Multimodal Propaganda Processing

Vincent Ng and Shengjie Li
Human Language Technology Research Institute
University of Texas at Dallas
Richardson, TX 75083-0688
{vince,sxl180006}@hlt.utdallas.edu

Abstract

Propaganda campaigns have long been used to influence public opinion via disseminating biased and/or misleading information. Despite the increasing prevalence of propaganda content on the Internet, few attempts have been made by AI researchers to analyze such content. We introduce the task of multimodal propaganda processing, where the goal is to automatically analyze propaganda content. We believe that this task presents a long-term challenge to AI researchers and that successful processing of propaganda could bring machine understanding one important step closer to human understanding. We discuss the technical challenges associated with this task and outline the steps that need to be taken to address it.

1 Introduction

Since the beginning of this century, significant progress has been made in the area of Sentiment Analysis and Opinion Mining on processing opinionated documents. Recent years have seen a surge of interest in processing a particular type of opinionated documents: persuasive documents. Work in this area is typically done under the umbrella of Argument Mining, in which the core task is to uncover the argumentative structure of a persuasive document. Specifically, the goal is to (1) identify the main claim, the claims, and the premises (i.e., supporting evidences) expressed in the given document, and (2) determine the relationships among them (e.g., identify which premises support which claim).

Work on argument mining has so far focused on processing legal text (Moens et al. 2007; Wyner et al. 2010; Walker et al. 2018), persuasive student essays (Persing and Ng 2016; Stab and Gurevych 2017), and Oxford-style debates (Orbach et al. 2020; Slonim et al. 2021). Although persuasive in nature, propagandistic articles (i.e., articles that aim to influence public opinion via disseminating biased and/or misleading information) have received relatively little attention in Natural Language Processing (NLP). This is somewhat surprising given the growing prevalence of *computational propaganda*, an "emergent form of political manipulation that occurs over the Internet" (Woolley and Howard 2018). From a research perspective, automatic processing of propaganda presents a number of challenges to AI researchers:

Multimodality. One characteristic that distinguishes propaganda content from other persuasive texts is that the former is often multimodal, comprising both text and images. As the saying goes, a picture is worth a thousand words. In multimodal propaganda, it is often the images that are most eye-catching and which create the biggest psychological impact on the reader. Although the text usually plays a supporting role, there are many cases where the image(s) could not be understood properly without the supporting text. How to combine the information derived from the two modalities to properly understand propaganda is an open question.

Deep understanding of text and images. Propaganda processing takes argument mining to the next level of complexity. As noted above, argument mining involves (1) extracting the claims and premises from the associated text and (2) identifying the relationships (e.g., support, attack) among them. For the kind of texts that NLP researchers have focused on so far (e.g., legal text, Oxford-style debates), the claims and premises are typically clearly stated. In contrast, the main claims and possibly some other supporting claims in propagandistic articles are often intentionally omitted, so we are faced with the additional challenge of recovering these hidden messages. Moreover, while the arguments in legal texts, essays, and debates can largely be interpreted literally, we often have to read between the lines when interpreting the text and images in propaganda content. For example, when given a picture of Russian soldiers killing Ukrainian civilians, current Computer Vision (CV) technologies would be able to produce a caption about this killing event, but if this picture appears in propagandistic articles, we probably need to infer the motive behind this picture (e.g., gaining the world's sympathy and support for Ukraine), which is currently beyond the reach of today's technology.

The need for background knowledge. Historical or cultural background knowledge may be needed to properly process propaganda content. For instance, given a propagandistic article with a picture showing a Palestinian social unrest event in the West Bank, the author may want to instill fear among the Israelis. However, without the knowledge of the long-standing conflict between the Palestinians and the Israelis, one may not be able to understand the author's intent.

Persuasion by deception. As noted above, argument mining researchers have focused on processing legal text, essays, and debates, where virtually all claims are established

Figure 1: An advertisement published by the White Motor Company in 1965.

using persuasion strategies like logos (i.e., through logical reasoning), pathos (i.e., through an emotional appeal), and ethos (i.e., through the speaker's authority or status). In contrast, the persuasion strategies used in propaganda are more sophisticated, often involving logical fallacies and framing.

