Theory-grounded NLP for generalizable framing detection

Julia Mendelsohn
University of Michigan

Introduction

Framing is the process of selecting particular aspects of an issue and making them salient in a communicating text in order to shape how people understand the issue, attribute responsibility, and endorse solutions (Entman, 1993; Iyengar, 1991); it thus has a significant impact on individual decision-making, public opinion, and policy (Tversky and Kahneman, 1981; Chong and Druckman, 2007). Traditionally, frames are identified in text with costly manual annotation, which limits analyses of framing and its effects to small unrepresentative data sets, single issues, and few distinct sources. Although recent NLP work has made significant progress in automated frame detection, existing models lack generalizability, so expanding to new frame types, issues, source domains, or spatiotemporal contexts still require massive annotation efforts. **I propose a new few-shot model, FRAMEDETECT, that supports generalizable frame detection across theoretically-motivated frame typologies, contexts, and modalities. With additional flexibility for handling user-specified frames, FRAMEDETECT will enable large-scale analysis of meaningful frames and their effects in all kinds of messages, and will thus be invaluable to stakeholders ranging from political scientists to product advertisers.**

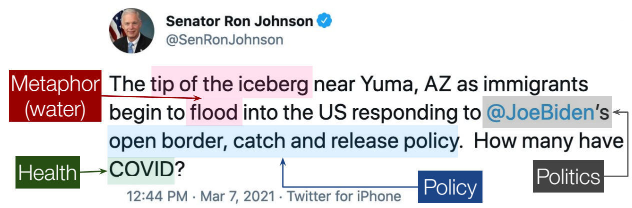

Figure 1: Example of frames invoked in a message about immigration

Prior work in social and cognitive sciences have identified numerous framing devices that can persuade audiences, including *issue-generic policy* frames (Boydstun et al., 2013), *issue-specific* frames (de Vreese, 2005; Benson, 2013), frames surrounding narrative and storytelling style (Iyengar, 1991; Semetko and Valkenburg, 2000), and metaphorical frames (Thibodeau and Boroditsky, 2011). Although framing analysis is most prominent in politically communication, it has widespread utility in other areas. For example, framing in "green advertising" influences consumers' preferences and intentions to purchase eco-friendly products (Chang et al., 2015; Cucchiara et al., 2015; Schmidt et al., 2017; Reczek et al., 2018). Recent public health research has found that framing in Covid-19 messaging impacts public attitudes towards mask-wearing, social-distancing, and vaccination (Chou and Budenz, 2020; Deslatte et al., 2020; Van Der Linden and Savoie, 2020; Ceylan and Hayran, 2021). Framing is even an essential consideration for the future of artificial intelligence. Frames in public-facing communication about technologies such as autonomous vehicles can impact if, when, and how the public adopts new technology (Cunneen et al., 2019). Because of such effects across many issues and fields, we urgently need to develop technologies to detect framing strategies at scale.

Beyond its potential in interdisciplinary applications, frame detection presents a unique NLP challenge because it requires sophisticated reasoning about both cognitive processes and social information. Fundamentally, framing is about relating an issue to other concepts. These conceptual mappings may be grounded in common sense, emotion, prior knowledge, cultural contexts, and stereotypes. Developing systems to automatically recognize frames is thus an important milestone for machines to achieve human-level language capabilities. Automated frame detection systems must be able to accurately identify **meaningful** frames in a **generalizable** way across issues, time, sources, and modalities. To be truly useful, models ought to be able to identify frames from different theoretically-informed typologies and flexibly handle any new user-specified frame of interest with minimal annotation efforts. In contrast to existing frame detection systems, **my proposed research will satisfy all of the above desiderata, and will thus further progress in both artificial intelligence and social science research**.

Despite tremendous progress in automated frame detection, existing models have yet to achieve this balance of identifying theoretically-meaningful frames, generalizing across contexts and frame types, and accommodating user-specified frames of interest. The majority of NLP work has focused on supervised detection of a single typology of *issue-generic policy frames* using the Media Frames Corpus which contains annotated news articles across several political issues (Boydstun et al., 2013; Card et al., 2015, 2016; Johnson et al., 2017; Naderi and Hirst, 2017; Field et al., 2018; Hartmann et al., 2019; Khanehzar et al., 2019; Kwak et al., 2020; Khanehzar et al., 2021). Although the categories present in the Media Frames Corpus are generalizable, these supervised models are not; language constituting the same frame varies across issues, sources, and time. Applying this frame typology to new contexts still requires massive annotation efforts. Furthermore, *issue-generic policy* frames may be underinformative and not sufficiently meaningful to be useful in practice. I have previously demonstrated that *issue-generic policy* frames obscure insights framing variation across subpopulations and effects on audience responses; these patterns were only revealed by *issue-specific* frames (Mendelsohn et al., 2021). While models for *issue-specific* frame detection have led to novel findings about political discourse (Morstatter et al., 2018; Liu et al., 2019a; Mendelsohn et al., 2021), they suffer from an even more severe lack of generalizability; one needs to collect a whole new annotated dataset to expand analysis just to include a new frame, another social media platform or a different time period. The full annotation process is time consuming and expensive, and undertaking such efforts is infeasible, particularly considering the volume and rapid evolution of online discussions.

