Journal of Simulation Engineering, Volume 3 (2022/2023). Article URL: https://articles.jsime.org/3/1

A Gamefied Synthetic Environment for Evaluation of Counter-Disinformation Solutions

Jesse Richman
jrichman@odu.edu

Lora Pitman ✉
lhadzhid@odu.edu

Girish S. Nandakumar
gnand002@odu.edu

Old Dominion University, Norfolk, VA, United States

ACM Subject Categories

- `Computing methodologies~Modeling and simulation`
- `Security and privacy~Social aspects of security and privacy`

Keywords

Disinformation, Simulation, Synthetic Environment

Abstract

This paper presents a simulation-based approach to developing strategies aimed at countering online disinformation and misinformation. This disruptive technology experiment incorporated a synthetic environment component, based on an adapted Susceptible-Infected-Recovered (SIR) epidemiological model to evaluate and visualize the effectiveness of suggested solutions to the issue. The participants in the simulation were given two realistic scenarios depicting a disinformation threat and were asked to select a number of solutions, described in Ideas-of-Systems (IoS) cards. During the event, the qualitative and quantitative characteristics of the IoS cards were tested in a synthetic environment, built after a SIR model. The participants, divided into teams, presented and justified their strategy which included three IoS card selections. A jury of subject matter experts, announced the winning team, based on the merits of the proposed strategies and the compatibility of the different cards, grouped together.

1 Introduction

Online disinformation (false information deliberately intended to mislead) has emerged as one of the most serious challenges in the era of digital information. For example, disinformation related to a pandemic, such as the COVID-19 one, can both exacerbate a health crisis and have implications for the cohesiveness and unity of international security organizations and institutions. Starting in early 2020, both state and non-state actors began carrying out disinformation campaigns aimed at exploiting the pandemic to instill fear, create distrust, and destabilize Western communities. Pandemic-related disinformation was used as a weapon to undermine NATO and U.S. forces in multiple countries such as Latvia, Poland, and Lithuania (BBC, 2020). Disinformation campaigns are slowing the response to the pandemic and weakening confidence in local authorities and international entities (e.g., WHO, NATO, EU). Examples of the harmful effects of these campaigns include fake letters and emails that aim to instill fear in communities which have a NATO presence.

The need for virtual environments or "synthetic environments" has been repeatedly recognized by NATO and by leading think tanks such as the Atlantic Council (Daw, 2005; Harper, 2020). Synthetic environments (henceforth referred to as SENs) such as flight simulators have also been in use continuously. Scenarios involving kinetic warfare can be modeled and simulated much more easily than scenarios involving non-kinetic aspects such as disinformation and strategic decision making. However, today's 'gray zone conflicts' (Chipman, 2018; Spitzack, 2018) have created a pressing need for simulation-based wargaming approaches to such non-kinetic topics. COVID-19 disinformation campaigns – the topic used in this experiment – is a suitable example for such an issue, requiring immediate attention. In the application reported here a SEN is adapted aimed at making people filter, refine, and combine the best solutions to the given problem (in the form of a scenario). Thus the virtual environment helps evaluate potential solutions to the disinformation problem being faced by NATO in a variety of domains.

This paper describes a successful application of SEN in the context of a wargame sponsored by NATO. It is the first study to describe the application of computational simulation methods to facilitate a virtual wargame in an international security context, with the application in this instance to strategies for combatting the spread of disinformation. Here the dynamics associated with COVID-19 disinformation served as a foundation for the scenarios used in the simulation. Much like a pandemic, disinformation and misinformation spread across communities and cast doubt in perceptions of security. Drawing on this parallel, a Susceptible-Infected-Resistant (SIR) model (Kermack & McKendrick, 1927) was chosen as the basis of the SEN for the war-simulation, described in this paper, to visualize and illustrate not only the detrimental and rapidly expanding consequences from disinformation, but also the potential solutions to this issue.

The study makes several contributions. First, it is a case study examining the implementation of SEN-based virtual war-game simulation that brought together participants in multiple NATO countries. Second, the SEN itself applies a novel SIR model customized to the problem of disinformation spread. Third, in the context of the SEN scenario case study, a series of new proposed technical strategies for combatting the spread of disinformation were tested through the wargame, providing a novel evaluation of these open-innovation-challenge sourced technological options. This paper's contributions thus include a case study evaluating the application of the SEN to multi-location virtual-wargaming by NATO, the modified SIR model which was the basis for the SEN, and the assessment and evaluation of the anti-disinformation technologies through the SEN-based virtual wargame.

