

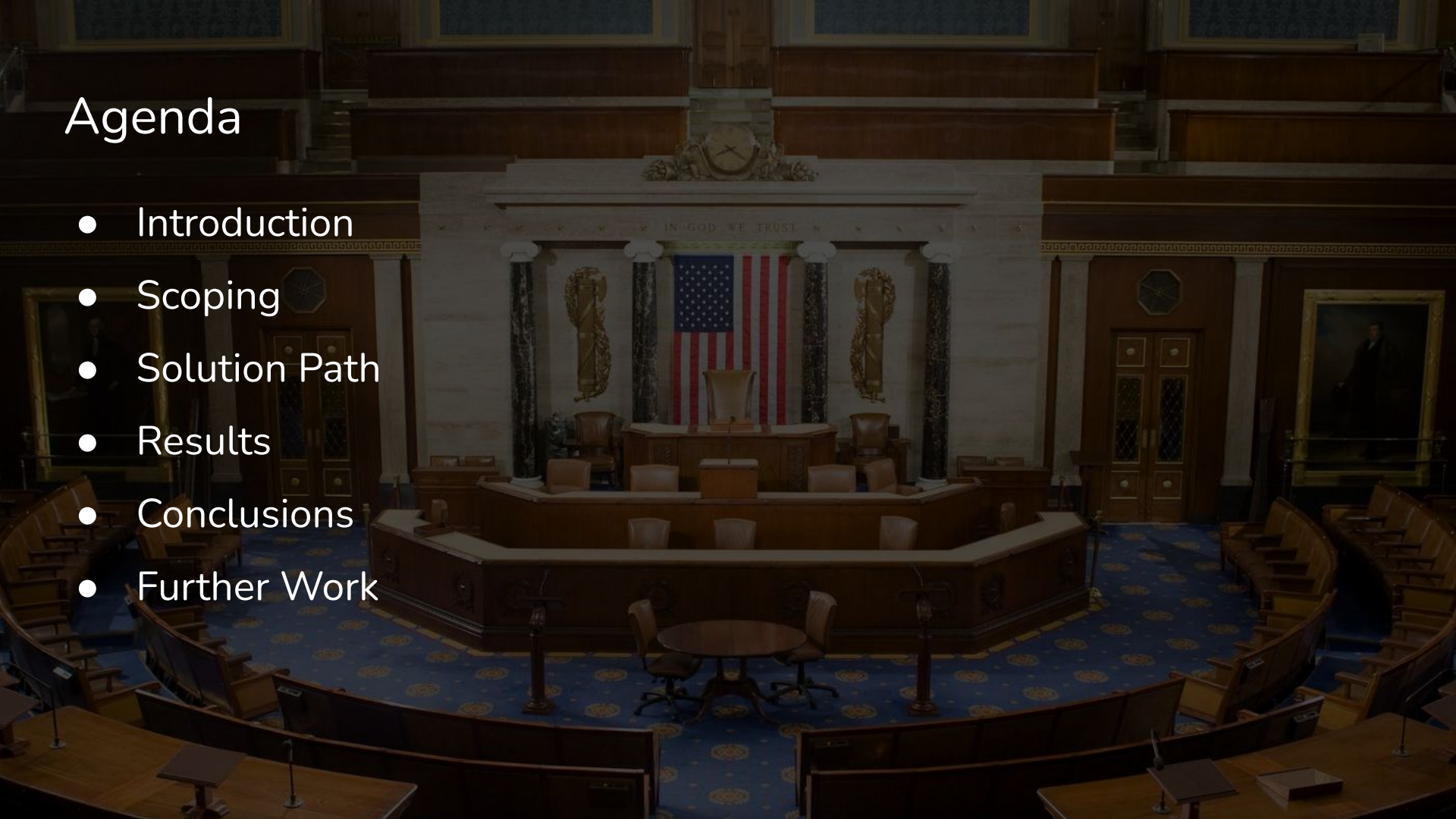


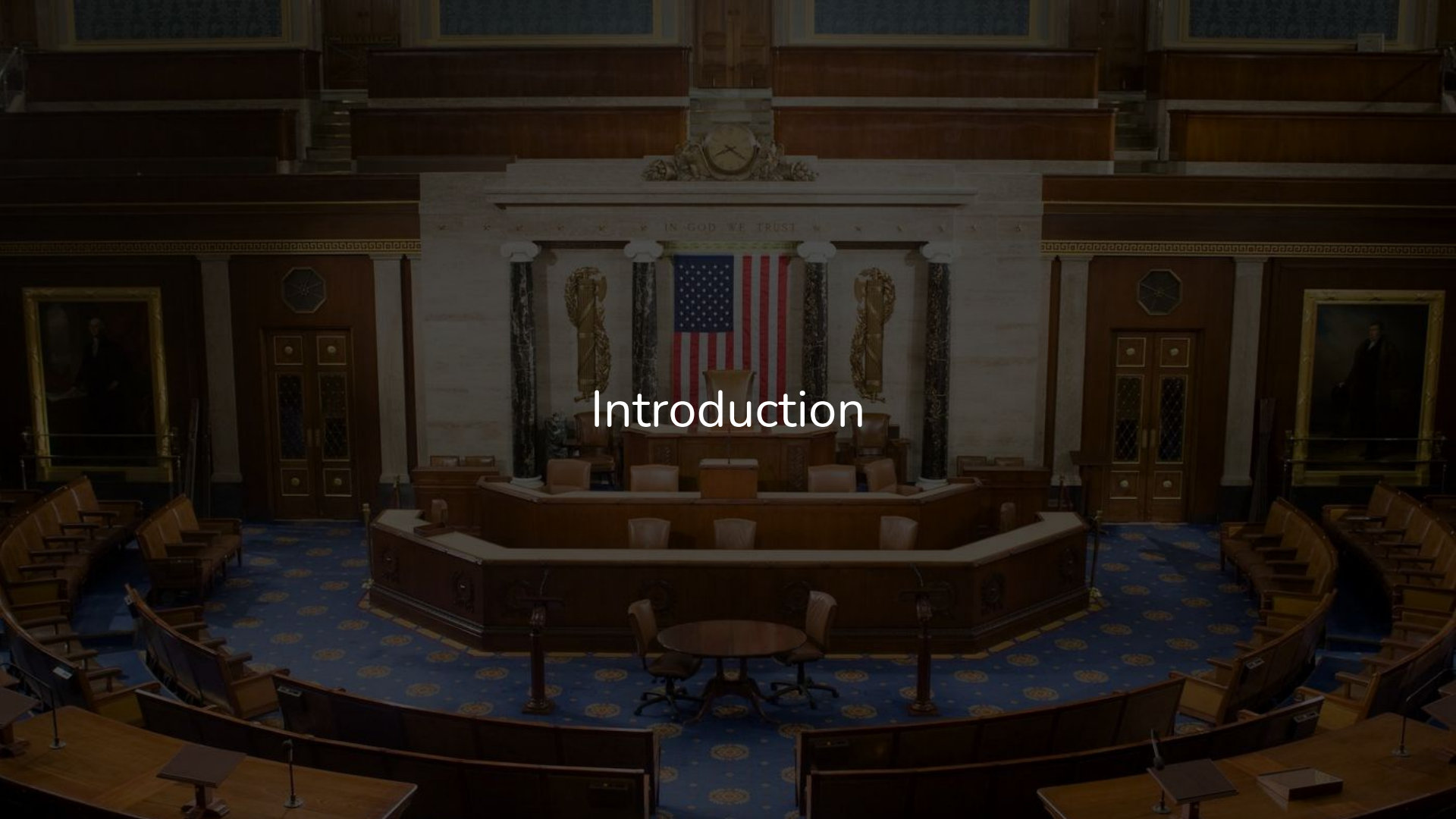
# Topic Modeling the State of the Union

Nate DiRenzo

# Agenda


- Introduction
- Scoping
- Solution Path
- Results
- Conclusions
- Further Work





# Introduction



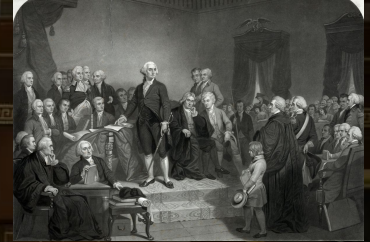


*“He shall, from time to time, give to the Congress information of the State of the Union and recommend to their Consideration such measures as he shall judge necessary and expedient.”*

-Article II, Section III  
United States Constitution

# What is the State of the Union Address

- First delivered in 1790 by George Washington
  - 1801-1913: Delivered in written form
  - 1923: Radio broadcast across the nation
  - 1947: First broadcast on TV
  - 1965: National evening TV broadcast
- Changes in purpose over time
  - Executive updates
  - Discuss minutia
  - Advocate legislation
  - Highlight Administration Accomplishments & Goals
- Importance in the modern political discourse





# Why is the SOTU worth analyzing?

- Identify the priority and goals of an administration
- Major initiative or legislative announcements
- Understand political messaging and positioning for a president's party
- Global, national zeitgeist
- Analysis over time shows changes in:
  - Word choice
  - Topics
  - Priorities

**NATIONAL  
YOUTH  
SUMMIT**  
**WAR ON  
POVERTY**





Scoping



# Project Scope

- Extract the top 10 topics from SOTU addresses 1790-2022
- Assign topic probabilities and identify top topic for each address
- Cluster speeches based on probabilities
- Examine topic prevalence and interpret clusters