Unsupervised methods have the opposite problem because they require zero annotation (Heidenreich et al., 2019; Kwak et al., 2021). However, frames are more complex than topics because they are defined by their role in shaping interpretations of issues (Entman, 1993). There are thus no guarantees that unsupervised topics or clusters produce meaningful cognitively-grounded frames, or that all frames are captured within topics. Although they are generalizable, unsupervised methods are also not necessarily replicable across datasets. This potential lack of consistency leads to an inability to compare or align framing strategies across datasets, which makes it infeasible to assess general attitudes or public opinion shifts.

Each phase of my proposed research addresses a major obstacle in frame identification. In Phase 1, I develop FRAMEDETECT, a Siamese network that uses transfer learning for few-shot frame detection. FRAMEDETECT is generalizable across issues, contexts, and flexibly allows domain experts to specify additional frames of interest. In Phase 2, I present a novel task, multimodal frame detection, motivated by cognitive and social science research on how visual and linguistic framing work in tandem to persuade audiences (Scheufele and Iyengar, 2012). I also propose a vision-language extension of FRAMEDETECT to identify multimodal frames. While existing systems focus on *explicitly-cued* frames that can be judged to be either present or absent in a given document, *implicitly-cued* metaphorical frames can also influence readers but rely on more subtle associations between issues and target concepts. In Phase 3, I introduce a new human-in-the-loop methodology that uses FRAMEDETECT for discovery and measurement of metaphorical frames. I will conclude by discussing how my research will advance NLP and other disciplines and highlight the broader impacts of the proposed work for both my academic community and society.

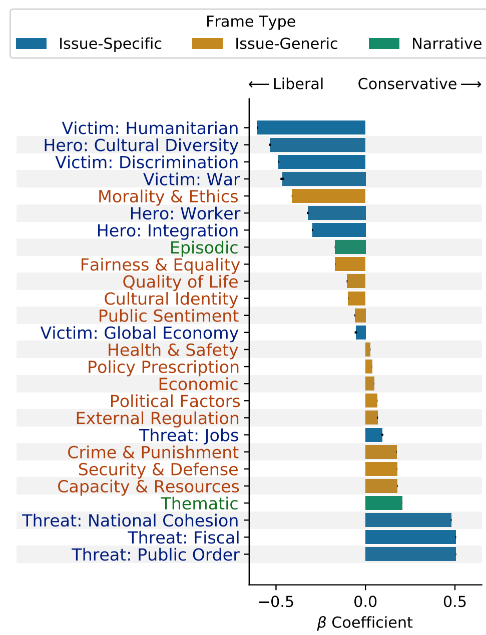

Figure 2: Figure from Mendelsohn et al., 2021 showing the relationship between political ideology and framing in immigration discourse. Although most NLP work focuses on *issue-generic policy frames* (shown in orange), these frames obscure sharp ideological stratification that is revealed by theoretically-grounded *issue-specific frames* (shown in blue).

Phase 1: Flexible frame detection with Siamese networks and multitask learning

FRAMEDETECT will detect meaningful frames across issues, sources, and spatiotemporal contexts. In contrast to existing models, FRAMEDETECT can flexibly leverage expert knowledge and theoretically-grounded frame typologies by allowing users to detect any frame of interest. For example, a political communication scholar may be interested in analyzing how online framing of immigrants as *victims of humanitarian crises* or *threats to public order* affects public opinion. Or a car company's marketing division wants to know whether emphasizing *affordability* or *power* would make consumers more likely to make a purchase. FRAMEDETECT can identify such frames while being robust to linguistic variation across different platforms and online communities (Lucy and Mendelsohn, 2019).

Methods FRAMEDETECT learns associations between three components: a set of *documents* D, linked *metadata* M containing contextual information about the data source (e.g. platform, community, time period), and a *frame bank* F containing the full set of built-in and user-specified frames with accompanying natural language descriptions. The model consists of two major components: *multitask learning* that uses text and metadata to learn enriched document representations and a *Siamese network* that classifies if any particular frame $f_i \in F$ is present in a given document $d \in D$. The overall architecture is shown in Figure 3.