Automatic processing of propaganda content could have important societal ramifications. In many cases, people are not aware that they are being brainwashed by propaganda campaigns, and this could lead to life-threatening consequences. One of the most compelling examples would be the ISIS propaganda and recruitment in 2014 (Farwell 2014), in which ISIS successfully recruited many people from all over the world, particularly those from the European Union, to go to Syria to serve as soldiers and sex slaves. Within the U.S. propaganda is typically manifested in the form of political manipulation campaigns with the goal of swaying public opinion. In fact, political manipulation campaigns have doubled since 2017 (Bradshaw and Howard 2018), and increased efforts of disinformation should be expected as the U.S. midterm elections draw near (DiResta et al. 2018).

In this paper, we present the novel task of multimodal propaganda processing, where the goal is to analyze propaganda content by extracting relevant information from the different modalities present in the content, identifying the persuasion devices and tactics that are used in different portions of the content, and eventually generating the message(s) being conveyed. We believe that time is ripe for AI researchers to work on this task. From a societal perspective, given the increasing influence that propaganda content has on our daily lives, it is more important than ever for us to be able to understand propagandistic articles. From an AI perspective, deep learning technologies have enabled revolutionary advances in machine understanding. It is time to examine how robust these technologies are when applied to a task as challenging as multimodal propaganda processing.

2 Examples

In this section, we explain why multimodal propaganda processing is interesting and challenging via two examples.

Example 1

Figure 1 presents an advertisement published by the White Motor Company in 1965 aiming to establish the superiority of the American truck transportation road networks to their Soviet counterparts. The advertisement shows on the left side the stone-surfaced roads being used for transportation in the USSR and the paucity of paved roads in the country, and it shows on the right side the modern highways along with a map of the USA that is full of road networks. The text states that (1) the distribution system by trucks was only one of the examples where the USA was superior to the USSR, and (2) the restrictions imposed by many states in the USA concerning truck sizes prevented motor truck transportation from operating to its full potential.

Human perception. Human readers need to possess certain knowledge in order to process the propaganda content in this advertisement. First, they have to have some geology knowledge to discover that the two maps are at different scales, which make the USA look larger than the USSR in terms of land mass. Second, they need to be aware of the oversimplifying language "SOVIET WAY - USA WAY", which implies that the Soviets adopted the American way of transportation. No evidence was provided to substantiate this claim, however. Third, they need to pay attention to the deceptive language in the text. While the advertisement contains a road map of the USSR and an estimated length of paved road in the country, the sources were never given and hence the information could be far from accurate.

Several propaganda devices and tactics are involved in this advertisement (see Section 3 for the list of devices and

Figure 2: Some pages of the *En Guardia* magazine. Figure 2(b) covers two pages.

tactics). First, the large font size associated with the title and the subtitle as well as the oversimplifying language in them signal the use of the *Binary Reduction* tactic, which employs the false-dilemma logical fallacy. Second, the use of language to depict that the Americans are superior in many ways (e.g., "Our higher standard of living", "for all the superiority of our American highways", etc.) signals the *All-encompassing* tactic, which is a sort of rhetoric that often appears as window dressing for a larger point. Third, the sentence "America has the advantage of maximum efficiency and economy" employs the *Cultural Signaling* tactic, calling on America's values of efficiency and success. Finally, the text gives "best estimates" of the USSR's road network, which is a case of the *Card Stacking* device, where only partial facts are used to defend a statement.

Challenges. It is non-trivial to automatically process the propaganda content in this advertisement. The key challenges stem from the need to (1) process multimodal information extracted from the images and the text; (2) exploit background knowledge to unveil hidden information; (3) interpret the use of different font styles and text sizes to highlight specific pieces of information; and (4) understand the hidden information conveyed in the images (e.g., the difference in scaling between the two maps).