The remainder of the paper proceeds as follows. Section 2 provides an overview of the structure and sequencing of components of the experiment. Section 3 introduces SIR epidemic models, and the history of their adaptation to the context of disinformation spread. Section 4 describes the integration of the virtual environment as a component of the virtual wargame, the purpose to which these were applied in this case: evaluating potential technological tools proposed to NATO for countering disinformation spread. Section 5 describes the results of the case study: how application of SEN as part of a virtual wargame played out, and the results of this application for the evaluation of the technology proposals. Section 6 outlines what was achieved with the simulation and the limitations of the experiment.

2 Project Structure

This study developed a SEN (Synthetic Environment) based on the SIR model as a core element of an internet-based virtual-wargaming exercise. The SEN was intended to use a distributed online format to help participants understand the problem of disinformation more deeply by modeling the dynamics that dictate the spread of both disinformation (i.e., false information intended to mislead) and misinformation (i.e., false information that is not spread with the intention to deceive) within social networks. At the same time it was also intended to help the organizers develop and evaluate solutions that can help counter such campaigns.

The simulation described in this paper presents an innovative approach that integrates a Disruptive Technology Experiment (DTEX). The Disruptive Technology Experiment (DTEX) is a NATO wargame designed by the NATO ACT Innovation Hub. DTEX is designed to test ideas and technologies that can solve problems for NATO. For this purpose, the simulation described in this paper was combined with the SEN that mimics the dynamics of disinformation and misinformation spread. The SEN, used in this simulation was an adaptation of an epidemiological SIR model used to understand the spread of diseases.

The overall experimental structure was as follows:

1. Building on a classic agent-based SIR model (Stonedahl & Wilensky, 2008), a model of the epidemic spread of disinformation in a network was created. This served as the SEN in the experiment.
2. Through an innovation challenge, proposed technological solutions to the challenge of disinformation spread were collected and summarized for experiment participants.
3. Experts rated the likely impact of the technological solutions for the parameters of the SEN.
4. Wargame participants were recruited, and two teams were created. Teams were briefed on the disinformation spread scenario and the technological solutions. Teams were given access to the SEN.
5. Teams communicated with each other using synchronous online communication to develop strategies involving selections of technological solutions.
6. Teams presented their solutions and were judged both on the basis of their presentation and on the duration of the disinformation 'infection' after the SEN parameters were modified based upon their proposed solution.
7. The research team evaluated the SEN and considered modifications to the model, and evaluated the proposed technological solutions.

This amalgamated approach has been used to test 46 suggested solutions to counter disinformation that were collected through an open innovation challenge – a competition between different individuals or entities intended to introduce a solution to a problem. The core activity in this simulation involved two teams which competed against each other to identify the best of the open-innovation-challenge sourced ideas that solved problems detailed in realistic scenarios. The teams in the disinformation wargame or DTEX assessed the merit of the ideas qualitatively and then quantitatively, using the SEN environment, to decide the best solutions for each scenario that they were given.

3 The SIR Model

3.1 Applicability of SIR Models to Disinformation Models

Susceptible-Infected-Recovered (SIR) models such as the one applied to create the synthetic environment (SEN) used in this study have a long history of application across many fields. These models began in epidemiology in the 1920s with work by William Ogilvy Kermack and Anderson Gray McKendrick (Kermack & McKendrick, 1927), but have also long been applied to study the transmission of ideas, narratives, and rumors (Goffman & Newill, 1964; Daley & Kendall, 1964). These models capture key aspects relevant to the spread of disinformation and misinformation in social networks, and provide a parsimonious way to characterize components of the strategic situation faced by those seeking to influence information spread.

SIR models include a population consisting of individuals or agents of at least three types: susceptible, infected, and recovered or resistant[1]. Transition probabilities in the SIR model govern the movement of agents from one state to another. Solutions for SIR models have been examined numerically, through simulations, and in recent years for specific parameter values exact analytical solutions have been computed as well (Harko et al., 2014).

The typical results of a SIR model run involve initial infection spread as infected individuals initially encounter mostly susceptible individuals. Then, a peak level of infection intensity as recovery and less availability of susceptible individuals balances new infections. Lastly, there is typically a decline in the number of infected individuals as recovery / resistance combine with diminished numbers of susceptible individuals to end the epidemic, often before all susceptible individuals have become exposed. There are several SIR model variants with alternative assumptions. For example, in the SIS model recovered agents remain susceptible, while in the SIR model, recovered agents are no longer susceptible. In the SIRS model, resistance to infection fades over time. The SEIHFR model has six categories, adding Exposed (but not yet symptomatic), Hospitalized (and thus perhaps less infectious), and Funeral (dead, not buried, and hence potentially still infectious) categories and has been used to model Ebola epidemics (Drake et al., 2015). The model variant used in this project allows for a possibility that infected agents who recover will transition to either the susceptible (S) or resistant (R) categories. Section 3.2 describes how we modified the standard SIR model to fit with the disinformation context.