To learn enriched document representations, word piece embeddings of d and its associated metadata m (if available) are concatenated and fed as inputs into a RoBERTa contextualized language model (Liu et al., 2019b). Enriched document representations are learned by fine-tuning RoBERTa on two objectives: masked language modeling and annotated framing strategies. Training RoBERTa to predict masked tokens from the concatenated (d, m) input allows the model to capture relationships between the text and metadata; this is necessary for generalizability, as it helps distinguish whether linguistic differences across documents are due to differences in framing or context. Additionally, enriched document representations will incorporate available information about framing by including frame prediction as an auxiliary task for the subset of already-labeled data. Even if the annotated data only contains labels for frames other than the particular frame of interest f_i, this will "nudge" the model to pay attention to parts of a document that are more likely to contain framing devices.

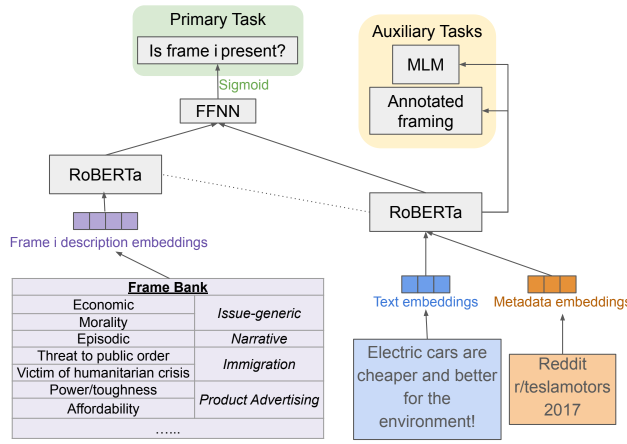

Figure 3: FRAMEDETECT architecture. Frames of interest are included along with brief descriptions in the frame bank on the left. As shown in this example, the frame bank can contain both *issue-generic* and *issue-specific* frames. The right side of the figure shows how enriched document representations are learned by fine-tuning joint text and metadata representations on auxiliary tasks. For the primary task, a Siamese network learns to detect frames from enriched document and frame description representations. frame description and enriched document representations.

The Siamese network jointly learns and compares representations for frames and documents via fine-tuning RoBERTa models with tied weights. The representation for f_i is learned from provided natural language descriptions. Similar to content analysis codebooks in social science research, these descriptions may contain a frame's definition and/or canonical examples. This design is inspired by Sentence-BERT, which uses Siamese networks to compare document similarity (Reimers and Gurevych, 2019), but unlike Sentence-BERT, FRAMEDETECT does not compare two documents from D; rather, it compares a document d with the description of the frame f_i. Finally, I formulate the primary frame detection task as a set of binary classifiers, where FRAMEDETECT determines if $f_i \in F$ is present in document d. This allows for d to contain multiple frames and provides flexibility because new frames can easily be added to F. Furthermore, this structure enables the practitioner to select a subset of frames of interest from the full frame bank F, which is necessary when F contains *issue-specific* frames from disparate issues.

Like standard supervised methods, FRAMEDETECT requires labeled data. However, because it uses multitask learning and frame representations, I posit that FRAMEDETECT requires less annotated data to achieve improvements over current state-of-the-art frame prediction methods, and will evaluate this hypothesis on existing datasets.

Data I will release a base version of FRAMEDETECT that can be used off-the-shelf with zero additional annotations. This model will be trained using the following datasets, which cover a range of issues and sources:

1. News articles from the Media Frames corpus about immigration, tobacco, and same-sex marriage labeled for *issue-generic policy frames* (Boydstun et al., 2013; Card et al., 2015)
2. Politicians' tweets about immigration, abortion, LGBTQ rights, guns, terrorism, and the Affordable Care Act labeled for *issue-generic policy frames* (Johnson et al., 2017)
3. News headlines about gun violence labeled for *issue-specific* frames (Liu et al., 2019a)
4. Immigration-related tweets labeled for *issue-generic policy*, *issue-specific*, and *narrative* (episodic and thematic) frames (Mendelsohn et al., 2021)

Finally, to illustrate FRAMEDETECT's potential to generalize beyond political discourse, I will create and include a novel dataset of vehicle advertisements from Twitter labeled for frames such as *affordability* and *power/toughness*.

Evaluation I will evaluate FRAMEDETECT's in-domain performance by calculating precision, recall, and F1-score for each frame on a randomly-sampled held out set. Because FRAMEDETECT is motivated by the need for generalizability, I will also evaluate out-of-domain performance for unseen issues, frame categories, and textual domains (e.g. evaluating a model trained with labeled tweets on news articles). I will compare FRAMEDETECT to prior work such as logistic regression and standard RoBERTa-based classifiers (Mendelsohn et al., 2021). I hypothesize that FRAMEDETECT requires less annotated data to achieve superior performance than existing models, so I will assess how the model's performance varies depending on the size of the training data. This evaluation will also inform guidelines for practitioners on how much annotateion is required for sufficient performance on new frames and contexts.

