SANDIA REPORT
SAND2022-13730
Printed September 2022

Identifying Disinformation Using Rhetorical Devices in Natural Language Models

Katrina J Ward, Hamilton Link, Kiril Avramov, Jean Goodwin

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@osti.gov
Online ordering:	http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.gov
Online order:	https://classic.ntis.gov/help/order-methods/

NNSA
National Nuclear Security Administration

ABSTRACT

Foreign disinformation campaigns are strategically organized, extended efforts using disinformation – false or misleading information deliberately placed by an adversary – to achieve some goal. Disinformation campaigns pose severe threats to our nation's security by misinforming decision makers and negatively influencing their actions when they are operating on limited amounts of evidence. Current efforts rely on subject matter experts to manually identify disinformation [1] [2], or on computers and traditional natural language processing algorithms to identify patterns in data to calculate the probability that something is disinformation or not. While both have their merits and successes, subject matter experts are unable to keep up with the high volumes of global information and traditional natural language algorithms do not do well in identifying "why" something is disinformation or not. Our hypothesis is that we can identify disinformation by looking at the way someone speaks, in the rhetorical devices they use. We have curated and annotated a dataset designed for multiple natural language processing tasks, but specifically useful for disinformation detection algorithms.

ACKNOWLEDGEMENTS

Supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

We want to thank our university partners whose subject matter expertise has been critical in this work. Dr. Kiril Avramov is a professor at University of Texas at Austin in the Center for Russian, East European, and Eurasian studies and is the lead for the Global Disinformation Lab. He and his graduate students, Ryan Williams, Taylor Ham, and Erica Buckland provided data collection skills and expertise on disinformation in Russia media. Dr. Jean Goodwin is a professor of Rhetoric & Technical Communication at NC State University. She provided subject matter expertise on rhetoric and how it is represented and labeled in speech.

CONTENTS

This page left blank

EXECUTIVE SUMMARY

Foreign disinformation campaigns are organized efforts meant to deceive, disorganize, and misinform with the intention of causing confusion, doubt, and influencing decisions. Traditional natural language processing falls short of being able to detect these campaigns due to current limitations in natural language processing tools and lack of subject matter expertise to train these systems properly. We believe that we can use rhetorical devices, or the way in which people speak, to detect and classify disinformation. In this work, we leverage the expertise of our partners at UT Austin and NC State to collect, annotate, and analyze a robust dataset containing both disinformation and neutral articles. The initial findings of this data show that there are significantly more occurrences of certain rhetorical devices in disinformation articles than neutral articles. We believe this means that additional work to create models to detect these rhetorical devices and identify disinformation is a worthwhile investment. Additionally, the collected and annotated dataset has been prepared for distribution and use by other academic research efforts.

ACRONYMS AND TERMS

Acronym/Term	Definition
Rhetoric	The art of effective or persuasive speaking or writing
Rhetorical Device	A specific category of persuasive speaking or writing used to achieve a specific outcome
Register	A variety of language use characteristic of a particular context; e.g., cable news
Disinformation	False or misleading information deliberately placed by an adversary [3]
NLP	Natural Language Processing
NNS	Neutral News Source
RSM	Russian State Media
KA	Kremlin Amplifiers
ML	Machine Learning

1. INTRODUCTION

Foreign disinformation campaigns are strategically organized, extended efforts using disinformation, false or misleading information deliberately placed by an adversary, to achieve some goal. Disinformation campaigns pose severe threats to our nation's security by misinforming decision makers and negatively influencing their actions when they are operating on limited amounts of evidence. Hyperconnectivity has transformed disinformation into a serious ongoing threat to the US and its allies, prompting many national security entities to seek solutions to identify, measure, and reduce the impact of disinformation.

Though our approach will be designed to work on any foreign disinformation campaigns, for a first proof of concept we had to choose known disinformation campaigns that we can annotate and test our hypothesis. For this reason and due to the abundance of examples, the logic used in this effort's methodology is based on our understanding of the specifics and constitutive elements of the contemporary Russian disinformation system; this affects the narratives we have selected, how we have sought to collect our data, and the rhetorical tactics we have chosen to code in the data. The Kremlin uses their associated ecology to execute influence operations at home and abroad against its perceived adversaries. As specified in State Department's Global Engagement Center's (GEC) Counter-Disinformation Dispatches [4], one of the main goals of these influence operations is "to weaken their perceived adversaries' political, economic, scientific, and technical and military positions. To do this, Russia's influence operations use four main tactics or techniques:

- Discredit
- Divide
- Disarm
- Demoralize.

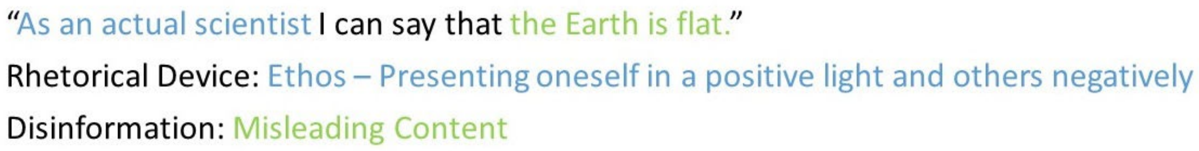
Russia tries to discredit adversaries to minimize their influence, divide targeted countries or groups to weaken them, disarm opponents and prevent them from mobilizing by downplaying the threat from Russia, and demoralize adversaries."

Based on this information, recent events, the availability of subject matter expertise, and the abundance of disinformation samples, our work focuses on Russian disinformation with the understanding and goal that the techniques developed will be applicable to any foreign disinformation campaign.

Previous work to detect and inform about disinformation have depended heavily on current natural language processing (NLP) techniques or human analyst forensics. [5] [6] While humans can more accurately detect and recognize disinformation, the proliferation of digital documents and the hyperconnectivity mentioned above make the task time consuming and nearly impossible to mitigate. On the other hand, traditional NLP approaches rely on attempts to match patterns in word sequences that may indicate disinformation and tend to perform poorly when wording changes, even though the devices used have not.

In this work, we hypothesize that, in sufficiently large groups, people's use of language will produce signals in publicly available digital media that can be harnessed to detect disinformation campaigns. For example, we as humans are adept at detecting conspiracy theory data that a computer, given enough data, will only understand the probability of the word sequences appearing together and not the context around them. The current state of the art struggles with linguistic nuances, adversarial tactics, and indirect consequences characteristic of disinformation campaigns. [7] [8] Specifically, we believe that the way people speak, or rhetorical devices, can indicate when disinformation is being used and how.

Rhetoric is persuasive discourse that is meant to sway, persuade, or convince a population towards a particular decision or opinion. Rhetoric itself is not malicious and is used in everyday conversation. However it is also used in malicious disinformation campaigns. Rhetoric can be broken down into categories, or rhetorical devices, where each category represents a specific purpose and way that purpose is represented in overt or covert ways. [9] We believe that specific rhetorical devices are present more frequently in particular ways when being used in disinformation campaigns when compared to neutral news sources or comments. In this work we focus on Doubt as a tactical goal of rhetoric to detect due to the abundance of use in both neutral and disinformation articles in our data. In the future, we believe we can extend our work to other types of rhetoric and provide a topology of rhetorical devices and how they are used when looking at disinformation campaigns.



"As an actual scientist I can say that the Earth is flat."

Rhetorical Device: Ethos – Presenting oneself in a positive light and others negatively

Disinformation: Misleading Content

Figure 1: Example of disinformation and rhetorical devices used.

Because disinformation campaigns vary widely in character, we have selected three Russian-attributed exemplars which will provide the data we need. These examples are events where known disinformation was used to negate or reduce the negative criticisms of Russia where data exists to contradict this disinformation. [10] In addition, given the number of disinformation types and the hundreds of rhetorical devices in use in disinformation, we focused on three categories in each to reduce our scope and form our proof of concept and decide on further work.

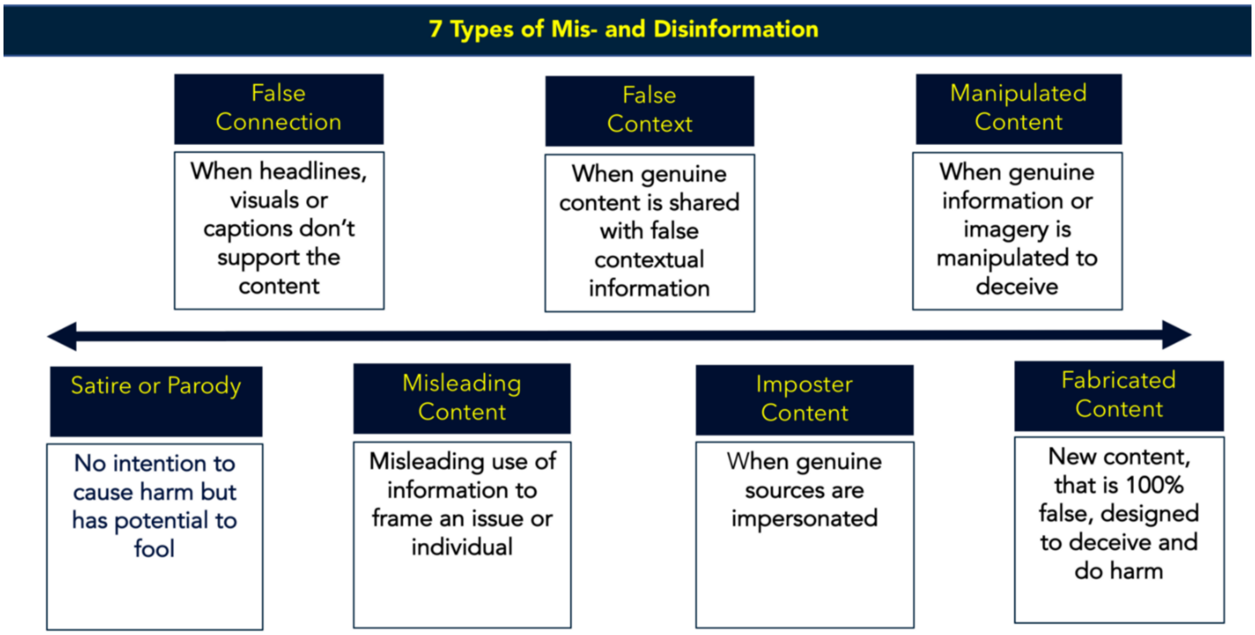
To approach our analysis, we require a data set with which we can train and test a model to detect disinformation. We worked with Dr. Kiril Avramov from the University of Texas at Austin as a subject matter expert on Russian media and disinformation. Dr. Avramov and his team curated both neutral and disinformation articles from our chosen events and provided context on the types of disinformation tactics in use for each article. To understand the language used in each article, we also worked with Dr. Jean Goodwin from NC State University who is a subject matter expert in linguistics and rhetorical communication. Dr. Goodwin and her team helped us understand the rhetorical devices in use in the curated data and provided annotation of our dataset to indicate where and what type of rhetorical devices were present. The data and annotations may be critical in training a model to detect these rhetorical devices and classify a document as disinformation or not. Additional information about the dataset will be covered below.

