Creating an Automated Ideological Transformer Using Moral Reframing

Ananya A. Joshi

Department of

Computer Science

Advisers: Prof. Christiane Fellbaum and Prof. Michael Guerzhoy

May 2019

Abstract

This work proposes a novel ideological transformer tool, called ShiftView, based on the moral reframing concept, by applying natural language processing tools and techniques.

Liberals and conservatives have different morals based on the Moral Foundations Theory [38]. For example, liberals are more likely to base their judgments on the moral value of harm while conservatives may base their judgments on the moral value of purity. Studies have shown that while people can convince others of the same political ideology, they use the same moral rhetoric to convince people with *different ideologies*, and ultimately fail to do so [29]. Mapping moral rhetoric to match a different ideology is a powerful persuasive tool called moral reframing. This thesis seeks to create ShiftView, an automated ideological transformer using moral reframing.

This project has four parts: generating original datasets from the records of the United States Congress and Twitter, designing a flexible moral classifier using topic modeling, creating a text generator using Recursive Neural Networks and transfer learning, and developing an ideological transformation framework. During the process, several moral reframing methods were prototyped and evaluated. Initial results are reviewed by Amazon Turk Workers. This research is the first step in the multi-year project of automating ideological transform and moral reframing.

Acknowledgements

This project would not have been possible without the support of the Princeton community. I thank my advisors, Prof. Christiane Fellbaum and Prof. Michael Guerzhoy, for their sustained guidance, support, and faith in this project. Several faculty across the Computer Science, Politics, and Entrepreneurship departments have shaped this inter-disciplinary project through their discussions and lectures. Finally, I thank my friends and family for their never-ending support and encouraging me to dream ambitiously.

Funding was provided by the Department of Computer Science, the Center for Statistics and Machine Learning, and the School of Engineering.

To my parents.

In search of greater understanding. In loving memory of Chester Lam.

Contents

sentences I examined, none were similar using cosine similarity (maximum 0.15 similarity, with 1 being a perfect match) to existing Tweets or speeches, again showing promise that this is approach creates effective generative models.

5.4 Moral Reframing Evaluation

Employing Amazon Turk Workers to Evaluate Moral Reframed Text To compare and evaluate the generated results for grammar, semantics, and persuasiveness, I designed a web form and selected the best experimental results from the basic models (4), RNN models (9), and OpenAI GPT models (6), on the topics of guns, environment, healthcare, and immigration.

Survey Design Twenty-six Amazon Turk Workers, who identified as conservative, were approved for the survey, named "Attitude Challenge!". The questions were designed to be fun and engaging. Each worker took an average of 12 minutes to complete the survey. Below, I explain each part of the form and the relevant results. My annotations, shown in italics, were not present to the Turkers.

Experiment 1: Political Ideology and Openness For the first question, Turkers briefly described their political ideology, their attitudes towards climate change policy, and if they agreed or disagreed with the following sentence using the 5 point Likert scale (1-disagree, 3-neutral, and 5-agree):

> Increasing energy efficiency and renewable energy will have dramatic effects on the environment. In fact, in this decade alone, as reported by the energy department, energy costs will rise by $8.7 billion a year due to this increase in power generation and industrial consumption. We are seeing more and more of these costs being passed on to consumers. The fact is that increasing

> energy efficiency is only possible and beneficial when the cost of energy is kept constant. *- generated by OpenAI GPT trained on moderate liberal speeches with prompt "Increasing energy efficiency..."*

At the end of the survey, Turkers were asked to evaluate other sentences generated via OpenAI GPT moderate liberal models and morally reframe the above sentence to match their ideology.

This experiment was designed to test if conservatives were open to moderate liberal points of view before and after they were exposed to morally reframed text generated from moderate liberals and conservatives. I summarized results by considering the strength of their conservative ideology (Strong, Typical, Weak), attitudes towards climate change policy (Con, Open, Pro), if they agree or disagree, on the 5 point Likert scale, with the above statement (Attitude Score), their moral rhetoric from open-ended questions (Morals), and the aggregate Likert scores from three OpenAI GPT generated moderate liberal sentences (Final Score). These are summarized in table 5.2. Openness is quantified by calculating Likert score increases before (Attitude Score) and after (Final Score) the test. A higher final score shows greater agreement with generated moderate liberal sentences. A paired, one-tailed t-test on this data yielded a high p-value of 0.34. Future studies on openness must control by topic, include an initial aggregate Likert attitude score, and increase the sample size.

