

Presidential Debate Themes Analysis

This tutorial explains how to analyze themes in presidential debates using the DocETL pipeline. We'll cover the pipeline structure, explain each operation, and discuss the importance of theme resolution.

Pipeline Overview

Our goal is to build a pipeline that will:

1. Extract key themes and viewpoints from presidential debate transcripts
2. Analyze how these themes have evolved over time, with references to specific debates and quotes

You can take a look at the raw data [here](#).

Let's examine the pipeline structure and its operations:

```
pipeline:
  steps:
    - name: debate_analysis
      input: debates
      operations:
        - extract_themes_and_viewpoints
        - unnest_themes
        - summarize_theme_evolution
```

**Full Pipeline Configuration**

```

datasets:
  debates:
    type: file
    path: "data.json"

system_prompt:
  dataset_description: a collection of transcripts of presidential debates
  persona: a political analyst

default_model: gpt-4o-mini

operations:
  - name: extract_themes_and_viewpoints
    type: map
    output:
      schema:
        themes: "list[{{theme: str, viewpoints: str}}]"
    prompt: |
      Analyze the following debate transcript for {{ input.title }} on {{
input.date }}:

      {{ input.content }}

      Extract the main themes discussed in this debate and the viewpoints of
the candidates on these themes.
      Return a list of themes and corresponding viewpoints (including the
specific quotes from the debate) in the following format:
      [
        {
          "theme": "Theme 1",
          "viewpoints": "Candidate A's viewpoint... Candidate B's viewpoint..."
        },
        {
          "theme": "Theme 2",
          "viewpoints": "Candidate A's viewpoint... Candidate B's viewpoint..."
        },
        ...
      ]

  - name: unnest_themes
    type: unnest
    unnest_key: themes
    recursive: true

  - name: summarize_theme_evolution
    type: reduce
    reduce_key: theme
    output:
      schema:
        theme: str
        report: str
    prompt: |
      Analyze the following viewpoints on the theme "{{ inputs[0].theme }}"
from various debates over the years:

      {% for item in inputs %}
      Year: {{ item.year }}

```

```
Date: {{ item.date }}
Title: {{ item.title }}
Viewpoints: {{ item.viewpoints }}

{% endfor %}
```

Generate a comprehensive summary of how Democratic and Republican viewpoints on this theme have evolved through the years. Include supporting quotes from the debates to illustrate key points or shifts in perspective.

Your summary should:

1. Identify *all* major trends or shifts in each party's stance over time
2. Highlight any significant agreements or disagreements between the parties
3. Note any external events or factors that may have influenced changes in viewpoints
4. Use specific quotes to support your analysis
5. The title should contain the start and end years of the analysis

Format your response as a well-structured report.

```
pipeline:
  steps:
    - name: debate_analysis
      input: debates
      operations:
        - extract_themes_and_viewpoints
        - unnest_themes
        - summarize_theme_evolution

  output:
    type: file
    path: "theme_evolution_analysis.json"
    intermediate_dir: "checkpoints"
```

Pipeline Operations

1. Extract Themes and Viewpoints

```
- name: extract_themes_and_viewpoints
  type: map
  output:
    schema:
      themes: "list[{{theme: str, viewpoints: str}}]"
  prompt: |
    Analyze the following debate transcript for {{ input.title }} on {{
input.date }}:

    {{ input.content }}

    Extract the main themes discussed in this debate and the viewpoints of the
    candidates on these themes.

    Return a list of themes and corresponding viewpoints in the following
    format:
    [
      {
```

```

    "theme": "Theme 1",
    "viewpoints": "Candidate A's viewpoint... Candidate B's viewpoint..."
  },
  {
    "theme": "Theme 2",
    "viewpoints": "Candidate A's viewpoint... Candidate B's viewpoint..."
  },
  ...
]

```

This operation processes each debate transcript to identify main themes and candidates' viewpoints. It uses AI to analyze the content and structure the output in a consistent format.

2. Unnest Themes

```

- name: unnest_themes
  type: unnest
  unnest_key: themes
  recursive: true

```

The unnest operation flattens the list of themes extracted from each debate. This step prepares the data for further analysis by creating individual entries for each theme.