In this report we talk about disinformation and the types of disinformation we chose to focus on for this work. We also define rhetoric and the rhetorical devices that we felt were strong indicators in our data. We detail our data collection methodology and the way the data is annotated. We present how the data is saved and organized for later use. Finally, we provide some initial statistics of the data that show our theory about certain rhetorical devices being more prevalent in disinformation is true, warranting potential future research to continue.

2. DISINFORMATION AND RHETORICAL DEVICES

Disinformation is false information intended to mislead, deliberately spread through overt or covert means. Often disinformation is meant to blend in with real information to confuse the consumer of the information or lead them to make incorrect conclusions. Countries will often use disinformation campaigns to change the world's opinion and perspective on events to put themselves in a more favorable light or to turn the world community against another nation. These campaigns can target wide audiences at world or national levels or be targeted at smaller focus groups. For example, the United States has alleged that Russia and China have targeted disinformation campaigns at US citizens to sway US policy. [11] [12] At a more targeted level, we have seen Saudi Arabia, China, and others target individual dissidents to discredit their online presence and the information they share.

Disinformation itself can be broken down into approximately seven types, though this is not a hard and fast rule, and these types are often broken down further into more categories to better distinguish between disinformation definitions. [13]



7 Types of Mis- and Disinformation

False Connection: When headlines, visuals or captions don't support the content

False Context: When genuine content is shared with false contextual information

Manipulated Content: When genuine information or imagery is manipulated to deceive

Satire or Parody: No intention to cause harm but has potential to fool

Misleading Content: Misleading use of information to frame an issue or individual

Imposter Content: When genuine sources are impersonated

Fabricated Content: New content, that is 100% false, designed to deceive and do harm

Given the number of types of disinformation and in the interest of limiting scope for a proof-of-concept project, we chose three disinformation types that were observed most often in the type of data we are interested in.

- Manipulated Content
  Example: "Russian soldiers aid wounded in Ukraine" – Real image shows aid given to other Russian soldiers, neglecting Ukrainian civilians, but the image presented has been altered to show soldiers aiding both.

- False Context
  Example: "Images of Russian and Ukrainian border show defenses in anticipation of a Ukrainian threat." – The images of the border are real; however, the real context is Russian buildup before the Ukrainian invasion.

- Misleading Content
  Example: "America sends civilians to Ukraine to help fight against Russia." – Though some Americans did go to Ukraine on their own, the information is changed to imply the US sent citizens on purpose.

Though we did see other disinformation types used, these were the most common; future work may expand to other types as they become prevalent in disinformation articles of interest to national security.

In addition to disinformation types, we also have rhetorical devices used in these disinformation articles. Rhetoric is the art of effective or persuasive speaking or writing including figures of speech and other compositional techniques. It is observed particularly in politics and/or speeches that are meant to convince an audience to take a side in an argument. Linguists have identified techniques useful for persuading resulting in a large number of defined rhetorical techniques, or rhetorical devices. These rhetorical devices are ubiquitous as all discourse contains one or more of these techniques. Different registers, intended audiences, and language nuances offer different contextual information that can affect the types of rhetorical devices used. Our hypothesis is that, given the type of disinformation used and the intended audience, certain rhetorical devices will be used. Or said a different way, given the frequency and type of rhetorical devices detected, we can detect and classify disinformation.

There are hundreds of different rhetorical devices. Therefore, based on the expertise of our university partners we identified three rhetorical devices that were seen frequently in our data to narrow our scope for this project.

- Ethos: A very broad category consisting of ways we use language to present ourselves and our allies in a positive light and others negatively.

  Example: "I was just defending myself and those guys are bullies."

- Euthymemes: Persuasive appeals resting on unstated assumptions.

  Example: "We can't trust Alice because she lied last week."

- Conspiracy Narratives: Alluding to the actions against an entity. Paranoia. Underdog complex.

  Example: "Alice bought a hammer. She's obviously going to break into her neighbor's house."

Continuation of this research would extend to other rhetorical devices to both refine our classifications of disinformation and to capture more of them. We believe we can train a machine learning model to detect these devices, determine if the article is disinformation, and classify what type if it is.

3. DATA COLLECTION AND ANNOTATION

Data collection was performed by UT Austin. The dataset contains 200 disinformation articles and 200 neutral articles for each event listed above. The coding scheme and annotation was performed by NCSU.

Our approach necessitates the gathering of well-documented antecedent-driven disinformation pieces emanating from various sources of disinformation "providers". We started our analysis with of well-documented disinformation outlets already identified by governmental agencies and scholars. To have varied samples of disinformation we have decided to select well-publicized events with international resonance in different time contexts. The approach to collection included the notion of various timeframes (time-periods) and in different contexts and with a different magnitude of impact or casualties (i.e., "gray zone" conflict [14, 15] vs. a "hot war") and gathering a "matching" "neutral" set of "control group" information that covers the same events or processes, to be able to contrast and capture weak signals at sentence level.

The final list of known Russian propagators was limited to what domains could be found in the searches of events. Below is the list of disinformation domains contained in the dataset and how they are qualified by the ecosystem as identified at the outset of the paper. Some include additional information.

Global Research – Kremlin Amplifier - Canadian – foreign state narrative amplification, Russian-aligned outlets with global reach - Canadian website, provides a Western voice

Zero Hedge – Kremlin Amplifier - Bulgarian - foreign state narrative amplification

Russia Insider – Kremlin Amplifier - Russian - witting proliferators of Russian narratives

Veterans Today – US – Intelligence, FSB - witting proliferators of Russian narratives - "Serving the Clandestine community," fuels right wing extremism

South Front – Crimea – Intelligence, SVR - witting proliferators of Russian narratives, foreign state narrative amplification -

Strategic Culture – Russia – Intelligence, SVR - Russian-aligned outlets with global reach - directed by SVR, affiliated with Russian Ministry of Foreign Affairs, targets Western audience and tries to obscure Russian origins

Oriental Review – Russia – Intelligence, SVR - International Russian socio-cultural - pseudo-academic journal from Russian Academy of Science's Institute of Oriental Studies, attempts to obscure links to state-funded institution

RT (Russia Today) – Overt - State-funded foreign facing mediaSFFFM - state-funded, state-directed global network

Sputnik International – Overt - State-funded foreign facing mediaSFFFM - Main foreign facing project of Russian international news agency created by presidential executive order

3.1. Data Collection

Collection instruments used in the process collection included:

- Search tools: Diffbot, Advanced Google search, Wayback Machine, manual searches via main search engines.
- VPN (Russian media searches)
- Cleaning of data collected - Excel, Python

Keywords used in the process collection included:

- Keywords: Skripal, Sergei Skripal, MH17, biolabs

Target timeframe used in the process collection included:

- Target timeframe: event to 6mos (to 9mos if not enough lines of data could be harvested from resulting search), for Biolabs event start is considered to be Russian invasion of Ukraine to last day of data collection (June 12 2022)
    - Skripal (Mar 4 2018 – Sept 4 2018)
    - MH17 (July 17 2014 – Jan 17 2015/Mar 17 2015)
    - Biolabs (Feb 24 2022 – termination of data collection June 12 2022)
        - Russian sources reported from event genesis to termination of data collection.
        - Neutral source reports on Biolabs could not be found after Mar 29 2022.

Entries were excluded if:

- the scraping tool pulled duplicates;
- the article mentioned the event in general news update but did not have substantive commentary on the actual event;
    - For example: "This week in British news, we will update you on the Skripal poisoning, a tea shortage, and prime minister shenanigans"
- the article was not in English. (Note: English production may have be done by non-native speakers.)

At the request of NCSU, a randomized sample was used for the articles collated. Diffbot returned over 5,000 results in the 6 month period for the Skripal poisoning so we used coding in Python to pull a random 200 articles from the larger csv file and added the first approximately 100 articles, for neutral and disinfo sources. Neutral sources were harder to find based on quantity and Diffbot results but the same process was used to randomly pull a smaller sample from the csv file returned from the scraping software.

Chosen were:

- neutral sources based on Ad Fontes Media's media bias chart (Jan 2022 edition) and extended comparative neutral parameters.
    - All sources were from "center" lower quadrants of pure fact based and upper quadrants of fact and analysis, ½ to 1 SD from middle.
    - All were written in English (including foreign press)

    - Sources used in final data set (ABC - 12, Al Jazeera - 32, AP - 34, Australian Broadcasting Network - 1 BBC - 64, Business Insider - 2, Business Mirror - 1, CBC News - 3, Chicago Tribune - 1, China Daily Asia - 2, DW News - 32, Financial Times - 2, France 24 - 18, Haaretz - 1, Jewish Business News - 1, La Monitor - 1, Le Monde - 1, Long Island Business News – 1, Los Angeles Times – 1, Mashable – 1, National Post – 1, NBC News – 15, NDTV – 1, New York Times – 4, Newsweek – 1, NPR – 1, PBS News – 9, People.com – 2, Philippine Daily – 1, Politico – 1, Reuters – 83, The Guardian – 19, The Independent – 1, The Journal.ie – 5, The Mercury News – 1, The Washington Post – 1, Toronto Star – 1, USA Today – 12, Wall Street Journal – 29.
  - Russian sources based on Clint Watts/State Dept Assessments in our typology:
    - Global Research – 78, New Eastern Outlook – 2, Oriental Review – 1, RT International – 136, Russia Insider – 25, South Front – 10, Sputnik International – 105, Strategic Culture – 11, Veterans Today – 6, Zero Hedge – 86
  - Final Total (*n* = 860)

No other topical entries were eliminated or skipped on any other basis than the reasons above. A randomized sample would be biased by the algorithm that pulled the sample but not a conscious decision by researchers.

3.2. Data Columns

In addition to columns for unique IDs, event category, title, date posted, source url, source title, source name, author, and content, we included:

- source_type - indicates the source category for the row
- source_detail - contextual information about the source
- engagement - engagement data, if available
- content_live - indicates whether the content is still up on the original page
- media_included - captures the kinds of media that appear alongside text content

3.3. Coding Scheme

Data was annotated at the sentence level. Rhetorical devices are developed through choices of vocabulary, syntactic structure, and thematic focus, so words and phrases provide insufficient context for assigning rhetorical devices. Some rhetorical devices are indeed developed over several sentences or even whole texts. However, we found that larger units of analysis failed to distinguish among rhetorical devices: many texts had them all. The sentence was also a convenient span for eventual use of the corpus to support natural language processing and automated detection of rhetorical devices. Finally, sentences were feasible: sentences could be reliably identified (as measured by inter-rater agreement of trained subject matter experts during coding) even in our real-world language data.