This experiment created a new dataset of morally reframed sentences and respective ideologies. In their reframing, many Turkers referenced the generated $8.7 billion dollar number, even though it is just generated text and not a fact. The moral generator ought to remain factual. Still, this shows that Turkers took care to reframe the sentence and reused the "same fact" in different ways.

Table 5.2: Raw Results for Experiment 1

Leaning	Attitude	Attitude Score	Morals	Final Score
Strong	Con	3	none	2.75
Strong	Con	2	Harm, Authority, Purity	3.5
Weak	Pro	3	none	3.75
Weak	Open	3	Fairness, Ingroup, Authority	2.75
Weak	Con	3	Harm, Morality	3
Strong	Pro	4	none	3
Weak	Con	4	Morality	3.75
Weak	Open	3	none	3
Typical	Open	3	Harm	2.5
Typical	Pro	5	none	3.25
Typical	Pro	2	Authority	3.5
Typical	Pro	3	none	3
Typical	Pro	3	none	3.25
Typical	Con	4	Harm, Authority, Purity	2.75
Strong	Open	3	Morality	2.75
Weak	Pro	4	Harm	2.5
Weak	Open	2	none	2.75
Typical	Open	2	none	2
Weak	Open	4	Fairness	4
Typical	Con	4	Fairness, Morality	2.75
Strong	Con	2	none	2.75
Weak	Open	4	none	2.5
Typical	Pro	5	Fair, Authority	4.25
Typical	Pro	4	Harm	3
Weak	Open	1	Harm	3
Weak	Open	1	Harm	2.75

Many respondents found the task enjoyable and one even wrote to me, "I rather enjoyed the task and I could have elaborated more [...] had I note been short on time." One problem was that the submit button on the form stopped working halfway through the experiment, causing many people to complete the survey but not get compensation. Further, even though the experiment was supposed to only be for conservatives, one liberal person filled out the form. Future trials should give Turkers 45 minutes, and all responses should be monitored in real time to avoid these issues.

Amazon Turk Workers can potentially provide a moral reframing dataset themselves. The data collected from the Turkers in this experiment is substantial and can be further analyzed.

Experiment 2: Comparing Point of View

1. Controlling for topic and method, I evaluated if Turkers resounded most with extreme or moderate liberals or conservatives using the basic model on speeches and Tweets for guns and crime. The four results are below:

   (a) Amendment blocks terrorists from buying guns AND protects due process. *-Moderate Conservative*

   (b) To call someone who slightly speeds a criminal is just deflecting from the real criminals *-Moderate Liberal*

   (c) Appreciate all the Tweets from people supporting criminal background checks. I very much agree with you.*-Extreme Liberal*

   (d) Taking guns from honest people= more likely criminals will attack them, that's why crime higher in gun controlled cities.*-Extreme Conservative*

2. I repeated above experiment using generated OpenAI GPT sentences from Tweets on the topic of healthcare.

1. We have some of the best health care in the world. In fact, our health care is top-notch in the world. *-Moderate Conservative*

2. #Obamacare is a scam. It does not lower costs, it increases them. #NoLargesse -*Extreme Conservative*

3. We are working to get the right kind of affordable coverage to the working poor, and today we are working to do so while maintaining the safety net for the millions of american children. This proposal will bring an end to the nightmare of a jobless child, and help over 4 million americans to participate in the health insurance market. *-Moderate Liberal*

4. The #GOPCare is about a year behind the $1.8 trillion cost reduction promises which will reduce the deficit by $9.2 trillion, and help Americans protect their health and freedom. *-Extreme Liberal*

Table 5.3: Results for Experiment 2

Method	**Extreme Liberal**	**Moderate Liberal**	**Moderate Conservative**	**Extreme Conservative**	**None**
Basic	0	0	0	21	5
GPT	3	9	5	7	2

Although preliminary, it is interesting to note that the generated speeches provided more variance, with the most people voting for the *moderate liberal* generated text. While this may be due to the difference in topic (healthcare for GPT vs. crime for basic), moral reframing may be persuade people to be open to a different ideology.

Experiment 3: Comparing Generation Methods 1. For the topic of immigration, I compared the persuasiveness of the RNN, and OpenAI GPT methods on extreme liberal data.

