Dangers and Solutions for Systematic Misinformation at Scale

J. Wesley Regian
PeopleTec, Inc.
Huntsville, AL
wes.regian@peopletec.com

David A. Noever
PeopleTec, Inc.
Huntsville, AL
david.noever@peopletec.com

ABSTRACT

Instantaneous sharing of information (fallacious or not) to large and sometimes uniquely receptive populations is routine. Cognizant entities (individuals, groups, religions, governments, software agents, bots, etc.) can selectively find information affirming what they already believe to be true by attending to information resources consistent with their existing preconceptions and ignoring or even filtering out information that is inconsistent. No specific entity class (e.g., Democratic vs. Republican) is more susceptible to misinformation vulnerabilities than another. Human individuals are particularly susceptible. Misinformation vulnerabilities are known to occur even unconsciously. This is well known to the intelligence community as confirmation bias (selective search, interpretation, recall). Equally important, hostile entities or propagandists can selectively push information to other entities to influence the targeted entities' world views, opinions, and behaviors. We argue that these dangers are potentially more severe in the cases of government officials, law enforcement, intelligence agencies, and military personnel – who daily make decisions affecting the safety, security, and the very lives of the general population. Our focus in this work is to address the misinformation problem among military personnel. Evidence suggests that misinformation proliferation and acceptance is proportionate among military personnel to that among the population in general (RAND, 2023). This research seeks to characterize the causes and consequences of systematic misinformation and explores 2 classes of technical solutions to address the dangers of unchecked misinformation proliferation and adoption. The first class we refer to as Information Flow Modeling (IFM), developing capabilities to model and visualize information pedigrees. Where did this information originate, what entities are pushing it, what entities believe it, and what is the agenda of the source entities? The second class of technical solutions is Training and Job Support for Critical Thinking (TJS-CT). Our solutions depend largely on recent advancements in Artificial Intelligence and Machine Learning (AI/ML).

ABOUT THE AUTHORS

J. Wesley Regian has 37 years of experience in cognitive performance modeling and knowledge-based software technology development, primarily for military application with AFRL, AFOSR, and DARPA. His work has supported over 50 fielded systems. He has published over 150 peer-reviewed papers on human terrain modeling, knowledge representation, knowledge management, human learning and memory, individual and developmental differences in human cognition, spatial ability and spatial information processing, cognitive modeling, skill acquisition, cognitive automaticity, psychometrics, artificial intelligence, machine learning, hypertext, hypermedia, computer-based training, intelligent computer-based training, virtual reality and augmented reality for training, intelligence analysis and multi-source intelligence fusion. Dr. Regian was a National Research Council research adviser for ten years while Senior Scientist for Knowledge-Based Systems at the US Air Force Armstrong Research Laboratory.

David Noever has 27 years of research experience with NASA and Department of Defense in machine learning and data mining. He received his Ph.D. from Oxford University, as a Rhodes Scholar, in theoretical physics and B.Sc. from Princeton University, summa cum laude, and Phi Beta Kappa. While at NASA, he was named 1998 Discover Magazine's "Inventor of the Year," for the novel development of computational biology software and internet search robots, culminating in co-founding the startup company cited by Nature Biotechnology as first in its technology class. He has authored more than 100 peer-reviewed scientific research articles and book chapters. He also received the Silver Medal of the Royal Society, London, and is a former Chevron Scholar, San Francisco. His primary research centers on machine learning, algorithms, and data mining for analytics, intelligence and novel metric generation.

Dangers and Solutions for Systematic Misinformation at Scale

J. Wesley Regian
PeopleTec, Inc.
Huntsville, AL
wes.regian@peopletec.com

David A. Noever
PeopleTec, Inc.
Huntsville, AL
david.noever@peopletec.com

INTRODUCTION

The rapid advancement of digital technology has transformed the way information is produced, disseminated, and consumed globally. While this digital revolution has undeniably led to a wealth of knowledge and facilitated seamless communication across borders, it has simultaneously given rise to a pervasive and pressing issue: the spread of misinformation at an unprecedented scale. The ease of sharing information—whether it is accurate or fallacious—coupled with the tendency of individuals and groups to selectively consume information that confirms their pre-existing beliefs, has created an environment ripe for the propagation of misinformation.

In this context, the issue of misinformation is not limited to any group, political affiliation, or belief system. Instead, it is a universal vulnerability that affects humans in general, with individuals often falling prey to confirmation bias. This phenomenon, which involves selectively searching, interpreting, and recalling information that aligns with one's preconceptions, is a well-established concern within the intelligence community (Whitesmith, 2020).

The potential dangers of misinformation are further amplified when considering the impact it can have on the decision-making processes of government officials, law enforcement, intelligence agencies, and military personnel. These individuals frequently make decisions that bear significant consequences on the safety, security, and well-being of the general population. It is crucial to recognize that the propensity for misinformation proliferation and acceptance is similarly prevalent among military personnel as it is among the wider public (RAND, 2023).

There is no shortage of theories, studies, and literature on misinformation and disinformation. A Google search on "misinformation and disinformation studies" returns over 2,050,000 results. Most of the work focuses on understanding causes and consequences, and a smaller subset focuses on solutions (Muhammed & Mathew, 2022). Those that focus on solutions tend to fall into 4 classes of solutions.

