Algorithmic Analysis of Text Propaganda by Artificial Intelligence Systems

A Master’s Degree Project Proposal

Chinmay Pai

Interactive Media and Game Development

Worcester Polytechnic Institute

Worcester, MA

Committee

**Advisor:** Brian Moriarty

**Reader:** Angela Rodriguez

**Reader:** Jennifer deWinter

2019

# Abstract

“Spin” is a form of propaganda that uses biased language to influence public opinion. This thesis will conduct research intended to inform the design of a prospective game which will allow players to explore the dynamics of spin. The game will employ a Web application which scrapes daily headlines from popular US new sources and submits them to AI systems such as Watson, Google Cloud and Amazon Comprehend for analysis of their emotional bias. The same headlines will be submitted to a demographically diverse population of American readers for similar evaluation. The results of these evaluations will be compared to identify the strengths and weaknesses of the AI analysis, and help determine which forms of bias should be used to improve the clarity and persuasiveness of the game.

# Introduction

“The conscious and intelligent manipulation of the organized habits and opinions of the masses is an important element in democratic society. Those who manipulate this unseen mechanism of society constitute an invisible government which is the true ruling power of our country.”

Edward Bernays, *Propaganda*

The role of propaganda in shaping democracies and our perceptions of the world and the constituents within it has been the focus of significant discussion and debate among both laypeople and academics regarding the reasons, methods and consequences of such propaganda in several cases, including but not limited to the rise of autocratic leaders all over the world, and political reactions to phenomena such as climate change, immigration both voluntary and forced, and majoritarianism. Particularly notable has been the method by which changing the speed of media transmission has affected the nature of the language used in influencing the public mind based on these issues.

One method used in the analysis of language is the field of critical discourse analysis. Cameron and Panović define critical discourse analysis as the approach to analyzing discourse which highlights its “social, ideological and political dimensions”. Discourse itself is defined by Leech and Short as “linguistic communication seen as a transaction between speaker and hearer, as an interpersonal activity whose form is determined by its social purpose”, and we can consider propaganda to be a type of discourse where a speaker speaks through any medium of communication to a large audience, crafted in a specific social context, with the purpose of influencing their perceptions or actions with regards to a particular person or issue.

This document intends to propose a thesis which analyses text propaganda using artificial intelligence. This research will be used to inform the design of a prospective game that allows players to select and manipulate real-time headlines from a variety of news outlets in order to achieve specific public opinion outcomes in a fictional context. My hope is that players experiencing the game will be able to see how propaganda operates, and how its effects might be inhibited.

# Prior Research

A large amount of research has been done regarding both the technical analysis and linguistic composition of propaganda. Perhaps the single closest to the proposed research is Holm’s research which uses natural language processing to analyze propaganda created by openly adversarial ideologies such as the Islamic State (IS). This paper uses natural language processing to analyze a corpus of IS propaganda pieces to determine with accuracy whether a piece of text is propaganda.

Most of the work in the field of the analysis of fake news has been either technical – understanding the spread of fake news through new mass media such as Twitter – or sociological – understanding the nature of fake news and how it influences the people it targets to achieve its ideological aims. Sobieraj and Berry have spoken about how the spread of outrage across all forms of media in the United States, such as television, talk radio, newspaper editorial columns, and political blogs, has been a significant measure of the success of these media in the form of viewership and clicks, which seems to correlate with the generally accepted idea that ‘outrage sells’, and have also measured so-called incivility from both sides of the political spectrum, concluding that both the left and the right use similar tactics and speak falsehoods at similar rates. However, Benkler, Faris and Roberts argue that this ignores the significant asymmetry between left-leaning and right-leaning citizens’ trust in their corresponding partisan media outlets, especially in the fields of television and talk radio. Langin has found that fake news spread is not, contrary to popular belief, primarily spread by bots, but by actual humans, whether for malicious or ignorant reasons.

A theoretical understanding of propaganda devices, methods used by propagandists to promote their product of choice, was codified by the Institute of Propaganda Analysis, noting these devices as:

1. Name-calling
2. Glittering generalities
3. Transfer
4. Testimonial
5. Plain folks
6. Card-stacking
7. Bandwagon

While the “card-stacking” device would be difficult to codify in code due to its somewhat ambiguous nature, the others seem likely to be relatively easily convertible to code.

There have also been several ideas regarding the methods by which fake news might be contained. These include the idea of “guardians”, proposed by Vo and Lee, users who can recommend verified facts to users in response to popular misconceptions/fake news about popular figures. Websites such as Politifact and Snopes have taken up this role with mixed success, with allegations of bias from both sides of the political spectrum having reduced their utility as trusted debunkers of false or misleading announcements. Baum also speaks about the possibility of weeding out fake news using algorithmic methods, which utilize bots to either correct or remove fake news, which would require collaboration between sociological academia and computer science experts, as well as careful balancing to ensure that governments do not subvert these systems to tyrannical ends.

# Research Design

The proposed game being researched for this thesis, tentatively titled The Foghorn, puts the player in control of a fictional small-town newspaper which aspires to expand into a cross-media empire, using the power of spin to increase their audience. The mechanics allow the player to work with current real-world news articles, exploiting or suppressing prevailing sentiments in order to propagate their preferred agenda (see Figure 1). The game will offer extra-diegetic observations regarding how the player’s choices may be affecting people, especially the less fortunate.

A screenshot of a cell phone

Description generated with very high confidence

*Figure 1. A preliminary mockup of the user interface for* The Foghorn*. Current news headlines (on the right) are dragged to the front-page newspaper layout on the left, with relative positioning indicating their importance. A timer (not shown) indicates how much time the player has left before the paper is sent to print.*

The game will draws its headlines from current real-world news organizations, scraped off the Web using method suggested by Robin. (Robin) Each headline will be analyzed to determine its conformity to professional standards of journalism, together with its potential to elicit strong emotions from readers. To mitigate accusations of personal bias in the game’s design, this research will attempt to determine a set of objective, psychologically valid and automatable criteria for characterizing the content of the headlines.

# Preliminary research

In preparation for this thesis, I have carried out preliminary research to examine the analytic data returned when submitting real-world headlines to several AI systems (such as Google’s Cloud, IBM’s Watson and Amazon’s Comprehend ) capable of evaluating the semantic and emotional content of news headlines. Sample headlines were drawn from a variety of sources such as CNN, Fox News, the *New York Times*, the *Washington Post*, *Mother Jones* and Breitbart. The raw results of this research are summarized in Appendix A.

In some cases, the results of the analysis were surprising. Even media sources considered highly partisan typically yielded headlines ranked as emotionally “neutral” by the AIs being tested. This suggests that simple text analysis may not be sufficient to distinguish the emotional content of headlines. Other strategies will need to be considered.

