Machine Learning-Enabled Narrative Search in the Information Environment

FORRESTER, Bruce
Defence Research and Development Canada
Quebec City
CANADA

bruce.forrester2@ecn.forces.gc.ca

GHAHAR-KHOSRAVI, Shadi
Defence Research and Development Canada
Toronto
CANADA

shadi.ghajar-khosravi@ecn.forces.gc.ca

WALDMAN, Suzanne
Defence Research and Development Canada
Ottawa
CANADA

suzanne.waldman@ecn.forces.gc.ca

ABSTRACT

The search for information superiority in the Information Age has often turned to open-source intelligence. Over the past 15 years, social media has increased in importance and now plays a key role in gaining and maintaining situational awareness needed for decision advantage. As social media (SM) is increasingly being used by NATO adversaries for influence and disinformation campaigns, research in this area continues to expand. More advanced and complex social media analytical techniques are being introduced by researchers utilizing artificial intelligence (AI) / machine learning (ML) – enabled techniques such as neural networks that can ingest and process enormous amounts of social media data, produce clusters of related artifacts, and help enable semantic narrative search. However, a key challenge for analysts adopting these technologies lies in detecting, monitoring, and searching for narratives that are typically not captured by any one piece of content but are distributed across a narrative-connected set of messages, memes, blogs, "news reports", videos, and more. Other key challenges for analysts lie in understanding and interpreting the output decisions or recommendations of opaque AI/ML-enabled analytics and using them to inform agile military decisions while assuring their credibility and traceability. This paper reports on a particular instantiation of ML engine and outlines a method for checking and validating its outputs. The ML engine represents social media content as vector embeddings and applying cluster analysis to identify topic clusters from the message vector space. In developing a cluster analysis validation technique, we will determine a set of characteristics to be considered when evaluating the outputs of the analysis and validating the algorithm, including apparent relationships amongst clusters as well as the ability of the analysis to provide analysts with narrative context and understanding. Such analytical assurances are necessary to produce valid and reliable intelligence reports and assessments for commanders and decision makers.

Keywords: Social Media, OSINT, Narrative, Machine Learning, Validation Model

1 ENABLING MILITARY CAPABILITY IN THE IE

New information technologies have significantly reduced the spatial, temporal and information gap between troops and command and controls. Frontal clashes of large groups of troops (forces) at the strategic and operational level are gradually a thing of the past. Remote non-contact impact on the enemy becomes the main way to achieve the goals of the battle and operation. The defeat of its objects is carried out to the entire depth of the territory. The distinctions between strategic, operational and tactical levels, offensive and defensive actions are being erased. -General Valeriy Gerasimov, The Value of Science in Foresight, 1983

1.1 The military importance of the IE

In the 2020s, there is no longer any operational “real world” separate from an information environment (IE). For mission success, militaries need to understand and engage expertly in this IE, defined as “the aggregate of the individuals, organizations, and systems that collect, process, disseminate or act on information, as well as the information itself” [1]. Making sense of this domain is challenging, requiring the structuring of massive and complex domains in real-time to identify relevant actions and other activities underway, extrapolate looming trends, and predict risks and opportunities [2]. A specific militarily important component of the IE is the open-source public IE, comprising “individuals, organizations, or systems that collect, process, and disseminate information for public consumption,” and including media as well as social media publications [3]. The public IE is understood as increasingly central to shaping actor and audience cognition and perceptions of military-relevant activities and events, and to an increasing degree, the outcomes of these activities and outcomes [4]. Maintaining situational awareness of emergent activities and dynamics in the open-source public IE is essential to achieving military objectives such as information superiority [5][6]. Commercial media and social media monitoring and intelligence extraction tools, some of which incorporate artificial intelligence (AI), can help military analysts obtain some degree of oversight of the public IE, however these tools are not yet sufficiently advanced to maintain a complete and lucid purview of the public IE capable of identifying important all emergent global trends and “unknown unknowns” in a timely fashion [7].