Example 2

During World War II, the U.S. propagandists sought support from the Latin Americas by publishing a high-quality Spanish periodical *En Guardia*. Figure 2 shows six pages from the first issue of *En Guardia*. Each page has its own propaganda messaging. Figure 2(a) is the cover of the magazine, which shows two naval ships moving fast in the ocean. The title "En Guardia" and the subtitle "Para la defensa de las Américas" translate into "On Guard" and "For the defense of the Americas" respectively. Figure 2(b) shows a naval ship in the ocean with a scope pointing at the ship. The main points in the caption of 2(b) translate into "The American navy must and will keep the seas free, and will protect merchant ships against the danger of bombardment". Figure 2(c) shows a merchant ship and discusses the importance of merchant ships in delivering goods and troops to all parts of the Americas. The boldfaced sub-caption translates into "America means 21 nations". Figure 2(d) shows a ship while Figure 2(e) focuses on the training and the sheer size of the U.S. Navy. Most importantly, these pages need to be considered as a *sequence* in order to obtain the full messaging, which is that "Maritime commerce in the Americas is under imminent threat, and protecting the oceans from the enemy is vital to western hemisphere interests. However, the U.S. navy has the best equipment and personnel to deal with such a threat."

Human perception. A human reader with the appropriate background would interpret these pages as follows. First, the cover points out the main theme of the magazine by using the eye-catching subtitle "For the defense of the Americas", which proposes a sense of shared identity and immediate danger. The warship depicted on the cover is moving fast in the ocean, as the water below it is splashing high, suggesting something is happening immediately. Second, Figure 2(b) shows an exaggerated scaling of a gun scope on the left and a ship that is being pointed to by the scope on the right. This would naturally take all the attention of the reader. Third, Figure 2(c) repeats the sense of a shared identity by saying "America means 21 nations". Finally, Figures 2(d) and 2(e) show that the U.S. has the equipment and the personnel to deal with the danger threatening free commerce.

Next, we analyze the propaganda devices and tactics used in these pages. Figures 2(a), 2(b), and 2(d) all use the *Visual Scaling* tactic, which is concerned with evoking emotional understanding (e.g., fear, power, etc.) by using images. These images also use the *Card Stacking* device, as they do not explicitly point out who is threatening the Americas. Figure 2(c) uses two devices: (1) *Band Wagon*, which implies that all of the countries in the Western Hemisphere are a collective and should work together; and (2) *Glittering Generalities*, where a "virtue word" (in this case, the impressiveness of U.S. merchant ships) is being used to create positive emotion and acceptance (of the U.S. military involvement) without examination of evidence.

Challenges. Automatically processing propaganda in this example is even harder than that in the first example since proper understanding depends heavily on visual clues rather than textual information. For instance, a machine needs to understand that (1) in Figure 2(b), the scope was enlarged to an exaggerated size and was pointing at the ship; (2) in Figure 2(d), the picture was taken at an angle that makes the ship look substantially larger than other objects in the background, with the intent of showing off the military might of the U.S.; and (3) these images need to be considered as a *sequence* in order to get the full messaging.

3 Corpus Creation and Annotation

In this section, we outline the initial steps needed to address the task of multimodal propaganda processing.

Corpus Creation

Given the recent advances in CV and NLP, we propose to approach this problem by building a machine learning (ML) system. Appropriately annotated corpora are critical to the successful application of any ML systems. Since the goal of the multimodal propaganda processing task is to analyze propaganda content, we need data instances that correspond to examples of propaganda. Since we do not have a system for automatically identifying propaganda content, it would be best for us to begin data collection by looking for websites or publications that are known to publish propagandistic materials. A possible source of historical propaganda would be the *En Guardia* magazine described in the previous section. So far, we have applied OCR to every page of every issue of this magazine and have used these articles to assemble the first version of our corpus.

In order to assemble a corpus that contains contemporary propagandistic articles, we propose to exploit the content published on some fact-checking websites. For example, Politifact[1] verifies the accuracy of claims made by elected officials. Those claims that were marked as inaccurate would constitute a good set of candidates of propagandistic articles. We can then manually go through these candidates to identify propaganda content. Similar websites include Full Fact[2], FactCheck[3], and Media Bias/Fact Check[4].

Annotation Tasks

Next, we define a set of annotation tasks that we believe would be helpful for analyzing propaganda content. The annotations we obtain via these tasks will provide the data needed to train models for processing propaganda content.

Task 1: Propaganda device and tactic detection and rationale generation. The first task concerns identifying the propaganda device(s) and tactic(s) used in propaganda content. The "Seven Propaganda Devices" (Childs 1936; Sproule 2001), a well-known propaganda theory, defines seven propaganda devices that represent the seven persuasion strategies commonly used in propaganda, including: *Band Wagon*, *Card Stacking*, *Glittering Generalities*, *Name-Calling*, *Plain Folks*, *Testimonial*, and *Transfer*.[5] We identify the device(s) used in the text portion and the image portion of the multimodal input separately. Note that more than one device can be used for a given piece of text or image.