1. Refugees who've been kicked out of their homes & death system. #KeepFamiliesTogether. *-RNN using Tweets and prompt "Refugees who've been kicked out of their homes"*

2. Thank you to have my office to hear from new refugees who have arrived in the U.S. from a country. *-OpenAI GPT using topic: refugees*

Table 5.4: Results for Experiment 3: Comparing Generation Methods

RNN	**OpenAI GPT**	**None**
4	2	20

The RNN and the OpenAI GPT were generated from extreme liberal data, so it makes sense that many Turkers found none persuasive. I was surprised that the RNN fared better than the OpenAI GPT despite grammatical issues. It may be because the sentence from OpenAI GPT isn't persuasive - it is just a statement. In the future, there should be an explanation box for Turkers to justify their choices.

Experiment 4: Comparing Coherency for RNN Turkers were asked to evaluate the following sentences, all generated from the RNN on using different prompts or data sources, for coherency.

1. Refugees who've been kicked out of their homes & death system. #KeepFamiliesTogether. *-RNN using Tweets and prompt "Refugees who've been kicked out of their homes"*

2. Immigrational security officials. Call this summer? Great to see the past 50 years, our community has stone. *-RNN using Tweet and topic: immigration"*

3. Background checks for all gun sales that as any committee on run attends temporis that they will remaining time-800 technicals today. *-RNN using speeches and prompt "Background checks for all gun sales"*

Table 5.5: Results for Experiment 4: Comparing Coherency for RNN

Tweet Prompt	**Tweet Topic**	**Speeches Prompt**	**None**
1	6	1	18

Many Turkers found the RNN largely incomprehensible. This was a significant challenge for this project. Even selecting these sentences from the RNN took substantial human effort because the RNNs spew out many incoherent sentences. It is promising that given a general topic and the much smaller Twitter dataset, the RNN was more understandable than when using speeches. More research is required to see if this trend holds across different topics.

Experiment 5: Comparing Data Methods Here, I compared the persuasiveness of all speech data, speech data filtered by topic, and Tweets for moderate liberals as data inputs to the OpenAI GPT and RNN models.

1. At least from the standpoint of economics, not very much pollution if you don't want to deal with it. In fact, in reality, global emissions of CO2 is declining. *-Base Sentence from untrained OpenAI GPT*

2. Over 70 areas around the country have been identified as failing to meet the NAAQS standards. Several cities face a construction ban and restriction on highway air and sewage treatment grants. One wants to see that happen, but if we delay sanctions, we reward those areas which have delayed adopting plans to clean their air and delayed adopting plans to bring themselves into compliance with the clean air act. *-Basic method*

3. I'm worried about our families. I'm worried about families and kids getting hurt. I'm worried about the security, public safety and health of our schools. *-OpenAI GPT Tweet topic*

4. At least from the standpoint of economics, not very much pollution investigational control their work, even if ip abuse today during their tripts *-RNN with prompt*

5. At least from the standpoint of economics, not very much pollution is allowed to take place. so it would seem that it would be a fair exchangeoff of credits if we were to give back at least $1 billion in credits to the companies. *-OpenAI GPT all speech with prompt*

6. At least from the standpoint of economics, not very much pollution does exist in the world. If we continue down this path, we are going to hit a new peak by the end of the century – the point where the climate is going to be truly dangerous. *-OpenAI GPT environmental speech with prompt*

7. If the Clean Power Plan's cost calculations are wrong, some argue that it would be irresponsible for states to spend tens of billions of dollars for one policy in order to keep another from going into effect. *-OpenAI GPT environmental speeches with topic*

Table 5.6: Results for Experiment 5: Comparing Data Methods

Name	**Counts**
Basic	11
BaseOpenAI GPT	0
OpenAI GPTTweetTopic	5
RNNPrompt	0
OpenAI GPTEnvSpeechPrompt	5
OpenAI GPTSpeechPrompt	1
OpenAI GPTSpeechTopic	2
None	2

The most successful model was the basic model. On inspection, the use of coherent facts and sustained argumentation could have made it the most persuasive. It is promising that for the generated sentences, the OpenAI GPT trained on environmental speeches was as persuasive as the OpenAI GPT trained on the Tweets, despite being smaller. It is interesting to note the continuing trend that the full speech trained models, which have more data, function worse than smaller models using only a subset of speech or Tweet data. More exploration is warranted as to why the speech dataset performs poorly and how to create the most persuasive datasets.