1.2 Narrative dynamics in the public IE

This article, and the work behind it, leverage the concept of “narrative” as a handle for creating meaning and utility out of the complex discourse dynamics that characterize the contemporary IE. Militaries have long signaled the relevance of the narrative to understanding and shaping the cognitive sphere [8]. A narrative is a mental pattern applied to events or activities that interprets how it relates to past, present and future based on analogous experiences [9]. Narratives operate as cognitive heuristics, helping people quickly assign meaning to events and provisionally predict outcomes [10]. Narratives also have strong social relevance and persuasive power, giving group influencers a means to characterize events and activities in relation to the shared values and goals of the group, and in turn to prescribe appropriate attitudes and behavior to other members [10]. Narratives have accordingly been described as the most impactive feature of all military cognitive-related activities [11].

Evidence of foreign narrative driven interference in numerous democratic processes, perpetrated using social media, is no longer surprising [12-15]. In fact, Bradshaw and Howard have found that in at least 70 countries there is evidence of social media manipulation to shape public attitudes [16]. NATO is also well aware of the 2013 “Gerasimov Doctrine”, Russian Chief of the General Staff that promotes indirect and asymmetric methods [17] that uses the information environment to attack human cognition centering on distraction and manipulation [18]. As an example of manipulation, according to a NATO study [12], Russia had four strategic narrative goals for its propaganda effort against Ukraine:

To promote Russia as a crucial player in the polycentric world in the process of international peace.
To claim Russia’s superiority over the US.

To prevent Ukraine's transformation into being part of the external border of NATO and the European union.

To soften and, in the nearest future, achieve the lifting of the sanctions regime against Russia.

Within Canada, for instance, the focus is more about distracting the population. Upsetting well-established values by increasing the divisions between citizens' with opposing views is an effective method; while the people of a country are busy "fighting" each other, Russia can move with greater freedom and less scrutiny.

One narrative dynamics in the IE relevant to militaries include narrative competition, whereby every new event is rapidly followed by numerous actors seeking to influence how that event should be interpreted by others in their cultural group and beyond it [18]. Particularly in operational contexts, militaries need to understand the state of play of narrative competition over relevant events in order to ensure how their own activities are being framed and interpreted by others and to insert their own framings and interpretations proactively [18].

Another military-relevant dynamic is that of weaponized narratives, which are targeted at audiences by hostile or malicious groups and incorporate tactics of disinformational and synthesized content, decontextualized discourse, false attribution, and automated spread [19]. Weaponized narratives are designed to ostensibly help make targeted audiences' make sense of events, but in fact advance own political and military agendas. Such agendas include the weakening or amplifying, the influence of target groups, or alternatively, wider publics and alliances; degrading public trust in national and institutional institutions; and fomenting social and political conflict and extremism. These objectives carry significant military implications if successfully achieved [20-25].

1.3 Narrative Capabilities

In operational environments, and on the world stage more broadly, militaries need to keep appraised of narratives pertinent to military activities in order to remain current and responsive in their definition of objectives as well as in their strategic as well as tactical engagement of audiences [5]. Specific requirements for understanding the play of narrative competition over military-relevant events and identifying and countering weaponized narratives are narrative search and discover, narrative analysis and sensemaking, and narrative monitoring techniques and tools enabling the anticipation, detection, contextualization, attribution, and assessment of military-relevant narratives emerging and circulating in the IE, whether allied or adversarial. Key challenges in achieving narrative capabilities in IE include:

Narrative search and discovery: identifying narratives that feature specific lexical, semantic, and affective elements as well as distribution and influential characteristics, despite how these are likely to drift widely across languages, social groupings and time.

Narrative analysis and sensemaking: quickly identifying and understanding emergent narratives that could become militarily relevant by influencing key audiences' cognition of military activities or inciting violent behaviour.

Narrative monitoring: following the trajectory or dynamics of relevant narratives to determine if and how they are changing and watching for important characteristics that are changing (such as an increased amount of sharing); distinguishing between organic narratives and weaponized narratives and attributing the latter to their sources so they can be effectively countered through appropriate means at all levels.