In addition, we extend our annotation scheme by including a set of propaganda *tactics*, which serve to underscore the methods of employing the devices. The set of tactics we have identified include: *Extremism*, *All-encompassing*, *Repetition*, *Visual scaling*, *Binary Reduction*, *Cultural Signaling*, *Prestige Signaling*, *Pandering*, and *Innuendo*.

Finally, we provide the rationale behind each device label and each tactic label we assign to the given propaganda content. A rationale is a natural language description of why the corresponding device/tactic label is assigned based on the information extracted from the input. As we will see, rationale generation could improve model interpretability.

Task 2: Domain-independent message detection with rationales. Inspired by existing propaganda theory regarding the content of an article (Ellul 1973; Altheide and Johnson 1980), we detect the types of the (possibly hidden) messages conveyed by the author in our second task. The messaging in this task is domain-independent and attempts to communicate a broad idea to provoke an emotional reaction. An initial set of message types we identified includes: *Might*, *Fear*, *Inspiration*, *Urgency*, *Unity*, *Teamwork*, *Patriotism*, *Superiority*, *Abundance*, *Reciprocity*, *Sacrifice*, *Masculinity*, *Ingenuity*, *Virtue*, *Progress*, *Security*, *Reassurance*, *Fun*, and *Sameness*. We expect this list to grow as we identify additional types. Multiple message types may be applicable to a given propaganda content. As in Task 1, in this task the rationale behind each annotated message will be annotated.

Task 3: Domain-specific message detection with rationales. Like Task 2, Task 3 also concerns identifying the types of messages conveyed by the author, but the message types in this task are domain-specific and therefore would need to be redesigned for each new domain. For wartime propaganda such as those that appear in *En Guardia*, the message types could include: *Military Strength*, *Industrial Production*, *US-Latin American Cooperation*, *US culture*, *US Leadership*, *WW2*, *Pan-Americanism*, *War preparation*, *Economic Interests*, *Gendered messaging*, *Civilian contributions*, and *Common Culture*. Again, multiple message types may be applicable to a given propaganda content. As in Tasks 1 and 2, in this task the rationale behind each annotated message will be annotated.

Task 4: Main message generation with rationales. This task concerns generating the main message conveyed by the author in natural language. As in the first three tasks, here the rationale behind the main message will be annotated.

Task 5: Background knowledge. As noted before, background knowledge may be needed to properly interpret propaganda content. The background knowledge needed will be annotated in the form of natural language sentences.

Task 6: Image captioning. Existing image encoders may fail to encode all the details of an image, particularly when the image contains abstract concepts. To mitigate the difficulty of accurately extracting information from images, we propose an auxiliary task, image captioning, where we annotate the information present in an image in natural language so that the resulting caption is an equivalent textual representation of the image. With these annotations, we can train a model to first caption an image and use the resulting caption in lieu of the original image for further processing.

Sample Annotation

We propose to annotate each propagandistic article in the form of an *argument tree*, which is the representation used by argument mining researchers to represent the argumentative structure of a persuasive document (Stab and Gurevych 2014). In an argument tree, the root node corresponds to the main claim of the document, and each child of a node corresponds to a piece of supporting evidence (which can be a claim or a premise) for the parent. In other words, each edge

[1]https://www.politifact.com/
[2]https://fullfact.org/
[3]https://www.factcheck.org/
[4]https://mediabiasfactcheck.com/
[5]Other propaganda theories can also be used.