SIR and related models have long been recognized as an effective framework for studying the spread of misinformation and disinformation. Key early work in the 1960s by Goffman and Newill (Goffman & Newill, 1964>) and Daley and Kendall (Daley & Kendall, 1964) pioneered the application of SIR and related models to the spread of information and rumors. These authors noted that the spread of ideas or information, like the spread of an infection, involved transmission from one individual to another, and that the SIR framework could provide a fruitful approach for modeling this process. At the same time, the models also account for a range of potential modifications such as effects of encountering other infected and/or resistant individuals.

The SIR model has been applied widely to information and idea transmission in fields including politics, economics, marketing, health, and communication. For example, recent work by Nobel prize winning economics professor Robert J. Schiller (Schiller, 2019), applies SIR epidemic models to understand the role of narratives in shaping economic behavior across a wide range of domains from speculation in Bitcoin to economic cycles, stock market bubbles, and many more. Work by Zhao, Weng and co-authors has expanded study of the spread of competing ideas and the dynamics of when and how ideas go 'viral' in social networks (Weng et al., 2012; Weng et al., 2013; Zhao et al., 2013). Bauckhage and colleagues examined attention to social media services (Bauckhage et al., 2014) and viral videos (Bauckhage et al., 2015). Internet memes can also be effectively modeled using an SIR framework, and Beskow and co-authors extended this work to study the evolution of political memes (Beskow et al., 2020). Across domains, epidemic models have provided useful insights into idea, information, and disinformation transmission.

One important distinction between the models involves whether agents assort at random or exist in a network structure. Random assortment is simpler to model for obvious reasons, but network structures often are particularly important for modeling transmission of ideas in realistic settings because they allow for differences in influence between actors. The most relevant models for the analysis of disinformation involve models with network effects and these models are often best analyzed using agent-based models in which the network structure can be directly analyzed (Ji et al., 2017). Infection of widely followed and trusted sources or sites has the potential to super-spread disinformation.

3.2 The SIR Synthetic Environment (SEN): Configuration and Settings

Because of the potential for greater realism in a network model, we model disinformation spread in an agent-based network. The networked disinformation spread model used to create the SIR based synthetic environment (SEN) in the wargame was developed by modifying and adapting the "virus on a network" SIR model presented by Stonedahl and Wilensky (2008). The model was programed in NetLogo, an open-source platform for agent-based modeling (Wilensky, 1999). For the purposes of the synthetic environment, the software was used to mimic and visualize the spread of disinformation. As mentioned previously in Section 3.1, related SIR models have a long history of application across many fields and in spite of their highly abstract and reductionist style, the SIR model can effectively capture the way in which disinformation spreads through a network of people.

Agents exist in a spatially clustered networked structure as in Stonedahl and Wilensky (2008). The configuration of our model, illustrated in Table 1, is made possible by initial settings which include the total `number-of-nodes` (agents in the SEN), the `average-node-degree`, showing how many other agents each agent in the SEN is connected to, the `initial-breakout-size`, depicting the scale of the disinformation spread, and by a series of transition or transmission probabilities which we describe below.

Each node can be in one of three states - susceptible (S), infected (I), and resistant (R) (Stonedahl & Wilensky, 2008). *Susceptible* (S) are vulnerable to disinformation due to low levels of awareness of the issue, lack of rational/critical thinking abilities, and/or other similar limitations. *Infected* (I) agents have been deceived by disinformation and perceive narratives spread by malicious actors as credible and trustworthy. Infected nodes tend to spread the information they have received and believed, thus becoming unwitting participants in the spread of disinformation or misinformation. Infected nodes are not always aware that they have been 'infected' at least until they 'fact-check'. Even those who do fact-check may still remain 'infected'. Therefore, not all infected nodes 'recover' from the condition of being infected. *Resistant* (R) agents are no longer vulnerable to disinformation due to fact-checking habits, high levels of awareness and rational/critical thinking abilities, and other cognitive and situational factors. The use of the term *resistant* which we adopt from Stonedahl and Wilensky (2008) is somewhat at variance with the use of the term *recovered* in some SIR models, but it is appropriate in our context as we distinguish between recovered agents.

Several parameters govern the transition of agents from one state to another. Infected agents spread disinformation to connected uninfected agents with a specified probability β. Infected agents also engage in fact checking with a specified frequency τ. When fact checking occurs, agents potentially recover (with a specified probability γ) with some failing to develop ongoing resistance to future infection by disinformation (returning to susceptible) and some developing resistance to future infection (with probability ρ.) Unlike most SIR models of disease, in the disinformation model, we also allow for the possibility that resistant agents connected with others infected with disinformation will push back, triggering additional fact checking. With a specified probability (ψ) a resistant agent may trigger fact checking among infected network connections and thereby potentially induce recovery to a susceptible state or the development of resistance.