I will further conduct ablation studies to understand the how various components of FRAMEDETECT contribute to the overall performance, such as the decision to incorporate additional metadata or fine-tuning RoBERTa on each auxiliary task. I will further probe the role of the natural language frame descriptions by addressing the following questions. Is just including the frame label (e.g. *affordability*) a sufficient input, or are additional descriptors necessary? Should descriptions be dictionary definitions or canonical examples of each frame, or some combination of both? How much does the precise wording or length of the description affect model performance?

Perhaps the most important sucess metric for FRAMEDETECT is its utility for practitioners. Thus, I will assess the interdisciplinary impact of this work through a qualitative user study of experts across disciplines.

Phase 2: Vision-language modeling for multimodal frame detection

Most information that people consume online is multimodal. On news websites and social media, text and images interact in intricate ways to form messages that shape people's understandings of important issues. It is thus urgent for computational systems to consider both linguistic and visual information when detecting frames. In contrast to current popular vision-language tasks such as image captioning (Xu et al., 2015), multimodal frame detection presents a unique challenge because the precise relationship between the text and image is unclear; they likely do not refer to the same scene, and may not even cue identical frames. Furthermore, both linguistic and visual framing rely on the audience's familiarity with an issue's background context, cultural artifacts, and stereotypical associations. **I introduce a new vision-language task: multimodal frame detection. Along with a novel dataset labeled for multimodal frames, I propose an extension of FRAMEDETECT using state-of-the-art joint vision-language models.**

Prior work in cognitive science, communication, and critical discourse analysis draws attention to the impacts of visual cues. By directing what audiences focus on, particular image-text combinations can reinforce textual frames, resolve semantic ambiguities that otherwise hinder cognitive accessibility of textual frames (Scheufele and Iyengar, 2012), alter audiences' interpretations, or invoke additional framing devices (Scheufele and Iyengar, 2012; Belmonte and Porto, 2020). Different camera angles can psychologically distance audiences from the subject matter or bring them closer, and create a sense of chaos or one of authority (Belmonte and Porto, 2020). Furthermore, visual and textual frames are processed via distinct mechanisms, and consequently can have different effects depending on the audience; readers who are more distracted or rapidly browsing may be more persuaded by images (Powell et al., 2019).

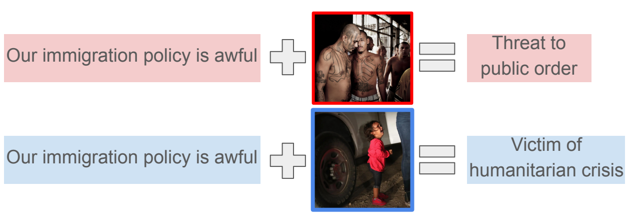

Figure 4: Motivating example for multimodal frame detection with images from the *New York Times*. Despite the top and bottom messages sharing the same text, the combination of text and image cue distinct frames.

Consider the two messages about immigration in Figure 4. Both messages have identical text, but the top image shows a group of tattooed men, while the bottom shows a crying young girl. These messages frame immigration differently due to the cultural values and stereotypes that the images surface. Based on my prior work with Twitter data, such scenarios are commonplace. We need multimodal frame detection to differentiate these messages and thus have a much better understanding about online discourses surrounding important issues.

Dataset Collection I will use the Twitter Academic Research API to collect tweets surrounding a variety of political and non-political issues, such as immigration and electric vehicles, and filter to posts containing accessible and downloadable images. Asking annotators to identify all frames at once is challenging. Instead, I will collect labels for each tweet by asking a binary question for each frame f_i: does the given message emphasize the f_i aspect of the issue? To directly facilitate comparisons between text-only and multimodal framing, I will randomly assign annotators to label messages with and without the accompanying image. I plan to recruit and train undergraduate research assistants to help with the data annotation. If the task is extremely easy and RAs achieve very high interannotator agreement, I will accelerate the data collection by crowdsourcing annotations on Amazon Mechanical Turk.