A. Education and media literacy
    1. Critical thinking skills
    2. Fact-checking and source evaluation
B. Technological interventions
    1. Algorithmic adjustments to limit the spread of misinformation
    2. Artificial Intelligence and Machine Learning (AI/ML) for detecting and flagging misinformation
C. Legal and regulatory approaches
    1. Strengthening accountability and transparency in information dissemination
    2. Developing international norms and agreements to counter misinformation
D. Collaboration between stakeholders
    1. Public-private partnerships
    2. Cooperation between tech companies, researchers, and civil society

This paper provides a concise overview of the causes and consequences of systematic misinformation, and a few case studies to illustrate both. Our ultimate objective is to suggest, with prototype examples, solutions that apply specifically to individual military personnel. We are prototyping technologies that address misinformation susceptibility for military personnel and that conform to the existing DoD education, training, and job aiding ecosystem. We are prototyping two overlapping classes of technical solutions to address the risks posed by unchecked misinformation proliferation and adoption among military personnel. The first class, referred to as Information Flow Modeling (IFM), involves the development of capabilities that enable the modeling and visualization of information pedigrees, which can help users trace the origins, dissemination channels, and agendas of various information sources.

The second class focuses on Training and Job Support for Critical Thinking (TJS-CT) to enhance individuals' ability to evaluate and assess the credibility of information. Through prototype development and demonstration of both classes of technical solutions, this paper seeks to offer actionable strategies for mitigating the risks associated with systematic misinformation at scale across DoD.

CAUSES OF SYSTEMATIC MISINFORMATION

The causes of systematic misinformation are complex and multifaceted, stemming from a combination of cognitive, social, psychological, economic, and technological factors. These factors interact in ways that create an environment conducive to the spread of misinformation and its subsequent acceptance by large populations. Cognitive biases and heuristics, such as confirmation bias, play a key role in shaping individuals' susceptibility to misinformation, while social and psychological aspects, including echo chambers and group polarization, amplify its dissemination. Economic incentives, such as clickbait and political agendas, further motivate the production of misleading content. Finally, software technology algorithms can inadvertently (or purposefully) contribute to the amplification of misinformation. Understanding these underlying causes is critical to developing effective strategies to counteract and mitigate the impact of systematic misinformation at scale.

Confirmation Bias

In an era of digital interconnectedness, the spread of misinformation by foreign adversaries poses a significant threat to the decision-making processes of government entities. One cognitive factor that exacerbates this issue is confirmation bias (Cherry, 2020), which refers to the tendency of individuals to selectively search, interpret, and recall information that confirms their pre-existing beliefs. In the context of government entities, confirmation bias can impair their ability to accurately assess and respond to situations, leading to potentially detrimental consequences for national security, diplomacy, and public trust. Foreign adversaries, recognizing this vulnerability, can exploit confirmation bias by strategically pushing misinformation that aligns with the predispositions of government officials, agencies, or departments. By doing so, they can manipulate the perceptions and actions of these entities, furthering their own objectives and undermining the targeted nation's interests. Understanding the role of confirmation bias in the context of misinformation spread by foreign adversaries is essential for developing effective countermeasures and safeguarding the integrity of government decision-making processes.

Social and psychological factors play a significant role in contributing to the spread and acceptance of misinformation. Two prominent phenomena in this context are echo chambers and filter bubbles, as well as group polarization. Both echo chambers and group polarization can create a feedback loop, where misinformation is not only spread but also further entrenched within communities. These dynamic underscores the importance of addressing social and psychological factors when designing interventions and strategies to combat the proliferation and acceptance of misinformation.

Confirmation bias can fuel agenda amplification by driving individuals to selectively consume and share information that supports their pre-existing beliefs, inadvertently magnifying the reach and impact of misinformation that aligns with those viewpoints. For instance, during the Red Scare in the 1950s, confirmation bias played a significant role in amplifying fears of communist infiltration, as individuals and authorities were more likely to accept and act on information that supported their pre-existing beliefs about the perceived threat. Indeed, the case of J. Robert Oppenheimer, the "father of the atomic bomb," is a concrete historical example of how confirmation bias played into agenda amplification during the Red Scare. In the early 1950s, Oppenheimer's security clearance was revoked due to accusations of disloyalty and communist sympathies, despite his significant contributions to the Manhattan Project. An atmosphere of paranoia and suspicion fueled the allegations, and the confirmation bias led many individuals to accept and act on information that aligned with the prevalent anti-communist sentiment, ultimately tarnishing Oppenheimer's reputation and career.

Echo Chambers and Filter Bubbles

Echo chambers (Del Vicario et al, 2016) refer to social environments in which individuals are exposed primarily to opinions and information that align with their own beliefs. In such settings, dissenting views are either excluded or marginalized. This phenomenon is exacerbated by the presence of filter bubbles (Eady et al, 2019), which are created by personalized algorithms on social media and search platforms. These algorithms curate content based on users' past

behavior, preferences, and interests, effectively insulating them from opposing viewpoints and reinforcing their existing beliefs. Consequently, echo chambers and filter bubbles can contribute to the spread of misinformation by creating an environment where inaccurate or misleading information is amplified and reinforced, while contradicting facts or perspectives are either dismissed or remain unseen. Entities leverage echo chambers and filter bubbles to ensure that their misinformation reaches and reinforces the beliefs of the target audience. By disseminating content within like-minded communities, they can create a self-reinforcing cycle where misinformation is amplified and opposing views are excluded or marginalized.