In every step of the simulation (represented by a *tick*), each infected agent, marked by a red node, attempts to infect all of its connections with the disinformation. As a consequence, susceptible connections, marked with green nodes, may or may not get infected. The probability of infection is determined by β the `dis-information-spread-chance` setting. This characteristic represents the real-world equivalent of falling prey to a misleading headline, or to propaganda designed to elicit an emotional response favoring the actor spreading the false information. People that are *resistant*, marked with gray nodes, do not get infected. This represents the real-world equivalent of highly-aware people who have fact-checked and/or critically analyzed the disinformation and are no longer susceptible to it.

As opposed to this, *infected* people, marked with red nodes, are not always aware that they have been 'infected' by false information. In this model, every person has the potential to conduct a fact-check with a probability, which is controlled by τ, the `fact-check-frequency` setting. This represents the real-world event of a learning process in which an individual is being told by a person or an outlet they trust, in verbal or written form, that a particular piece of information is false.

If an agent successfully discovers through a fact check that they have indeed been 'infected', there is a chance that they might 'recover', i.e., get reliable and credible information. The probability of such a recovery is controlled by γ, the `recovery-chance` setting in the model. At the same time, a person's 'recovery' does not mean they will never get infected again. An appropriate analogy would be that one single human can get scammed or fall victim of phishing attacks many times. Therefore, some nodes may get infected again (modeled by a return to the susceptible group), some may not.

The probability of gaining this 'resistance' or 'immunity' is controlled by ρ, the `gain-resistance-chance` setting. When a person becomes resistant, the

Table 1. SEN Baseline Inputs constructing the environment in which DTEX was executed.

Variable Reference Code	Variable (SEN Slider)	Baseline Value	Explanation
Q	Total `number-of-nodes`	200	Represents the total number of "people" in the virtual world. This number will remain the same throughout the experiment. Everyone is interconnected and shares information constantly, i.e., during every tick. The tick is the only unit of time in this SEN. In the beginning of each simulation, every node is treated as being susceptible to disinformation. Susceptible nodes are represented as blue stick figures.
N	`average-node-degree`	20	The average number of 'people' each person is connected to. This number will remain the same throughout the experiment.
A	`initial-outbreak-size`	15	The initial number of 'bad actors' who have opinions that are factually incorrect. Bad actors are represented as yellow stick figures.
B	`disinformation-spread-chance`	5%	Represents the probability of yellow nodes spreading their opinions to their nodes in each tick.
T	`fact-check-frequency`	10 ticks	Represents how often each node fact-checks information before sharing it with others connected to that node. The baseline value indicates that, on average, each node fact-checks only 1 out of 10 times.
Γ	`recovery-chance`	5%	Represents the probability of a yellow node recovering from disinformation.
P	`gain-resistance-chance`	5%	Represents the probability of a node becoming immune to future disinformation altogether. Immune nodes are represented as green stick figures.
Ψ	`resistance-fact-check-chance`	0%	Represents the probability that a node which has become immune will 'push back' against disinformation by causing connected infected nodes to fact check.

links between them and their connections are darkened, since they are no longer possible vectors for spreading misinformation. Figure 1 shows a screenshot of the simulation in its final stage.

As a result of feedback concerning the match between epidemiological models and the disinformation context in the SEN, we also modified the SIR model to allow for the potential that resistant individuals might actively resist the spread of disinformation triggering fact checks by connected infected agents with probability ψ.

Figure 2 illustrates the impact of differences in the model parameters in the simulation. The key point is that the outcome of a model run is highly contingent upon the parameters. With the same starting values except for the frequency of fact checking (τ), the panel on the left follows a trajectory in which a severe infection develops (fact checking occurs only every 10 ticks). The panel on the right follows a trajectory in which a more rapid development of resistance more rapidly ends the spread of disinformation and prevents it from ever simultaneously attracting a majority of the population (fact checking occurs every tick).

All simulation parameters could potentially be influenced by the teams playing the DTEX wargame through their strategic choices, as will be discussed in Section 4. This modification of parameters was one of the two ways the wargame-based test of the implementation of the anti-disinformation-spread technologies was evaluated. One half of the choice of the winning team was based upon which team's SEN inputs led to the most rapid elimination of the disinformation in the model (the lowest number of ticks at the end of the simulation).