2 CHALLENGES EXPLOITING AI

Given the complexity of the IE as a battlespace requiring complex analytics for sensemaking [26-29, 30, 31] in its ever-changing dynamics, analysts are easily overwhelmed with information and cognitive overload and require special tools and analytics to deal with this chaotic space. At the same time, users of social media are increasingly aware of privacy and legal issues, making attribution and identification of individuals a challenge. Using narratives as a 'unit of measure' helps to eliminate privacy issues but complicates analysis. Classical natural language processing (NLP) methods using statistical methods, and 'bag of words' is generally not powerful enough to process the narrative level. As a result, the use of Machine Learning (ML) engines (algorithms) using deep neural networks (DNNs) is being explored to allow for narrative discovery and search.

An interesting challenge arises with AI/ML enabled capabilities. A key promise of semantic narrative search capabilities enabled by ML engines is to support analysts' sensemaking and situational awareness of emergent activities and dynamics in the IE by extracting semantically similar content relevant to analysts' search criteria. However, a critical challenge in using these technologies, particularly with the advent of "black box" or opaque models such as DNNs [32,33], with opaque neural feedback and optimization cycles, is the explainability of the outputs to lay-users such as analysts, or operators lacking data science skills, as well as commanders. Models with low explainability (a.k.a interpretability or human comprehension) such as DNNs are often highly accurate but difficult, if not impossible, to explain or to understand (e.g., through a causal mechanism) [34,35] what is happening as the algorithm is processing inputs and forming clusters of 'like things'.

Without well-designed explanation techniques built into AI/ML enabled technologies, it will be challenging for analysts to build trust in the detections, assessments, and analytics provided by these tools, and ultimately to provide the analyst with the level of credibility and traceability they will require to inform agile and confident military decisions by commanders. Such explanations will additionally be important to ensure the algorithms are performing as expected, ensure algorithms' fairness, identify potential biases in their processes or data, and also improve the usability, usefulness, and trustworthiness of these technologies for analysts. The importance of Explainable AI (XAI) is emphasized by the emergence of recent programs and institutes such as DARPA's XAI program [35] Stanford University's Human-Centered AI institute, DoD's AI Ethical Principles [36], and the Joint Artificial Intelligence Center Responsible AI Champions Pilot [37]. Yet AI systems designers have thus far been challenged to identify and directly address users' needs for understanding AI, to identify ways of measuring the suitability of XAI techniques, and to design usable systems with interpretable outputs that can be trusted and understood by non-expert users [38].

Consequently, a clear understanding of the relationship between inputs and outputs is required to increase the interpretability of ML, to validate the clustering decisions made by the algorithms, and to allow a high degree of analyst' confidence in the reporting of results.

3 TOWARDS VALIDATION OF A MACHINE LEARNING ENGINE

A validation mechanism is required to confirm to the analyst that the algorithm 'does what it says it does'. Initial attempts at narrative search [7, 39], using commercially available tools that employed traditional NLP methods, were unsuccessful in producing a usable search string with corresponding results that highlighted narratives. However, ongoing innovation resulting from Defense Research and Development Canada (DRDC) Innovation for Defence Excellence and Security (IDEaS) program will allow for state-of-the-art AI to be exploited. The innovators involved in the IDEaS Program Challenge 15 "Making Sense of the Chatter" produced state-of-the-art capabilities in the form of ML, that now allow for analysis of social media (SM) in new ways. One of the new ways being researched is 'Narrative Search and Discovery'. Such analysis would exploit the ML clustering, summarization, and entity analysis algorithms, and be visualized via narrative visualization. The initial goals of this research would be to allow NATO to:

Understand what comprises a narrative and what are the characteristics that are assessable in the

IE/SM. Search topic areas to discover the main narratives in an operational environment and their characteristics to understand existing discourse associated with a narrative and what influences, in the aggregate, are affecting populations of interest.

Baseline existing discourse clusters around NATO narratives, as well as around alternative narratives, to monitor changes over time.

Search discourse clusters around NATO narratives, as well as around alternative narratives, and identify evolving characteristics in them and in surrounding discourses.

Below we will outline the characteristics and tests that analysts can use to validate an ML engine that will be consequently used as a baseline tool enabling the validation of future analytical algorithms using ML or other types of AI.