3. Summarize Theme Evolution

```

- name: summarize_theme_evolution
  type: reduce
  reduce_key: theme
  output:
    schema:
      theme: str
      report: str
  prompt: |
    Analyze the following viewpoints on the theme "{{ inputs[0].theme }}" from
    various debates over the years:

    {% for item in inputs %}
    Year: {{ item.year }}
    Date: {{ item.date }}
    Title: {{ item.title }}
    Viewpoints: {{ item.viewpoints }}

    {% endfor %}

    Generate a comprehensive summary of how Democratic and Republican
    viewpoints on this theme have evolved through the years. Include supporting
    quotes from the debates to illustrate key points or shifts in perspective.

    Your summary should:
    1. Identify all major trends or shifts in each party's stance over time

```

2. Highlight any significant agreements or disagreements between the parties
 3. Note any external events or factors that may have influenced changes in viewpoints
 4. Use specific quotes to support your analysis
 5. The title should contain the start and end years of the analysis
- Format your response as a well-structured report.

This operation analyzes how each theme has evolved over time. It considers viewpoints from multiple debates, identifies trends, and generates a comprehensive summary of the theme's evolution.

The Need for Theme Resolution

An important consideration in this pipeline is the potential for similar themes to be generated with slightly different wording (e.g., "Climate Change Policy" vs. "Environmental Regulations"). To address this, we need to add a resolve operation before the summarization step.

To synthesize a resolve operation, we can use the `docetl build` command:

```
docetl build pipeline.yaml
```

This command adds a resolve operation to our pipeline, resulting in an optimized version:

```
operations:
  ...
  - name: synthesized_resolve_0
    type: resolve
    blocking_keys:
      - theme
    blocking_threshold: 0.6465
    comparison_model: gpt-4o-mini
    comparison_prompt: |
      Compare the following two debate themes:

      [Entity 1]:
      {{ input1.theme }}

      [Entity 2]:
      {{ input2.theme }}

      Are these themes likely referring to the same concept? Consider the
      following attributes:
      - The core subject matter being discussed
      - The context in which the theme is presented
      - The viewpoints of the candidates associated with each theme

      Respond with "True" if they are likely the same theme, or "False" if
      they are likely different themes.
```

```

embedding_model: text-embedding-3-small
compare_batch_size: 1000
output:
  schema:
    theme: string
resolution_model: gpt-4o-mini
resolution_prompt: |
  Analyze the following duplicate themes:

  {% for key in inputs %}
  Entry {{ loop.index }}:
  {{ key.theme }}

  {% endfor %}

  Create a single, consolidated key that combines the information from
  all duplicate entries. When merging, follow these guidelines:
  1. Prioritize the most comprehensive and detailed viewpoint available
  among the duplicates. If multiple entries discuss the same theme with varying
  details, select the entry that includes the most information.
  2. Ensure clarity and coherence in the merged key; if key terms or
  phrases are duplicated, synthesize them into a single statement or a cohesive
  description that accurately represents the theme.

  Ensure that the merged key conforms to the following schema:
  {
    "theme": "string"
  }

  Return the consolidated key as a single JSON object.

pipeline:
  steps:
    - name: debate_analysis
      input: debates
      operations:
        - extract_themes_and_viewpoints
        - unnest_themes
        - synthesized_resolve_0
        - summarize_theme_evolution

```

The new `synthesized_resolve_0` operation groups similar themes together, ensuring a more accurate and comprehensive analysis of each theme's evolution.

Running the Optimized Pipeline

With the resolve operation in place, we can now run our optimized pipeline:

```
docetl run pipeline_opt.yaml
```

This command processes the debate transcripts, extracts themes, resolves similar themes, and generates summaries of theme evolution over time. The results will be

saved in `theme_evolution_analysis.json`, providing insights into the changing landscape of topics discussed in presidential debates. Since we've also set an `intermediate_dir` in our pipeline configuration, intermediate results will be saved in the `intermediate_dir` directory.