An observation of the Web sites of the selected news organizations, samples of which are provided in Appendix X, suggests two other factors that may be useful for analyzing headlines: positioning and frequency. Positioning headlines on the parts of the Web page most likely to be seen by readers affects their perceived importance. Placing a story at the top of the homepage effectively magnifies its subjective importance, while stories placed beneath it may be considered less consequential.

For example, the December 8, 2018 announcement that John F. Kelly was stepping down from his post as White House Chief of Staff appeared at significantly different positions on various Web sites. The *New York Times* and *Washington Post* placed it front and center. The *Wall Street Journal* placed it at the top along with other headlines. Fox News relegated it to a secondary headline, which required scrolling down the page to view, underneath their lead story about James Comey’s testimony to the House Oversight and Judiciary Committee. Breitbart placed it beneath an article about how Bill and Hillary Clinton had to sell tickets to their speaking engagements through Groupon.

# Evaluation

This thesis will attempt to identify an automated strategy for characterizing news headlines in a way that will be perceived as “fair” by a demographically diverse group of Americans.

Since the non-trivial analysis of any text is, to an extent, subjective, I will use human evaluation as the primary means of determining the validity of headline analysis. I do not believe it would be sufficient to limit my research subjects to the WPI community, as there is likely to be significant bias in such a narrow demographic. To obtain greater variety, I intend to distribute the survey instrument nationally using Amazon’s Mechanical Turk infrastructure, with filters only allowing for US-based IP addresses to participate. To ensure that test subjects are appropriately compensated for their effort, they will be paid the minimum wage for the state of Massachusetts (currently $12 per hour).

Evaluation will be carried out by showing a sample of headlines and news articles from a wide variety of news publications of different ideological bent, and asking the test subject to rate the headlines on different parameters, such as:

1. Emotional appeal and subjective urgency (shock value or “outrageousness”).
2. “Clickbait” (the perception that the headline has been written in such a way so as to entice the user to explicitly interact with it, usually by clicking or touching it).
3. Factual accuracy.
4. Bias, or the degree to which it attempts to push a particular agenda.

A comparison of the perception of the sample group and the analysis of the AI algorithms applied to the same headlines will form the central part of my report.

I am aware of the possibility of bots interfering with accurate data collection, but I believe this problem can be minimized by using pictorial response buttons, samples of which are shown in Appendix B.

# Timetable

|  |  |
| --- | --- |
| Event | Deadline |
| Thesis proposal to IMGD steering | 18 September 2019 |
| Submission of first draft to thesis committee | On or before 19 October 2019 |
| Final thesis presentation | On or before 22 November 2019 |
| Final thesis report submission | On or before 13 December 2019 |

# Works Cited

Sarah Sobieraj & Jeffrey M. Berry (2011) “From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News.” *Political Communication,* 28:1, 19-41, DOI: 10.1080/10584609.2010.542360.

Critical discourse analysis. (2014). In Cameron, D., & Panović, I. Working with written discourse (pp. 66-80). 55 City Road, London: SAGE Publications, Ltd doi: 10.4135/9781473921917

Baum, M., Greenhill, K., Metzger, M., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S., et al. (n.d.). “The science of fake news.” *Science,* 359(6380), 1094–1096. doi:10.1126/science.aao2998.

Mills, S. (2004). Discourse. Retrieved from https://ebookcentral-proquest-com.ezproxy.wpi.edu

N. Vo, and K. Lee. “The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News.” *SIGIR,* July 2018.

Rose-Stockwell, Tobias (2017) “This is How Your Fear and Outrage Are Being Sold For Profit.” *Medium.* Web. URL: https://medium.com/@tobiasrose/the-enemy-in-our-feeds-e86511488de

Langin, K. (n.d.). “Fake news spreads faster than true news on Twitter — Thanks to people, not bots.” *Science.* doi:10.1126/science.aat5350.

Benkler, Yochai, Robert Faris, and Hal Roberts. "The Propaganda Feedback Loop." Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics. : Oxford University Press, October 18, 2018. Oxford Scholarship Online

Kumar, Ravish (2018). *The Free Voice: On Democracy, Culture and the Nation.* New Delhi, India: Speaking Tiger.

How to Detect Propaganda. (1938). Bulletin of the American Association of University Professors (1915-1955), 24(1), 49-55. Retrieved from http://www.jstor.org/stable/40219502

Robin, Alexander (2017) “How I Scraped 7000+ Articles from a Newspaper Website Using Node.” *FreeCodeCamp.* Web. URL: <https://medium.freecodecamp.org/how-i-scraped-7000-articles-from-a-newspaper-website-using-node-1309133a5070>

# Resources for future research

Jang, S., & Kim, J. (n.d.). Third person effects of fake news: Fake news regulation and media literacy interventions. Computers in Human Behavior, 80, 295–302. doi:10.1016/j.chb.2017.11.034

Mihailidis, P., Viotty, S., & Payne, J. (n.d.). Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in “Post-Fact” Society. American Behavioral Scientist, 61(4), 441–454. doi:10.1177/0002764217701217

Vargo, C., Guo, L., & Amazeen, M. (n.d.). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media & Society, 20(5), 2028–2049. doi:10.1177/1461444817712086

Brigida, M., & Pratt, W. (n.d.). Fake news. North American Journal of Economics and Finance, 42, 564–573. doi:10.1016/j.najef.2017.08.012

M. Rajdev, and K. Lee. Fake and Spam Messages: Detecting Misinformation during Natural Disasters on Social Media (short paper). WI. December 2015.

Mullen, A. (2010). Twenty years on: the second-order prediction of the Herman-Chomsky Propaganda Model. Media, Culture & Society, 32(4), 673–690. <https://doi.org/10.1177/0163443710367714>

Herman, E. S. & Chomsky, N. (2002). *Manufacturing Consent: The Political Economy of the Mass Media*. New York, NY: Pantheon.

Sinha, P., Shaikh, S. & Sidharth, A. (2019). *India Misinformed: The True Story.* Noida, India: HarperCollins India.

Holm, R. R. (2017). Natural Language Processing of Online Propaganda as a Means of Passively Monitoring an Adversarial Ideology. http://hdl.handle.net/10945/52993

# Appendix A: Semantic Analysis of News Headlines

The following headlines were semantically analyzed by IBM Watson, Amazon Comprehend, and Google Cloud Services. Both the headlines themselves as well as the entire article were analyzed when possible, but at minimum the headlines were submitted.