Group Polarization

Group polarization (Iyengar & Westwood, 2014) is a psychological phenomenon that occurs when individuals' attitudes and opinions become more extreme after discussions with like-minded individuals. This process can intensify pre-existing beliefs and foster a more rigid and uncompromising stance on various issues. In the context of misinformation, group polarization can exacerbate the problem by causing individuals to become more resistant to fact-checking, counterarguments, or evidence that challenges their views. Furthermore, it may lead them to actively seek and disseminate misinformation that supports their extreme positions, amplifying the reach and impact of misleading or false information.

Selective Pushes and Agenda Amplification

Selective agenda-pushing (DiResta et al, 2019) through systematic misinformation is a method employed by various entities, such as political groups, foreign adversaries, or even businesses, to influence public opinion, decision-making processes, or the perception of specific issues. The key elements involved in this strategy include the adversary discovering a susceptible target audience, crafting a compelling narrative, amplifying and disseminating misinformation, then exploiting the echo chambers and filter bubbles. Depending on the adversary, the endpoint strategy may also try to hide its intention or develop plausible deniability. Understanding these elements is crucial for developing effective strategies to counteract the influence of systematically pushed misinformation and protect the integrity of public discourse and decision-making processes.

Target Audience Identification

To effectively push an agenda, entities must first identify the target audience that they aim to influence. This profile may include specific demographics, communities, or individuals with particular beliefs, preferences, or vulnerabilities that make them more susceptible to misinformation.

Crafting Tailored Narratives

Entities create and disseminate misinformation that aligns with their agenda and resonates with the target audience's pre-existing beliefs or concerns. These narratives may exploit cognitive biases, emotional triggers, or societal divisions to maximize their impact and appeal.

Obfuscation and Plausible Deniability

To evade detection or accountability, entities involved in systematically pushing misinformation may use tactics such as hiding behind anonymous online personas, using proxies, or employing disinformation techniques that blend truth with falsehoods, making it challenging to debunk the misinformation definitively.

Monitoring and Adapting

Entities may continuously monitor the spread and impact of their misinformation campaigns, gauging the success of their efforts and adjusting their tactics accordingly. This message shaping might involve refining narratives, altering dissemination channels, or employing new techniques to stay ahead of countermeasures and maintain the effectiveness of their agenda-pushing efforts. Research into misinformation strategies has highlighted several cases that feature elements of systematic efforts at scale, each with a known target audience and relatively successful outcomes from the adversaries. In the last part of the paper, we will examine case studies surrounding tampering in elections, referendum sentiment, and anti-vaccine campaigns.

Technical Foundations

Technological advancements have played a significant role in rapidly scaling up misinformation dissemination. First, the proliferation of social media platforms, such as Facebook, Twitter, and Instagram, has enabled the swift and far-reaching dissemination of information, including misinformation. These platforms allow users to share content with thousands or even millions of individuals instantaneously, creating a highly efficient mechanism for misinformation to spread quickly. Secondly, the algorithms employed by search engines and social media platforms curate content based on users' preferences, past behavior, and interests, creating filter bubbles. These algorithms can inadvertently amplify misinformation by presenting users with content that aligns with their beliefs or piques their curiosity, regardless of accuracy.

To ensure the misinformation reaches its intended audience, propagandists employ various strategies for amplification and dissemination, including social media, fake news websites, and in some cases, even mainstream media outlets. Techniques such as astroturfing, bots, or coordinated campaigns may be employed to boost the visibility and credibility of the misleading content artificially. The use of bots and automated accounts on social media platforms can significantly amplify the spread of misinformation. These accounts can rapidly share and promote misleading content, making it appear more credible, popular, or widely accepted than it is. While bots received attention on Facebook during the 2016 election, deeper analysis has shown that the actual number of bots may not be that large (<100), but if they support a target audience to amplify the message, then the effect or consequence of a few initiators becomes large. Similarly, Twitter CEO Elon Musk sought to rid Twitter of the estimated 5% robotic personas, and evidence suggests that while the number of bot posts may be that large, the actual number of accounts tends to be fewer.

Misinformation purveyors often use Search Engine Optimization (SEO) techniques to increase the visibility of their content on search engines. By optimizing their websites or articles to rank higher in search results, they can attract more attention and drive traffic to their misleading content, further scaling up their reach. This technique calls back to the early days of email, where spam could amplify a single person or marketer's ability to reach millions in a single message.

Finally, misinformation technologies include content generation, which an adversary can tailor to entice the target audience and capture their attention. Creating viral content or clickbait headlines designed to attract attention and encourage users to share or click on the content can contribute to rapidly scaling up misinformation. Such strategies exploit human curiosity, emotions, or cognitive biases to maximize engagement and dissemination. Artificial intelligence and machine learning (AI/ML) advancements have led to deepfakes and other AI-generated content, which can convincingly impersonate real individuals, manipulate media, or fabricate information. This technology makes creating and disseminating convincing misinformation easier, making it more challenging for users to discern between fact and fiction.

CONSEQUENCES OF SYSTEMATIC MISINFORMATION

Systematic misinformation has far-reaching consequences that permeate various aspects of society, undermining the very foundations of democratic processes, public discourse, and decision-making. The erosion of trust, polarization, threats to national security, and misinformed decision-making in government service collectively undermine democratic processes, institutions' effectiveness, and individuals' well-being.