3.1 Description of our Machine Learning Engine

DRDC is conducting research using an ML engine that uses DNNs to represent social media text (tweets or other text-based messages) and binary content (including images, videos, and audio tracks) as high-dimensional numerical vectors, also known as "vector embeddings" or "encodings". These vectors form a high dimension, message vector space, a key property of which is that messages with semantically similar content are represented by vectors that are close to each other in this vector space. Figure 1 shows a 3D representation of some search strings (diamonds) and a small number of messages (circles) forming topic clusters within the message space. The search strings are shown in the legend on the right-hand side. We can search for messages similar to the search strings by encoding the search strings into the message space and finding their nearest neighbors within the space. Therefore, a key capability provided by this ML engine not found in traditional data retrieval systems is sentential semantic search enabling semantic search by statements in addition to words and phrases, hence, potentially instrumental for the narrative search goals presented above. While a 3D representation of this 3000+ dimensional space is rudimentary, it helps to visualize the validation challenge.

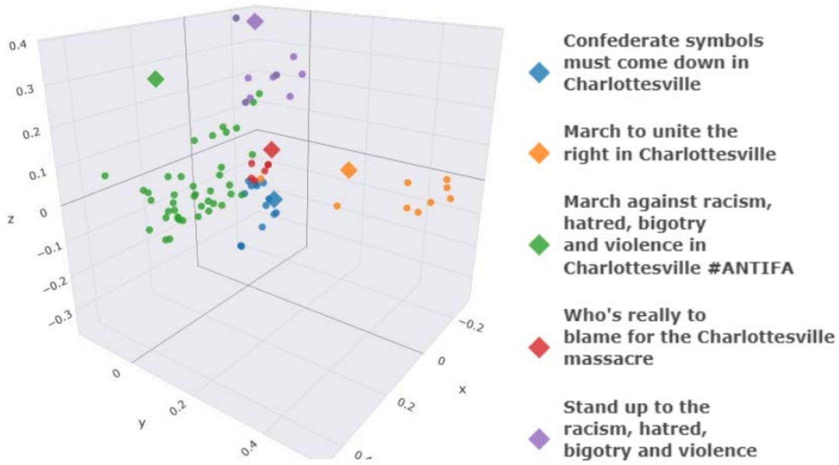

Figure 1. Search strings and messages form topic clusters.

For narrative search purposes, a narrative statement becomes the search string and clusters of semantically similar messages and binary content are formed. The analytical aim is to form a deep understanding of cluster characteristics such that an analyst can determine the nature of the discourse clustering around a particular narrative. Public discourse of military relevance can be monitored and compared over time across key identified characteristics to see if adversary and our own narratives change over time, how are they changing, why are they changing, and what is behind such a change. With a better understanding of the nature of ‘narratives in the wild’, one can discover the following: narratives that support selected strategic narratives, narratives that counter or try to dilute these narratives (possibly as part of an influence campaign), baseline the messages that cluster around these narratives to monitor trends such as increases or decreases in positive or negative sentiment, changes in authors, changes in credibility of messages or reliability of their sources, and changes in circulation and flow within the IE. Some of the key SM analytical capabilities resulting from DRDC’s IDEaS program Challenge 15 enabling narrative search and discovery as described above include:

- Representing social media text (tweets or other text-based messages) and binary content (including images, videos, and audio tracks) as “vector embeddings” or “encodings” to form a high dimension, message vector space.
- Sentential semantic search enabling analysts to search for statements or messages with similar meanings supporting a “narrative”.
- Semantic search capability by keywords or phrases
- Selection of messages from a specific author using author characteristics such as writing style and topics of interest
- Selection of messages by affect including the sentiments and emotions expressed in them
- Evaluation of the credibility of messages based on their consistency with relevant independent observations or other messages with assessed credibility
- Evaluation of the reliability of message sources based on author detection and the credibility level of their authors evaluated as above
- Semantic alignment (sorting) of messages by unstructured message features such as their topics, affect, or style fused with traditional structured features such as user identifiers or location.