Methods I will extend the FRAMEDETECT model from Phase 1 with OSCAR, a state-of-the-art vision-language fusion model (Li et al., 2020). OSCAR leverages natural language tags from object detection on the image (Zhang et al., 2021), which occupy an intermediary space between language and vision that facilitates learning alignments between the two modalities. However, the image-text pairs for multimodal frame detection are more complex than the typical datasets for vision-language tasks; although there is some relationship between the text and image, the text doesn't directly describe the image. To address this issue, I will adapt Oscar by fine-tuning it to learn framing strategies that are identifiable through the joint analysis of visual and linguistic cues. Even though objects detected in images may not directly map onto text spans, my model will be able to learn information from both modalities to identify frames. The other proposed modification to OSCAR is including metadata about when and where a message comes from. As in Phase 1, learning metadata representations can improve generalizability by disentangling what signals are relevant for framing, and what are artifacts of the source domain or spatiotemporal context.

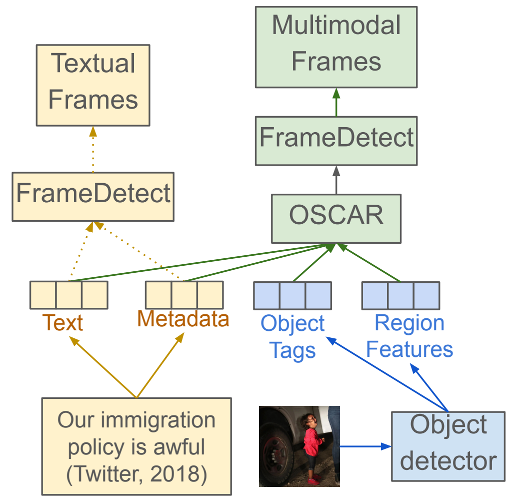

Figure 5: Extension of FRAMEDETECT for multimodal frame detection. This model employs object detection to learn natural language object tags region features, which are provided inputs to OSCAR along with text and metadata to learn enriched document representations.

Evaluation I will first design an experiment to quantitatively evaluate the value of the proposed multimodal frame detection task. In one version, I will show participants text from a particular message and four candidate images. In the control condition, the participants will select which image actually comes from the same message based on the text alone. In the treatment condition, participants will be asked to select the image based on both the text and the multimodal frame labels. If multimodal frames are meaningful, I would expect humans' to have greater accuracy selecting the corresponding image in the treatment condition. In the second version, I will switch the text and images, so participants are provided an image and asked to select the corresponding text.

I will then compare my FRAMEDETECT + OSCAR model to the text-only FRAMEDETECT on a held-out sample to assess how much visual information helps in multimodal frame detection. Additionally, prior work has shown that training models on combined vision and language representations can even improve performance for detecting complex rhetorical devices such as sarcasm that only emerge within text (Schifanella et al., 2016). Analogously, I would also expect the FRAMEDETECT + OSCAR model to outperform FRAMEDETECT on textual frames detection.

Phase 3: FRAMEDETECT for human-in-the-loop discovery of metaphorical framing

Existing datasets and models for frame detection have solely focused on explicit framing strategies that can readily be judged to be "present" or "absent" in a given document. However, a growing body of work in cognitive and social sciences has emphasized the persuasive impact of subtler implicit strategies, especially metaphorical framing. For example, experimental work has demonstrated that subtly manipulating text to cue metaphorical associations with concepts such as vermin, beasts, and viruses impacts participants' attitudes and support for particular policies (Thibodeau and Boroditsky, 2011; Marshall and Shapiro, 2018; Van Stee, 2018). I address this gap by proposing a pipeline that uses FRAMEDETECT for human-in-the-loop discovery and measurement of metaphorical frames.

To the best of my knowledge, no NLP work has attempted to computationally identify specific metaphorical frames and measure their prominence in a given message. In earlier work, I trained static word embeddings to quantify implicit associations between social groups and the dehumanizing vermin metaphor Mendelsohn et al. (2020), but this technique produced measurements at a corpus-level rather than a document-level. A more well-established NLP challenge is metaphor detection. Not to be confused with the proposed task of *metaphorical frame detection*, metaphor detection is a binary classification task where models predict whether or not each word in a document is being used in a metaphorical sense. Despite differences in the tasks, I draw upon cognitively-informed approaches that detect metaphorical word usage based on if that word is unlikely to appear in the given document's context, or if the word has a more basic or well-known meaning in other contexts (Crisp et al., 2007; Choi et al., 2021; Tong et al., 2021).