Analysis of Turk Experiment Based on the current results, there are no conclusions if exposure to morally reframed arguments had any impact on openness in conservatives. More respondents and more questions need to be asked to ensure that the results are due to persuasiveness and not population variation, grammar, charged language, or specific facts used. Using this preliminary information, investing more time into bolstering the Twitter dataset and splitting the speech dataset by topic could yield better results for the OpenAI GPT and RNN methods. A metric of success is if the generated reframed sentence is more persuasive than a sentence from the basic model.

Here are some results from the Turkers who morally reframed the statement from Experiment 1: "Increasing energy efficiency":

Participant 23: Typical Conservative

> Increasing energy efficiency and renewable energy will have dramatic effects on the environment In fact, in this decade alone, as reported by the energy department, energy costs will rise by $8.7 billion a year which will make the cost increase to the consumer in the long run. *As the world moves toward more efficient and renewable sources, power suppliers must also have the option of using cost-effective resources to prices affordable for the poor and for developing countries. Otherwise, the progress will be slower and keep more people and nations reliant on traditional sources which pollute more.*

This participant retained the structure and topic of the initial statement and appealed to the morals of fairness and harm. It is a different argument than the initial statement so this may not be a moral reframing.

Participant 21: Weak Conservative

> Increasing energy efficiency and renewable energy are not going to have dramatic effects on the environment. *Consumers can't afford the costs that trying to force renewable energy on the population will bring. They can't afford it with their wallets and we can't afford as a society the jobs that will be lost trying to implement something like this. Improve the economy first and make sure are children are being properly educated so that we can more forward and look for a real feasible solution to this problem...*

This participant chose a generic argument that could be used for any topic. It appeals to the morals of human harm, but is vastly different from the Participant 23's sentence. It would be a weak example of moral reframing.

Participant 14: Typical Conservative

> Increasing energy efficiency and renewable energy will possibly slow down climate change, *but what if it doesn't? Do we really want to tax our citizens, who are already taxed to the max, on a policy that may not even help?*

Participant 14 chose to reframe by using rhetorical questions, a powerful persuasive technique. As I had preprocessed out most of the punctuation in my datasets, I lost all rhetorical questions. Generating rhetorical questions can help with moral reframing.

Others who reframed the statement added facts about greenhouse gasses, used emotional language when describing the cost, or reworded the existing paragraph logically and concisely. Their responses give insight into the type of language that they individually find persuasive.

Bibliography

[1] Afl-cio scorecard. https://aflcio.org/legislative-scorecard.

[2] Afscme congressional scorecards. https://www.afscme.org/issues/congressional-scorecards.

[3] A beginner's guide to neural networks and deep learning. https://skymind.ai/wiki/neural-network.

[4] gensim. https://pypi.org/project/gensim/.

[5] Immigration-reduction grades | numbersusa - for lower immigration levels. https://www.numbersusa.com/content/my/tools/grades.

[6] Nra grades archive. https://everytown.org/nra-grades-archive/.

[7] Socialseer. https://www.socialseer.com/resources/us-senator-twitter-accounts/.

[8] spacy industrial-strength natural language processing in python. https://spacy.io/.

[9] Text generation using a rnn with eager execution | tensorflow core | tensorflow. https://www.tensorflow.org/tutorials/sequences/text_generation.

[10] What is machine translation? rule based machine translation vs. statistical machine translation. http://www.systransoft.com/systran/translation-technology/what-is-machine-translation/.

[11] How language 'framing' influences decision-making. Jun 2017. https://www.psychologicalscience.org/publications/observer/obsonline/how-language-influences-decision-making.html.

[12] All member of congress scores. Apr 2018. http://scorecard.lcv.org/members-of-congress.

[13] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg,

Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org.

[14] Jay Alammar. The illustrated bert, elmo, and co. (how nlp cracked transfer learning). http://jalammar.github.io/illustrated-bert/.

[15] Jay Alammar. The illustrated transformer. http://jalammar.github.io/illustrated-transformer/.

[16] Jay Alammar. Visualizing a neural machine translation model (mechanics of seq2seq models with attention).