To develop the coding scheme, a provisional coding approach was employed at the start. Based on the literature on fallacies, propaganda devices and communication of uncertainty in news media, a preliminary set of nine codes was identified:

1. Abusive Question
2. Assert Baseless Speculation
3. Conspiracy Accusation
4. Red Herring
5. Pile Up Points
6. Accuse Adversary of Bias
7. Accuse Adversary of Speculation
8. Accuse Adversary of Deploying Propaganda
9. Use of "Downgrading" Evidentials

The NC State team performed repeated rounds of coding of texts in the Training Data Set with frequent reviews of disagreements and uncertainties among coders and report-outs with opportunities for feedback to the full project team. As expected, the coding scheme was significantly revised. Some of the provisional codes proved to be inapplicable at the sentence level, including Conspiracy Accusation, Red Herring and Pile Up Points, and so were eliminated. Others were found to overlap significantly, and so were combined (#1+2, #6-8). Additional rhetorical devices related to Doubt were found in the texts, and the coding scheme expanded to include them.

One overall code for DOUBT with four codes for Rhetorical Devices emerged from this iterated process:

1. DOUBT. Does this sentence invite the reader to be doubtful, skeptical, distrustful and/or confused about some information?
2. Rhetorical Device Codes.
    1. SPECULATIVE TROLLING. Does the sentence speculate about alternative theories related to the core events? This code captures sentences that convey conspiracy theories without taking responsibility for backing them with evidence. For example:

        While Sergei and Yulia were comatose in a secured hospital wing, it could have been possible for their blood samples to be doctored with a chemical weapon, the notorious Novichok, which was subsequently and hastily attributed to Russia.

        Yes, "it could have been possible" that the blood samples were doctored–it could even be possible that it was done by Martians. But reasonable people do not consider every theory–only those which have some support in evidence, including evidence of ordinary practices. The communicator is trying here to introduce a wild theory without providing any such evidence.
    2. UNTRUSTWORTHY COMMUNICATOR. Does the sentence directly attack the trustworthiness of a communicator? This code captures ad hominem attacks on a communicator's credibility. For example:

Now, Britain, the U.S. and a few countries blindly following them have dropped all decorum and engaged in blatant lies and disinformation. The USA, putting the blame on the self-defense forces, has yet refused to release any intelligence material.

Our review revealed many types of credibility attacks: open declarations that a communicator was untrustworthy or of a dishonest character; attacks on them as biased or interested; characterizing their activities as lying, disinformation or propaganda; and criticizing them for speaking recklessly, without adequate investigation and with no basis in facts.

3. UNCERTAINTY LANGUAGE. Does the sentence suggest that we cannot or should not be certain about some fact? This code captures three ways of undermining the certainty of information: through direct denial, through declaring the topic to be open, disagreed upon or in need of investigation, or through use of language that both conveys the information and conveys potential doubts about it. For example:

   That information has been shown by other investigators to be based on fabricated video and audio material.

   It's still not clear when or where this took place. For example, media circulated a video supposedly showing a Buk system being moved from Ukraine to Russia.

4. NONE OF THE ABOVE. If a sentence coded as DOUBT could not be assigned any of the three Rhetorical Device codes above, it was coded NONE OF THE ABOVE.

Although even single sentences may have two or even three of the rhetorical devices present, we decided to adopt an "exclusive" approach to coding, in which each sentence would receive one and only one Rhetorical Device code. This both allowed for inter-rater agreement testing and restricted the temptation to find and code hints of rhetorical devices. The Coding Manual, detailed in Appendix B; advised coders to proceed in the order listed above; as soon as one rhetorical device had been detected, the appropriate code was to be assigned and the coder was to move on to the next sentence.

Finally, a code OF INTEREST was added, to allow coders to flag whole documents for further study at some point in the future. This might include particularly vivid conspiracy theories, or creation of Doubt through rhetorical techniques that extend beyond the sentence level.

Coding of this corpus was performed using the Atlas.ti qualitative data analysis software package. After a training period using the data set, members of the NC State team who had not been involved in the development of the coding scheme performed interrater agreement testing on 10% of the corpus. A reliability rating of .787 was reached on Krippendorf's c-a-binary [16] as implemented in Atlas.ti, a measure of whether coders agree on the the presence or absence of Doubt in sentences. In addition, a rating of .679 was achieved on Krippendorf's cu-alpha, a measure of whether coders agree on differentiation among the four Rhetorical Device codes. Both of these ratings indicated an acceptable level of agreement. Disagreements were resolved through discussion and the remaining corpus was split for independent coding between two of the authors.

3.4. Technical Description of Database

Language of Doubt Database

While our peers in qualitative analytics are using ATLAS.ti, this is not a common data format used across natural language processing, machine learning, or computer science generally. In addition to this complication, we anticipated there would be some common preprocessing steps required by any NLP and ML practitioners. To encourage academic interest in our data set, we have taken steps to reformat and preprocess the data to make it more approachable.

We…

- exported and then converted the data through the qualitative data project exchange format (QDPX) to an SQL Server database
- run initial tokenization on various text selections to create "spaCy" Doc objects (see https://spacy.io/api/doc) using an English byte pair encoding (BPE) reference language model
- plan to distribute code through GitHub for loading and preprocessing the raw files, for easy extension if other organizations build on this data set
- plan to store the JSON representation of these (see https://spacy.io/api/doc#to_json and https://spacy.io/api/doc#from_json) in the same database; and
- plan to export the database for redistribution, initially considering Kaggle to host the data

For reproducibility we are also redistributing the BPE language model and its hash, while referring users back to the original distributor of this model and the library package with which it is used.

QDPX

The qualitative data project exchange (QDPX) format is an XML schema (available here: https://www.qdasoftware.org/wp-content/uploads/2019/03/Project-mrt2019.xsd) designed to facilitate the annotation and sharing of a variety of information by qualitative analysts (**Users**). For a computer science audience, this means that the "database" represented in QDPX tabulates the selections being annotated, and the codes assigned by the SMEs. Source records correspond to data files that may be text, images, audio, video, or PDFs; each selection (specifically in our data set each **PlainTextSelection** of a **TextSource**) is a delimited portion of these files. Selections may have multiple annotations, each of which is a reference to a type code (or simply a **Code**) from the **CodeBook** for a **Project**, such as "article content", or "misleading language". In the XML schema, a Selection will directly contain a list of **Coding** objects wrapping a reference to a code (i.e. a **CodeRef**); by contrast in a relational database, these are captured in a pivot table linking selections and codes by their unique ids. As a side note, sources are organized by **Set** and referred to indirectly from other objects (i.e. by **SourceRef**), much as codes are.

QDPX allows representation of a number of other pieces of information, and our published code is intended to support loading of any XML data that complies with the QDPX schema, but the tables called out above (Users, PlainTextSelections, TextSources, etc.) are the only ones that contain data in our published data set.

Supplemental Tables

Alongside the XML metadata, ATLAS.ti will export the source documents as files referenced by path from the XML. We have added two tables, **Contents_Raw** and **Contents_Tokenized**, so that the database is self-contained. The table of raw contents is, simply, a BLOB (binary large object) corresponding to each file along with its corresponding source GUID (globally unique identifier). The tokenized content is a text string in JSON (javascript object notation) corresponding to each selection's tokenized spaCy Doc. This contains both the numeric token under the BPE model and the flags signifying whether each token has a subsequent space character. It is important to reiterate that the tokenized contents correspond to the annotated *selections*, not the raw contents; this was decided to allow for circumstances in which content was deliberately annotated at unusual boundaries in the text that would not have been consistent with a default tokenization.

Example Queries

Along with the software for loading a fresh database from new QDPX results, we have included SQL queries for cross-referencing the tables in the database and returning example content for each code. This is intended to help the audience understand the organization of the database and quickly examine the original published content.

These queries follow a consistent formula (see Appendix):

- for a given type code in the code book, pull all the code references with a matching target GUID;
- use the "codings" pivot table to retrieve the plain text selections by GUID;
- pull the source file content corresponding to the selection; and
- construct a table of results with the desired supplemental information (in our case, the source GUID and path, the source file's short preview, the selected content, and the start position, end position, and length of the selection).

4. INITIAL STATISTICS

The goal of this work is to have a proof of concept and published data set, to facilitate future research on narratives and natural language processing. To this end, we gathered some initial statistics to see how often rhetorical devices are used in both neutral and disinformation news sources. We gathered data from three different types of news sources: neutral news sources (NNS), Russian state media (RSM), and Kremlin amplifiers (KA).

As far as balance, roughly 40-50% of the documents from each type of source are on Skripal events, the same for MH17, and the remaining 10-20% are on the Biolabs event. We also found the largest skew in the data occurred in the Biolabs documents. 18% of the KA documents focused on Biolabs while only 12% of RSM and 8% NNS sources did so.

We now look at how often rhetorical devices on *doubt* are used in disinformation sources versus neutral sources. Doubt has legitimate uses in general rhetoric and is not unique to disinformation, however the frequency of its use could be used as an indicator. In our data, a user who read one article would on average encounter:

- 7.6 Doubt Rhetorical Devices in NNS
- 9.1 Doubt Rhetorical Devices in RSM
- 19.0 Doubt Rhetorical Devices in KA

We did notice that KA sources were significantly longer than other articles, so they may present more doubt rhetorical devices as there is more opportunity in the article to do so. When we normalized the number of Doubt rhetorical devices per 1K words, we found the following:

- 11.5 in NNS
- 16.1 in RSM
- 16.7 in KA

Based on this, we conclude that disinformation sources do use Doubt Rhetorical Devices more frequently than neutral news sources and this can be an indicator of such disinformation.

Breaking Doubt into the three rhetorical devices we coded for in the data set, we can look at more fine grain detail to further separate disinformation from real information. As mentioned above, we are looking at "Speculative Trolling", "Untrustworthy Communicator", and "Uncertainty Language" as three sub-parts of Doubt. Here are some key findings from this initial screening.

- NNS tend to use uncertainty language more frequently than Speculative Trolling or Untrustworthy Communicator techniques.
- RSM uses untrustworthy communicator techniques much more frequent than NNS in most cases except when Russia is the one making the accusations as is the case of the Biolabs where uncertainty language is preferred. At the same time, NNS did not cover the Biolabs topic as much and erred on the side of skepticism when it did.
- KA focus very strongly on speculative trolling and when this is not possible, their profile aligns closely to RSM.

What this leads us to believe is that there are distinctive patterns in the use of certain types of rhetorical devices in disinformation that can be leveraged in a machine learning approach to automatic detection. With the data we have collected along with their annotations and these initial findings, we believe we are in a good place to continue this work further and that automatic detection of this type of disinformation in this register is possible.

Table: Overview of results

	NNS			RSM			KA		
	Skripal	MH17	Biolabs	Skripal	MH17	Biolabs	Skripal	MH17	Biolabs
SPECULATIVE TROLLING	.2	.1	1.2	.7	.8	4.7	1.7	1.3	5.1
UNTRUSTWORTHY COMMUNICATOR	2.0	1.5	9.4	7.5	5.0	2.6	6.5	5.7	3.9
UNCERTAINTY LANGUAGE	8.7	8.3	10.8	10.6	8.9	5.1	11.3	7.5	6.1
Total DOUBT	10.9	9.8	21.4	18.9	14.7	12.4	19.5	14.5	15.0

Codings per 1000 words; columns may not sum due to rounding

5. CONCLUSION AND NEXT STEPS

Based on our initial findings, we know that for Doubt rhetoric, we see a higher frequency of these specific rhetorical devices being used in Russian disinformation than in neutral sources. Therefore, at least manually, we can see trends that indicate disinformation campaigns. This gives us a promising proof of concept to move forward with developing an ML model to recognize rhetorical devices and automatically detect disinformation. In addition, we have curated an annotated dataset of disinformation and neutral news articles that we plan to release as open source.