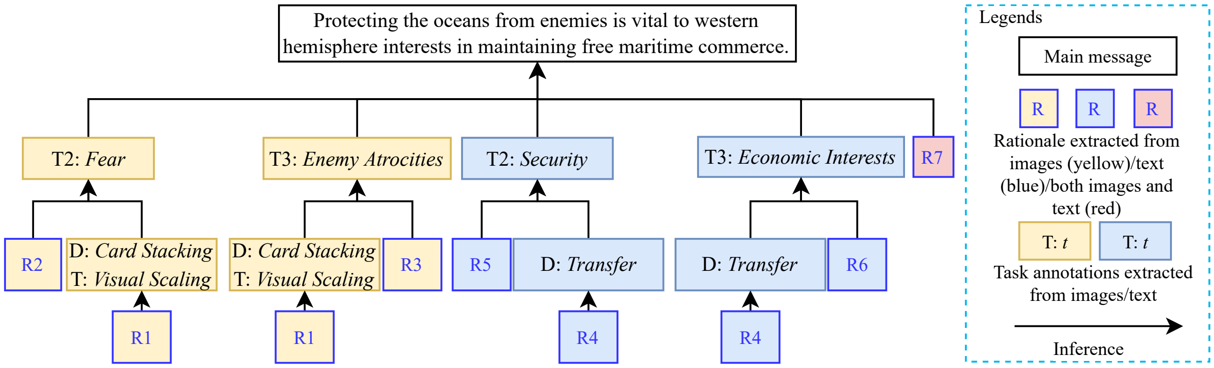

Figure 3: Sample argument tree for the input taken from Figure 2(b). Given this figure, a possible caption generated for Task 6 is "A ship is sailing on the sea, while a large gun scope is pointing at it". R1: A gun scope is pointing at a ship, which creates a sense of danger, hence *Visual Scaling*. However, who is holding the gun is unclear (partial information), hence *Card Stacking*. R2: A sense of *Fear* created by the gun scope. R3: the image depicts a (potential) crime of bombardment, hence *Enemy Atrocities*. R4: Use the idea of the U.S. protecting maritime commerce to justify other U.S. military involvements, hence *Transfer*. R5: "La armada norteamericana ... protegerá a los buques mercantes" means "the American army will protect merchant vessels", which suggests a sense of *Security*. R6: "Los océanos no son barreras, sino las amplias rutas del comercio mundial" means "the oceans are the routes of world trades", which focuses on *Economic Interests*. R7: Since the image suggests threats to maritime trades, while the text suggests the U.S. will protect maritime trades, we can get the overall message by combining them.

denotes a support relation. A leaf node always corresponds to a premise, which by definition does not need any support.

To enable the reader to understand how to annotate a propagandistic article as an argument tree, we show in Figure 3 the argument tree that should be produced for input taken from Figure 2(b). As we can see, the root node contains the main message (see Task 4). It has five children, which implies that it is supported by five pieces of evidence, including the domain-independent and domain-dependent messages derived from the image, the domain-independent and domain-dependent messages derived from the text and the rationale associated with the main message, which is also derived from the text. For each of the first four children, there are two child nodes, one corresponding to its rationale and the other corresponding to the device(s) and tactic(s) used. The fifth child, which is a rationale, is a leaf node. Note that a rationale always appears in a leaf node, the reason being that rationales are derived directly from either the image or the text (or both) and therefore do not need any support. The remaining nodes in the tree can be interpreted in a similar fashion. Note that the annotations for Tasks 1–5 will always appear as nodes in the tree.

4 Models

Given a dataset annotated using our annotation scheme, we can train a model to perform the six annotation tasks. Given the recent successes of neural models in NLP, we propose to employ neural models for our task. As a first step, we can employ existing models and design new models for this task if needed. There are several considerations.

Multimodal vs. unimodal models. Since our input is multimodal and composed of text and image(s), it would be natural to train a multimodal model assuming three inputs: two of them correspond to the two modalities and the remaining one encodes the background knowledge base (assembled using the background knowledge annotated for each training instance, for example). The images can be encoded using a visual encoder such as ResNet (He et al. 2016) and ViLBERT (Lu et al. 2019), whereas the text inputs (including the background knowledge base) could be encoded using a neural encoder such as SpanBERT (Joshi et al. 2020). The outputs from the encoders can then be concatenated together for further processing.

Alternatively, one can employ a unimodal model where we caption the image first (Task 6) with the help of an object detection system (e.g., YOLO (Redmon et al. 2016)) and possibly an off-the-shelf image captioning system (e.g., X-Transformer (Pan et al. 2020), VinVL (Zhang et al. 2021)). As noted before, the caption is supposed to be an equivalent textual representation of the corresponding image. The caption can then be encoded by a text encoder, and the resulting representation can be concatenated with the encoded outputs from the text side for further processing.