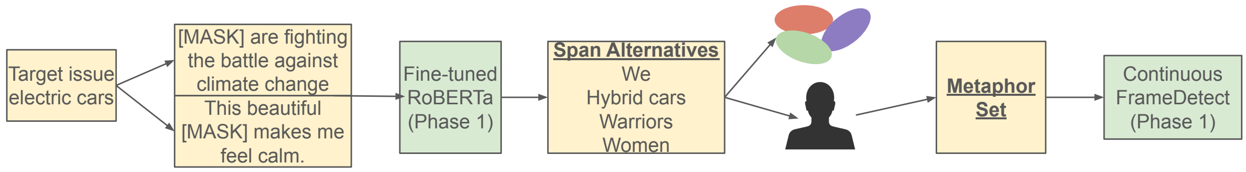

Figure 6: Human-in-the-loop pipeline for discovering and measuring implicit metaphorical frames. Potential metaphors are surfaced with masked span prediction using the fine-tuned RoBERTa model from FRAMEDETECT. Candidate span alternatives are then automatically clustered and analyzed by a human to determine the set of metaphorical frames. These frames can then be measured in documents using a modified FRAMEDETECT model with continuous output.

Methods The procedure for metaphorical frame discovery is shown in Figure 6. Consider some issue or entity of interest E (such as *electric cars* in Figure 6). Intuitively, if a message cues a metaphorical frame f_m, then replacing E with f_m should yield a plausible message (i.e. with low perplexity). Specifically, I propose identifying metaphorical frame candidates for E based on its most likely alternatives in a fill-in-the-blank masked span prediction task.

Although implicit metaphorical framing is distinct from the explicit frames discussed in Phases 1 and 2, I posit that tuning language models to recognize linguistic patterns in explicit framing can improve the models' ability to discover metaphorical frames. Thus, I will use the fine-tuned RoBERTa model from FRAMEDETECT to identify the most likely alternative text spans for for E. This logic is supported by prior work that has shown that jointly learning both metaphor and emotion detection improves performance on both tasks (Dankers et al., 2019), and that models trained for metaphor detection have higher performance on detecting *issue-generic policy frames* (Cabot et al., 2020).

I will uncover distinct concepts from the set of span alternatives by clustering based on vector representations. However, the resulting clusters are not guaranteed to correspond to unique metaphors. Some clusters may correspond to pronouns or synonyms of E rather than other concepts, or multiple clusters could correspond to the same metaphor. To produce a high-quality set of distinctive metaphorical frames, I propose human-in-the-loop curation of the surfaced span alternatives and clusters. Finally, the resulting set of metaphorical frames can be added to FRAMEDETECT's frame bank. With a slight modification to yield continuous rather than binary outputs, FRAMEDETECT can then be used to measure the extent to which a specific metaphorical frame is implicitly invoked in any given document.

As with explicit frames, FRAMEDETECT will require a small amount of data labeled with ground truth annotations. This presents an additional challenge for implicit frames because humans cannot easily assign scores to documents based on the degree of metaphorical associations. I thus propose collecting labels for metaphorical framing using Best-Worst Scaling (Kiritchenko and Mohammad, 2017), where each document's score is calculated based on a pairwise labeling task that asks annotators to simply determine which of two documents (d_1, d_2) is more evocative of the metaphorical frame f_m.

Evaluation As in Phase 1, I will evaluate the utility of FRAMEDETECT and the fill-in-the-blank task for surfacing metaphorical frames by recruiting experts in a variety of domains to use this system. These experts will assess ease of using this pipeline, its ability to retrieve established theoretically-informed metaphorical frames for particular issues, and the quality of any new metaphors discovered by the system. I will specifically evaluate the extent to which the fine-tuned RoBERTa model from FRAMEDETECT aids in metaphorical frame discovery compared to off-she-shelf contextualized language models. I also plan to evaluate FRAMEDETECT's continuous predictions of metaphorical framing by comparing it to held-out samples of annotated data. As in Phase 1, these held-out samples will include both random samples from the training distribution as well as on unseen issues, frame types, and source domains to assess the model's generalizability.

Research Impacts

My proposed research program will greatly benefit both the computer science and interdisciplinary research communities as well as society more broadly. By developing computational technologies to capture diverse types of theoretically-informed framing strategies across issues, domains, and modalities, I will advance machines' capabilities to reason about cognition and the social world. In Phases 2 and 3, I introduce novel challenges of multimodal frame detection and metaphorical frame measurement. Building models to perform well on these tasks will dramatically increase our ability to understand online discourses at a large scale. To encourage future research on generalizable frame detection, I will publicly share datasets that I create throughout this work as well as pre-trained models that can readily be expanded to novel settings. Furthermore, this ambitious agenda provides ample mentorship opportunities for undergraduates and master's students interested in getting started with NLP and computational social science research.