The headlines selected as samples are as follows:

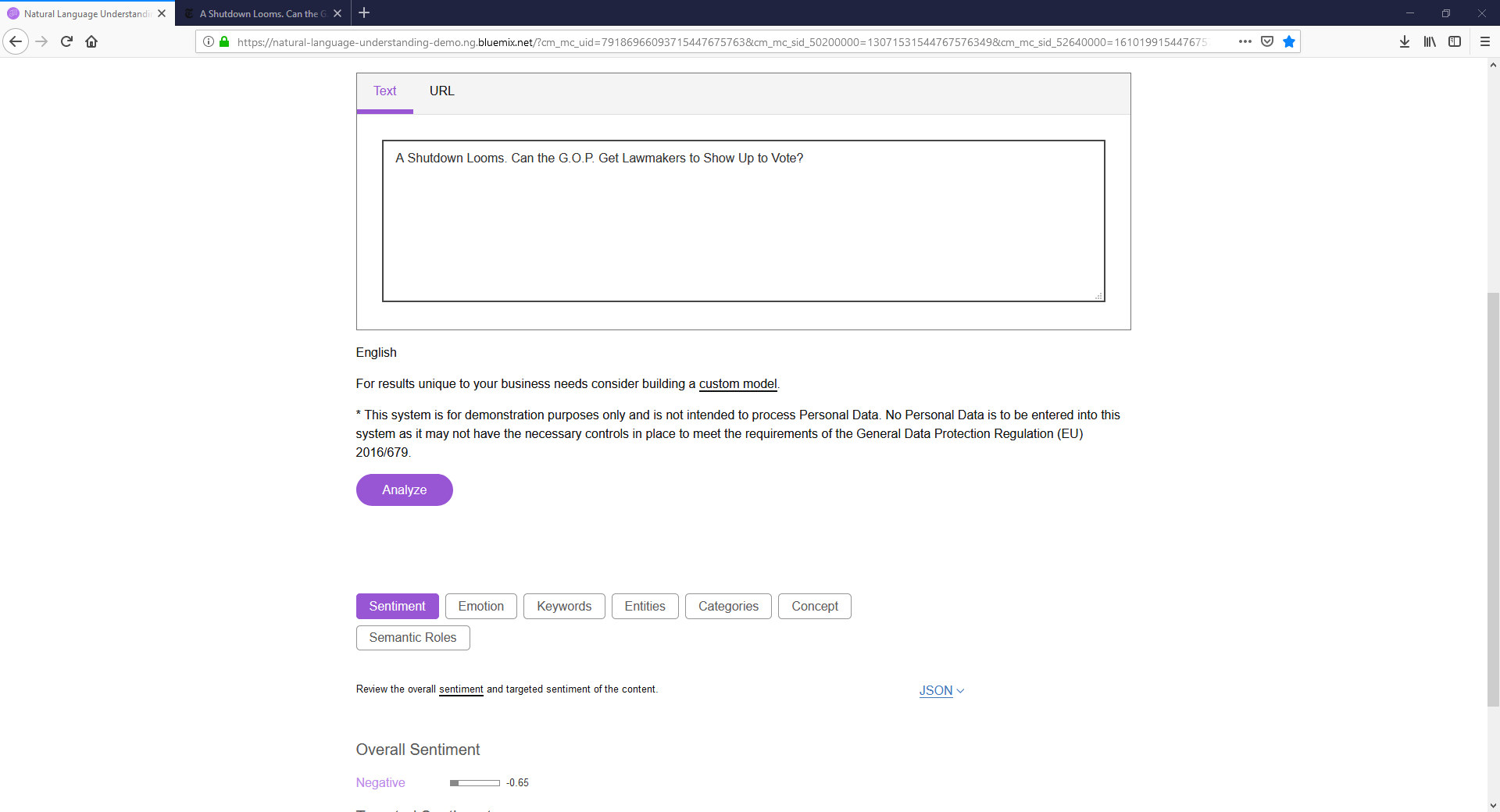
Hirschfeld Davis, Julie and Cochrane, Emily; “*A Shutdown Looms. Can the G.O.P. Get Lawmakers to Show Up to Vote?”*, The New York Times, December 16, 2018: <https://www.nytimes.com/2018/12/16/us/politics/congress-trump-shutdown.html?action=click&module=Top%20Stories&pgtype=Homepage>. This headline is hereafter referred to as **NYT**.

McCarthy, Tyler; Donald Trump Tweets NBC, ‘SNL’ should be tested by courts after Christmas parody sketch, Fox News, December 16, 2018: <https://www.foxnews.com/entertainment/donald-trump-tweets-nbc-snl-should-be-tested-by-courts-after-christmas-parody-sketch>. This article is hereafter referred to as **FOX.**

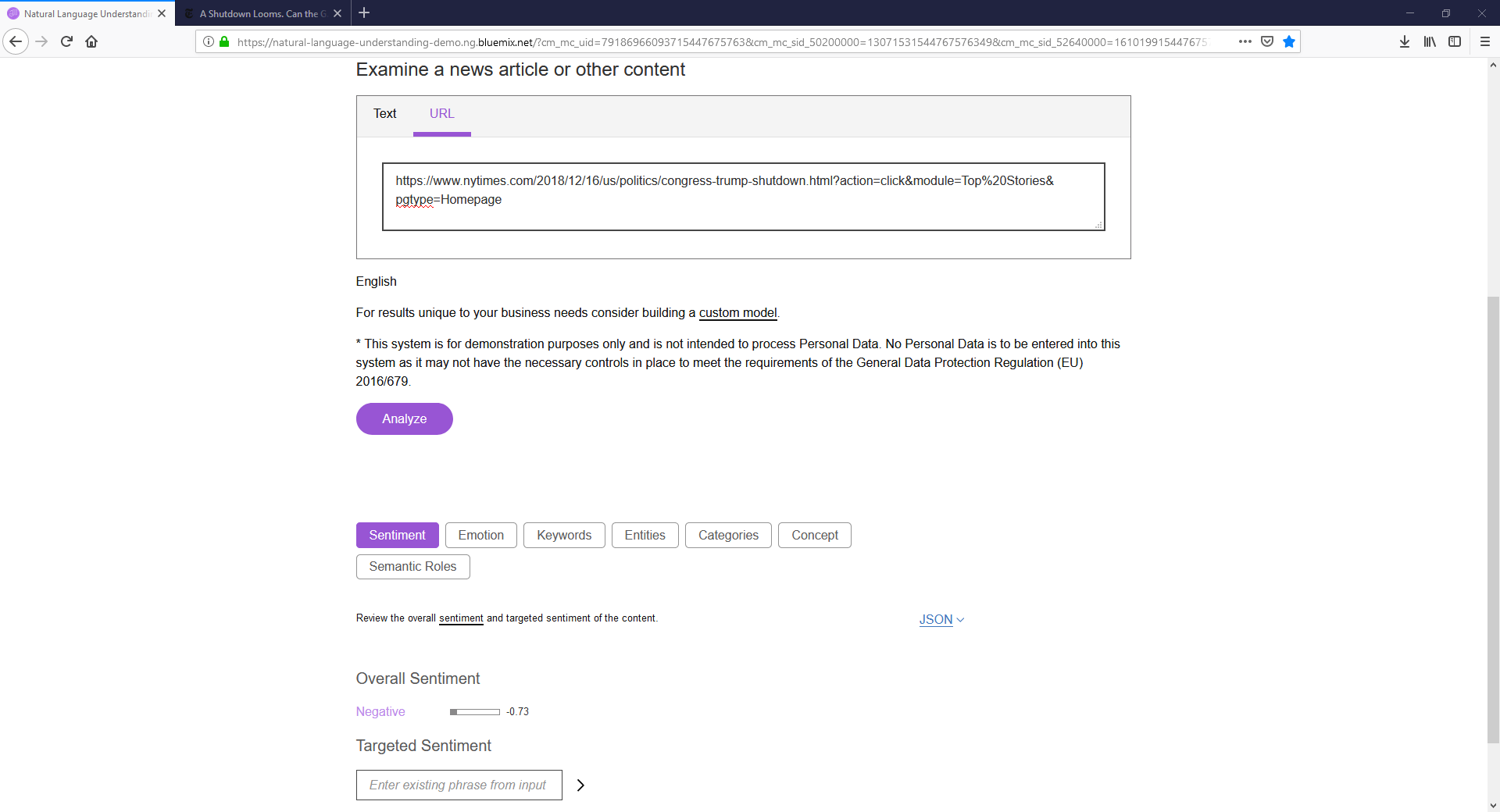
The responses returned by each AI service are shown below:

## IBM Watson

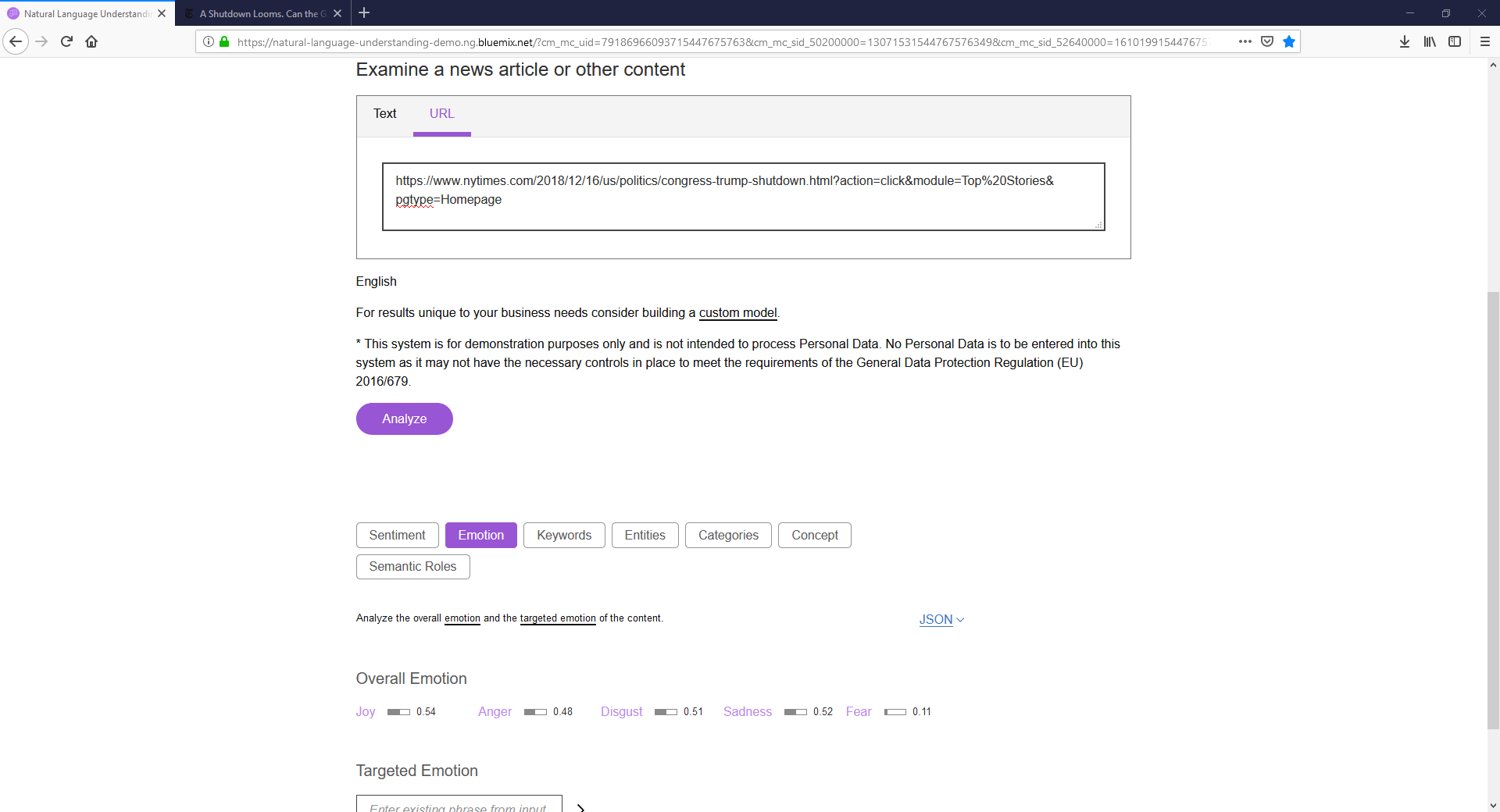
### NYT, Headline Sentiment Analysis



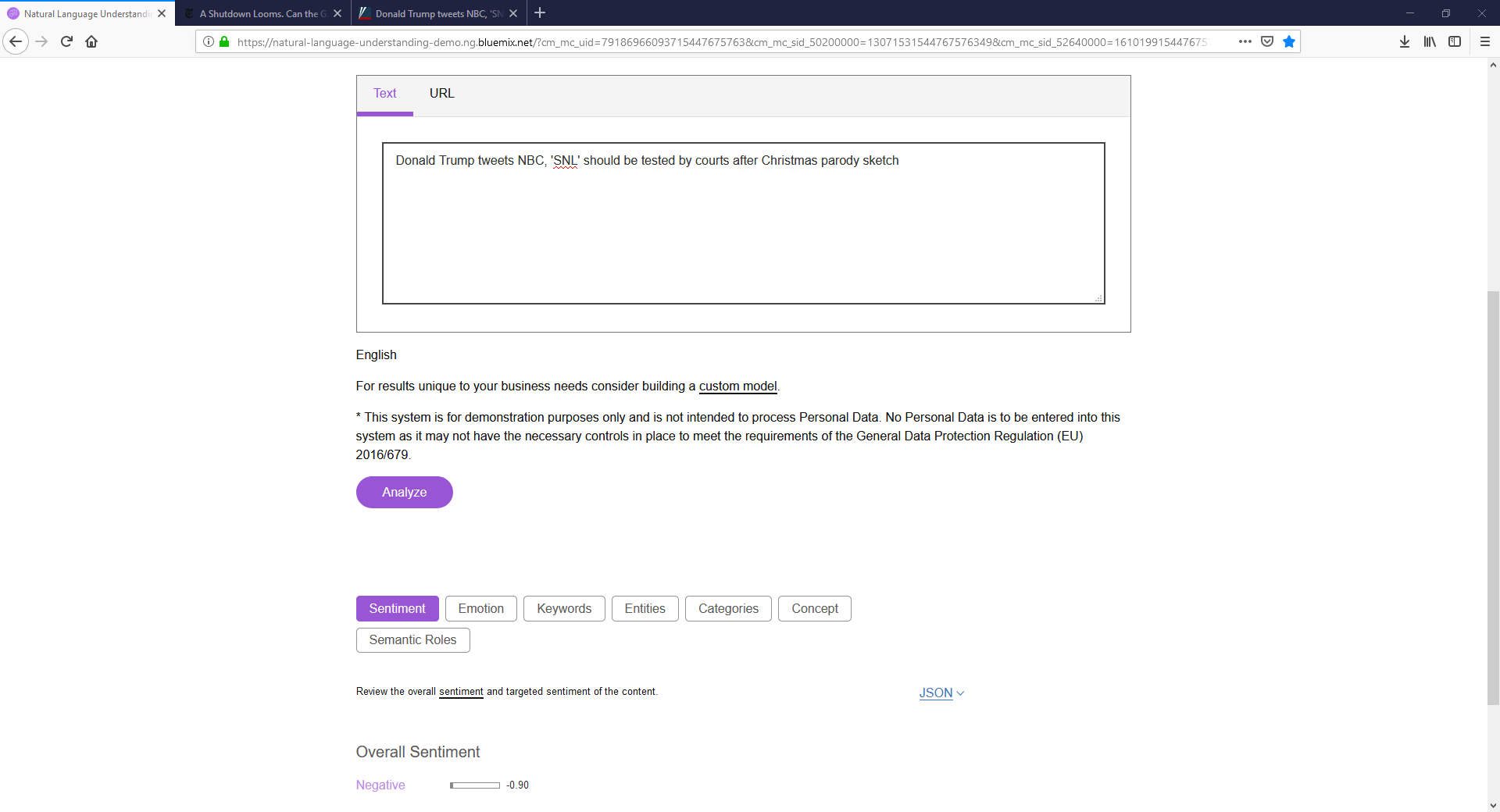
### NYT, Article Sentiment Analysis



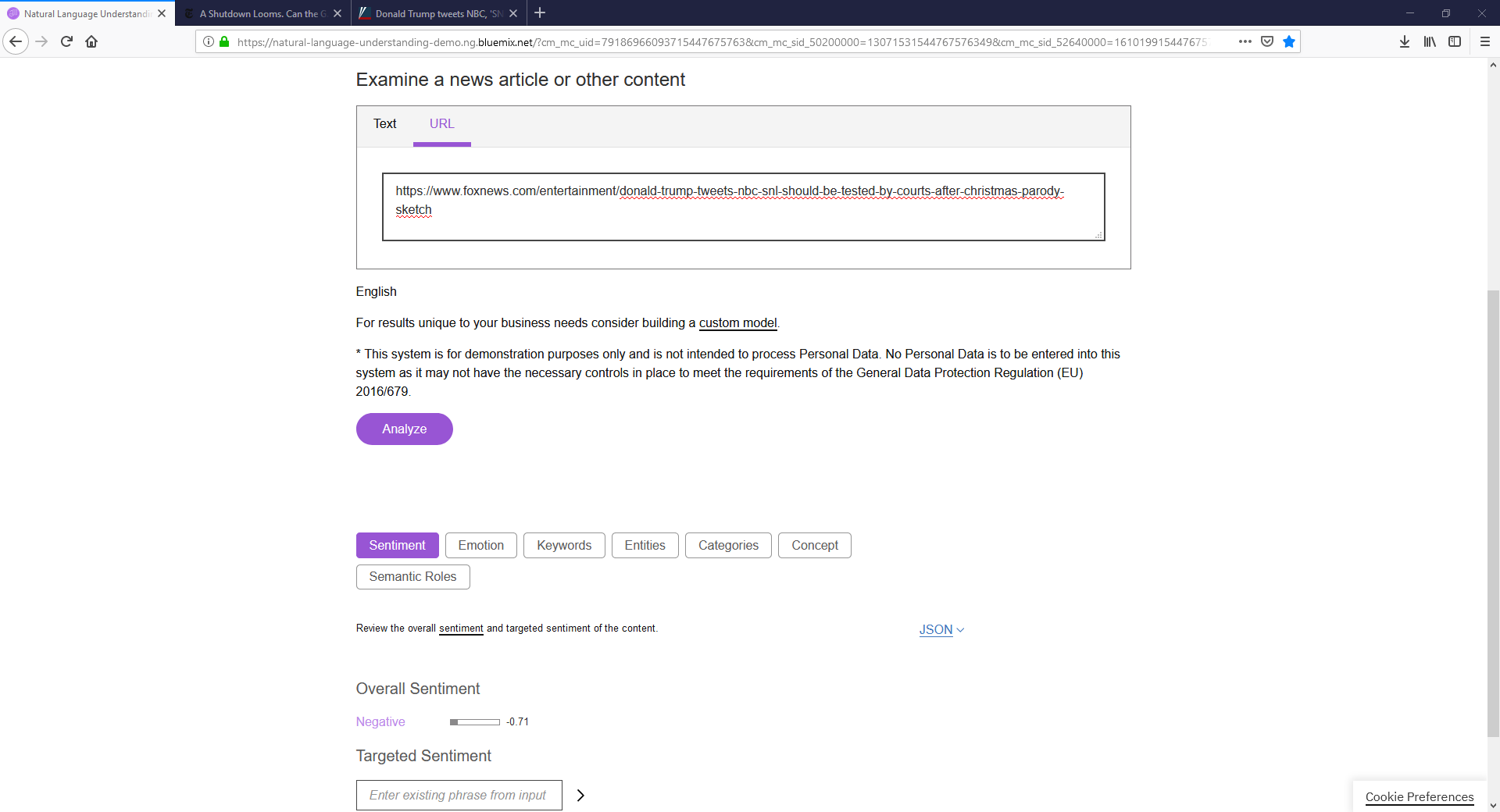
### NYT, Article Emotion Analysis



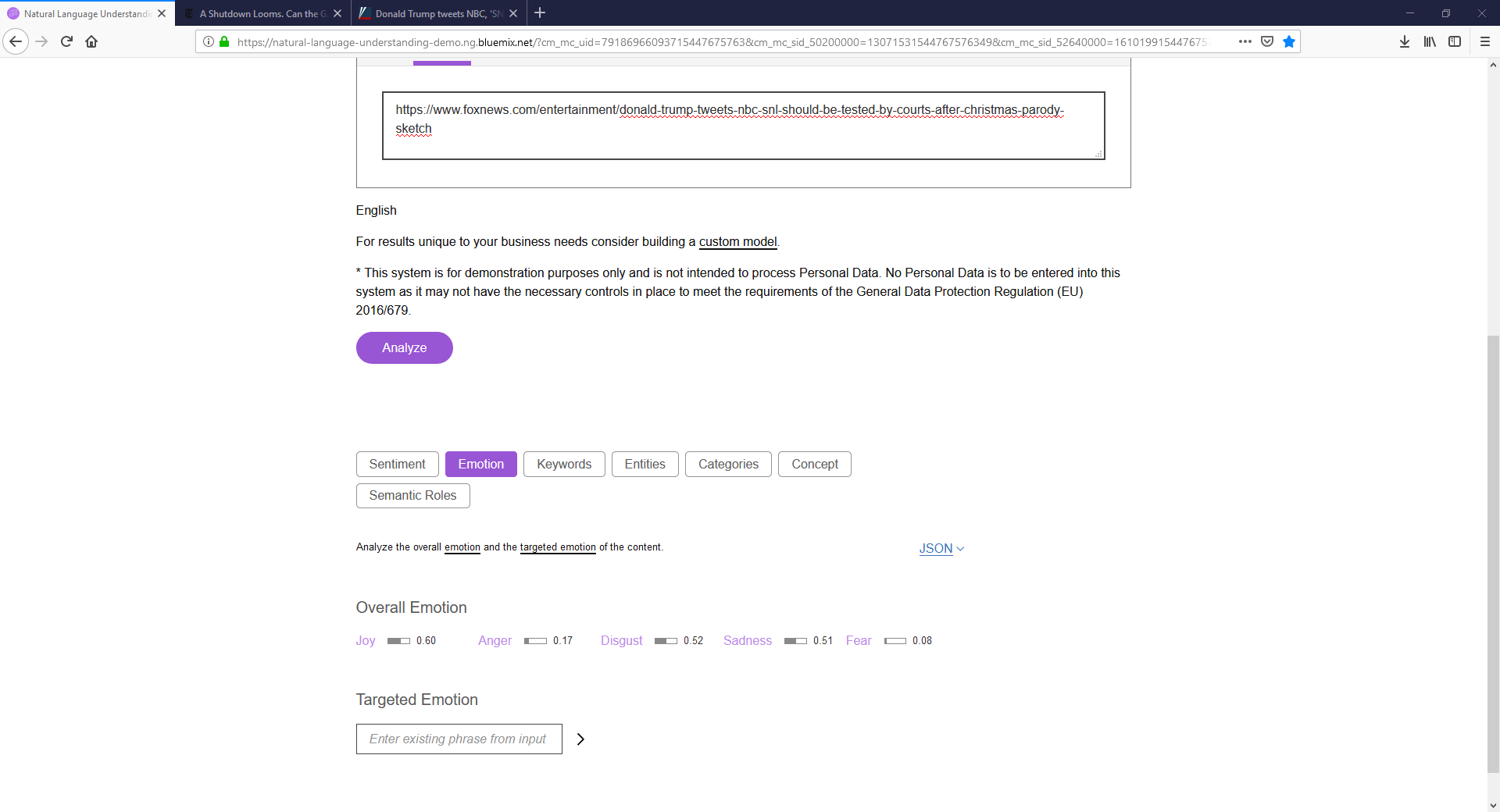