Joint vs. pipeline models. Should we adopt a pipeline architecture where we first train a model for each task independently of the others and then apply the resulting models in a pipeline fashion? For instance, given multimodal propagandistic articles, we can first train a model to caption the image (Task 6), as described above. After that, we can train a model to identify the device(s) and another model to identify the tactic(s) (Task 1). To improve model interpretability, the rationales can be predicted jointly with the corresponding device(s)/tactic(s). We can similarly train models to predict the domain-independent and domain-specific labels (Tasks 2 and 3) jointly with their rationales by using all of the available information predicted so far (e.g., the tactics and devices, the caption). Finally, we can train a model to predict the main message (Task 4).

Recall that pipeline models are prone to error propagation, where errors made by an upstream model will propagate to a downstream model. To mitigate error propagation, we can consider joint models. Specifically, we can train *one* model to perform all of the six tasks jointly. Joint models allow the different tasks involved to benefit each other via a shared input representation layer. The major downside of a joint model is that the resulting network (and hence the corresponding learning task) is typically very complex.

Pre-trained models. A key challenge in the automatic processing of propaganda is the need for background knowledge. While we have access to background knowledge through the manual annotations obtained as part of Task 5, it is conceivable that the amount of background knowledge we need will far exceed what these annotations can provide. A potential solution to this background knowledge acquisition bottleneck is *pre-training*. More specifically, researchers in NLP have shown that a vast amount of general knowledge about language, including both linguistic and commonsense knowledge, can be acquired by (pre-)training a language model in a *task-agnostic* manner using *self-supervised* learning tasks. Self-supervised learning tasks are NLP tasks for which the label associated with a training instance can be derived automatically from the text itself.[6] Because no human annotation is needed, a language model can be pre-trained on a large amount of labeled data that can be automatically generated, thereby acquiring a potentially vast amount of knowledge about language. Many pre-trained language models have been developed and widely used in NLP, such as BERT (Devlin et al. 2019), XLNet (Yang et al. 2019), RoBERTa (Liu et al. 2019), ELECTRA (Clark et al. 2020), GPT-2 (Radford et al. 2019), T5 (Raffel et al. 2020), and BART (Lewis et al. 2020). These models have been shown to offer considerable improvements on a variety of NLP tasks.

To acquire the background knowledge needed for processing the articles in *En Guardia*, for instance, we can pre-train a language model on as many *unannotated* articles in *En Guardia* as we can collect. The resulting model can then be optimized for a specific task by *fine-tuning* its parameters using the task-specific labeled data we obtained via our annotation process in the standard supervised fashion. While there has been a lot of work on developing pre-trained language models, the development of *multimodal* pre-trained models that can understand both text and images, which is what we need for multimodal propaganda processing, is a relatively unexplored area of research.

5 Related Work

Memes. Memes are user-created pictures, often accompanied by text, that are used to express opinions on all kinds of topics. Similar to propaganda messaging, memes typically require background knowledge for proper interpretation. Memes are widely used in political manipulation campaigns as a tool for conveying propaganda messaging (Farwell 2014; Forelle et al. 2015; Bradshaw and Howard 2018; DiResta et al. 2018). Hence, unveiling hidden information from memes is highly related to processing propaganda messaging from images and text. There has been recent work that aims to build a model to detect a rich set of propaganda techniques in memes (Dimitrov et al. 2021).

Document-level unimodal misinformation analysis. Several publicly-available datasets are composed of news articles labeled with whether they contain misinformation. For example, in the `TSHP-17` dataset (Rashkin et al. 2017), each article is labeled with one of four classes: *trusted*, *satire*, *hoax*, and *propaganda*, whereas in the `QProp` dataset (Barrón-Cedeño et al. 2019), only two labels are used: *propaganda* and *non-propaganda*. Da San Martino et al. (2019), on the other hand, develop a corpus of news articles labeled with the propaganda techniques used. Their corpus enables the study of multi-label multi-class classification task in propaganda identification.

Multimodal misinformation classification. Some researchers have examined the task of multimodal propaganda identification. For instance, Volkova et al. (2019) construct a dataset consisting of 500,000 Twitter posts classified into six categories (*disinformation*, *propaganda*, *hoaxes*, *conspiracies*, *clickbait*, and *satire*) and build models to detect misleading information in images and text. While this attempt seeks to perform a shallow analysis of tweets, we propose to perform a deep analysis of propaganda content, which would lead to the generation of the hidden messages conveyed.