3.2 Method for Cluster Validation

For research purposes, we have developed an operational concept of narrative. Thus, a narrative statement consists of a claim (X was/is/will be Y; X did/is doing/will do Y) or a call to action (X should do Y, X should be done to Y) or may be expressed as a shorthand (“lock her up”). Narratives are characterized by transformations (changes over time) of actions and characters that are brought together in a plotline, often triggered by an event. Event Narratives create a shared belief about something that happened. Likewise, Topic Narratives are a shared belief about a person, place or thing. Narratives must have a point which often takes the form of a moral message. Narratives are supported by stories (pieces of content about something that happened such as news items, social media posts, videos, memes, or user comments). Finally, a Master Narrative is a shared general belief or plan. At this point, we will be exploring our concept of a narrative statement (event and topic) and related stories.

Understanding the nature of a narrative allows the formulation of a list of important characteristics that can be used as indicators to compare SM content that clusters together within a particular search string. This, in turn, allows a concreate way to determine the make-up of clusters and possible way to fine tune the ML engine to represent narratives as clusters.

The following steps are proposed as a method of cluster validation:

- Step 1: Use a well-known dataset or single topic area if possible; one that has already undergone detailed analysis. This will aid in establishing ground truth.
- Step 2: Input the data set into the ML engine to produce a message vector space of multi-

dimensional vectors.

- Step 3: Identify a first cluster of messages that are within a specified cosine distance to each other.
- Step 4: Identify a second cluster that is ‘close’ to the first and a third cluster that is ‘far away’ from the first and second clusters.
- Step 5: Perform a detailed analysis, with a known analysis tool or using a qualitative analysis method, of each identified cluster using the following candidate characteristics and metrics:
  - Entities within a message (people, places, objects, location, organizations, etc.)
  - Authors (individuals, official accounts, BOTs, influencers)
  - URL content
  - Hashtags used
  - Words used
  - Sentiment
  - Metadata (timestamp, frequency counts, shares, mentions, likes, number of followers, etc.)
  - Images or videos present
  - Other characteristics may be found depending on the format and type of inputs. Further, some may be determined as irrelevant.
- Step 6: Compare the analysis in step 5 noting the similarities and differences.
- Step 7: Repeat steps 3-6, looking for similarity of characteristics in close clusters and differences in characteristics for far clusters. A pattern should emerge that will allow for fine tuning of the candidate characteristics and metrics.
- Step 8: Repeat the above steps using data based on a single event or wider topic area, until an understanding is acquired.

Can we say a single cluster represents a single narrative? At this point we may need to experiment with the variable parameters within the ML engine in order to tune the sensitivity (the size of clusters or cosine distance grouping). How does the training set used to train the ML engine affect the type of narratives identified? It is likely that several instances of the ML engine, using different training data sets for special topics, will be needed to meet the requirements of clients served by intelligence analysts.

Once an understanding of how the ML engine is clustering messages, a validation process optimization will be conducted. The aim of optimization is to determine the minimum sample size, and which of the characteristics (number, type) are needed in order to produce a valid conclusion for an entire data set. The statistical understanding of social media populations is found in [6].

4 DISCUSSION AND FOLLOW-ON RESEARCH

The results of the validation steps will be an in-depth understanding of how a narrative can be represented by a cluster or set of clusters. This will allow for a narrative search capability within a limited dataset. These results will provide indications of how the engine needs to be tuned, and how the type and size of the input data set effects the clusters. We expect to make progress, in our research, in answering:

How similar are messages within a single cluster?
What are the differences with individual messages that make them appear in a different cluster?
What are the users’ explanation requirements for a better understanding of how the clustering engine

is working
What is the separation between clusters?
What happens near the edge of clusters?
How do discourse clusters tend to relate to each other?

It is likely that our current operational definition of narrative will evolve as will the characteristics used to define the 'intent' of a particular narrative. Narrative discovery, useful for exploring the information space of a country of interest, will involve careful identification datasets. Suitable methods for identifying topic areas are discussed in [27, 28, 29, 40]. For example, one could start by searching using geo-tagged social media content to identify topic areas and hashtags used in a region, then expanding the search to include other topics and hashtags found through the initial messages. Authors, organizations, and news agencies associated with a particular country can also be used to build a suitable dataset for narrative discovery.