Erosion of Trust

Misinformation erodes public trust in institutions, the media, and even other individuals. People exposed to contradictory and misleading information may become increasingly skeptical of news sources and struggle to identify trustworthy outlets. This distrust can extend to government institutions, healthcare organizations, and scientific bodies, impairing their ability to serve and communicate with the public effectively. An erosion of trust can destabilize societies and hinder the ability to collectively address pressing issues, such as public health crises, climate change, or socioeconomic inequalities.

Polarization

Systematic misinformation fuels societal polarization by amplifying divisive narratives and exploiting pre-existing social, political, or cultural divides. Misinformation campaigns often target specific groups or demographics, reinforcing their beliefs and driving them further apart from those with opposing views. As a result, individuals become more entrenched in their positions, and the public discourse becomes increasingly fragmented and antagonistic. This polarization can impede constructive dialogue and consensus-building, impairing the capacity to find common ground on critical issues.

Threats to National Security

Misinformation poses a significant threat to national security, as foreign adversaries can use it to manipulate public opinion, interfere in electoral processes, or undermine the credibility of government institutions. These adversaries can destabilize nations, create chaos, and distract from their own activities or objectives by spreading disinformation. Misinformation can also exacerbate tensions between countries, increasing the risk of conflict or miscalculation. In this way, misinformation becomes a tool for state actors to pursue their strategic goals at the expense of the targeted countries.

Misinformed Decision-making in Government Service

Misinformation can infiltrate decision-making processes at various levels of government, leading to policies or actions based on inaccurate or misleading information. Government officials and agencies may be influenced by misinformation campaigns, resulting in decisions that are not in the public's best interest or that fail to address the actual problems at hand. Moreover, systematic misinformation can impair the ability of government entities to accurately assess and respond to threats or crises, as they may be working with incomplete or incorrect information.

CASE STUDIES

Case 1: 2016 United States Presidential Election

During the 2016 US Presidential Election, Russian operatives conducted a systematic misinformation campaign to sow discord among the American public and influence the election outcome. This influence effort involved creating and disseminating fake news, divisive content, and politically charged advertisements on social media platforms. By leveraging the cycle of systematic misinformation, they amplified their foreign agenda and effectively reached millions of Americans. The campaign exploited echo chambers and filter bubbles, with misinformation being shared and reinforced within like-minded communities. This example illustrates how foreign adversaries can manipulate the public discourse and political landscape through systematic misinformation campaigns.

During the 2016 and 2020 US Presidential elections, researchers and analysts used IFM-like techniques to identify, monitor, and understand the flow of disinformation across social media platforms, news websites, and other channels. This monitoring helped to uncover the extent of foreign interference in the elections and the strategies used by malicious actors to amplify divisive content and disinformation. Such misinformation forensics has been instrumental in understanding the origins, dissemination, and amplification of conspiracy theories, such as QAnon or Pizzagate. By monitoring the flow of information related to these conspiracy theories, researchers can identify the key influencers and networks responsible for their spread and design targeted interventions to counteract their impact.

Case 2: The Brexit Referendum

In the lead-up to the Brexit referendum 2016, various misinformation campaigns were employed to push specific agendas, some allegedly originating from foreign sources. Misleading information was disseminated regarding the economic impact of Brexit, immigration, and the benefits of leaving the European Union. These campaigns exploited the existing divisions within the UK population and amplified the narratives by targeting echo chambers and reinforcing pre-existing beliefs. As a result, public opinion was influenced, and the discourse surrounding Brexit was significantly affected by misinformation, which may have impacted the referendum's outcome.

Case 3: Anti-vaccine Misinformation Campaigns

Foreign adversaries have been known to exploit the issue of vaccine hesitancy and public health concerns by pushing anti-vaccine misinformation campaigns. These campaigns spread false or misleading information about vaccine safety, efficacy, or side effects, creating fear and distrust among the target population. Using social media, fake news websites, and other digital channels, the propagators of misinformation can reach a large audience and be amplified within the echo chambers of like-minded individuals. The objective is often to destabilize public trust in health institutions, foster social unrest, and undermine the targeted country's efforts to manage public health crises, such as the COVID-19 pandemic.

IFM-like techniques have been employed to track the spread of anti-vaccine misinformation, especially during the COVID-19 pandemic. Researchers have used IFM tools to identify the sources of anti-vaccine content, the networks responsible for disseminating this misinformation, and the target audiences most susceptible to it. This information has been invaluable in developing targeted public health messaging to counter vaccine hesitancy.

Further case studies provide an emerging research growth area. With the rise of deep fakes and AI-generated content, IFM techniques have been used to trace the origins of manipulated media and identify the actors responsible for creating and disseminating them. This information can help to hold those responsible accountable and inform efforts to develop technological solutions for detecting and preventing the spread of deep fakes. Both sides of the 2022 Ukraine-Russia conflict have spread deep fakes of leadership in unflattering ways, such as Putin being arrested or kowtowing to China. Notably, in the chaotic early days when Russia was on the verge of entering Kyiv, relatively naïve deep fakes were published of Ukrainian leader, Vladimir Zelensky, declaring surrender and telling his fighters to put down their weapons. This propaganda, while not new in the history of warfare, carries a particularly high threat given the rapid advance of video technology and the potential rapid dissemination of actionable (and potentially disastrous) misinformation.