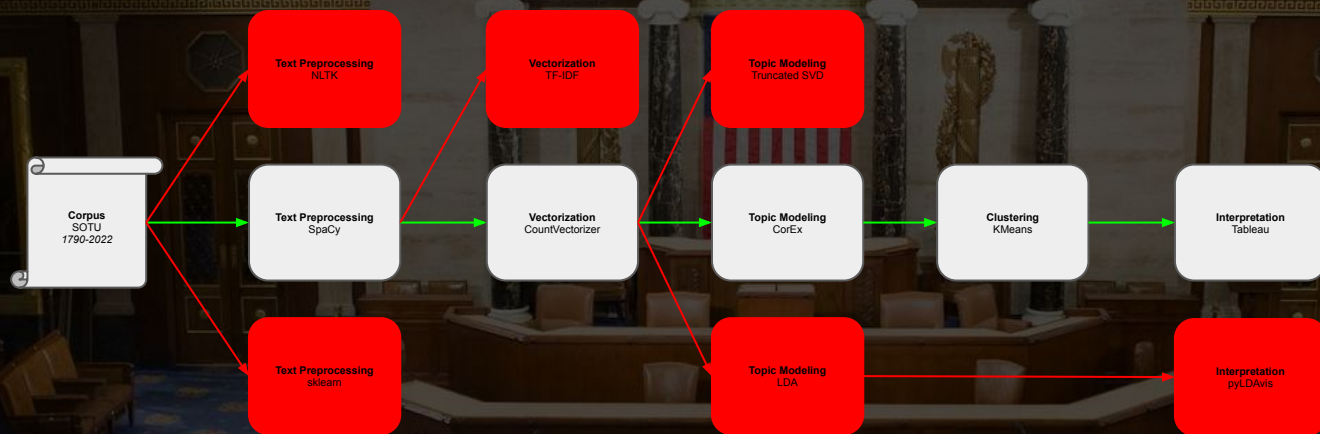







Solution Path

# Solution Path







Results

# CorEx Topic Modeling - Unsupervised

- 1: custom,make,rigidly,postage,postmaster,benefit,favorable consideration,expend,postmaster general,aggregate
- 2: clause,convinced,organization,value,product,outlet,investigation,topic,absence,condemn
- 3: far concern,await,dangerous,domestic,announce,imminent,administration justice,duty place,attention direct,head
- 4: tonight,decade,speaker,deficit,consideration,help,job,nuclear,provision,technology
- 5: opinion,vest,network,coalition,iran,regard,practicable,iraq,arrangement,modernize
- 6: court district,thrift,guard,increase funding,crush,express opinion,insufficient,remain quarter,age,cut cost
- 7: big,remember,report postmaster,report postmaster general,hard,ahead,sharing,worker,develop,new technology
- 8: useful,small business,article,missouri,salutary,role,permanent,beneficial result,occur,satisfactory
- 9: accede,sword,early action,time recommend,sagacity,legislate,invalid,england france,necessity exist,difficult
- 10: ought,scourge,commonly,view,perspective,straits,effect,secure citizen,long war,evinced



# CorEx Topic Modeling - Semi-Supervised

Labor: program, **job**, **economic**, today, budget, help, goal, ahead, basic, **worker**

Manufacturing: crop, farmer, **agriculture**, farm, method, production, business, department **commerce**, type, interstate commerce

Government: constitution, outbreak, **federal**, **supreme**, amendment, service law, countryman, election, money order, total

Finance: cent, tariff, **fiscal**, expenditure, **currency**, treasury, end june, suggestion, postmaster, fiscal end

Social Issues: mexico, **civil**, protest, movement, **immigration**, slight, indian, **citizenship**, purely, lay senate

Foreign Policy: **cuba**, navy, army, **japanese**, consent, france, spain, **rico**, officer, **panama**

Science & Technology: research, technology, decade, achieve, challenge, **environment**, **small business**, **cancer**, leader, big

Law/Jurisprudence: court, hold, dangerous, perpetuity, commissioners district, department end, british north, fictitious, department end june, end june estimate

Education: school, student, **education**, college, university, math, street, teacher, **medicare**, **tax credit**

Infrastructure: interstate, **railway**, criminal, **highway**, commission appoint, colorado, consular service, naturalization, sign, **bridge**


# Clustering

[Tableau](#)





# Interpretation



Cluster 1: 1913-1978, Labor, Manufacturing, Government, Science & Technology

Cluster 2: 1790-2022, Labor

Cluster 3: 1817-1909, Government, Manufacturing, Law

Cluster 4: 1845-1981, Labor

Cluster 5: 1827-1898, Government, Manufacturing, Finances

Cluster 6: 1843-1888, Labor

Cluster 7: 1790-1917, Government, Law

Cluster 8: 1825-1906, Manufacturing, Finance

Cluster 9: 1874-1956, Manufacturing, Labor

Cluster 10: 1841-1955, Labor



Conclusions



# Conclusions

- Labor, Manufacturing dominate
- Finance, Government, Law are also popular
- Science and Technology shows up once (Roosevelt, 1936)
- Clusters follow epochs/prevaling topics?
- Almost all post 1943 SOTU's in Cluster 2, Economic focus
- Computers can't do everything (yet)!