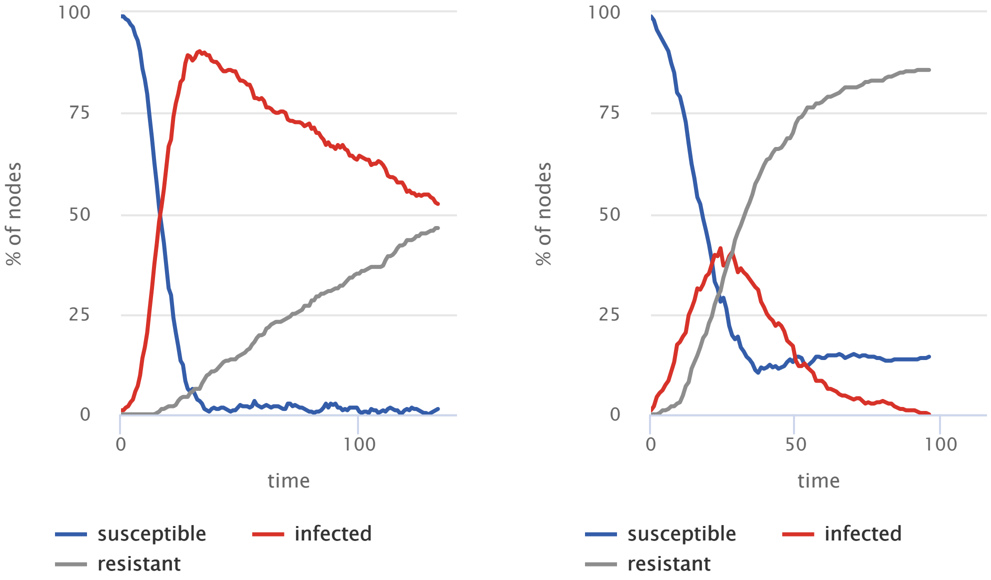
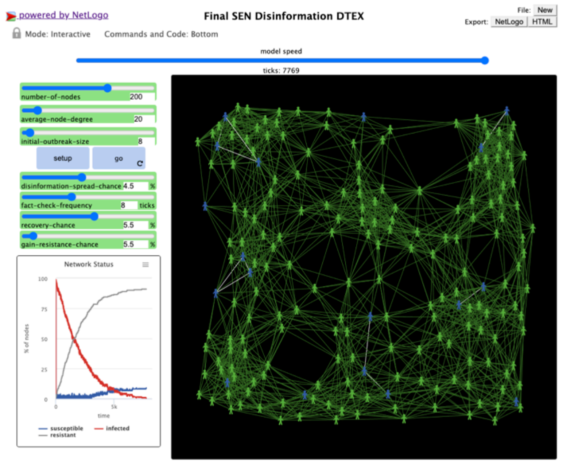


Figure 2. Example model runs with different fact check frequencies.



Final SEN Disinformation DTEX

Mode: Interactive   Commands and Code: Bottom

model speed

ticks: 7769

number-of-nodes 200

average-node-degree 20

initial-outbreak-size 8

disinformation-spread-chance 4.5 %

fact-check-frequency 8 ticks

recovery-chance 5.5 %

gain-resistance-chance 5.5 %

Figure 1. The Synthetic Environment (SEN) used in the simulation.

Teams were also judged on their argument concerning the choice of technologies and the strategy for deploying them.

4 DTEX War Game

4.1 DTEX Process

The DTEX Process used in this simulation was adapted from NATO's Disruptive Technology Assessment Game (DTAG) structure. The latter "is a table-top seminar wargame, used to assess potential future technologies and their impact on military operations and operating environment" (NATO ACT, 2010). Similarly to DTAG, DTEX also adopts the seminar wargame core, but reveals some more nuances in the way the simulation was conducted - in a fully online, synchronous environment.

The DTEX Process, illustrated in Figure 3, incorporated five steps, as follows. First, the participants studied the scenario and the issues described in it. The exact text of the scenarios can be found in Appendix 2. They were also given some supplementary materials and had the opportunity to receive guidance about the scenario and the solutions from a facilitator. Second, the participants reviewed the IoS cards (see Appendix 3) with proposed solutions. Third, each participant individually made a choice of three IoS cards which they found suitable to resolve the issues at hand. Fourth, participants discussed their choices with their teams and debated the rationales behind their choices. Fifth, each of the two teams deliberated on a final selection of IoS cards, based on the merits of the suggested solutions, their combined, synergetic effects, and the impact of the entire set of cards, as tested in SEN. After this process was completed, the participants prepared one-slide presentations with their choices, defended their strategy, and the winner was announced by a subject matter expert, who served as a judge.