Generalizable models that can flexibly incorporate diverse and meaningful theoretically-informed frames will serve as crucial tools to answer longstanding open questions across disciplines. For example, studies of the effects of a message's framing on its audience's emotions, attitudes, and opinions have been restricted to small-scale experiments in contrived settings. However, social media platforms and news websites provide insight into framing effects via interactive signals such as likes, shares, and comments. Sophisticated frame detection models will allow practitioners to leverage these signals in large-scale observational studies of how framing can influence people (Mendelsohn et al., 2020). Furthermore, the generalizability of FRAMEDETECT will facilitate unified analyses of framing across contexts. Political communication scholars, for example, are interested in how particular frames emerge, evolve over time, and how the public's framing of political issues online affects offline media coverage and politicians' attitudes (Russell Neuman et al., 2014). These questions are not fully answerable with existing methodologies, but can be easily addressed using FRAMEDETECT. To ensure that this work realizes its full potential beyond NLP, I plan to engage in interdisciplinary collaborations and publish in both NLP venues and social science journals.

Finally, this work has implications for societal benefit. Users' exposure to frames on social media can have immense consequences. FRAMEDETECT will help us better understand the content is spread throughout social networks and can guide social media platform policies surrounding major events, such as elections, as well as advertising regulations. Understanding framing can also inform social media algorithms; for example, a platform may seek to mitigate political polarization by recommending content to users that highlight a diversity of frames. My research can also be applied in the pursuit of social justice. Many frames about sociopolitical issues that center around marginalized groups could be harmful and prejudiced but are too subtle for explicit hate speech detection systems. FRAMEDETECT can be used to monitor such discourses in order to make online spaces safer and more welcoming for the most vulnerable members of our communities.

References

Belmonte, I. A. and Porto, M. D. (2020). Multimodal framing devices in european online news. *Language & Communication*, 71:55–71.

Benson, R. (2013). *Shaping immigration news*. Cambridge University Press Cambridge.

Boydstun, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2013). Identifying Media Frames and Frame Dynamics Within and Across Policy Issues. *New Directions in Analyzing Text as Data*, pages 1–13.

Cabot, P.-L. H., Dankers, V., Abadi, D., Fischer, A., and Shutova, E. (2020). The pragmatics behind politics: Modelling metaphor, framing and emotion in political discourse. *ACL Anthology*.

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. In *ACL-IJCNLP*, volume 2, pages 438–444.

Card, D., Gross, J. H., Boydstun, A. E., and Smith, N. A. (2016). Analyzing framing through the casts of characters in the news. In *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 1410–1420.

Ceylan, M. and Hayran, C. (2021). Message framing effects on individuals' social distancing and helping behavior during the covid-19 pandemic. *Frontiers in psychology*, 12:663.

Chang, H., Zhang, L., and Xie, G.-X. (2015). Message framing in green advertising: The effect of construal level and consumer environmental concern. *International Journal of Advertising*, 34(1):158–176.

Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D., and Lee, J. (2021). Melbert: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1763–1773.

Chong, D. and Druckman, J. (2007). Framing Theory. *Annual Review of Political Science*, 10(1):103–126.

Chou, W.-Y. S. and Budenz, A. (2020). Considering emotion in covid-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. *Health communication*, 35(14):1718–1722.

Crisp, P., Gibbs, R., Deignan, A., Low, G., Steen, G., Cameron, L., Semino, E., Grady, J., Cienki, A., Kövecses, Z., et al. (2007). Mip: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.

Cucchiara, C., Kwon, S., and Ha, S. (2015). Message framing and consumer responses to organic seafood labeling. *British Food Journal*.

Cunneen, M., Mullins, M., and Murphy, F. (2019). Autonomous vehicles and embedded artificial intelligence: The challenges of framing machine driving decisions. *Applied Artificial Intelligence*, 33(8):706–731.

Dankers, V., Rei, M., Lewis, M., and Shutova, E. (2019). Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2218–2229, Hong Kong, China. Association for Computational Linguistics.

de Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal*, 13(1):51–62.

Deslatte, A. et al. (2020). To shop or shelter? issue framing effects and social-distancing preferences in the covid-19 pandemic. *Journal of Behavioral Public Administration*, 3(1).

Entman, R. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Field, A., Kliger, D., Wintner, S., Pan, J., Jurafsky, D., and Tsvetkov, Y. (2018). Framing and Agenda-setting in Russian News: A computational analysis of intricate political strategies. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580.

Hartmann, M., Jansen, T., Augenstein, I., and Søgaard, A. (2019). Issue Framing in Online Discussion Fora. pages 1401–1407. Association for Computational Linguistics (ACL).

Heidenreich, T., Lind, F., Eberl, J.-M., and Boomgaarden, H. G. (2019). Media Framing Dynamics of the 'European Refugee Crisis': A Comparative Topic Modelling Approach. *Journal of Refugee Studies*, 32(Special_Issue_1):i172–i182.

Iyengar, S. (1991). *Is anyone responsible? How television frames political issues.* U. of Chicago Press.