A second concerning growth area is when the immediate need for accurate information can cost lives before any error correction is possible. In the aftermath of natural disasters, misinformation can spread rapidly and hamper relief efforts. IFM-like techniques can be used to track the flow of false information during such events, enabling emergency responders and government agencies to debunk false claims and provide accurate, timely information to the public.

SOLUTIONS TO SYSTEMATIC MISINFORMATION

As previously stated, our interest is developing misinformation solutions that can readily apply to military personnel and easily plug into the DoD education, training, and job-aiding ecosystem. Feasible solutions include two overlapping classes. The first class we refer to as Information Flow Modeling (IFM), developing capabilities to model and visualize information pedigrees. Where did this information originate, what entities are pushing it, what entities believe it, and what is the agenda of the source entities? The second class of technical solutions is Training and Job Support for Critical Thinking (TJS-CT). There are various definitions of Critical Thinking (CT), but generally, they include open-mindedness, respecting evidence and reasoning, considering different perspectives and points of view, cognitive flexibility, not being stuck in one position, withholding judgement until various sources have been considered, skepticism, clarity, and precision.

Information Flow Modeling (IFM)

We are designing IFM to identify the origin of information, track its dissemination, understand the beliefs and agendas of publishing entities, and analyze amplification toward the target audience. IFM relies on Artificial Intelligence and Machine Learning (AI/ML) to automatically construct and display an information pedigree in real time to support soldiers during training and operations and evaluate the provenance of information. Unlike journalist sites like Snopes.com, which rely on human verifiers and focus on headline stories, IFM looks with an objective and auditable filter that includes non-traditional newspaper sources, social media, forums, and other online assets that seed future headlines or influence campaigns. Essential to assessing the potential impact of systematic misinformation, IFM emphasizes visualizing information flows and amplification metrics like "retweets," sentiment, and follower counts. IFM can also be set up to alert users when new misinformation or agenda-driven content is detected. These alerts summarize the content, its origin, the entities involved in its dissemination, and its potential impact on the target audience.

IFM incorporates a range of features implemented as methods to pragmatically identify and assess misinformation using automated technical means. These methods include:

1. Origin Identification: IFM employs advanced algorithms to trace the origin of information, leveraging digital forensics techniques to identify the original source and track its dissemination across various platforms and channels.
2. Dissemination Tracking: By monitoring the spread of information, IFM analyzes the trajectory and pathways through which misinformation propagates, allowing for a comprehensive understanding of its reach and potential impact.
3. Belief and Agenda Analysis: IFM utilizes natural language processing and machine learning algorithms to analyze the beliefs and underlying agendas of publishing entities, providing insights into the motivations behind the dissemination of misinformation.
4. Amplification Analysis: IFM focuses on visualizing information flows and employs metrics such as "retweets," sentiment analysis, follower counts, and other amplification measures to assess the level of reach and influence of misinformation campaigns.
5. Inclusion of Diverse Sources: Unlike traditional fact-checking platforms, IFM casts a wide net by including non-traditional newspaper sources, social media platforms, forums, and other online assets that play a crucial role in shaping future headlines or driving influence campaigns.
6. Real-Time Information Pedigree: Leveraging AI/ML capabilities, IFM constructs and displays an information pedigree in real time. This pedigree provides a comprehensive overview of the information's provenance, credibility, and dissemination history, aiding users in evaluating the trustworthiness of the content.
7. Alert System: IFM can be configured to send proactive alerts to users when new instances of misinformation or agenda-driven content are detected. These alerts summarize the content, its origin, the entities involved in its dissemination, and its potential impact on the target audience, empowering users with timely and actionable information.
8. Objective and Auditable Filtering: IFM applies an objective and auditable filtering mechanism to assess the veracity of information. By employing automated algorithms, it mitigates the reliance on subjective human verification, enabling consistent and scalable evaluation of misinformation.

By implementing these methods (Figure 1), IFM provides a comprehensive and efficient means to identify, track, and evaluate the spread of misinformation. It supports soldiers during training and operations, enhances situational awareness, and enables users to make informed decisions while combating the influence of deceptive or misleading information.

