

The Impact Project was designed to show the effects of government funding cuts via an interactive map. Historically, the platform relied on manual coding to track metrics like funding amounts and affected populations, an unscalable process when facing a backlog of over 60,000 potential news sources. To resolve this bottleneck, an end-to-end automated pipeline was engineered to streamline data acquisition and extraction.

A custom web scraping model successfully processed an initial batch of 16,000 articles, followed by a cleaning script that filtered out approximately 10,000 irrelevant pages. Utilizing this refined dataset, a hybrid extraction model combining pattern recognition and Large Language Models (LLMs) was constructed to automatically identify critical data points including monetary figures, geographic locations, and population impacts. Finally, a prototyped chatbot interface was implemented to allow users to query the database directly. This comprehensive system eliminated the manual data entry workload, establishing a scalable foundation for real-time policy tracking. Below is the HTML wireframe for this chatbot interface. It will augment interactions with data visualizations on its left.

[About](#) [Data](#) [Help](#)

Impact ChatBot

Ask me about federal worker impact data

Hello! I'm here to help you explore federal worker impact data. You can ask me about specific counties, states, or search for trends and statistics.

Tell me about Washington County Colorado

Washington County, Colorado

Key Statistics:

- Probationary Federal Workers: 4,023
- Total Federal Workers: 39,910
- Washington County Workers: 32 (2.72%)

Show neighboring counties

Compare to Colorado average

Historical data

Ask about impact data...

Send