6 Concluding Remarks

We presented the task of multimodal propaganda processing and discussed the key challenges. Below we conclude with several other issues that are also relevant to the task.

Propaganda identification. While we have focused on analyzing propaganda content, it is equally important to identify such content. Although we did not explicitly discuss how such content can be identified, a system that can analyze propaganda content could also be used for identifying such content. More specifically, if the system determines that no persuasion devices and tactics were being used in the given content, it could imply that the content is not propagandistic. Another possibility would be to train a model to distinguish propaganda content from non-propaganda content on our corpus of propaganda articles and other non-propaganda articles collected from the Internet.

Domain transferability. Since the models described thus far are trained on domain-specific annotations (i.e., the background knowledge from Task 5 and the domain-specific labels from Task 4), they are necessarily domain-specific. To facilitate their application to a new domain, especially when labeled training data in the new domain is scarce, we can explore domain adaptation techniques.

Ethical considerations. Care should be taken to ensure that propaganda processing technologies would not be misused by people to attack their political opponents by intentionally using a propaganda processing system to draw wrong conclusions or generate propaganda content aiming to achieve their personal agenda, for instance.

[6] A well-known self-supervised learning task is Masked Language Modeling (MLM) (Devlin et al. 2019). Given a sequence of word tokens in which a certain percentage of tokens is *masked* randomly, the goal of MLM is to predict the masked tokens.

References

Altheide, D. L.; and Johnson, J. M. 1980. *Bureaucratic Propaganda*. Allyn and Bacon.

Barrón-Cedeño, A.; Jaradat, I.; Da San Martino, G.; and Nakov, P. 2019. Proppy: Organizing the News Based on Their Propagandistic Content. *Information Processing & Management*, 56(5): 1849–1864.

Bradshaw, S.; and Howard, P. N. 2018. Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. Technical report, Working Paper 2018.1. Oxford, UK: The Computational Propaganda Project.

Childs, H. L. 1936. Propaganda; Its Psychology and Technique. By Leonard W. Doob. (New York: Henry Holt and Company. 1935. Pp. 417.). *American Political Science Review*, 30(2): 389–390.

Clark, K.; Luong, M.; Le, Q. V.; and Manning, C. D. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Da San Martino, G.; Yu, S.; Barrón-Cedeño, A.; Petrov, R.; and Nakov, P. 2019. Fine-Grained Analysis of Propaganda in News Article. 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)*, 5636–5646. Hong Kong, China: Association for Computational Linguistics.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Dimitrov, D.; Bin Ali, B.; Shaar, S.; Alam, F.; Silvestri, F.; Firooz, H.; Nakov, P.; and Da San Martino, G. 2021. Detecting Propaganda Techniques in Memes. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 6603–6617. Online: Association for Computational Linguistics.

DiResta, R.; Shaffer, K.; Ruppel, B.; Sullivan, D.; Matney, R.; Fox, R.; Albright, J.; and Johnson, B. 2018. The Tactics & Tropes of the Internet Research Agency. https://apo.org.au/node/211296. Accessed: 2019-09-14.

Ellul, J. 1973. *Propaganda: The Formation of Men's Attitudes*. New York, NY: Vintage Books.

Farwell, J. P. 2014. The Media Strategy of ISIS. *Survival*, 56(6): 49–55.

Forelle, M.; Howard, P.; Monroy-Hernández, A.; and Savage, S. 2015. Political Bots and the Manipulation of Public Opinion in Venezuela. *SSRN Electronic Journal*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. Las Vegas, NV, USA.

Joshi, M.; Chen, D.; Liu, Y.; Weld, D. S.; Zettlemoyer, L.; and Levy, O. 2020. SpanBERT: Improving Pre-training by Representing and Predicting Spans. *Transactions of the Association for Computational Linguistics*, 8: 64–77.

Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; and Zettlemoyer, L. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871–7880. Online: Association for Computational Linguistics.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *Computing Research Repository, arXiv:1907.11692*.

Lu, J.; Batra, D.; Parikh, D.; and Lee, S. 2019. ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks. In *Advances in Neural Information Processing Systems 32*, 13–23. Vancouver, Canada.