This initial effort was funded to collect and curate a data set, and run some baseline algorithms against the content as time allowed. This report thus represents the first step of a longer-term effort to characterize and automate the detection of these types of malicious narratives. Next steps fall in to two rough efforts: expanding upon the data available to the community for research, and experimenting on the data with analytics:

- Extend to more disinformation and rhetorical device types for a more comprehensive detection
- Extend to other regions of known disinformation including the use of other languages
- Develop machine learning models to recognized rhetorical devices and detect disinformation
- Identify patterns of disinformation and rhetoric used to assist with attribution of disinformation sources

As we are able our team will continue to expand upon this data set, and we invite others to do so as well. The essential elements are the content, the rhetorical characteristics of that content, and documented processes for annotating the data that give reproducible results (here, measured by reviewer agreement on the annotations). Defining clear and tested definitions of the rhetorical techniques we see in disinformation and propaganda will, we believe, increase the robustness of subsequent analytics. We feel this represents continued refinement of the approach represented by e.g. the PTC-SemEval20 corpus [17] where informal, intuitive language characterization is made.

Given high quality data sets, we will be able to characterize the gross statistical patterns between the presence of language of doubt and the assessment that the content represents the introduction or continuation of a disinformation narrative. While the language of doubt is not necessary and sufficient in Russian disinformation, we have observed in our data set the language of doubt is more frequently (per word) used than in fact-based narratives. Further computer analysis will be needed to automatically annotate text by its use of these rhetorical devices, and then capitalize on this new feature in the data to make such assessments.

6. REFERENCES

[1] K. Shu, S. dumais, A. Ahmed and L. Huan, "Detecting Fake News with Weak Social Supervision," *IEEE Intelligent Systems,* vol. 36, pp. 96-103, 2021.

[2] A. Nabozny, B. Balcerzak, M. Morzy and A. Wierzbicki, "Focus on Misinformation: Improving Medical Experts' Efficiency of Misinformation Detection," in *International Conference on Web Information Systems Engineering*, 2021.

[3] C. Wardle and H. Derakhshan, Information Disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe, 2017.

[4] U. S. Department, "Global Engagement Center," State Department, 23 April 2021. [Online]. Available: https://e.america.gov/t/ViewEmail/i/CD46E76EEAD07F9E2540EF23F30FEDED. [Accessed 01 07 2022].

[5] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Computing Surveys,* vol. 109, no. 53, pp. 1-40, 2021.

[6] J. Khan, M. Khondaker, T. I. A. Islam and S. Afroz, "A benchmark study on machine learning methods for fake news detection," *arXiv preprint arXiv:1905.04749,* pp. 1-14, 2019.

[7] J. C. S. a. C. A. a. M. F. a. V. A. a. B. F. Reis, "Supervised Learning for Fake News Detection," *IEEE Intelligent Systems,* vol. 34, pp. 76-81, 2019.

[8] X. a. Z. R. a. S. K. a. L. H. Zhou, "Fake News: Fundamental Theories, Detection Strategies, and Challenges," in *Association for Computing Machinery WSDM'19*, New York, NY, 2019.

[9] Merriam-Webster, "31 Useful Rhetorical Devices," Meriam-Webster, [Online]. Available: https://www.merriam-webster.com/words-at-play/rhetorical-devices-list-examples. [Accessed 01 August 2022].

[10] K. Avramov, T. Ham, L. Newsom, M. Simanovskyy and R. Williams, "Literature Brief: Characteristics and Goals of Disinformation," UT Global Disinformation Lab, Austin, TX, 2021.

[11] J. M. Lee, "How Fake News Affects US Elections," University of Central Florida, 26 October 2020. [Online]. Available: https://www.ucf.edu/news/how-fake-news-affects-u-s-elections/. [Accessed 01 04 2022].

[12] M. Repnikova and B. Schafer, "How the People's Republic of China Amplifies Russian Disinformation," US Department of State, 27 April 2022. [Online]. Available: https://www.state.gov/briefings-foreign-press-centers/how-the-prc-amplifies-russian-disinformation. [Accessed 1 August 2022].

[13] C. Wardle, "Fake News. It's Complicated," First Draft, 16 February 2017. [Online]. Available: https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79. [Accessed 01 07 2022].

[14] S. L. Pettyjohn and B. Wasser, "Competing in the Gray Zone: Russian Tactics and Western Responses," *RAND,* 12 2019.

[15] B. Wasser, J. Oberholtzer, S. L. Pettyjohn and W. Mackenzie, "Gaming Gray Zone Tactics: Design Considerations for a Structured Strategic Game," *RAND,* 12 2019.

[16] K. Krippendorf, "Computing Krippendorff's Alpha-Reliability," University of Pennsylvania Library, 2011.

[17] G. Da San Martino, A. Barrón-Cedeno, H. Wachsmuth, R. Petrov and P. Nakov, "SemEval-2020 task 11: Detection of propaganda techniques in news articles," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 2020.

[18] B. Nimmo, "Anatomy of an Info-War: How Russia's Propaganda Machine Works, and How to Counter It," *Central European Policy Institute 15,* 2015.

[19] P. Pomerantsev and M. Weiss, "The menace of unreality: How the Kremlin weaponizes information, culture, and money," *The Institute for Modern Russia,* p. 6, 2014.

[20] A. Dawson and M. Innes, "How Russia's internet research agency built its disinformation campaign," *The Political Quarterly,* pp. 245-256, 2019.

[21] R. Oshikawa, J. Qian and W. Yang Wang, "A Survey on Natural Language Processing for Fake News Detection," *CORR,* vol. abs/1811.00770, 2018.

APPENDIX A. DATABASE CONTENT SELECTION BY CODE

```
-- This function takes a string code, and pulls a table of annotated content from the database.
CREATE FUNCTION content_from_code (@codename nvarchar(max))
  RETURNS TABLE AS RETURN
  SELECT [PlainTextSelectionGuid]
        ,[TextSourceGuid]
        ,[PlainTextPath]
        ,[StartPosition]
        ,[EndPosition]
        ,[Preview]
        -- StartPosition is 0-based, but substring() takes a ONE-BASED starting position and a
        -- length.  Thus args start+1, and end-start.  Furthermore, concat will coerce from
        -- UTF-8-encoded binary to strings that display and copy properly, but only with the
        -- proper collation.
        ,SUBSTRING(concat([Data], '' COLLATE Latin1_General_100_CI_AS_SC_UTF8),
                   [StartPosition]+1,
                   [EndPosition]-[StartPosition]) [Content]
        ,LEN(SUBSTRING(concat([Data], '' COLLATE Latin1_General_100_CI_AS_SC_UTF8),
                       [StartPosition]+1,
                       [EndPosition]-[StartPosition])) [ContentLength]
    FROM [NLP_Data].[dbo].[TextSources] source
    INNER JOIN (
      SELECT [TextSourceGuid]
            ,[Guid] as PlainTextSelectionGuid
            ,cast([StartPosition] as INT) [StartPosition]
            ,cast([EndPosition] as INT) [EndPosition]
            ,[Name] as [Preview]
      FROM [NLP_Data].[dbo].[PlainTextSelections]
      WHERE [Guid] IN (SELECT [PlainTextSelectionGuid] /* ,[CodeRefId],[Guid] */
                       FROM [NLP_Data].[dbo].[Codings]
                       WHERE [CodeRefId] IN (SELECT [Id] as [ArticleContent_CodeRefId]
                                             FROM [NLP_Data].[dbo].[CodeRefs]
                                             WHERE [TargetGUID]=(SELECT [Guid]
                                                                 FROM [NLP_Data].[dbo].[Codes]
                                                                 WHERE [Name]=@codename)
                                             )
                       )
      ) content
      ON source.[Guid]=content.[TextSourceGuid]
    INNER JOIN [NLP_Data].[dbo].[Contents_Raw] raw
      ON source.[Guid]=raw.[Guid]
;
```

APPENDIX B. CODING MANUAL

Introduction

Our target: DOUBT

The goal of this coding project is to identify and classify a particular set of rhetorical devices frequently used in Russian disinformation: rhetorical devices that focus on *increasing doubt*.

A good place to start is to recognize that these rhetorical devices are not unique to disinformation. Conveying information while also conveying the possible limitations or weaknesses of that information is an ordinary communicative task. News reporters in particular need to be careful both to convey newsworthy items while also making sure that their readers can assess how much credibility they have. For example, right now in the Ukraine war reporters are having to report somewhat speculative and possibly unreliable information, because that's all anyone has. But when they report it, they also make clear *that* it is somewhat speculative and unreliable. Here are some recent texts from NPR, a reputable source:

> [1] Russian missiles struck the outskirts of the southern port city of Odesa overnight. [2] Ukrainian authorities said seven missiles were fired and hit a shopping center and warehouse, killing at least one person. [3] They suggested Russian troops were relying on old missiles with faulty targeting systems.

Sentences [1] and [2] convey information with no indication–not even a hint–that there is any reason to disbelieve it. But notice that in [3], the author wants both to convey information and to caution the reader against outright accepting it. The speakers are only *suggesting* a possibility, not *saying* that it is true (as in [2]). The journalist who wrote this sentence is not committing themselves to the missiles being old; instead, they are bringing to the readers' attention information that is useful and interesting, but may not turn out to be true.

Disinformation exploits this basic communicative method. Traditional propaganda aimed to persuade audiences to believe false things–for example, that Stalin was a great military leader who was responsible for Communist Party victories in the civil war and WWII. Some contemporary disinformation, by contrast, aims not to persuade but simply to arouse doubt: creating *confusion* about what's really happening, *distrust* in media and experts, pervasive *skepticism* about the possibility of finding the truth, and *exhaustion* that causes people to turn away from politics. If they become widespread, confusion, distrust, skepticism and exhaustion can undermine the democratic process just as effectively as the falsehoods of traditional propaganda. Confused, distrustful, skeptical and exhausted citizens don't listen to each other, don't believe mainstream media, and feel that the ordinary democratic process of debate is worthless.

The rhetorical techniques that we will be identifying and labeling all fall into this new category of propaganda, which we'll call DOUBT. But always keep in mind: our goal in coding these documents is not to detect *disinformation*, but to detect *DOUBT* rhetorical techniques, whether they are legitimate or illegitimate.

Our events

Our documents come from news and analysis related to three events:

MH17. In July, 2014 Russian-backed Ukrainian separatists shot down a Malaysian Airlines commercial flight from Amsterdam, using a Buk missile provided by Russia and killing all 298 people on board. See the Wikipedia article for full details.