Of particular importance for military use-cases is understanding how narratives are changing over time. NATO is interested in learning how its narratives are being received and viewed. Currently, metrics such as reach, number of likes, and sentiment of related comments are used. Narrative search analysis would add another significant dimension to our understanding. It would allow analysts to determine how reactions to and potential influence campaigns are used to shape the narratives. For this we need to track narratives over time. This involves research into cluster size and the relations between close clusters. Do narratives tend to change radically, or do they slowly change, due to changes in message content, over time? How do events cause narratives to change? Research on baselining and frequency of sampling will need to be conducted. It is possible to miss important indications of change if a period is too small or too large.

Additional questions of interest, and further research include:

How well do discourse clusters reflect narrative?
What is the size of clusters needed to define a narrative?
How will the ML engine handle messages that cover two or more narratives?
What explanation techniques are most suitable for supporting users' understanding of narrative clusters?
Can we identify the geographical influence of a narrative?
To what degree do narratives tend to remain stable or change when linguistic and/or geographical barriers are crossed?
How well can narratives be attributed to a particular source?
Is it possible to discover entirely unknown narratives?

Can such an ML engine be used for predicting the success of NATO narratives and messages before they are released? It might be possible to feed NATO messages and artifacts into a particular dataset, say based on a topic or location, to see where they cluster. How does one measure the relative importance of various narratives? Not all narratives will need to be followed and militaries will need to determine which narratives are important. NATO priorities and geopolitical events will dictate the narratives that need to be followed. At first glance, our own narratives and others that support them would provide valuable feedback. Narratives that subvert or try to undermine democracy will also be important to discover and follow.

There are additional challenges in the field of AI that may need to be addressed including addressing users' AI explanation requirements. Two general strategies to improve the explainability of AI systems include: designing explainable models such as the third wave AI built around the concept of contextual adaptation [33] with learning processes easier to understand by human and designing effective explanation interfaces and/or interactions [35]. Limited XAI design guidelines and recommendations have been proposed in theoretical literature based on psychology of explanations and human decision making. Example guidelines include the need for question-driven, contrastive, interactive, and causal explanations and transparency to facilitate trust [38,41-44]. However, empirical research suggested transparency and full user control (e.g., over models' processes, features, outputs) could reduce trust if the user became aware of errors in the system and could undesirably increase user's cognitive workload and inhibit decision making if presented with too much information [45-47]. Given the recipient-centric (or user-centric) nature of the explainable AI problem, XAI strategies need to be designed from a 'human factors perspective', focusing on the issue of user trust in AI rather an 'AI-centric perspective' only focusing on predictive and descriptive accuracy measures [48]. While predictive accuracies will be important in building user trust, they will need to be accompanied by causal explanations for improved trust and a more effective sensemaking and improved situational awareness of emergent narratives and their dynamics within the IE space.

5 Conclusion

The information environment is an active battlespace (interference in elections, influence campaigns, disinformation about covid-19 and vaccines) in which NATO members at home and NATO operations abroad are being targeted by influence operations and disinformation campaigns to influence attitudes, beliefs and behaviour so they are more favourable to adversarial strategic agendas, as well as to undermine social cohesion and the autonomy of democratic processes and institutions. Protecting NATO countries interests against information operations in the IE requires a concerted effort by not only Whole of Government in partnership with Industry and Academia but also between NATO member countries. In the Defence Community, key requirements include maintaining understanding and situational awareness of adversarial strategies, tactics, and vulnerabilities in the IE; as well as the developing effective approaches to countering and building resilience against grey zone adversarial activities in the IE. Understanding the effect of narratives in this space is an essential asset.

The affordance provided by AI, and in particular ML engines and DNNs, has just recently enabled research focused on designing and developing narrative search and discovery. The validation of message and binary content clustering is an important first step in developing this state-ot-the-art analytical tool that can be used by analysts to search and discover narratives of importance, and to understand how these narratives are influenced and change over time.

There are many remaining research questions and important analytical techniques and methods to be developed. Used in a predictive way, such a tool could help NATO commanders to understand the likely effects of strategic, operational and tactical messaging. Using such understanding to guide NATO discourse, instead of kinetic actions, to influence a target audience's master narratives is indeed a noble aim.