Here's the output from running our optimized pipeline:

```
$ docetl run pipeline_opt.yaml
[09:28:17] Performing syntax check on all operations...
        Syntax check passed for all operations.
        Running Operation:
            Type: map
            Name: extract_themes_and_viewpoints
.: Running step debate_analysis...
[09:28:36] Intermediate saved for operation 'extract_themes_and_viewpoints'
        Running Operation:
            Type: unnest
            Name: unnest_themes
        Intermediate saved for operation 'unnest_themes'
        Running Operation:
            Type: resolve
            Name: synthesized_resolve_0
[09:28:38] Comparisons saved by blocking: 56002 (97.75%)
": Running step debate_analysis...
[09:29:02] Number of keys before resolution: 339
        Number of distinct keys after resolution: 152
": Running step debate_analysis...
[09:29:04] Self-join selectivity: 0.0390
        Intermediate saved for operation 'synthesized_resolve_0'
        Running Operation:
            Type: reduce
            Name: summarize_theme_evolution
.: Running step debate_analysis...
[09:29:54] Intermediate saved for operation 'summarize_theme_evolution'
        Flushing cache to disk...
        Cache flushed to disk.
Step debate_analysis completed. Cost: $0.29
Operation extract_themes_and_viewpoints completed. Cost: $0.16
Operation unnest_themes completed. Cost: $0.00
Operation synthesized_resolve_0 completed. Cost: $0.04
Operation summarize_theme_evolution completed. Cost: $0.09
📁 Output saved to theme_evolution_analysis_baseline.json
Total cost: $0.29
Total time: 97.25 seconds
```

This output shows the progress of our pipeline execution, including the different operations performed, intermediate saves, and the final results. Note the total cost was only \$0.29!

Initial Results

Our pipeline generated reports on various themes discussed in the presidential debates. We've put the results up [here](#). However, upon inspection, we found that these reports were lacking in depth and recency. Let's look at a few examples:



Example Reports Lacking in Recent Quotes

Infrastructure Development

Infrastructure Development: A Comparative Analysis of Democratic and Republican Viewpoints from 1992 to 2023

Introduction

Infrastructure development has long been a pivotal theme in American political discourse, with varying perspectives presented by major party candidates. This report analyzes shifts and trends in Democratic and Republican viewpoints from 1992, during the second presidential debate between George Bush, Bill Clinton, and Ross Perot, to 2023.

Republican Viewpoints

Early 1990s

In 1992, President George Bush emphasized a forward-looking approach to infrastructure, stating, "We passed this year the most furthest looking transportation bill in the history of this country...\$150 billion for improving the infrastructure." This statement indicated a commitment to substantial federal investment in infrastructure aimed at enhancing transportation networks.

2000s

Moving into the early 2000s, the Republican party maintained a focus on infrastructure but began to frame it within the context of economic growth and public-private partnerships. However, after the 2008 financial crisis, there was a noticeable shift. The party emphasized tax cuts and reducing regulation over large public investments in infrastructure.

Recent Years

By 2020 and 2021, under the Trump administration, the emphasis returned to infrastructure. However, the tone shifted towards emphasizing private sector involvement and deregulation rather than large public spending. The Republican approach became more fragmented, with some factions calling for aggressive infrastructure investment, while others remained cautious about expenditures.

Democratic Viewpoints

Early 1990s

In contrast, Governor Bill Clinton in 1992 proposed a more systematic investment strategy, noting, "My plan would dedicate \$20 billion a year in each of the next 4 years for investments in new transportation." This highlighted a stronger emphasis on direct federal involvement in infrastructure as a means of fostering economic opportunity and job creation.

2000s

Through the late 1990s and early 2000s, the Democratic party continued to push for comprehensive federal infrastructure plans, often attached to broader economic initiatives aimed at reducing inequality and spurring job growth. The party emphasized sustainable infrastructure and investments that address climate change.

Recent Years

By 2020, under the Biden administration, the Democrat viewpoint strongly advocated for significant infrastructure investments, combining traditional infrastructure with climate resilience. The American Jobs Plan symbolized this shift, proposing vast funds for transit systems, renewable energy projects, and rural broadband internet. The framing increasingly included social equity as a

core component of infrastructure, influenced by movements advocating for racial and economic justice.

Agreements and Disagreements

Agreements

Despite inherent differences, both parties have historically acknowledged the necessity of infrastructure investments for economic growth. Both Bush and Clinton in 1992 recognized infrastructure as vital for job creation, but diverged on the scope and funding mechanisms.