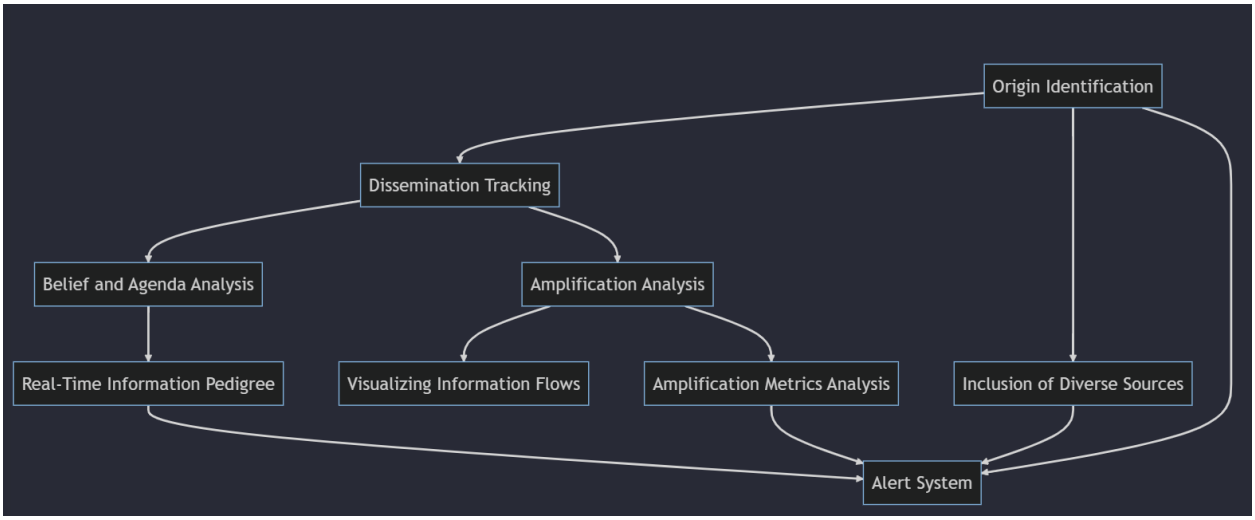

Figure 1. Workflow Diagram Identifies IFM Components to Identify and Counter Misinformation at Scale

Training and Job Support for Critical Thinking (TJS-CT)

While Critical Thinking (CT) is among the essential cognitive capabilities for understanding the world and anticipating the future, CT is rarely taught. A limited conceptualization of CT is often taught in 7th-grade curricula (USA) as the Scientific Method, but this is typically taught narrowly as experimental design rather than as generalizable CT. Over the past four decades, DoD automated training has improved and is now quite good in many cases. For example, training for Insider Threat awareness and Cyber Security is now efficient and effective (almost enjoyable). Except for Selection Bias modules for DoD intelligence professionals, we are unaware of any online CT training – particularly in the domain of information provenance. There are, however, many good DoD textual sources about CT (for example, https://irp.fas.org/agency/army/mipb/2022_01.pdf). We have developed and demonstrated a capability to automatically convert military textual/graphical training manuals into high-quality multimedia/online training (Noever & Regian 2022a, b, c, d).

For a job training course on critical thinking, the following components can be considered as essential elements to cover in the curriculum:

1. Understanding Logical Reasoning: Provide an overview of logical reasoning and its significance in critical thinking. Teach students how to recognize logical fallacies, evaluate arguments, and distinguish between valid and invalid reasoning.
2. Developing Analytical Skills: Foster analytical thinking by teaching students how to break down complex problems into manageable components, identify patterns, and draw logical conclusions based on evidence.
3. Enhancing Problem-Solving Abilities: Equip students with problem-solving techniques, such as brainstorming, root cause analysis, and decision-making frameworks. Emphasize the importance of systematic and structured approaches to tackle challenges effectively.
4. Evaluating Information Sources: Teach students how to assess the credibility and reliability of information sources. Focus on identifying bias, recognizing misinformation, and differentiating between fact and opinion.
5. Strengthening Evidence-Based Reasoning: Educate students on the importance of using evidence to support arguments and make informed judgments. Teach them how to gather relevant data, evaluate its quality, and apply it effectively in their reasoning.
6. Encouraging Open-Mindedness and Perspective Taking: Foster an environment of intellectual curiosity and open-mindedness. Teach students to consider diverse perspectives, challenge their own biases, and engage in respectful dialogue to broaden their understanding of complex issues.
7. Developing Effective Communication Skills: Emphasize the significance of clear and concise communication in critical thinking. Teach students how to articulate their thoughts, ask probing questions, and engage in constructive discussions.
8. Promoting Creative and Innovative Thinking: Cultivate creative thinking skills by encouraging students to explore alternative solutions, think outside the box, and challenge conventional wisdom. Teach them techniques for generating novel ideas and embracing innovation. By incorporating these components into the job training course on critical thinking (Figure 2), learners can develop essential skills and mindsets that enable them to approach problems and decision-making with a critical and analytical mindset.