4.2 Scenarios

The scenarios with which the participants in the simu-

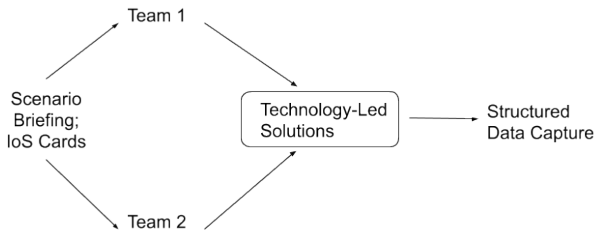


Figure 3. The DTEX Process.

lation were presented focused on social media disinformation. They presupposed that the Supreme Allied Commander Transformation (SACT) formed a small task force that will assist an Allied Command Operations (ACO) team in the ongoing fight against disinformation and the participants were a part of it. Next, they were asked to select three IoS cards (described in Section 4.3) which addressed the various specific issues underscored in both scenarios. The teams qualitatively evaluated the merits of each IoS card (and the combined impact of the chosen cards) and after they made their final choice of IoS cards, the quantitative effects of their choice of IoS cards, based on the expert ratings, was also tested in the SEN provided to them and their facilitator. Teams did not have direct access to the expert ratings of the cards. The faster the SEN eliminates the spread of dis/misinformation (fewer ticks to elimination), the better. The winning team was chosen based on both their rationale for their IoS card choices and on the temporal impact of their choices within the SEN. Equal weight was given to these two criteria to make sure that the solution is supported by qualitative and quantitative factors.

4.3 DTEX IoS Cards

As mentioned in Section 4.1 and Section 4.2, scenario play involved a choice of IoS cards by participants. As for the structure of the IoS cards, as shown in Figure 4, each card consists of various sections describing the technology intended to serve as a solution to the problem of disinformation on social media. In the first one, called *offerings*, the objectives of the technology are outlined, and then the technology itself is introduced through a brief overview. Next, the second section of the cards summarizes the input, the output, the process the technology is using to achieve its goals, and the supported technologies in which it will operate. The third and last section of the cards highlights advantages and limitations of the technology. The purpose of this section is to guide participants in their choices, as they could not obtain information about the proposed technologies directly from the contributors in the NATO Innovation Challenge through which these ideas were gathered. Description of the features of all IoS cards is available in the Appendix 3.

In addition to the content summary of each card, the subject matter experts invited to contribute to this simulation assigned each IoS card a specific impact. The latter was expressed in numerical value calculated as the average of the expert ratings and contributed to visualizing the solutions in SEN. Figure 5 shows the worksheet with all of the IoS cards' SEN inputs that was compiled and used by the facilitators to coordinate the team's activities and to process the inputs in SEN for the participants during the simulation.

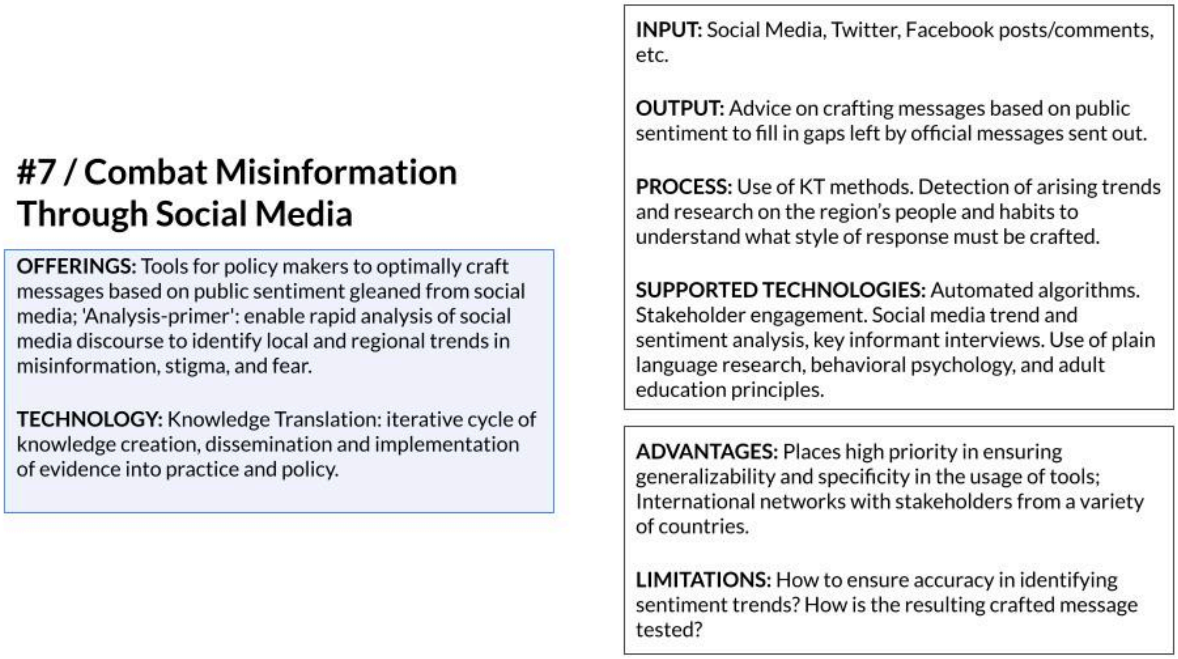
Each of the categories of impact on the SEN (A through E) shapes elements of the simulation environment (e.g., fact check frequency τ, probability of disinformation spread β, etc.). Participants did not directly receive information about the ratings on the cards they received, but the ratings informed the way in which the simulated SIR model in the scenarios was modified as a result of group choices. The rated impacts of the cards are discussed in Section 5.2.