Skripal poisoning. In March, 2018 Russian agents poisoned ex-Russian-spy Sergei Skripal and his daughter, who were living in the UK, using the nerve agent Novichok. See the Wikipedia article for full details.

Biolabs. In March, 2022, Russian-affiliated sources (and others) began spreading false information that the United States had been funding a biological weapons lab(s) in Ukraine, intended to attack Russia. See the *Guardian* explainer for more details.

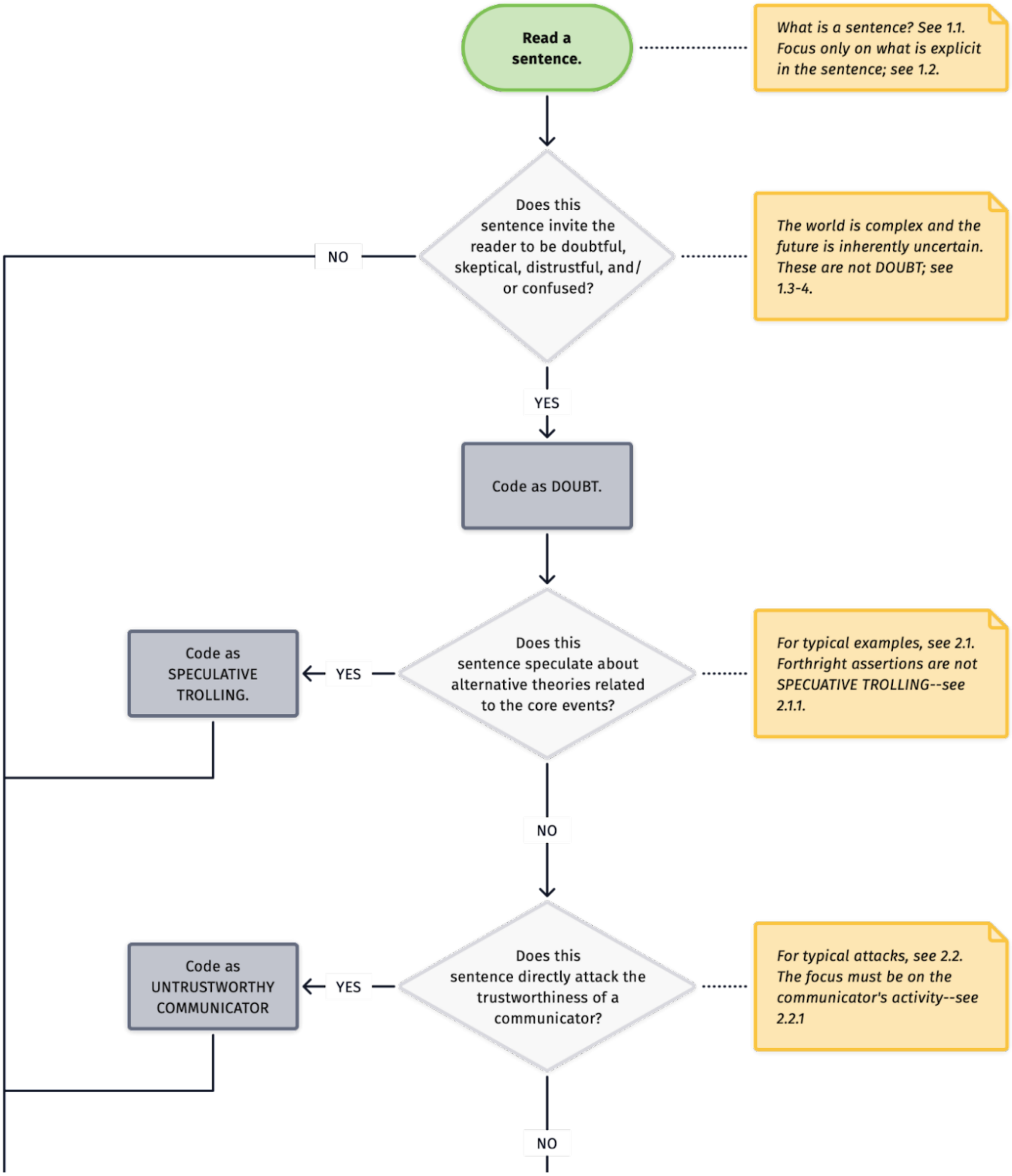
Overall, there will be about 300-400 documents for each of these three events, taken from Russian State Media, commentators that have been identified as associated with Russia, and legitimate news media.

Our coding scheme: overview

You will be applying two main sets of codes to quotations in our documents. The most basic code, DOUBT, will identify sentences that convey to the reader that they shouldn't trust some information. Next, a set of rhetorical device codes will identify what specific kind of technique the sentence is using.

See the following diagram for a decision tree describing the overall process. The remainder of this Coding Manual discusses each code in detail, giving examples of typical sentences receiving the code and noting potentially misleading cases where the code does not apply.

Decision Tree

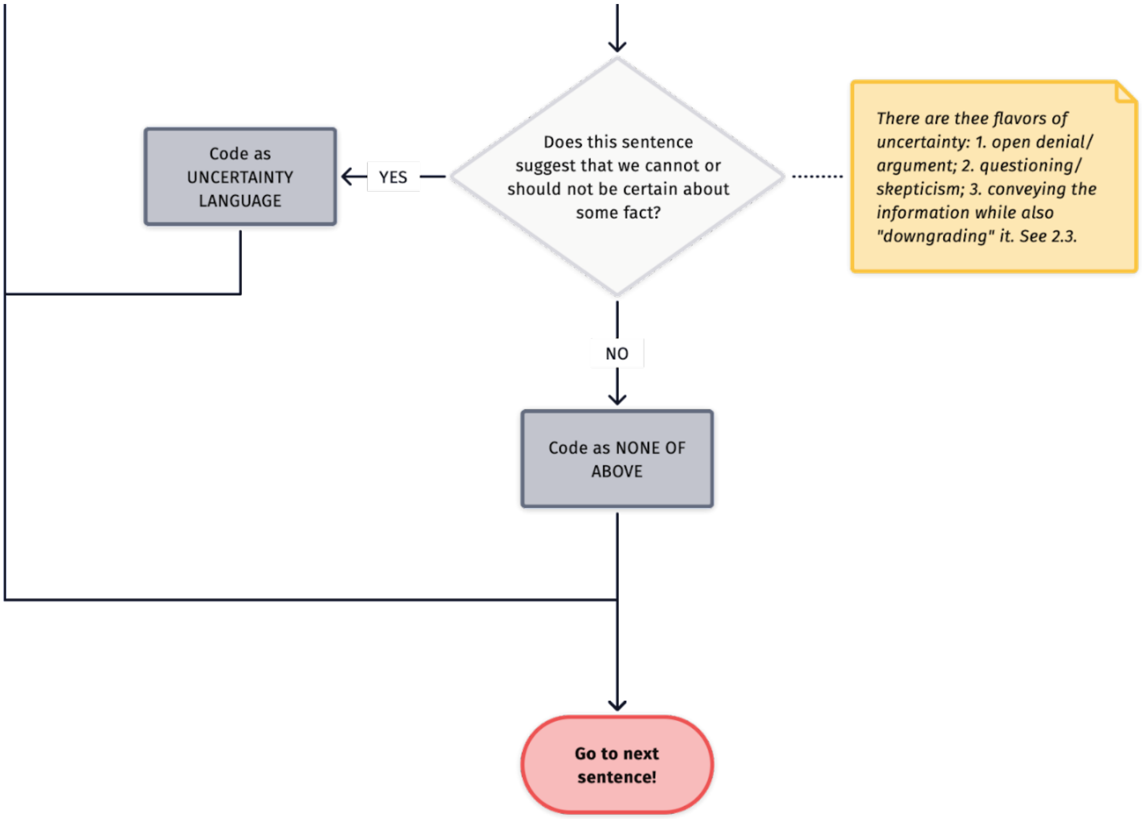

Read a sentence.
Does this sentence invite the reader to be doubtful, skeptical, distrustful, and/ or confused?
NO
YES
Code as DOUBT.
Does this sentence speculate about alternative theories related to the core events?
YES
Code as SPECULATIVE TROLLING.
NO
Does this sentence directly attack the trustworthiness of a communicator?
YES
Code as UNTRUSTWORTHY COMMUNICATOR
NO

What is a sentence? See 1.1. Focus only on what is explicit in the sentence; see 1.2.

The world is complex and the future is inherently uncertain. These are not DOUBT; see 1.3-4.

For typical examples, see 2.1. Forthright assertions are not SPECUATIVE TROLLING--see 2.1.1.

For typical attacks, see 2.2. The focus must be on the communicator's activity--see 2.2.1


Code as UNCERTAINTY LANGUAGE
YES
Does this sentence suggest that we cannot or should not be certain about some fact?
There are thee flavors of uncertainty: 1. open denial/ argument; 2. questioning/ skepticism; 3. conveying the information while also "downgrading" it. See 2.3.
NO
Code as NONE OF ABOVE
Go to next sentence!

Code: DOUBT

Use this code on *sentences* in the content section of the document.

This code identifies every sentence that tries to get you, the reader, to feel a bit doubtful, uncertain, or skeptical of some information. Each time you come to a new sentence, you should ask yourself, is the language used–whether it's from the author or someone else–trying to get me to feel doubtful, uncertain, or skeptical of some information? *If the answer to this question is NO, move on to the next sentence.*

Here are some common ways that DOUBT gets articulated:

- Perhaps the information itself is sketchy, uncertain or even outright false.
- Perhaps there's little or no evidence that it's true.
- Perhaps it came from an untrustworthy source.
- Perhaps there are alternative perspectives that haven't been adequately considered.
- Perhaps the reasons for doubt are openly stated.
- Perhaps the suspiciousness of the information is hinted at in subtle word choices.

Note this is not an exhaustive list. There are many, many ways that information can be presented as suspicious and not worthy of trust. This code labels them all, without making any distinctions about what flavor they come in.

Remember that it is good journalistic practice not only to convey information, but also to convey to readers how much credibility they should give to it. So you will find DOUBT techniques in *both* apparently legitimate and apparently disinformation documents. At the same time, there are plenty of propaganda techniques in addition to DOUBT–including outright lying. So you may find *no* DOUBT even in the most blatant disinformation documents. In sum: you should look for DOUBT, not disinformation.

In coding a sentence, focus primarily on *that sentence*. It is impossible to "forget" all the contextual information you've absorbed by reading other sentences in the document. But strive to limit your use of that information when coding *this sentence*. If there are subtle themes that you feel are getting left out, use the miscellaneous code X-LARGE SCALE DOUBT.

Coding *sentences*.