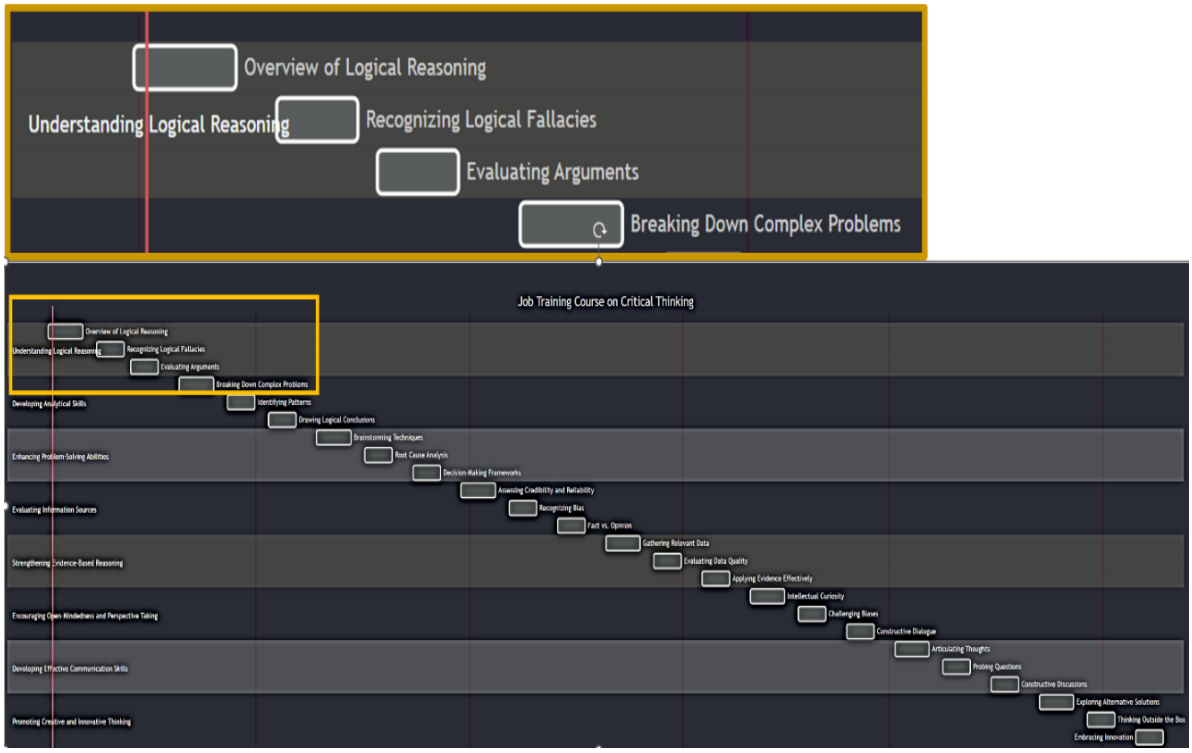

Figure 2. Notional Five-Month Lesson Plan Implementing Critical Thinking Skills for Soldier Training

Fake News Detector

To demonstrate the potential of AI/ML for detecting misinformation at scale, we built an automated fake news detector. The detector employs natural language processing techniques (Ahmed, Traore & Saad, 2018) to isolate the features that distinguish a corpus of "real" and "fake" news, with a high testable accuracy of 99.8% on previously unseen (or untrained) examples. An example fake news headline reads "*WATCH: Paul Ryan Just Told Us He Doesn't Care About Struggling Families Living in Blue States*". The training news consisted of 17,903 articles split in half between true and false news (39%), politics (29%) and other (32%) topics authored during the 2016 Trump administration's first year. The fake news was collated from unreliable websites flagged by Politifact and Wikipedia, while the real articles appeared on Reuters. The University of Victoria, Canada, (ISOT Fake News Dataset, 2023) has since updated a similar compilation. The detector itself uses a linear classifier based on extracted word counts called term frequency-inverse document frequency (TF-IDF) matrix. The resulting vectors align the commonly used terms found in fake news compared to true news in the dataset. Figure 3 shows diagonal dominance of true positives and

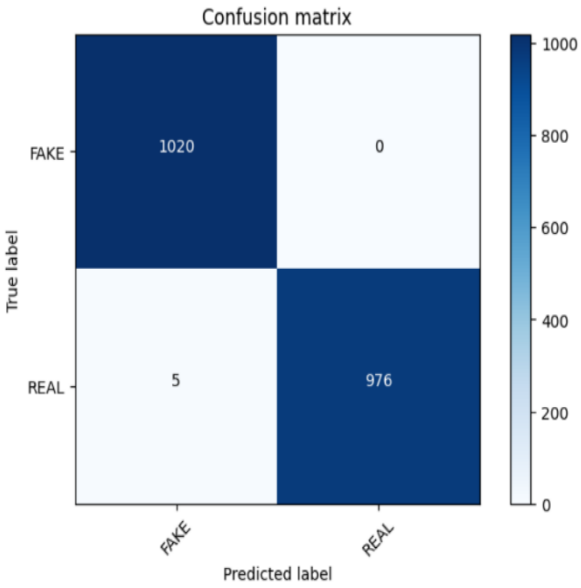

Figure 3. Error Matrix Showing True Positives and Negatives

negatives with only 5 of the approximately 2000 sub-sampled test set as real, but labeled as fake (false positive for a detector of fake news). Table 1 shows the complete classifier results for a test set of articles not used for the training of linguistic features commonly used by fake news articles. Figure 4 shows a method for visualizing and discriminating between probably true versus probably fake news articles in high dimensions (principal components). It makes the misinformation understandable and plausibly solvable by soldiers. While not a complete answer to identifying misinformation, this initial automated screening capability offers a way to lower the noise and focus on the likely candidates for further downstream processing.

Table 1. Class Report for Test News Articles Automatically Flagged as Fake Based on Language Characteristics

classification report:	precision	recall	f1-score	support
0	1.00	1.00	1.00	1020
1	1.00	0.99	1.00	981
accuracy			1.00	2001
macro avg	1.00	1.00	1.00	2001
weighted avg	1.00	1.00	1.00	2001