Disagreements

Over the years, major disagreements have surfaced, particularly in funding approaches. The Republican party has increasingly favored private sector involvement and tax incentives, while Democrats have consistently pushed for robust federal spending and the incorporation of progressive values into infrastructure projects.

Influencing Factors

The evolution of viewpoints has often mirrored external events such as economic recessions, technological advancement, and social movements. The post-9/11 era and the 2008 financial crisis notably shifted priorities, with bipartisan discussions centered around recovery through infrastructure spending. Additionally, increasing awareness of climate change and social justice has over the years significantly influenced Democratic priorities, leading to a more inclusive and sustainable approach to infrastructure development.

Conclusion

The comparative analysis of Democratic and Republican viewpoints on infrastructure development from 1992 to 2023 reveals significant shifts in priorities and strategies. While both parties agree on the need for infrastructure improvements, their approaches and underlying philosophies continue to diverge, influenced by economic, social, and environmental factors.

Crime and Gun Control

The Evolution of Democratic and Republican Viewpoints on Crime and Gun Control: 1992-2023

Introduction

This report analyzes the shifting perspectives of the Democratic and Republican parties on the theme of "Crime and Gun Control" from 1992 to 2023. The exploration encompasses key debates, significant shifts in stance, party alignments, and influences from external events that shaped these viewpoints.

Democratic Party Viewpoints

1. **Initial Stance (1992)**: In the early 1990s, the Democratic viewpoint, as exemplified by Governor Bill Clinton during the Second Presidential Debate in 1992, supported individual gun ownership but emphasized the necessity of regulation: "I support the right to keep and bear arms...but I believe we have to have some way of checking handguns before they're sold."

- **Trend**: This reflects a moderate position seeking to balance gun rights with public safety—a common theme in Democratic rhetoric during this era that resonated with many constituents.

2. **Shift Towards Stricter Gun Control (Late 1990s - 2000s)**: Following events such as the Columbine High School shooting in 1999, the Democratic Party increasingly advocated for more stringent gun control measures. The passing of the Brady Bill and the Assault Weapons Ban in 1994 marked a peak in regulatory measures supported by Democrats, emphasizing public safety over gun ownership rights.

- **Quote Impact**: During this time, Democratic leaders often highlighted the need for legislative action to combat rising gun violence.

3. **Response to Mass Shootings (2010s)**: The tragic events of the Sandy Hook Elementary School shooting in 2012 ignited a renewed push for gun control from leading Democrats, including then-President Barack Obama. His call for "common-sense gun laws" marked a decisive moment in Democrat advocacy, focusing on background checks and bans on assault weapons.

- **Quote**: Obama stated, "We can't tolerate this anymore. These tragedies must end. And to end them, we must change."

4. **Current Stance (2020s)**: The Democratic viewpoint has continued to become increasingly aligned with comprehensive gun control measures, including calls for universal background checks and red flag laws. In the wake of ongoing gun violence, this approach highlights a commitment to addressing systemic issues related to crime and public safety.

Republican Party Viewpoints

1. **Consistent Support for Gun Rights (1992)**: In the same 1992 debate, President George Bush emphasized the rights of gun owners, stating, "I'm a sportsman and I don't think you ought to eliminate all kinds of weapons." This illustrates a steadfast commitment to Second Amendment rights that has characterized Republican positions over the years.

- **Trend**: The Republican Party has traditionally promoted a pro-gun agenda, often resisting calls for stricter regulations or bans.

2. **Response to Gun Control Advocacy (2000s)**: As Democrats pushed for stricter gun laws, Republicans increasingly framed these measures as infringements on individual rights. The response to high-profile shootings tended to focus on mental health and crime prevention rather than gun regulation.

- **Disagreement**: Republicans consistently argued against the effectiveness of gun control, indicating belief in personal responsibility and the right to self-defense.

3. **Shift to Increased Resistance (2010s)**: In the wake of prominent mass shootings, the Republican party maintained its focus on supporting gun rights, opposing federal gun control initiatives. Notable figures, such as former NRA spokesperson Dana Loesch, articulated this resistance by stating, "We are not going to let tragedies be used to violate our rights."

- **Impact of External Events**: The rise of organizations like the NRA and increased gun ownership among constituents have fortified this pro-gun stance.