Moens, M.-F.; Boiy, E.; Palau, R. M.; and Reed, C. 2007. Automatic Detection of Arguments in Legal Texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, ICAIL '07, 225–230. New York, NY, USA: Association for Computing Machinery.

Orbach, M.; Bilu, Y.; Toledo, A.; Lahav, D.; Jacovi, M.; Aharonov, R.; and Slonim, N. 2020. Out of the Echo Chamber: Detecting Countering Debate Speeches. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7073–7086. Online: Association for Computational Linguistics.

Pan, Y.; Yao, T.; Li, Y.; and Mei, T. 2020. X-Linear Attention Networks for Image Captioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10968–10977. Seattle, WA, USA.

Persing, I.; and Ng, V. 2016. End-to-End Argumentation Mining in Student Essays. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1384–1394. San Diego, California: Association for Computational Linguistics.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners. https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf. Accessed: 2022-09-10.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21: 140:1–140:67.

Rashkin, H.; Choi, E.; Jang, J. Y.; Volkova, S.; and Choi, Y. 2017. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of* *the 2017 Conference on Empirical Methods in Natural Language Processing*, 2931–2937. Copenhagen, Denmark: Association for Computational Linguistics.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. Las Vegas, NV, USA.

Slonim, N.; Bilu, Y.; Alzate, C.; Bar-Haim, R.; Bogin, B.; Bonin, F.; Choshen, L.; Cohen-Karlik, E.; Dankin, L.; Edelstein, L.; Ein-Dor, L.; Friedman-Melamed, R.; Gavron, A.; Gera, A.; Gleize, M.; Gretz, S.; Gutfreund, D.; Halfon, A.; Hershcovich, D.; Hoory, R.; Hou, Y.; Hummel, S.; Jacovi, M.; Jochim, C.; Kantor, Y.; Katz, Y.; Konopnicki, D.; Kons, Z.; Kotlerman, L.; Krieger, D.; Lahav, D.; Lavee, T.; Levy, R.; Liberman, N.; Mass, Y.; Menczel, A.; Mirkin, S.; Moshkowich, G.; Ofek-Koifman, S.; Orbach, M.; Rabinovich, E.; Rinott, R.; Shechtman, S.; Sheinwald, D.; Shnarch, E.; Shnayderman, I.; Soffer, A.; Spector, A.; Sznajder, B.; Toledo, A.; Toledo-Ronen, O.; Venezian, E.; and Aharonov, R. 2021. An Autonomous Debating system. *Nature*, 591(7850): 379–384.

Sproule, J. 2001. Authorship and Origins of the Seven Propaganda Devices: A Research Note. *Rhetoric & Public Affairs*, 4: 135–143.

Stab, C.; and Gurevych, I. 2014. Annotating Argument Components and Relations in Persuasive Essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1501–1510. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Stab, C.; and Gurevych, I. 2017. Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, 43(3): 619–659.

Volkova, S.; Ayton, E.; Arendt, D. L.; Huang, Z.; and Hutchinson, B. 2019. Explaining Multimodal Deceptive News Prediction Models. In *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media*, 659–662. Munich, Germany: AAAI Press.

Walker, V. R.; Foerster, D.; Ponce, J. M.; and Rosen, M. 2018. Evidence Types, Credibility Factors, and Patterns or Soft Rules for Weighing Conflicting Evidence: Argument Mining in the Context of Legal Rules Governing Evidence Assessment. In *Proceedings of the 5th Workshop on Argument Mining*, 68–78. Brussels, Belgium: Association for Computational Linguistics.

Woolley, S. C.; and Howard, P. N., eds. 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press.

Wyner, A.; Mochales-Palau, R.; Moens, M.-F.; and Milward, D. 2010. *Approaches to Text Mining Arguments from Legal Cases*, 60–79. Berlin, Heidelberg: Springer Berlin Heidelberg.

Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R. R.; and Le, Q. V. 2019. XLNET: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems 32*, 5754–5764. Vancouver, Canada.

Zhang, P.; Li, X.; Hu, X.; Yang, J.; Zhang, L.; Wang, L.; Choi, Y.; and Gao, J. 2021. VinVL: Revisiting Visual Representations in Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5579–5588. Virtual.