[17] Quinn Albaugh, Julie Sevenans, Stuart Soroka, and Peter John Loewen. The automated coding of policy agendas: A dictionary-based approach. In *6th Annual Comparative Agendas Conference, Atnwerp, Beligum*, 2013.

[18] Quinn Albaugh, Julie Sevenans, Stuart Soroka, and Peter John Loewen. Lexicoder topic dictionaries. In *6th Annual Comparative Agendas Conference, Atnwerp, Beligum*, 2013.

[19] Amirbar. amirbar/rnn.wgan. Apr 2018. https://github.com/amirbar/rnn.wgan.

[20] Sarah A. Binder. Going nowhere: A gridlocked congress. Jul 2016. https://www.brookings.edu/articles/going-nowhere-a-gridlocked-congress/.

[21] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. 3(Jan):993–1022, 2003.

[22] Reily BourneReily Bourne and Renaud. How to compute the similarity between two text documents? https://stackoverflow.com/a/24129170.

[23] Jason Brownlee. How to develop an encoder-decoder model for sequence-to-sequence prediction in keras. Jul 2018.

[24] Keith Carlson, Allen Riddell, and Daniel Rockmore. Evaluating prose style transfer with the bible. 5(10):171920, 2018.

[25] François Chollet et al. Keras. 2015.

[26] Franziska Cvar and Agnieszka Ewa Tytus. Moral judgement and foreign language effect: when the foreign language becomes the second language. 39(1):17–28, 2018.

[27] Mark Daku, Stuart Soroka, and Lori Young. Lexicoder, version 3.0. 2015.

[28] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018.

[29] Matthew Feinberg and Robb Willer. The moral roots of environmental attitudes. 24(1):56–62, 2012.

[30] Pew Research Center for the People and the Press. Political polarization in the american public. Oct 2016. https://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/.

[31] Kendall Fortney. Pre-processing in natural language machine learning. Nov 2017.

[32] Yonah Bromberg Gaber. A list of twitter handles for members of congress. May 2017. https://gwu-libraries.github.io/sfm-ui/posts/2017-05-23-congress-seed-list.

[33] Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. Allennlp: A deep semantic natural language processing platform. 2018.

[34] Justin Garten, Reihane Boghrati, Joe Hoover, Kate M Johnson, and Morteza Dehghani. Morality between the lines: Detecting moral sentiment in text. In *Proceedings of IJCAI 2016 workshop on Computational Modeling of Attitudes*, 2016.

[35] Jesse M. Shapiro Gentzkow, Matthew and Matt Taddy. Congressional record for the 43rd-114th congresses: Parsed speeches and phrase counts. 2018.

[36] Jesse Graham and Johnathan Haidt. Moral foundations dictionary.

[37] Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. Chapter two - moral foundations theory: The pragmatic validity of moral pluralism. 47:55–130, 2013.

[38] Jesse Graham, Jonathan Haidt, and Brian A. Nosek. Liberals and conservatives rely on different sets of moral foundations. 96(5):1029–1046, 2009.

[39] Masato Hagiwara. Building seq2seq machine translation models using allennlp. Dec 2018.

[40] Melanie Hoff. melaniehoff/partisan_thesaurus. Mar 2017. https://github.com/melaniehoff/.

[41] Huggingface. transfer-learning-conv-ai. https://github.com/huggingface/transfer-learning-conv-ai/blob/master/convai_evaluation.py.

[42] Huggingface. pytorch-pretrained-bert. May 2019. https://github.com/huggingface/pytorch-pretrained-BERT.

[43] Ravi Iyer, Spassena Koleva, Jesse Graham, Peter Ditto, and Jonathan Haidt. Understanding libertarian morality: The psychological dispositions of self-identified libertarians. 7(8):e42366, 2012.

[44] Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1113–1122, 2014.

[45] Yangfeng Ji and Noah Smith. Neural discourse structure for text categorization. 2017.

[46] Bradley Jones. The morality of representation: Constituent moral foundations and position-taking in congress. 2012.

[47] Jad Kabbara and Jackie Chi Kit Cheung. Stylistic transfer in natural language generation systems using recurrent neural networks. In *Proceedings of the Workshop on Uphill Battles in Language Processing: Scaling Early Achievements to Robust Methods*, pages 43–47, 2016.

[48] Dhruvil Karani. Introduction to word embedding and word2vec. Sep 2018.