6 REFERENCES

[1] U.S. Department of Defence, Joint Publication (JP) 1-02, Department of Defense Dictionary of Military and Associated Terms, (November 2019) https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf

[2] D. Ancona, "Sensemaking: Framing and Acting in the Unknown," The Handbook for Teaching Leadership, New York, NY, Sage, 2011.

[3] U.S. Air Force, Air Force glossary [electronic resource] U.S. Air Force, Washington, D.C., 2007.

[4] UK Ministry of Defence, Joint Doctrine Note 2/19 Defence Strategic Communication: an Approach to Formulating and Executing Strategy, 2019.

[5] W. Marcellino, M. L. Smith, C. Paul and L. Skrabala, Monitoring Social Media: Lessons for Future Department of Defense Social Media Analysis in Support of Information Operations, 2017.

[6] U.S. Department of the Navy, Information Superiority Vision, February 2020.

[7] S. Waldman, B. Forrester, J. McInnis, S. Ghajar-Khosravi & S. Gibbon, Enabling Narrative Sensemaking in the Information Environment. An Evaluation of Commercially-Available Media and Social Media Monitoring and Analysis Tools for Enabling Situational Awareness of Military-Relevant Narratives in the Information Environment (Protected B), 2020.

[8] U.S. Joint Forces, Commander's Handbook for Strategic Communications and Communication Strategy, 3.0, 2010.

[9] P. Ricoeur, "On Narrative," Critical Inquiry, vol. 7, no. 1, pp. 169-190, Autumn, 1980.

[10] W. R. Fisher, "Narration as a human communication paradigm: The case of public moral argument," Communication Monographs, Vol. 51, Issue 1. p 1-22, 1984.

[11] U.S. Joint Forces, Commander's Handbook for Strategic Communications and Communication Strategy, 3.0, 2010.

[12] NATO, The dynamics of Russia's information activities against Ukraine during the Syria campaign. NATO Strategic Communications COE, Riga, 2016.

[13] J. Berzins, Russia's new generation warfare in Ukraine: Implications for Latvian defense policy, C.f.S.a.S. Research, Editor. National Defence Academy of Latvia. p. 15, 2014.

[14] C. Paul. and M. Matthews, The Russian "Firehose of Falsehood" Propaganda Model. Rand Corporation, 2016.

[15] S. Shane, "The Fake Americans Russia Created to Influence the Election," The New York Times. New York, 2017.

[16] S. Bradshaw and P.N. Howard, Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation. Working Paper 2019.2, in Project on Computational Propaganda. Oxford, UK, 2019.

[17] C. Bartles, "Getting Gerasimov Right." Military Review, vol. 96, pp. 30-38, 2016.

[18] T. Nissen, "Narrative-Led Operations," Militaert Tiddskrift, pp. 67-77. 2013.

[19] B. Decker, Adversarial Narratives, Global Disinformation Index, 2019.

[20] A. Maan, Weaponized Narratives, CreateSpace, 2018.

[21] C.Wardle, "Fake news. It's complicated," [Online]. Available: https://firstdraftnews.org/latest/fake-news-complicated/. February 2017 [Accessed 16 August 2020]

[22] C. Wardle & H. Derakhshan, "Information Disorder: Toward an Interdisciplinary Framework for Research and PolicyMaking,"Council of Europe, 2017.

[23] S. Woolley, "The Reality Game: How the Next Wave of Technology Will Break the Truth," Public Affairs, 2020.

[24] NATO, "NATO's approach to countering disinformation: a focus on COVID-19," 2020.

[25] CSIS Academic Outreach, "Who Said What? The Security Challenges of Modern Disinformation," Ottawa, CA, 2018.

[26] B. Forrester, Deep Understanding of Social Media Content: Making Sense of the Chatter (U), DRDC-RDDC-2019-D166, December 2019.

[27] P. Kwantes, S. Ghajar-Khosravi, and B. Forrester, Semantic Dimension Analysis: Data-Driven Content Analysis of Text Generated in Twitter (U), DRDC-RDDC-2018-D006, February 2018.

