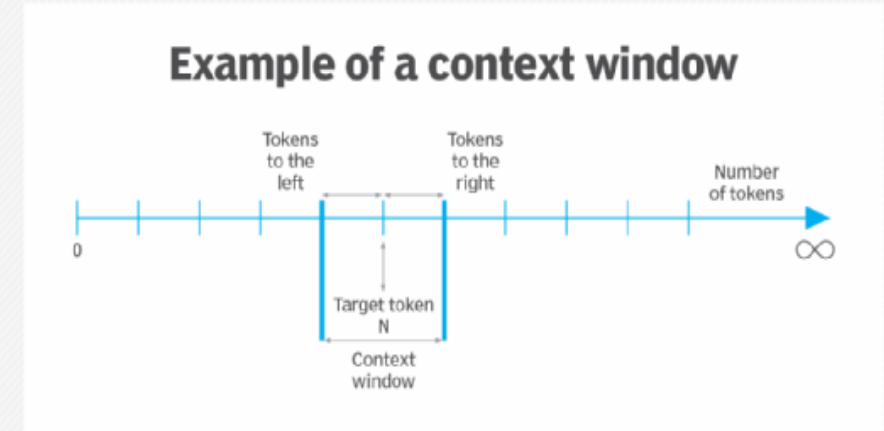
Analysis Part 2: Context of Words

Q3

How do political parties and presidential administrations contextualize evaluation and data differently?

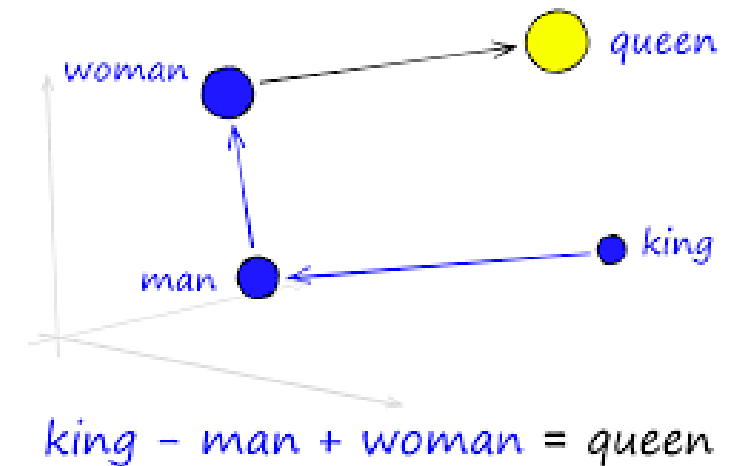
Context of Words

- “You shall know a word by the company it keeps” (Firth, 1957)
- Previous analysis showed how often terms are used not how they are being used.
- Context Window: X number of tokens to the left and right of a target token



Word Embeddings

- GloVe (Global Vectors for Word Representation) model
 - An unsupervised learning algorithm that capture information about tokens in a corpus by obtaining a multi-dimensional vector representation of the words
 - Utilizes feature co-occurrence (2 tokens appearing together in a context window) to measure the linguistic or semantic similarity of words in a corpus
 - Uses an optimized cost function to learn and improve embeddings iteratively



Word Embeddings - Methodology

- Subsetted corpus by presidential administration
 - 6 groups: Bush-1, Bush-2, Obama-1, Obama-2, Trump, Biden
- Subsetted corpus by president's political party
 - 2 groups: Democrat and Republican
- With each corpus:
 - Created feature co-occurrence matrix based on a context window of 5 with a weighted count
 - Fit GloVe model with 300 dimensions with 12 iterations to minimize error
- Examined bigrams within Democrat and Republican corpuses

Insight 9: Context of evaluation is relatively consistent across administrations

Nearest Neighbors for “Evaluation”