Johnson, K., Jin, D., and Goldwasser, D. (2017). Modeling of Political Discourse Framing on Twitter. Technical report.

Khanehzar, S., Cohn, T., Mikolajczak, G., Turpin, A., and Frermann, L. (2021). Framing unpacked: A semi-supervised interpretable multi-view model of media frames. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2154–2166.

Khanehzar, S., Turpin, A., and Mikolajczak, G. (2019). Predicting Political Frames Across Policy Issues and Contexts. In *Proceedings of the 17th Workshop of the Australasian Language Technology Association*, pages 101–106.

Kiritchenko, S. and Mohammad, S. (2017). Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 465–470.

Kwak, H., An, J., and Ahn, Y.-Y. (2020). A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. *arXiv preprint arXiv:2005.01803*.

Kwak, H., An, J., Jing, E., and Ahn, Y.-Y. (2021). Frameaxis: characterizing microframe bias and intensity with word embedding. *PeerJ Computer Science*, 7:e644.

Li, X., Yin, X., Li, C., Hu, X., Zhang, P., Zhang, L., Wang, L., Hu, H., Dong, L., Wei, F., Choi, Y., and Gao, J. (2020). Oscar: Object-semantics aligned pre-training for vision-language tasks. *ECCV 2020*.

Liu, S., Guo, L., Mays, K., Betke, M., and Wijaya, D. T. (2019a). Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019b). Roberta: A robustly optimized bert pretraining approach. *arXiv*.

Lucy, L. and Mendelsohn, J. (2019). Using sentiment induction to understand variation in gendered online communities. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, pages 156–166.

Marshall, S. R. and Shapiro, J. R. (2018). When "Scurry" vs. "Hurry" Makes the Difference: Vermin Metaphors, Disgust, and Anti-Immigrant Attitudes. *Journal of Social Issues*, 74(4):774–789.

Mendelsohn, J., Budak, C., and Jurgens, D. (2021). Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263.

Mendelsohn, J., Tsvetkov, Y., and Jurafsky, D. (2020). A framework for the computational linguistic analysis of dehumanization. *Frontiers in artificial intelligence*, 3:55.

Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R., and Liu, H. (2018). Identifying Framing Bias in Online News. *ACM Transactions on Social Computing*, 1(2):1–18.

Naderi, N. and Hirst, G. (2017). Classifying frames at the sentence level in news articles. *International Conference Recent Advances in Natural Language Processing, RANLP*, 2017-Septe:536–542.

Powell, T. E., Boomgaarden, H. G., De Swert, K., and de Vreese, C. H. (2019). Framing fast and slow: A dual processing account of multimodal framing effects. *Media Psychology*, 22(4):572–600.

Reczek, R. W., Trudel, R., and White, K. (2018). Focusing on the forest or the trees: How abstract versus concrete construal level predicts responses to eco-friendly products. *Journal of environmental psychology*, 57:87–98.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Russell Neuman, W., Guggenheim, L., Mo Jang, S., and Bae, S. Y. (2014). The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication*, 64(2):193–214.

Scheufele, D. A. and Iyengar, S. (2012). The state of framing research: A call for new directions. *The Oxford handbook of political communication theories*, pages 1–26.

Schifanella, R., de Juan, P., Tetreault, J., and Cao, L. (2016). Detecting sarcasm in multimodal social platforms. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1136–1145.

Schmidt, S., Langner, S., Hennigs, N., Wiedmann, K.-P., Karampournioti, E., and Lischka, G. (2017). The green brand: Explicit and implicit framing effects of ecolabelling on brand knowledge. *Cogent Psychology*, 4(1):1329191.

Semetko, H. A. and Valkenburg, P. M. (2000). Framing european politics: A content analysis of press and television news. *Journal of communication*, 50(2):93–109.

Thibodeau, P. H. and Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PloS one*, 6(2):e16782.

Tong, X., Shutova, E., and Lewis, M. (2021). Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4673–4686, Online. Association for Computational Linguistics.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.

Van Der Linden, C. and Savoie, J. (2020). Does collective interest or self-interest motivate mask usage as a preventive measure against covid-19? *Canadian Journal of Political Science/Revue canadienne de science politique*, 53(2):391–397.

Van Stee, S. K. (2018). Meta-analysis of the persuasive effects of metaphorical vs. literal messages. *Communication Studies*, 69(5):545–566.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., and Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In Bach, F. and Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2048–2057, Lille, France. PMLR.

Zhang, P., Li, X., Hu, X., Yang, J., Zhang, L., Wang, L., Choi, Y., and Gao, J. (2021). Vinvl: Making visual representations matter in vision-language models. *CVPR 2021*.