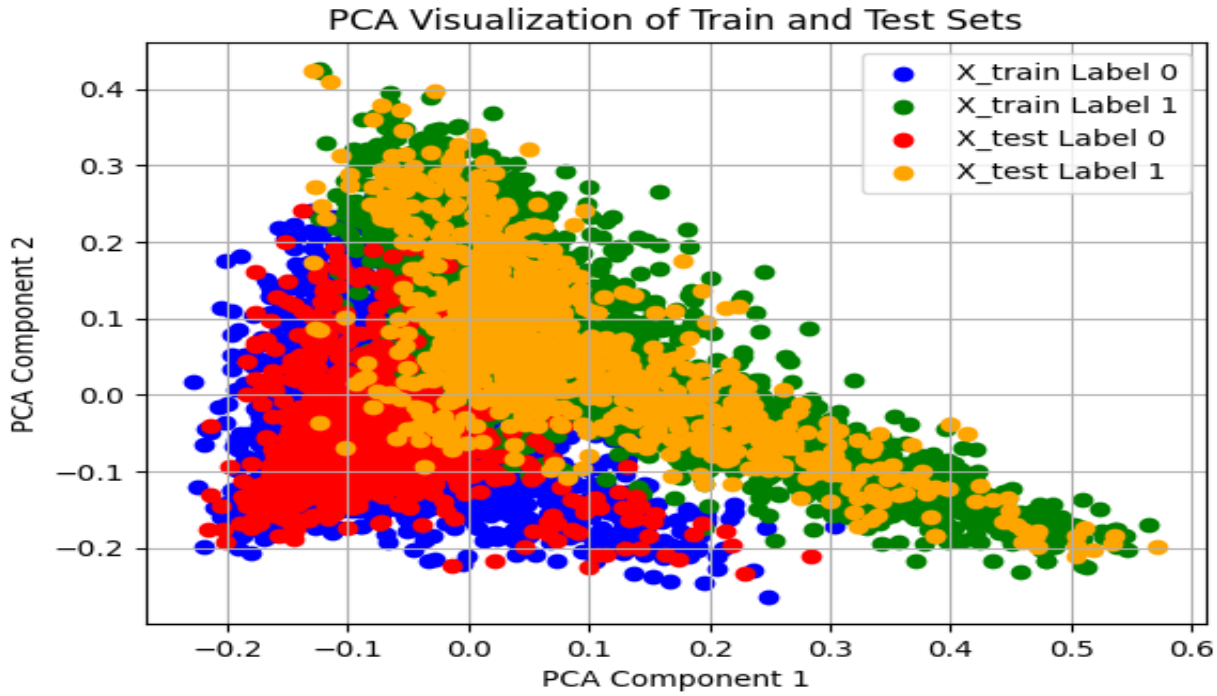

Figure 4. Using Principal Components, the linguistic elements of fake news (blue) form a cluster compared to true news (green).

CONCLUSIONS

This paper has outlined the dangers of systematic misinformation at scale, focusing on causes (e.g., technological changes, social and psychological factors, and the selective pushing of agendas by foreign adversaries) and consequences (e.g., erosion of trust, polarization, threats to national security, and misinformed decision-making in government service). To mitigate the consequences among military personnel, we recommended implementing an AI/ML system for automated and immediate evaluation of information provenance (IFM) and online training/job support for critical thinking (TJS-CT).

ACKNOWLEDGEMENTS

The authors would like to thank the PeopleTec Technical Fellows program for encouragement and project assistance. This research benefited from encouragement from US Army Space and Missile Defense Command.

REFERENCES

Ahmed H, Traore I, & Saad S. (2018) "Detecting opinion spams and fake news using text classification", Journal of Security and Privacy, Volume 1, Issue 1, Wiley, January/February 2018.

Cherry, K. (2020). *Why Do We Favor Information That Confirms Our Existing Beliefs?* https://www.verywellmind.com/what-is-a-confirmation-bias-2795024

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). *Echo chambers: Emotional contagion and group polarization on Facebook*. Scientific Reports, 6(1), 37825.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J., & Johnson, B. (2019) *The Tactics & Tropes of the Internet Research Agency.* DigitalCommons@University of Nebraska - Lincoln, October. https://digitalcommons.unl.edu/senatedocs/2/.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). *How many people live in political bubbles on social media? Evidence from linked survey and Twitter data*. SAGE Open, 9(1).

ISOT Fake News Dataset, (2023), https://onlineacademiccommunity.uvic.ca/isot/wp-content/uploads/sites/7295/2023/02/ISOT_Fake_News_Dataset_ReadMe.pdf

Iyengar, Shanto, & Westwood, Sean (2014). "Fear and Loathing Across Party Lines: New Evidence on Group Polarization". *American Journal of Political Science*. **59** (3): 690–707.

Muhammed T S, Mathew SK. (2022) *The disaster of misinformation: a review of research in social media*. Int J Data Sci Anal. 2022;13(4):271-285. doi: 10.1007/s41060-022-00311-6. Epub 2022 Feb 15. PMID: 35194559; PMCID: PMC8853081.

Noever, D., Regian, J. W. (2022a). "The AI Director: From Document to Documentary?" accepted paper and presentation for IITSEC 2022, Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL 11/28-12/2/2022

Noever, D., Regian, J. W. (2022b). Chemical, Biological, Radiological, And Nuclear Operations, CBRNE Field Manual FM 3-11, https://deeperbrain.com/challenge/cbrnv1.mp4, based on 2019 Field Manual, https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1007035

Noever, D., Regian, J. W. (2022c). Cyber (CEMA) Field Manual FM 3-12, https://deeperbrain.com/challenge/cyberv3.mp4 , based on 2021 Field Manual, https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1022713

Noever, D., Regian, J. W. (2022d). Video Human Intelligence Collector Operations, https://deeperbrain.com/challenge/trainv6.mp4, based on 2006 Field Manual, https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=82535

RAND (2023). *Support for Extremism Among US Military Veterans Is Similar to Public at Large.* RAND News Release, 23 May. https://www.rand.org/news/press/2023/05/23.html

Whitesmith, M. (2020). *Cognitive Bias in Intelligence Analysis: Testing the Analysis of Competing Hypotheses Method*. Edinburgh University Press. http://www.jstor.org/stable/10.3366/j.ctv182jrtn