4. **Contemporary Stance (2020s)**: Currently, the Republican viewpoint remains largely unchanged with an emphasis on individual rights to bear arms and skepticism regarding the effectiveness of gun control laws. Recent discussions around gun violence often focus on addressing crime through law enforcement and community safety programs instead of legislative gun restrictions.

Key Agreements and Disagreements

- **Common Ground**: Both parties, at different times, have recognized the necessity for addressing gun-related violence but diverge on methods—Democrats typically advocate for regulations while Republicans focus on rights preservation.

- **Disagreements**: A significant divide exists on whether stricter gun laws equate to reduced crime rates, with Republicans consistently refuting this correlation, arguing instead that law-abiding citizens need access to firearms for self-defense.

Conclusion

The evolution of viewpoints on crime and gun control from 1992 to 2023

highlights a pronounced divergence between the Democratic and Republican parties. While Democrats have increasingly pursued stricter regulatory measures focused on public safety, Republicans have maintained a consistent advocacy for gun rights, underscoring a broader ideological conflict over individual freedoms and collective responsibility for public safety. The trajectories of both parties reflect their core values and responses to notable events impacting society.

Drug Policy

Evolution of Drug Policy Viewpoints: 1996 to 2023

Summary of Democratic and Republican Perspectives on Drug Policy

Over the years, drug policy has been a contentious issue in American politics, reflecting profound changes within both the Democratic and Republican parties. This report examines how views have evolved, highlighting significant trends, areas of agreement and disagreement, and influential external factors utilizing debates as primary reference points.

Democratic Party Trends:

1. **Increased Emphasis on Comprehensive Approaches:**

- In the 1996 Clinton-Dole debate, President Bill Clinton stated, "We have dramatically increased control and enforcement at the border." This reflects a focus on enforcement as part of drug policy. However, Clinton's later years signaled a shift towards recognizing the need for treatment and prevention.

2. **Emergence of Harm Reduction and Decriminalization:**

- Moving into the 2000s and beyond, Democrats began to embrace harm reduction strategies. For instance, during the 2020 Democratic primary debates, candidates such as Bernie Sanders and Elizabeth Warren emphasized decriminalization and treatment over incarceration, signifying a notable shift from punitive measures.

3. **Growing Advocacy for Social Justice:**

- Recent years have seen an alignment with social justice movements, arguing that drug policy disproportionately affects marginalized communities. Kamala Harris in 2020 stated, "We need to truly decriminalize marijuana and address the impact of the War on Drugs on communities of color."

Republican Party Trends:

1. **Strong Focus on Law and Order:**

- During the 1996 debate, Senator Bob Dole reflected a traditional Republican stance, highlighting concerns over drugs without markedly addressing the social implications. "The President doesn't want to politicize drugs, but it's already politicized Medicare..." displays a defensive posture toward the political ramifications of drug issues.

2. **Shift Towards Treatment and Prevention:**

- By the mid-2010s, there was a growing recognition of the opioid crisis, leading to a bipartisan approach promoting treatment. For example, former President Donald Trump, in addressing the opioid epidemic, stated, "We have to take care of our people. We can't just lock them up."

3. **Conflict Between Hardline Stance and Pragmatism:**

- Despite some shifts, many Republicans still emphasize law enforcement solutions. This tension was evident in the polarizing responses to marijuana legalization across states, with figures like former Attorney General Jeff Sessions taking a hardline stance against marijuana legalization, contrasting with more progressive approaches adopted by some Republican governors.

```

### Areas of Agreement:
1. Opioid Crisis:
    - Both parties acknowledged the severity of the opioid epidemic, leading to legislation aimed at addressing addiction and treatment, indicating a rare consensus on the need for health-focused solutions.

### Areas of Disagreement:
1. Approach to Drug Policy:
    - The Democratic party's shift towards decriminalization and harm reduction contrasts sharply with segments of the Republican party that still advocate for strict enforcement and criminalization of certain drugs.

### Influential External Events and Factors:
1. The Opioid Crisis:
    - The rise of the opioid epidemic has forced both parties to reevaluate their positions on drug policy, pushing them towards more compassionate approaches focusing on addiction treatment.

2. Social Justice Movements:
    - The Black Lives Matter movement and other social justice efforts have altered the discourse surrounding drug policies, with increased focus on the need to rectify injustices in enforcement practices, particularly among minorities.

### Conclusion:
Through the years, drug policy viewpoints within the Democratic and Republican parties have experienced significant evolution, characterized by complex layers of agreement and disagreement. As social dynamics shift, both parties continue to grapple with finding a balanced approach towards a more effective drug policy that prioritizes health and social justice.