[28] [28] B. Forrester., "Authentic Chatter", Computational and Mathematical Organization Theory (2020) Vol. 26, pp. 382–411, DRDC-RDDC-2020-P211, November 2020.

[29] M. Jobidon and B. Forrester, "Understanding Social Media Networks through Trending Analysis," Proceeding of the ICCRTS London, UK, DRDC-RDDC-2016-P114, September 2016.

[30] B. Forrester, A. Bacovcin, Z. Devereaux, and S. Bedoya, Propaganda Filters; Tracking Malign Foreign Interventions on Social Media in Real Time, Proceedings of the NATO IST-178 inter-panel/inter-group workshop on Big Data Challenges: Situation Awareness and Decision Support, STO-MP-IST-178, Budapest, DRDC-RDDC-2020-N062, October 2019.

[31] B. Forrester, C. J. den Hollander, M. Floris, A. Pritzkau, U. Franke, H. McVeigh, M. Rosell, J. A. Juhlin, and J. Richardson, Intelligence Exploitation of social media, DRDC-RDDC-2019-N215, STO Technical Report SAS-IST-RTG-102(U), June 2018.

[32] L. Deng, "Artificial intelligence in the rising wave of deep learning: The historical path and future outlook," IEEE Signal Processing Magazine, vol. 35, 180-177, 2018.

[33] J. Launchbury, A DARPA Perspective on Artificial Intelligence, March 19, 2017, Retrieved from: https://machinelearning.technicacuriosa.com/2017/03/19/a-darpaperspective-on-artificial-intelligence/,

[34] G. Allen, Understanding AI Technology, Joint Artificial Intelligence Center (JAIC), The Pentagon, United States, 2020.

[35] D. Gunning and D. W. Aha, "DARPA's explainable artificial intelligence program," AI Magazine, vol. 40(2), pp. 44-58, 2019.

[36] U.S. Department of Defence. "Department of Defence Adopts Ethical Principles for Artificial Intelligence," U.S. Department of Defense, 24 February 2020, Retrieved from: https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

[37] U.S. Department of Defence, Department of Defense Joint Artificial Intelligence Center Responsible AI Champions Pilot, 2020. Retrieved from: https://www.ai.mil/docs/08_21_20_responsible_ai_champions_pilot.pdf.

[38] Q. V. Liao, D. Gruen, and S. Miller, Questioning the AI: Informing Design Practices for Explainable AI User Experiences, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-15, April 2020.

[39] S. Waldman, B. Forrester, J. McInnis, and S. Ghajar-Khosravi, Evaluation and Test of Free Tools for

Disinformation Verification and Tracking, DRDC-RDDC-2020-L178, November 2020.

[40] B. Forrester and M. Gagnon, Sampling Social Networks: Principles and Methods, DRDC-RDDC-2020-R053, September 2020.

[41] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, P. & E. Horvitz, “Guidelines for human-AI interaction,” Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems, pp. 1-13, May 2019.

[42] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences.” Artificial Intelligence, 267, pp. 1-38, 2019.

[43] T. Lombrozo, “The structure and function of explanations.” Trends in cognitive sciences, 10(10), pp. 464-470, 2019.

[44] M. Ribera and A. Lapedriza. “Can we do better explanations? A proposal of user-centered explainable AI”, IUI Workshops, 2019, March.

[45] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan and H. Wallach, “Manipulating and measuring model interpretability”, arXiv preprint arXiv:1802.07810, 2018.

[46] B. J. Dietvorst, J. P. Simmons and C. Massey, “Algorithm aversion: People erroneously avoid algorithms after seeing them err”, Journal of Experimental Psychology: General, vol. 144, no. 1, p. 114, 2015.

[47] B. J. Dietvorst, J. P. Simmons and C. Massey, “Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them.” Management Science, vol. 64, no. 3, pp. 1155-1170, 2018.

[48] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, R., & B Yu “Definitions, methods, and applications in interpretable machine learning.” Proceedings of the National Academy of Sciences, vol. 116, no. 44, pp. 22071-22080, (2018).