4.4 The role of the participants

As mentioned in Section 4.2, the participants in the experiment were asked to select three IoS cards and explain why they are the best choices to address the issues highlighted in the scenario. The participants also had to identify the priorities to which they adhered when choosing the cards. These priorities included five different objectives - *identification* of malicious communication material online, *categorization* of information (real vs. fake), *attribution* (finding sources of fake information), *additional analyses* (processing and analysis of collected information to fulfill other objectives), *visualization* of analyses, and *mitigation of effects* (countering disinformation and their effects by shielding the audience being targeted, disseminating counternarratives, etc.) After completing the selection of IoS cards, the participants were invited to test their choices in the SEN, where both the individual effects of their choices and their combined synergetic effects were visualized and assessed. Lastly, during a *confrontation session* between the different teams, the participants presented their proposed plan to the jury, which consisted of subject matter experts on the topic of disinformation.

5 Results

This section discusses the results of the DTEX simulation. The DTEX event was well organized, the basic structure of the simulation worked well, and participants found the SEN a useful component in conjunction with their deliberations. Participants used the SEN during their deliberations to visualize the consequences of different strategies. The SEN was also used as one component of judging team decision-making. It also helped or-



#7 / Combat Misinformation Through Social Media

OFFERINGS: Tools for policy makers to optimally craft messages based on public sentiment gleaned from social media; 'Analysis-primer': enable rapid analysis of social media discourse to identify local and regional trends in misinformation, stigma, and fear.

TECHNOLOGY: Knowledge Translation: iterative cycle of knowledge creation, dissemination and implementation of evidence into practice and policy.

INPUT: Social Media, Twitter, Facebook posts/comments, etc.

OUTPUT: Advice on crafting messages based on public sentiment to fill in gaps left by official messages sent out.

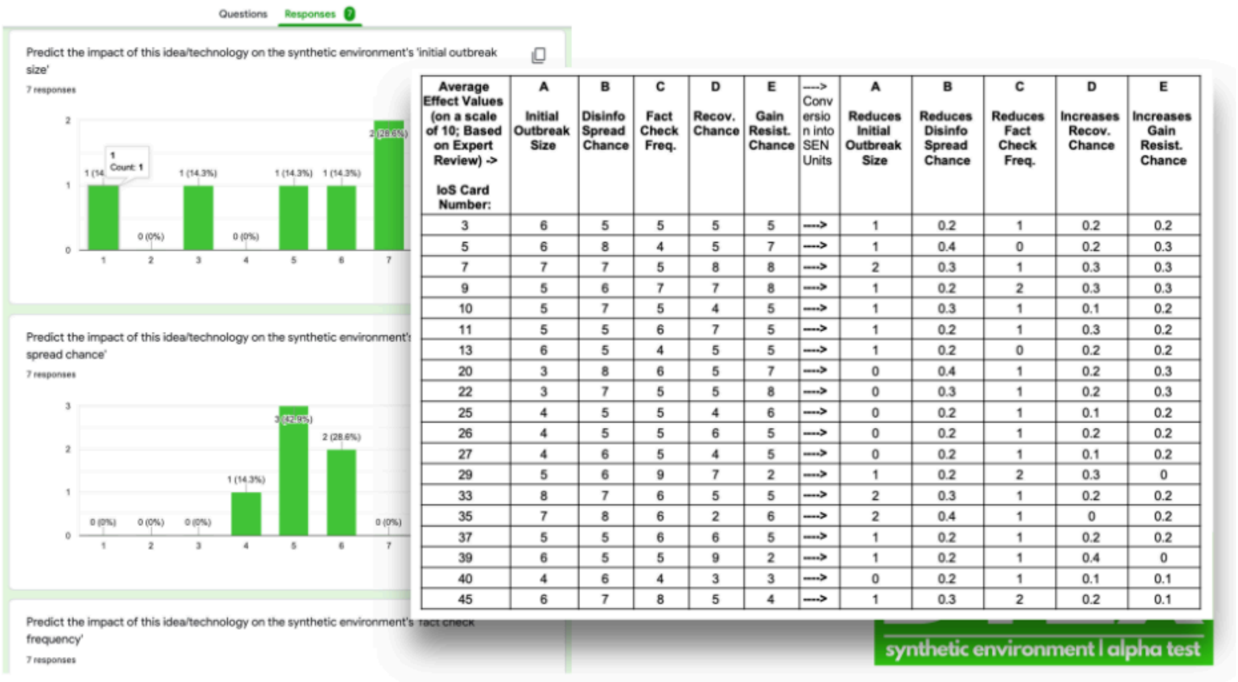
PROCESS: Use of KT methods. Detection of arising trends and research on the region's people and habits to understand what style of response must be crafted.