### FOX, Headline Sentiment Analysis



### FOX, Article Sentiment Analysis

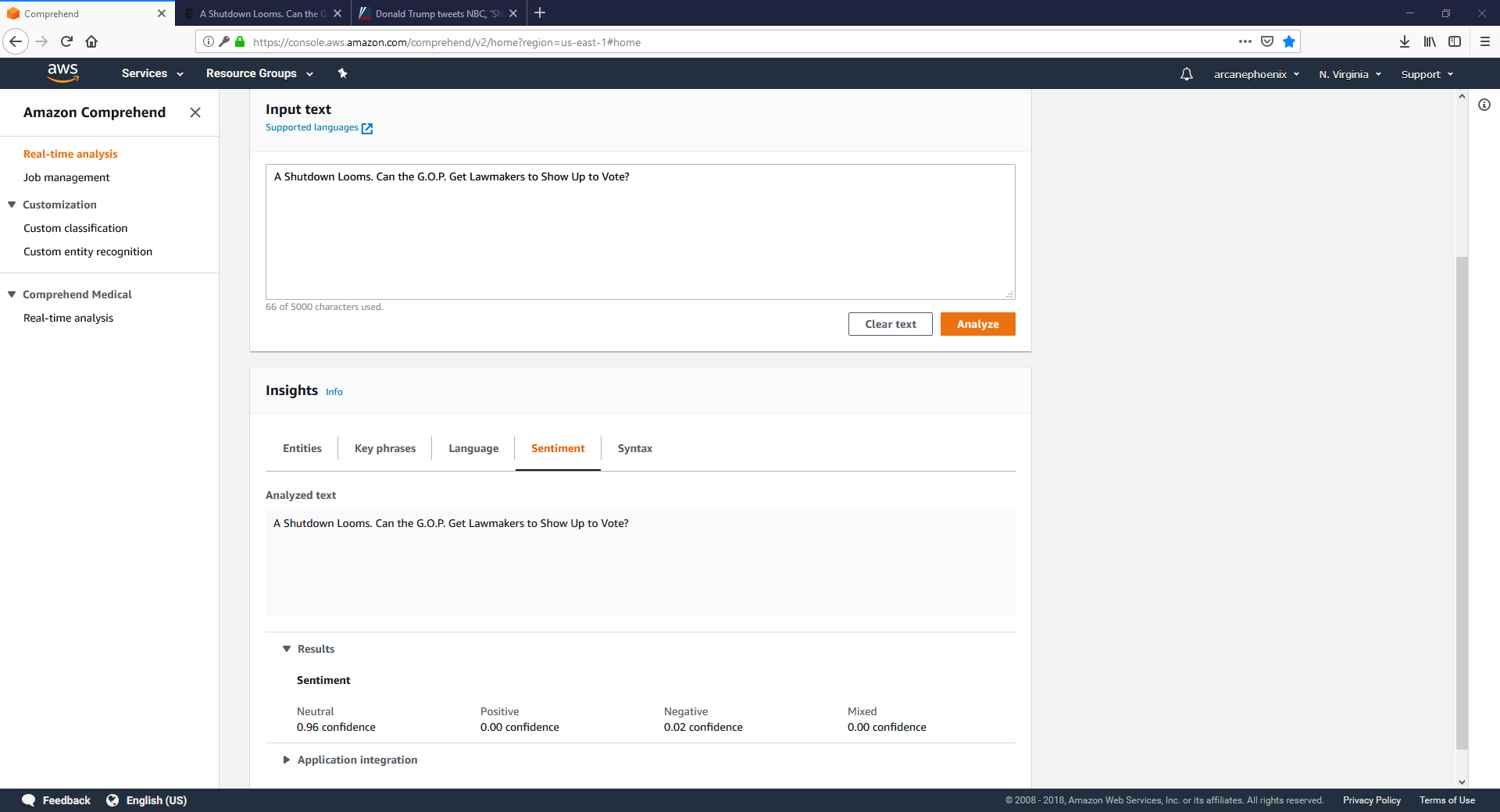


### FOX, Article Emotion Analysis

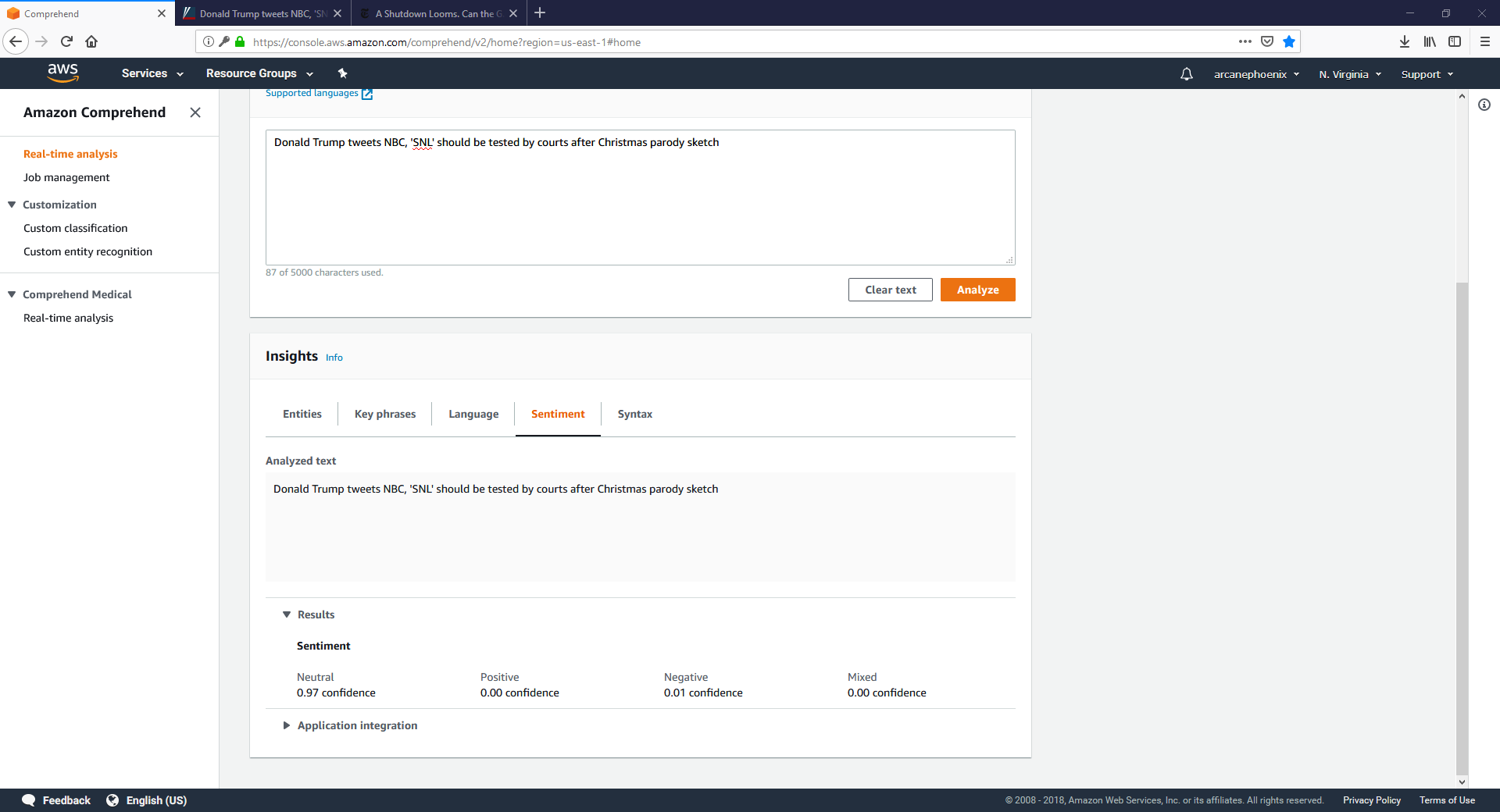


## Amazon Comprehend

### NYT, Headline Sentiment Analysis

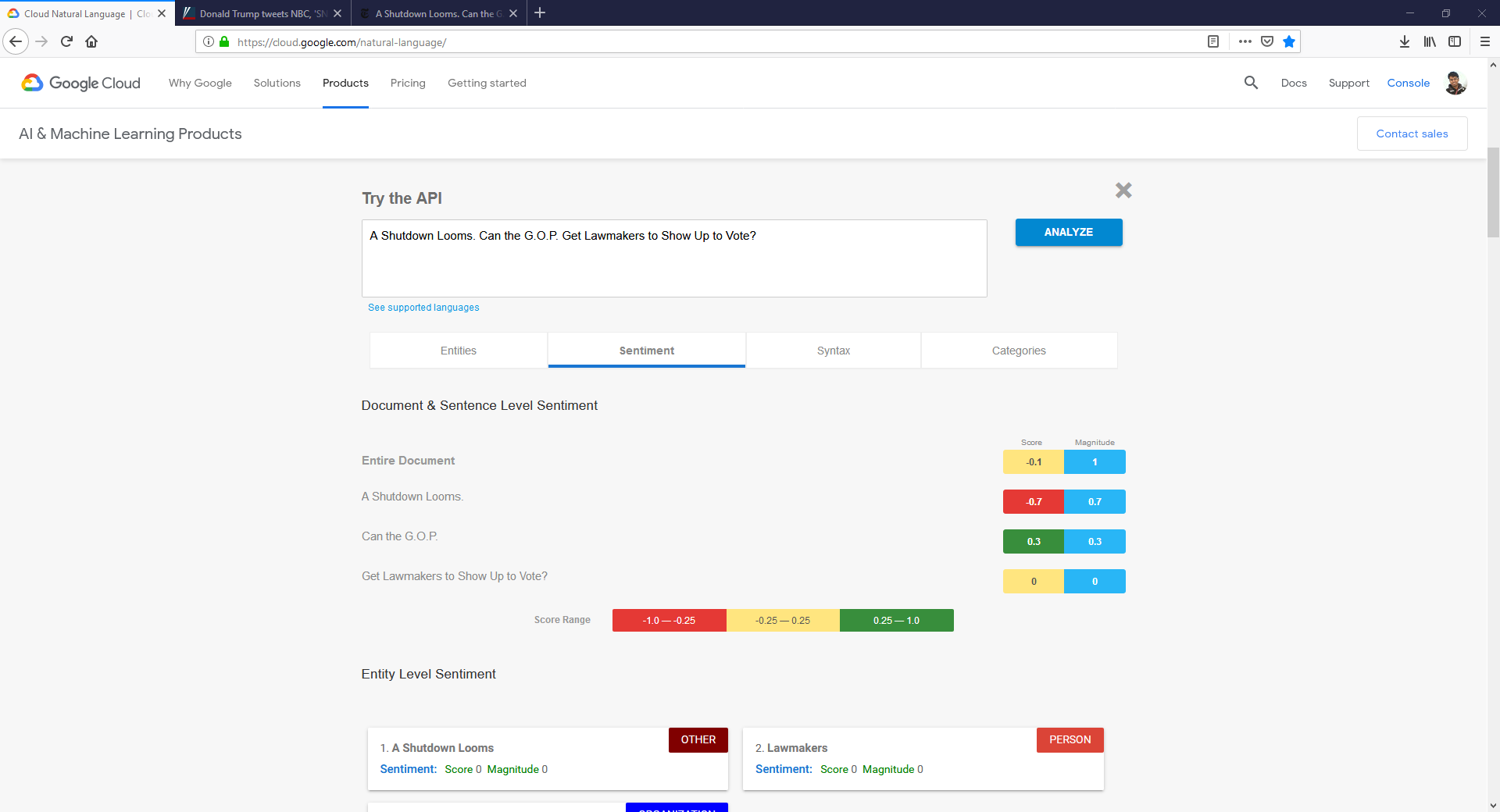