[49] Andrej Karpathy. The unreasonable effectiveness of recurrent neural networks. May 2015.

[50] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M Rush. Opennmt: Open-source toolkit for neural machine translation. 2017.

[51] Matthew Kugler, John T Jost, and Sharareh Noorbaloochi. Another look at moral foundations theory: Do authoritarianism and social dominance orientation explain liberal-conservative differences in "moral" intuitions? 27(4):413–431, 2014.

[52] Keith Poole Howard Rosenthal Adam Boche Aaron Rudkin Lewis, Jeffrey B. and Luke Sonnet. Voteview: Congressional roll-call votes database. 2019. https://voteview.com/.

[53] Ying Lin, Joe Hoover, Gwenyth Portillo-Wightman, Christina Park, Morteza Dehghani, and Heng Ji. Acquiring background knowledge to improve moral value prediction. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 552–559. IEEE, 2018.

[54] Edward Loper and Steven Bird. Nltk: The natural language toolkit. In *In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. Philadelphia: Association for Computational Linguistics*, 2002.

[55] Scott Martelle. What were they thinking? america's enduring culture of argument. Apr 2013.

[56] Nolan McCarty, Keith T Poole, and Howard Rosenthal. *Polarized America: The dance of ideology and unequal riches.* mit Press, 2016.

[57] Minimaxir. minimaxir/textgenrnn. Jan 2019. https://github.com/minimaxir/textgenrnn.

[58] MoralFoundations. Moral foundations questionnaire. https://www.moralfoundations.org/questionnaires.

[59] John J Nay. Predicting and understanding law-making with word vectors and an ensemble model. 12(5):e0176999, 2017.

[60] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[61] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. 2018.

[62] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 1:8, 2019.

[63] Eyal Sagi and Morteza Dehghani. Measuring moral rhetoric in text. 32(2):132–144, 2014.

[64] Benjamin Sherman and Zayd Hammoudeh. Make deep learning great again: Character-level rnn speech generation in the style of donald trump.

[65] Carson Sievert and Kenneth Shirley. Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70, 2014.

[66] Joanna Sterling and John T Jost. Moral discourse in the twitterverse. 17(2):195–221, 2018.

[67] Christopher L Suhler and Patricia Churchland. Can innate, modular "foundations" explain morality? challenges for haidt's moral foundations theory. 23(9):2103–2116, 2011.

[68] Gilbert Tanner and Gilbert Tanner. Generating text using a recurrent neural network. Oct 2018.

[69] Taspinar. taspinar/twitterscraper. Feb 2019. https://github.com/taspinar/twitterscraper.

[70] TheBerkin. Theberkin/rant. Feb 2019. https://github.com/TheBerkin/rant.

[71] Svilen Todorov. Generating fake conversations by fine-tuning openai's gpt-2 on data from facebook messenger. Mar 2019. https://svilentodorov.xyz/blog/gpt-finetune.

[72] Unitedstates. unitedstates/congress. May 2019.

[73] Unitedstates. unitedstates/congress-legislators. May 2019. https://github.com/unitedstates/congress-legislators.

[74] Vlraik. vlraik/word-level-rnn-keras. Jul 2016. https://github.com/vlraik/word-level-rnn-keras.

[75] Heather Winskel, Theeraporn Ratitamkul, Victoria Brambley, Tulaya Nagarachinda, and Sutheemar Tiencharoen. Decision-making and the framing effect in a foreign and native language. 28(4):427–436, 2016.

[76] Thomas Wolf. The current best of universal word embeddings and sentence embeddings. May 2018.

[77] Thomas Wolf, Victor Sanh, Julien Chaumond, and Clement Delangue. Transfertransfo: A transfer learning approach for neural network based conversational agents. 2019.

[78] Dani Yogatama, Manaal Faruqui, Chris Dyer, and Noah Smith. Learning word representations with hierarchical sparse coding. In *International Conference on Machine Learning*, pages 87–96, 2015.

[79] Zhenye-Na. Zhenye-na/da-rnn. Apr 2019. https://github.com/Zhenye-Na/DA-RNN.

This work includes sections of my SML 310 final project on Flexible Moral Dictionaries with Prof. Michael Guerzhoy.

I pledge my honor that this work is in line with University guidelines.
Ananya A. Joshi