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	analysis	review	analysis
results	initial	analyses	consideration	response	consideration
appropriate	results	criteria	<i>process</i>	<i>process</i>	development
<i>final</i>	included	response	response	consideration	results
performance	conducted	initial	include	determination	criteria
page	<i>final</i>	<i>based</i>	evaluations	appropriate	include
<i>comprehensive</i>	<i>comprehensive</i>	described	results	testing	report
initial	provided	research	<i>based</i>	include	following

Bold = unique word

Italicized = appears in two adjacent administrations

Insight 9: Context of evaluation is relatively consistent across administrations

Nearest Neighbors for “Evidence”

Bush-1	Bush-2	Obama-1	Obama-2	Trump	Biden
showing	show	indicating	studies	lack	lack
indicate	whether	studies	lack	studies	record
sufficient	studies	lack	indicating	sufficient	supporting
results	upon	supporting	record	whether	showing
lack	<i>data</i>	indicate	indicate	results	studies
studies	documentation	showing	provided	relevant	indicating
adequate	showing	documentation	<i>fact</i>	supporting	whether
<i>data</i>	presented	show	documentation	demonstrate	documentation
whether	provided	sufficient	whether	<i>fact</i>	show
provide	sufficient	indicates	supporting	record	sufficient

Bold = unique word

Italicized = appears in two adjacent administrations

Insight 10: Party-level results provide more nuance into different contexts of terms

GloVe Embedding Model – Party level (no Bigrams)

Evaluation		Evidence	
Democrat	Republican	Democrat	Republican
assessment	review	indicating	lack
review	assessment	lack	show
analysis	analysis	absence	studies
conducted	testing	studies	relevant
results	comprehensive	showing	showing
consideration	results	record	whether
criteria	criteria	show	indicate
analyses	initial	fact	results
based	conducted	indicates	fact
appropriate	complete	supporting	sufficient

Insight 11: Comparing the context of bigrams is more revealing than comparing unigrams

GloVe Embedding Model – Party Level (with Bigrams)

Data		Analysis		Data Analysis		Data Collection	
Democrat	Republican	Democrat	Republican	Democrat	Republican	Democrat	Republican
available	using	analyses	analyses	Data_collection	calculations	reporting	reporting
using	results	results	results	extension	extensive	data	improve
report	available	estimates	assessment	shows	results	implementation	ongoing
based	used	based	estimates	data	performed	collection	facilitate
results	based	estimate	based	calculations	tests	monitoring	implementation
used	report	final	impact	detailed	validation	measures	measures
reported	indicate	presented	data	showed	data	believe	efforts
reports	print	impact	discussed	results	laboratory	improve	data
addition	addition	discussed	economic	tool	sampling	additional	submission
page	reports	assessment	final	surveys	testing	ongoing	process



Summary & Looking Ahead

Summary

- Use of evaluation-related terminology generally (but not continuously) increases over time and presidential administration.
- Context varies slightly across administrations, but it is more pronounced when comparing bigrams.



Summary

- While not a causal study, this examination illuminates some of the nuances in how the federal government incorporates new data-related terminology into its written documents.



Applications

- Applying text analysis methods to federal government documents can help policymakers:
 - Identify shifts in policy priorities
 - Identify possible topics for training and development
 - Detect possible emerging trends or technologies in this policy area
 - Guide resource allocation within program/agency

Extensions

- Implement a large language model (LLM) like Bidirectional Encoder Representation from Transformer (BERT) model to create a contextualized word embedding on the whole corpus
- While this is a case study on specific evaluation and data-related terminology, this methodology could be applied to other key terminology or policy concerns such as:
 - Cybersecurity/cyberspace
 - Information Technology



Thank You!

Acknowledgement & Contact Information

Special thanks to Text as Data Professor Nejla Asimovic and Professors Tiago Ventura and Rebecca Johnson at Georgetown University who provided support on this presentation.

- GitHub Repository with code and replication materials:
 - https://github.com/kloweth84/PPOL6801_TextAsData_FinalProject
- Contact information:
 - ksl84@Georgetown.edu
 - Katharyn.Loweth@gmail.com

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/>