What is "a sentence"? Normally, a sentence starts with a capital letter and ends with a period ;-). But sometimes fragments have been included in the documents–for example, section headers, leads (short summaries), captions from images or titles of other articles (originally, links). Or formatting and punctuation may have gotten lost in the process of scraping the data from the online source. In these cases, use the available punctuation, line breaks and your understanding of online news article structure to determine where sentences start and end. For example:

> The meeting was due to take place around 1530 GMT, he added.// British officials demanded at a meeting with Russia's charge d'affaires on Wednesday that those responsible for the poisoning of the Skripals were brought to justice, the spokesperson said.// '**Numerous questions**' //Russia on Wednesday questioned the charges.// "The names published by the media, like their photographs, mean nothing to us," Maria Zakharova, the foreign ministry's spokeswoman, told TASS news agency. //

In this case, all formatting got stripped from the document we have. Although there is no line break before/after the fragment "'Numerous questions,'" the starts and ends of the sentences around it suggest that it's a header that should be coded as its own sentence.

The author's punctuation may be highly unconventional by our standards, but we will respect it. The following lines are all one sentence, although a normal reading might be to stop at the colon.

> On this same day, December 21st, Craig Murray headlined "British Government Covert Anti-Russian Propaganda and the Skripal Case", and he described the connections between the Integrity Initiative, and the Russiagate campaign by America's Democratic Party, to restore that Party's version (replacing the Republican Party's version) of the Cecil-Rhodes-founded operation for an all-encompassing U.S.-UK global empire:
>
> Now let us tie that in with the notorious name further down the list; Pablo Miller, the long-term MI6 handler of Sergei Skripal, who lived in Salisbury with Skripal.

When coding a sentence as DOUBT, the quotation should include the full sentence, including the final period, but no spaces, or symmetrical quotation marks or parentheses before or after. When quotations are used in a sentence, do your best to include either zero quotation marks or two quotation marks. There will inevitably be some instances where it is only possible to include one, but whenever possible, try to avoid this.

Focus only on what is explicit in the sentence

In this project, our unit of analysis is the sentence. This means that in coding, we want to look just for what is happening *in this sentence*. This will sometimes be challenging.

Why is it challenging? Humans actually don't produce sentences–they produce discourse: extended stretches of language use glued together by being on the same topic, exploring the same theme, etc. Your human brain, therefore, is going to be very good at detecting how *this* sentence is continuing some topic or theme from previous sentences–even if that topic or theme isn't explicit.

Of course, we cannot read each sentence entirely on its own. Understanding the basic semantic meaning of *this* sentence often requires some contextual information. Among other things, it is necessary to draw on contextual information to figure out the meaning of pronouns like "it, this, that" etc. For example:

> "From the very beginning, it became a political issue on how to accuse Russia of the wrongdoing," the prime minister said.

The meaning of the sentence above cannot be understood without figuring out what "it" refers to. Here, it refers back ("anaphora") to things mentioned in previous sentences, most likely an investigation. To interpret pronouns, you will need to draw on information from other sentences.

Irony also requires contextual information. Again, the meaning of *this* sentence can't be understood unless we construct a theory about the author's intentions, based on other sentences. Consider, for example, an author putting "investigation" in quotations. With context, it may be clear that these are ironic scare quotes. In other words, the author actually means that the "investigation" is NOT a legitimate one, which is a classic DOUBT statement.

Beyond this, be very cautious about filling in additional information. For example, authors sometimes follow up a DOUBT statement with a short sentence or phrase that drives it home. However, only code the follow-up as DOUBT if that follow-up explicitly contains indicators of DOUBT. If the follow-up sentence on its own contains too little information to make a judgment, do not code it as DOUBT. For example:

> So, it's these great investigators (journalists and scholars), against the UK-&-U.S.A.'s Deep State (including its 'news'-media).--DOUBT (NOTE: The "Deep State" is the billionaires.)--NO DOUBT.
>
> Isn't it because the real poison was not Novichok but some other agent developed at Porton Down?--DOUBT Could be.--NO DOUBT You never know.--NO DOUBT

Distinguishing epistemic DOUBT from intrinsic uncertainty.

There are two broad classes of uncertainty in the world. First, we can be uncertain because the world is indeterminate or random–we can call these "intrinsic uncertainties." We will never overcome intrinsic uncertainties (although we will continue to try to limit their scope). For example, the weather tomorrow will always be uncertain. Second, we can be uncertain because of limitations of knowledge–we can call these "epistemic uncertainties.". These uncertainties can be eliminated by new information. It is this second type of uncertainty, epistemic uncertainty, that we want to target in our coding of DOUBT.

Here is an example:

> Did Putin authorize the attempted murder of the Skripals?

More evidence could resolve this question–e.g., a tape recording of a conversation between Putin and some advisor. This question represents an epistemic uncertainty and should be coded DOUBT. Alternatively, there are other questions which cannot be conclusively resolved by more information:

> What will Russia do if sanctions are imposed?
> What effect will sanctions on Russia have on Europe?

When it comes to the future, we can make better or worse founded predictions based on past actions and present circumstances. But future choices and impacts are fundamentally unknowable–they are intrinsically uncertain. Sentences with only intrinsic uncertainty should NOT be coded as DOUBT.

A general rule of thumb is to focus on whether a sentence focuses on the past, present, or future. Uncertainty about the past and present is often (but not always) epistemic uncertainty, coded as DOUBT. Uncertainty about the future is often intrinsic uncertainty, not coded as DOUBT. There are exceptions–of course! The following are (some) examples of when there *could* be epistemic DOUBT about the future:

- Declarations that some prediction is wrong–"denial."
- Attacks on experts making predictions.
- Attacks on the reasoning/evidence underlying predictions.

Distinguishing DOUBT from imprecision and vagueness.

The world is complex, so generalizations made in public communication are often necessarily partial, or vague. For our purposes, statements about a messy, fuzzy world should NOT be coded as DOUBT unless there is additional language that invites the reader to question the epistemic certainty of the statement.

What makes coding difficult is that such statements about a messy world are often positioned very close to DOUBT sentences which invite epistemic questioning. For example:

> When a flight is identified on a radar it is often **assumed** it comes up as "friend" or "foe". [DOUBT] O'Brien said, in fact, it will come up as friend, but not necessarily as foe. [NO DOUBT]... The flight doesn't come up as a friend on radar, but you can determine that it was probably a commercial flight [NO DOUBT].

The first sentence above identifies an *assumption*, which by definition means to suppose something without proof. The author/speaker does not state that things *are* a certain way, but instead states that things are *assumed to be* a certain way, which makes this a sentence which should be coded DOUBT. In contrast, the second and third sentences contain firm assertions about an ambiguous world, but give us no reason to question or be skeptical.

Here's another example:

> "If you were to give a dose of something radioactive, that could take many weeks" to become clear, he said - unless it were highly radioactive. In that case, those who came into contact with you would also be affected.--NOT DOUBT

This is the way the world is, according to the expert: a radioactive poison takes a while to cause injury, unless it is in a high dose, in which case others are also injured. The whole thing is in the subjunctive ("if..were..could..were..would"), but it is NOT a DOUBT sentence because the expert is stating it as an abstract, general rule, rather than a description of what happened in the real-life poisoning case.

Example #3:

> [1] "Novichok agents significantly extend the range of possibility for nerve agents," said Andrea Sella, a professor of inorganic chemistry at University College London.--NOT DOUBT [2] There are five known nerve agents, which are mostly colorless liquids that can kill within minutes, if ingested.--NOT DOUBT [3] "With Novichok, you have the potential for a slower-release agent, which gives you much more control," Sella said.--NOT DOUBT [4] "Using Novichok <u>makes it pretty clear</u> that it was <u>likely</u> Russia that was behind this."--DOUBT

Explanation: The first three sentences express complex facts somewhat vaguely, and abstractly, not unlike the previous example (as is common for newspaper reports). By contrast, the final statement has two DOUBT indicators about what actually happened in the poisoning case.

Below you will find a table that summarizes this section which can help you make distinctions between DOUBT and imprecision/vagueness. While the right-hand column lists a series of words which should arouse your "DOUBT" sensor, it is too simplistic to just code the sentence as DOUBT as soon as you see them. These cases are among the trickiest, so be sure to not lose sight of the first question in your decision tree: is this sentence designed to make me feel doubtful about something? If not, move on to the next sentence.

Ways of describing the world that may be somewhat imprecise or vague. (NO DOUBT)	**Ways of instilling DOUBT about that information. (DOUBT)**
Nerve agents are colorless. All nerve agents are colorless. Three nerve agents are colorless, two have colors. The majority of nerve agents are colorless. Most nerve agents are colorless. Nerve agents are mostly colorless. Some nerve agents are colorless.	"are" → "may, might, could" (if appropriate) Apparently, … It appeared to X that… Allegedly, … X alleged that …. I suggest that…

A few nerve agents aren’t colorless.	X speculated that… It is said/reported that… Perhaps…. It’s possible that… It’s not clear that… Is it really true that…? We don’t really know whether… There’s no evidence that… It’s false that….
Nerve agents are capable of killing people. Nerve agents have the power to kill. Nerve agents have the potential to kill. Nerve agents can kill. Nerve agents, if ingested, can kill. If ingested, a nerve agent would kill. Some nerve agents, if ingested in a large enough dose, can kill.	
Russia attempted to kill the Skripals. Russia was partly responsible for the attempt against the Skripals. Russia played some part in the attempt against the Skripals. Russia provided the nerve agent for the attempt against the Skripals.	

Mainstream Theory	**Typical Alternative Theories**
Russian-backed Ukrainian separatists shot down MH17.	The Ukrainians shot down MH17, either with a Buk missile from the ground, or from a fighter jet.
Russia attempted to poison Mr. Skripal.	The UK/US attempted to poison Mr. Skripal.
The US is funding some helpful, disease-tracking and prevention research in the Ukraine.	The US is maintaining biological warfare facilities in the Ukraine in order to attack Russia. NOTE: the labs must be identified as related to war, weapons, harm, the Pentagon or other marker of *military purpose* in order to count as an “alternative theory.”

CODES: RHETORICAL DEVICES

If a sentence is coded as DOUBT, the next step is to place it into one and only one of the following four categories that identify the strategy the sentence uses for creating DOUBT.

Only one category can be applied to any given sentence. Work through the four options in order, stopping as soon as one is applicable. In other words, if a DOUBT sentence can be classified as UNTRUSTWORTHY COMMUNICATOR, there is no need to consider whether it also could qualify as UNCERTAINTY LANGUAGE.

SPECULATIVE TROLLING

Use this code when a communicator raises some alternative possibility *for the core events in question* (MH 17 downing, Skripal poisoning, existence of biolabs) without committing to defend it. The alternative possibilities will be off the wall, outlandish, well outside of accepted hypotheses–in general, a "conspiracy theory." For example:

However, merely mentioning, asserting or arguing for a conspiracy theory does not in itself constitute SPECULATIVE TROLLING. To be SPECULATIVE TROLLING, a communicator must be pretending that it is worth serious consideration, or trying to sneak the conspiracy theory in without defending it. Such sneaky behavior is a fallacy. Among other things, speculations distract attention from the real arguments, and eventually exhausts the reader who follows the communicator down too many rabbit holes.