### FOX, Headline Sentiment Analysis

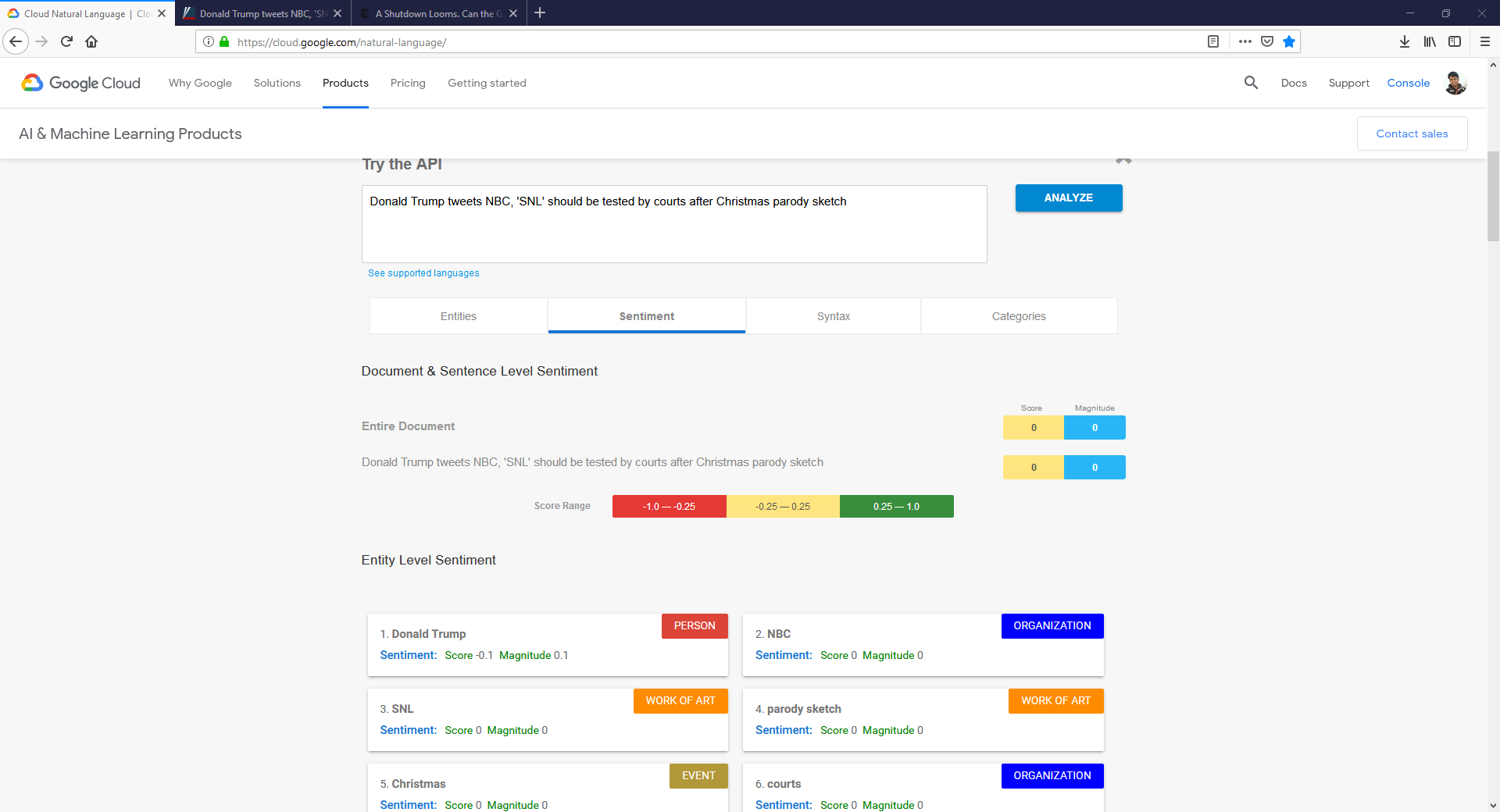


## Google Cloud Services

### NYT, Headline Sentiment Analysis



### FOX, Headline Sentiment Analysis



# Appendix B: Pictorial Representation of Survey Questions

In order to solve the problem of bots on the Mechanical Turk skewing the data received from the survey evaluation, I will be using pictorial representation to represent the emotions which the survey-taker must interpret in order to answer the question accurately. I believe that this approach will be difficult for bots taking the survey to understand and perceive, and they would prefer to go for simpler alternatives. The same question can also have the solution order changed to further confuse any bots.

The following is a representation of the same question attempting to gauge emotional reaction from the survey taker, but in two different iterations.