```

Upon inspecting the intermediates, it appears that the map operation is doing a good job at extracting relevant information. The issue seems to lie in the reduce operation, which is ignoring some of the analysis.

It's possible that trying to summarize all the insights across all debates for a topic in a single LLM call is too ambitious. To address this, we can set `optimize: true` for the reduce operation.

Let's update our `summarize_theme_evolution` operation in the pipeline:

```

- name: summarize_theme_evolution
  type: reduce
  reduce_key: theme
  optimize: true
  output:
    schema:
      theme: str
      report: str
  prompt: |
    [... existing prompt ...]

```

Rerunning the build command `docetl build pipeline.yaml` will synthesize the optimized reduce operation (make sure `pipeline.yaml` is the pipeline you want to optimize).

Running the Pipeline (with Optimized Reduce)

In our optimized pipeline, we see that DocETL added a `gleaning` configuration to the reduce operation:

```
- name: summarize_theme_evolution
  type: reduce
  reduce_key: theme
  ...
  gleaning:
    num_rounds: 1
    validation_prompt: |
      1. Does the output adequately summarize the evolution of viewpoints on
the theme based on the
      provided debate texts? Are all critical shifts and trends mentioned?
      2. Are there any crucial quotes or data points missing from the output
that
      were present in the debate transcripts that could reinforce the
analysis?
      3. Is the output well-structured and easy to follow, following any
formatting guidelines specified in the prompt, such as using headings
for
      sections or maintaining a coherent narrative flow?
```

Tip

Note that you should always feel free to edit the `validation_prompt` to better suit your specific needs! The optimizer uses LLMs to write all prompts, but you have the best context on your task and what you're looking for in the output, so you should adjust anything accordingly.

And when running the pipeline, we can observe the impact of this optimization; for example, one of the outputs gets amended to include more recent quotes:

```
Validator improvements (gleaning round 1):
1. Coverage of Critical Shifts and Trends: While the output captures
several significant trends and shifts in Democratic and Republican viewpoints,
it could benefit from a more thorough overview, especially about the changing
perceptions and proposals related to economic downturns and recoveries across
the decades. For instance, it could include how the response to the 2008
financial crisis tied back to historical precedents, linking back to earlier
debates about the importance of jobs and middle-class stability (like in the
1976 or 1992 debates).

2. Missing Quotes and Data Points: The response could further bolster the
analysis by incorporating additional quotes that illustrate the evolving
narrative, particularly surrounding tax fairness and job creation. For
example, quotes like George Bush in 1984 calling out "working man"
perspectives against seemingly progressive taxation could enhance the depth.
Additionally, quotes from debates emphasizing the impact of tax cuts on
```

economic recovery and job growth—such as Obama's focus on the automotive industry's recovery in 2012 or McCain's 'putting homeowners first'—could provide essential context to the arguments for both parties.

3. **Structure and Flow**: Overall, the output is fairly well-structured and maintains a logical flow, using headings appropriately to signal key sections. However, it may benefit from clearer subsections under each party's overview to delineate specific key points, such as 'Tax Policy', 'Job Creation', and 'Response to Economic Crises'. This would enhance readability and assist the reader in navigating the shifts in viewpoints. For example, adding bullet points or more vivid transitions between historical periods could clarify the evolution timeline. Moreover, resolving any redundancy (such as multiple mentions of similar themes across years) would streamline the narrative.

Check out the new output [here](#) to see the improvements made by the optimized pipeline! Of course, we can probably optimize the initial map operation too, do prompt engineering, and more to further enhance the pipeline.



Interactive Pipeline Creation

We're currently building interfaces to interactively create and iterate on these pipelines after seeing outputs. This will allow for more intuitive pipeline development and refinement based on real-time results. If you're interested in this feature or would like to provide input, please reach out to us! Your feedback can help shape the future of DocETL's user experience.