Some leading mechanisms for SPECULATIVE TROLLING include:

- Asking a ridiculous question or expressing curiosity or confusion about something, hinting at some type of alternative view. For instance, "But what about…?", or "I'm just asking questions…"
- Hypothetical language such as "it could be…" "could have…" "if so, it would…" coupled with some alternative story of what happened. (Note: be careful with these and refer back to section 1.4 on distinguishing DOUBT from imprecision/vagueness).
- Raising purportedly interesting/relevant evidence for an alternative theory that is highly speculative and vague

It is very hard to present arguments for such alternative theories (although disinformers will try). Instead, disinformers will *act as if* the speculative theory is worth considering.

Here are some classic examples of SPECULATIVE TROLLING:

> **One is left to wonder**, what kind of irrational psychopaths would order such a murderous act, what brain-dead automatons would carry it out, and what supine and servile corporate mainstream media would try to deflect attention from the perpetrators and blame other parties.
>
> While Sergei and Yulia were comatose in a secured hospital wing, **it could have been possible** for their blood samples to be doctored with a chemical weapon, the notorious Novichok, which was subsequently and hastily attributed to Russia.
>
> And while we are in "who done it?" mode, another important **possible lead** is this: **if** Venomous Agent X (VX) was used to harm the former Russian spy, the perpetrators [the British] **would have had** a convenient source by which to carry out their deed.
>
> He in turn is **connected** to other scientists at Porton Down who have died under questionable circumstances, for instance, Dr. Richard Holmes, whose body was found in the same woods as Dr. Kelly, in 2012, two days after going for a walk, and one month after resigning from Porton Down, and to Vladimir Pasechnik's death in November 2001, another Russian defector, who allegedly died of a stroke.

> What is especially **interesting**, is that shortly after the Skripals were hospitalized, the UK government issued a "D" Notice, the effect of which was to prevent the publication of Miller's name in the UK media.
>
> Russia's own significant evidence into the air disaster – and **what could have really happened** – has also been continually and unreasonably repudiated by the JIT.

Keep in mind that the alternative theories can be stated in highly vague and seemingly benign terms. For example:

> But today we have **a dozen plausible theories** of what happened.

Sometimes there will be overly vague references to some "evidence" which is said to support the "fact" of an alternative theory.

> That evidence reportedly includes radar and air traffic control data which puts the onus of responsibility for the crash on the Ukrainian authorities in Kiev.

Sometimes the alternative theory will be put forward as a question:

> What was a military plane doing on the route intended for civilian flights? Why was the military jet flying at so close to a passenger plane?

Here, the questions presuppose that there was a military plane nearby; this is similar to the old fallacy, "When did you stop beating your wife?"

Finally, keep in mind that SPECULATIVE TROLLING can either be asserted by the author of the document, or quoted by the author from another source. It could even be that the author is *quoting* the alternative in order to *refute* it. If the source's statement meets the criteria for SPECULATIVE TROLLING, code the entire sentence as SPECULATIVE TROLLING. For example:

> Echoing a theory floated in Russian state media, Rink said the British **could have been** behind the attack.
>
> In July, state-run channel RT — whose motto "Question more" perfectly encapsulates Moscow's muddy-the-waters approach — reported that **MH17 could have been downed by an Israeli Python air-to-air missile**, even though Almaz-Antey had reported the month before that the missile used "could only have been" a Buk 9M38M1 (a model it said Russia didn't have).--SPECULATIVE TROLLING

[NOT] SPECULATIVE TROLLING

Often there will be sentences which immediately feel like SPECULATIVE TROLLING, but are not. These are described below.

<u>1. Firm assertions of an alternative theory, stated as fact, are *not* SPECULATIVE TROLLING</u>. They are the beginnings of counter-argument, which is a non-fallacious move in public discourse, not a fallacy. "The British poisoned the Skripals" should NOT be coded as

SPECULATIVE TROLLING, while “did the Brits do it?” and “the Brits could have done it” should. For example:

> **Ukraine should bear responsibility** for Malaysian airliner tragedy, said the head of the Russian state at a meeting on economic issues, which he proposed to start with a minute of silence in memory of the victims of the disaster. – NO DOUBT

Note in particular that this may occur in the biolabs cases, where it is often stated as fact that the U.S. is aiding Ukraine in the development of bioweaons.

2. Counter-arguments against a mainstream theory are *not* SPECULATIVE TROLLING. Often sentences will position the mainstream theory as wild, baseless, irrational, etc. But this alone–without mentioning or implying some alternative–is not SPECULATIVE TROLLING. For example:

> Rink told RIA it **would have been absurd** for Russian spies to have used Novichok to try to kill the Skripals because of its obviously Russian origin and Russian name.--UNCERTAINTY LANGUAGE (see section 2.3)
>
> The timing of such an **alleged** plot **would be ludicrous** from a Russian point of view. **Why** would a has-been Russian agent who has been living quietly and undisturbed for nearly a decade in England be targeted on the eve of Russia's presidential elections by Kremlin avengers?--UNCERTAINTY LANGUAGE

3. Accusations of a conspiracy to smear or a conspiracy to do other actions unrelated to Skripal/MH17/biolabs are *not* SPECULATIVE TROLLING. The “conspiracy” claim makes it look like a speculative troll, but the conspiracy is not *about the core events*. Consider what is different about the two examples below:

> I believe, today more than ever, that **this affair** [the Skripal poisoning itself] **is a carefully constructed drama** to push Russia in[to] a corner and justify Western foreign policies in various places such as Ukraine, Iran and Syria.--SPECULATIVE TROLLING
>
> There clearly is **an organized gang**, if not an organized gang of gangs (including media-heads), (and let’s call this **a “conspiracy,”** because it certainly is that) **which is behind all of this smearing against Russia** and against any government that is not hostile toward Russia–UNTRUSTWORTHY COMMUNICATOR.

5. Assertions of truly outlandish ideas. The code SPECULATIVE TROLLING applies only to single sentences with irresponsible suggestions of alternative theories about the core events. When you hit vivid conspiracy theories that don’t meet these three criteria, feel free to apply the OF INTEREST code to flag them for further study.

UNTRUSTWORTHY COMMUNICATOR

There is a well-known fallacy technically called *poisoning the well ad hominem* where one creates DOUBT by attacking the credibility of the communicator who is responsible for conveying information. We will be coding such ad hominem attacks as UNTRUSTWORTHY COMMUNICATOR.

There are a wide variety of such attacks, including:

- Statements that a communicator is untrustworthy or that the author of the document does not trust them.
- Direct attacks on the communicator as a liar, or for lying, propagandizing or disinforming.
- Attacking a communicator for doing a hasty, sloppy or inadequate investigation.
- Attacking a communicator for failing to share evidence in their possession, which suggests that they are covering something up.
- Direct attacks on the communicator as biased or prejudiced, or being motivated by something other than the truth.
- Indirect attacks on the communicator because they were trying too hard or moving too fast, suggesting that they were pursuing some plan more complicated than simply telling the truth; "The lady doth protest too much."
- Direct attacks on a communicator for coming forward with an accusation although they had no evidence evidence.

Below are some classic examples of UNTRUSTWORTHY COMMUNICATOR:

> We sincerely **don't trust** the U.S.

> **We cannot believe our governments** anymore like we used to, just because a prime minister stands up and says, 'the security services have told us so.'

> Now, Britain, the U.S. and a few countries blindly following them have dropped all decorum and **engaged in blatant lies and disinformation**.

> It is a **hopelessly flawed investigation** based on **prejudice** and **preconceived notions of guilt**.

> **Within hours** of the MH-17 plane crash, the United States **pinned the blame** on Russia generally, and Putin particularly.

> In the meantime, **very important questions** surrounding the shoot-down **have gone entirely unaddressed by** US officials and the western media.

> The USA, putting the blame on the self-defense forces, **has yet refused to release any intelligence material**.

> Lavrov, speaking on a trip to Belarus, said British officials have engaged in "**frantic and convulsive efforts** to find arguments to support their indefensible position" instead of producing evidence.

The communicators being attacked may be individuals (Boris Johnson, unnamed officials), organizations (the embassy, NATO), governments or nations (the Kremlin, the UK), institutions (the media, the investigation, the inquiry)--any entity presented as communicating publicly. Furthermore, even if the recipient of the attack is vague and unnamed, any reference to “they/he/she/it” should similarly be coded as UNTRUSTWORTHY COMMUNICATOR. For example, all sentences below would receive the code:

> Despite its grand-sounding legal title, the **JIT [Joint Investigation Team]** is a mockery of jurisprudence. It has, for example, included Ukrainian police in its “fact finding” while excluding Russia. That has ensured bias in the investigation in favor of a party – the Ukrainian state – which should have been treated as a suspect.
>
> Unlike Soviet propaganda, which sought to convince viewers of an alternate truth, the new **Kremlin** strategy seeks to convince viewers that nothing is true at all, experts say.
>
> The announcement came a day after Russia's foreign minister accused **Britain** and the **United States** of spreading "lies and disinformation" about the poisoning of ex-spy Sergei Skripal and his daughter in the English city of Salisbury on March 4.
>
> And, yet, **Western ‘news’-media** routinely lie (usually by stenographically reporting the lies by officials, and then having the gall to call that their ‘journalism’, instead of their “propaganda” — and such allegations are then propaganda about propaganda, or meta-propaganda).

[NOT] UNTRUSTWORTHY COMMUNICATOR

1. Attacks made solely on authors’ *output* (report, publication, website, photo, video, etc.) should not be coded UNTRUSTWORTHY COMMUNICATOR. Reserve UNTRUSTWORTHY COMMUNICATOR for sentences that focus on the communicator and what *they* are *doing* (lying, repeating, promoting, spreading, making baseless accusations). If the focus of an attack is elsewhere (on statements, evidence, claims, etc.), code as UNCERTAINTY LANGUAGE.

Making this distinction can be tricky, since an *explicit* attack on output is an *implicit* attack on the communicator. But keep in mind that we focus only on what is *explicit* in the sentence. In fact, legislatures and other formal settings for debate generally prohibit members making personal, ad hominem attacks on other members. The standard work-around is for debaters to focus criticism on the information, not the person. Although the difference seems small, it is important: to say that someone is a liar is “fighting words,” but to say that the information they conveyed is false is a potentially legitimate criticism. That is the distinction we are trying to capture with UNTRUSTWORTHY COMMUNICATOR.

Compare the following examples:

> **Russia** has **routinely spread misinformation** about the biolabs. – UNTRUSTWORTHY COMMUNICATOR
>
> The **statements** from Russia about biolabs in the Ukraine are absurd.--UNCERTAINTY LANGUAGE
>
> Meanwhile, the **UK hasn't offered even a shred of proof to back up its claim** that Russia was behind the Skripal attack, even as their European peers have questioned

> their claim that "Novichok" was only manufactured in Russia, where it was developed under the Soviet Union.--UNTRUSTWORTHY COMMUNICATOR
>
> The Russian Foreign Ministry has called Washington's **accusations unsubstantiated innuendos**.--UNCERTAINTY LANGUAGE
>
> Any guesses as to who the **British authorities have ruled – without a trial, evidence or motivating factor** – is the main culprit?--UNTRUSTWORTHY COMMUNICATOR
>
> The latest so-called **report did not bring any new "evidence"** to back up previous claims that Russia is culpable for the alleged shoot-down of the Boeing 777 over eastern Ukraine.