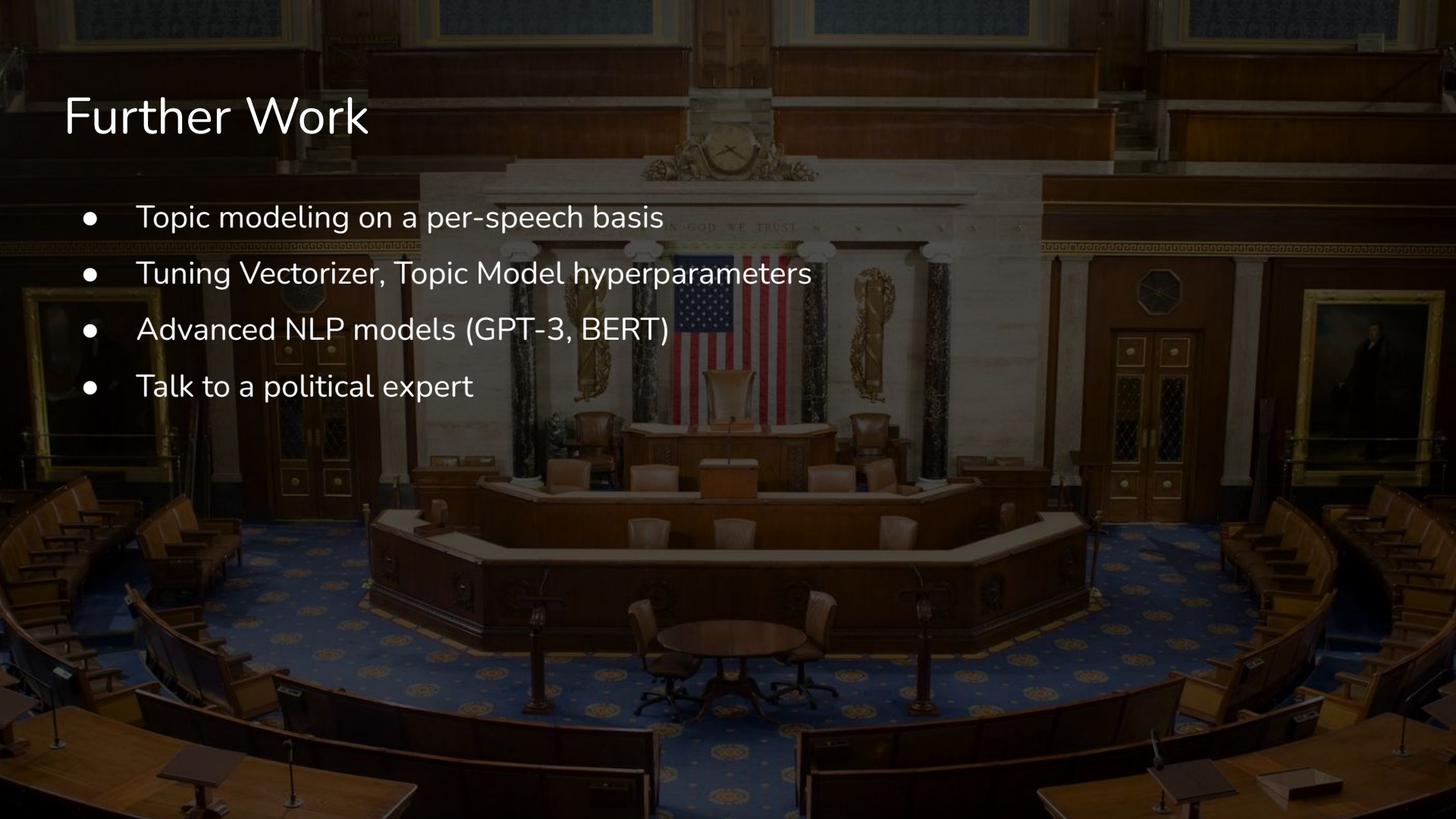


Further Work



# Further Work

- Topic modeling on a per-speech basis
- Tuning Vectorizer, Topic Model hyperparameters
- Advanced NLP models (GPT-3, BERT)
- Talk to a political expert





Thank You!