2. Attacks on a subject for bad non-communication behavior are *not* UNTRUSTWORTHY COMMUNICATOR. A document may be discussing all sorts of negative activities by a person or organization. However, code as UNTRUSTWORTHY COMMUNICATOR only attacks that focus on their perceiving, thinking, and communicating. For example:

> He said Russia had a proven record of state-sponsored assassinations and had tested ways of delivering chemical weapons, including using door handles to spread nerve agents. – NO DOUBT
>
> Donald Trump and Boris Johnson want Assange dead and his fiancé and two young children to be deprived of his love and of his very presence.--NO DOUBT
>
> Lavrov also assailed the British authorities for stonewalling Russia's request for consular access to Yulia Skripal, 33, a Russian citizen whose condition has improved since she and her father fell critically ill on March 4.--NO DOUBT

The passages above are blatant character attacks, so it's tempting to code as UNTRUSTWORTHY COMMUNICATOR. However, the characteristic being targeted here is not related to their *credibility as a communicator*, so they should not be coded as such. All these accuse persons of being *bad*, but are not particularly targeted to make them *unbelievable*.

3. As always, do not code simply for Russian disinformation. It's likely that many UNTRUSTWORTHY COMMUNICATOR attacks are indeed Russian attempts at disinformation, but watch out–even a legitimate press report may occasionally impugn a communicator's character. As always, code as UNTRUSTWORTHY COMMUNICATOR anything that would lead the reader to be suspicious of the source of some information, whether that suspicion is well-founded or not. For example:

> Prominent social media users and conservative voices have **amplified a baseless theory promoted by Russian state media** accusing the United States of funding biological weapons laboratories in Ukraine. – UNTRUSTWORTHY COMMUNICATOR

Amplifying baseless theories is indeed bad communication behavior, which is here being called out by the legitimate news media.

UNCERTAINTY LANGUAGE

The DOUBT sentence either contains a direct statement of that some information is false or questionable, or is phrased in a way to suggest that the information is uncertain. UNCERTAINTY LANGUAGE commonly is found in one of three main types:

1. Direct attacks on arguments, statements, or evidence.

> Even though ‘involvement’ **had yet to be proven**.--UNCERTAINTY LANGUAGE
>
> The Ukraine crisis has prompted so much **exaggeration** and **falsehoods** in Russian media that a group of students started the site StopFake.org last year to debunk them.--UNCERTAINTY LANGUAGE
>
> Russia and the Ukrainian separatist militia have both **denied** any involvement. They reject the JIT **claims** as “**baseless**”.--UNCERTAINTY LANGUAGE
>
> Meanwhile, the Swedish foreign minister dismissed Russian **claims** that Sweden was the source of the nerve agent as “**ridiculous and totally unfounded**”. --UNCERTAINTY LANGUAGE
>
> That information has been shown by other investigators to be based on **fabricated** video and audio material.--UNCERTAINTY LANGUAGE
>
> Like previous JIT reports, the **so-called “evidence”** is **vague** and relies more on **innuendo** of guilt.--UNCERTAINTY LANGUAGE
>
> Charles Shoebridge, a former UK counterterrorism intelligence officer, said the **material produced** had been of a “**pretty basic nature**”, however.--UNCERTAINTY LANGUAGE
>
> The narrative put out by the Metropolitan Police is not simply **questionable**, it is **plain impossible**.--UNCERTAINTY LANGUAGE
>
> The JIT report this week into the crash in eastern Ukraine on July 17, 2014 again draws on Bellingcat **“information”**.--UNCERTAINTY LANGUAGE

2. Making clear that some issue is still open, that there are questions or disagreements about the topic.

> You state Russians did this – fine then, **where** are the persons that administered it, **how** did they do it, **where** did they do it and **when** did they do it?--UNCERTAINTY LANGUAGE
>
> USA TODAY Network lists what we know now about the downed plane as the countries **seek answers**:--UNCERTAINTY LANGUAGE
>
> Wilbert Paulissen, head of the national investigative department of the Dutch police, said **the investigation has much further to go**, according to Radio Free Europe reporting.--UNCERTAINTY LANGUAGE

It's still not clear when or where this took place.--UNCERTAINTY LANGUAGE

The failed attack sparked an international diplomactic crisis with **Russia being accused** by several countries – **allegations Moscow has repeatedly denied**.--UNCERTAINTY LANGUAGE

3. Making evident that some information has potential weaknesses or limitations by using "downgrading" language when expressing it. The English language has multiple resources communicators can use both to convey some information, but also to convey DOUBTS or uncertainties about that information. Consult the following table for some leading techniques.

Some common linguistic methods that potentially indicate 'downgrading' information		
	"Neutral" or even upgraded	*"Downgrading"* **--UNCERTAINTY LANGUAGE**
Verbs used to express the information	Is/was Novichok was placed on the Skripals' door knob.	modal verbs: e.g., may, could, would–*under the right circumstances–see 1.3* Novichok may have been placed on the door knob. It could be… It was likely that…
[Stance] adverbs used to qualify the information	certainly	perhaps… possibly… maybe… as far as we know… we do not exclude that…
Scare quotes	The evidence	The "evidence"
Status of the information	fact [simple declarative statement]	theory hypothesis scenario
Inferences	the evidence **indicates** the conclusion demonstrates shows supports is consistent with	the evidence **suggests** the conclusion

Some common linguistic methods that potentially indicate ‘downgrading’ information		
People’s perceptions of the world	[simple declarative statement] is evident…	appears… looks/looks like… seems… apparently… ostensibly…
Attributing information to a source	Named source Police **say** that… The ambassador **considers** that… The investigation **reported…**	General, unnamed source It is said… Some consider that… It is widely reported that… …is thought/said to be… Allegedly…
Reports about someone’s thinking	The police **concluded** that... know decide think	believe speculate wonder
Reports about someone’s communicating	Police **say** that… accuse announce report declare assert maintain argue	allege claim profess purport suggest theorize suppose assume/presume

Remember, UNCERTAINTY LANGUAGE is common in both legitimate news reporting and in disinformation. Legitimate journalists use UNCERTAINTY LANGUAGE as a cue to their readers that they shouldn’t fully trust the information being conveyed. As a result, it is not uncommon to apply this code multiple times throughout piece of high-quality journalism.

[NOT] UNTRUSTWORTHY COMM

1. Use good judgment! While “downgrading” words like the above *can* signal DOUBT, they also can serve other functions. The same goes for the “neutral” terms, which are generally less indicative of DOUBT but nevertheless *can* signal DOUBT. Try not to become too fixated on the words themselves, but on what work the words are doing in the sentence.

Here are a few examples of why we cannot simply scan and code for DOUBT words alone:

> The arbitration panel in the Netherlands said it had awarded shareholders in the GML group just under half of their $114 billion **claim**, going some way to covering the money they lost when the Kremlin seized Yukos, once controlled by Mikhail Khodorkovsky.--NO DOUBT
>
> A pending application from Moscow for three more diplomatic posts in Canada is being **denied**.--NO DOUBT
>
> We would like to reassure members of the public that this incident is being taken extremely seriously and we currently do not **believe** there is any risk to the wider public.--NO DOUBT

The sentences above all contain “downgrading” terms, as we’ve labeled them. However, the sentences are not DOUBT-inducing, and should therefore be left uncoded (NO DOUBT).

2. Words related to accusation and blame, on their own, are not UNCERTAINTY LANGUAGE. “Accusation” has negative connotations, but does not say anything about whether the events are certain or uncertain. If anything, accusing someone requires a high burden of proof, suggesting that the accusation is in fact well-evidenced and certain. For example:

> Britain blamed Russia for the March 4 nerve agent attack on former double agent Sergei Skripal and his daughter in the city of Salisbury..--NO DOUBT
>
> Ukraine accused pro-Russian separatists of shooting down a Malaysian jetliner with 298 people aboard, sharply escalating the crisis and threatening to draw both East and West deeper into the conflict..--NO DOUBT

Note that what the blame is *for* may raise DOUBT. In these cases (below), it is not the accusation but the action that is the focus of the Rhetorical Device.

> The UK accused Russia **of a massive disinformation campaign**. --UNTRUSTWORTHY COMMUNICATOR.
>
> Russia accused the UK of **anti-Russian bias**.--UNTRUSTWORTHY COMMUNICATOR.

By contrast, *denying* an accusation raises DOUBTS about the accusation’s truth and should be coded UNCERTAINTY LANGUAGE.

> The Russian government has **denied** any responsibility.--UNCERTAINTY LANGUAGE
>
> Russia and the Ukrainian separatist militia have both **denied** any involvement.--UNCERTAINTY LANGUAGE
>
> They **reject** the JIT claims as “**baseless**”--UNCERTAINTY LANGUAGE
>
> Britain said the attack was almost certainly approved “at a senior level of the Russian state,” an allegation that Moscow has vehemently **denied**.--UNCERTAINTY LANGUAGE

3. Statements that someone has no evidence are not UNCERTAINTY LANGUAGE. Remember that statements that someone is *suppressing* evidence are an UNTRUSTWORTHY COMMUNICATOR attack on the information source. By contrast, statements that evidence is absent or needs to be provided are UNCERTAINTY LANGUAGE.

NONE OF ABOVE.

If a sentence is coded DOUBT, and you cannot code it as SPECULATIVE TROLLING, UNTRUSTWORTHY COMMUNICATOR, or UNCERTAINTY, then use this final, catch-all category.

A productive practice in coding is to use NONE OF ABOVE when you're not sure about what code to apply, or even if you cannot decide whether what you're looking at is a DOUBT sentence at all. Rather than spending time and energy dwelling, apply the NONE OF THE ABOVE code and you can then revisit these sentences later with fresh eyes.

CODE: OF INTEREST

There are many ways to produce doubt, confusion, distrust and exhaustion that won't show up in our sentence-based coding. For example, there may be extended conspiracy theories. Or detailed considerations of historical events that aren't directly relevant to our three cases. Or knock-downs of mainstream views that are developed in an almost literary fashion across multiple sentences.

If you have an intuition that there are significant DOUBT or other disinformation techniques in a document that you aren't able to code as DOUBT, apply the OF INTEREST code *to the full CONTENT quotation*. This code will flag the document as one that we should revisit as we continue to study Russian disinformation.

Use this code freely! We are not testing it for agreement. You have become an expert in disinformation, and this code captures your sense that something suspicious is going on here.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

Email—External

Name	Company Email Address	Company Name
Kiril Avramov	Kiril.avramov@austin.utexas.edu	UT Austin
Jean Goodwin	jegoodwi@ncsu.edu	NCSU

This page left blank

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.