SUPPORTED TECHNOLOGIES: Automated algorithms. Stakeholder engagement. Social media trend and sentiment analysis, key informant interviews. Use of plain language research, behavioral psychology, and adult education principles.

ADVANTAGES: Places high priority in ensuring generalizability and specificity in the usage of tools; International networks with stakeholders from a variety of countries.

LIMITATIONS: How to ensure accuracy in identifying sentiment trends? How is the resulting crafted message tested?

Figure 4. Outline of an IoS Card.



Predict the impact of this idea/technology on the synthetic environment's 'initial outbreak size'

7 responses

Predict the impact of this idea/technology on the synthetic environment's spread chance'

7 responses

Predict the impact of this idea/technology on the synthetic environment's fact check frequency'

7 responses

Average Effect Values (on a scale of 10; Based on Expert Review) -> IoS Card Number:	A Initial Outbreak Size	B Disinfo Spread Chance	C Fact Check Freq.	D Recov. Chance	E Gain Resist. Chance	---> Conversion into SEN Units	A Reduces Initial Outbreak Size	B Reduces Disinfo Spread Chance	C Reduces Fact Check Freq.	D Increases Recov. Chance	E Increases Gain Resist. Chance
3	6	5	5	5	5	--->	1	0.2	1	0.2	0.2
5	6	8	4	5	7	--->	1	0.4	0	0.2	0.3
7	7	7	5	8	8	--->	2	0.3	1	0.3	0.3
9	5	6	7	7	8	--->	1	0.2	2	0.3	0.3
10	5	7	5	4	5	--->	1	0.3	1	0.1	0.2
11	5	5	6	7	5	--->	1	0.2	1	0.3	0.2
13	6	5	4	5	5	--->	1	0.2	0	0.2	0.2
20	3	8	6	5	7	--->	0	0.4	1	0.2	0.3
22	3	7	5	5	8	--->	0	0.3	1	0.2	0.3
25	4	5	5	4	6	--->	0	0.2	1	0.1	0.2
26	4	5	5	6	5	--->	0	0.2	1	0.2	0.2
27	4	6	5	4	5	--->	0	0.2	1	0.1	0.2
29	5	6	9	7	2	--->	1	0.2	2	0.3	0
33	8	7	6	5	5	--->	2	0.3	1	0.2	0.2
35	7	8	6	2	6	--->	2	0.4	1	0	0.2
37	5	5	6	6	5	--->	1	0.2	1	0.2	0.2
39	6	5	5	9	2	--->	1	0.2	1	0.4	0
40	4	6	4	3	3	--->	0	0.2	1	0.1	0.1
45	6	7	8	5	4	--->	1	0.3	2	0.2	0.1

Figure 5. Final SEN impacts for IoS Cards.

ganize and structure discussion of the merits of different technologies aimed at combatting the spread of disinformation. A framework of two scenarios (see Appendix 2 for details) of increasing complexity was deemed appropriate, and seemed to help engender participant interest, engagement, thought, and analysis.

5.1 Group Dynamics Qualitative Observations

In the first scenario, Group 2 seemed less organized than Group 1. Group 1 used screen share capabilities more effectively to help ground discussion of alternative cards, while Group 2 seemed to struggle a bit more to reach consensus, and as a result did not develop as effective and clear a set of plans for how to address the challenges in the scenario, nor how to present their plans.

In the second scenario, one of the members of Group 2 opened the discussion with a proposal that helped set the tone for a more productive deliberative process which set the stage for the Group 2 win in scenario 2. With her leadership they identified goals and reached consensus about them. Then they developed a combination of technology cards that would allow them to effectively achieve those goals. The structure of the deliberations could have potentially benefitted from more involvement by the moderators and a division of the cards into different categories (e.g., dashboards versus tools for intervention). By the second scenario, Group 2 seemed to have begun to do this kind of sorting of cards into categories on its own, and that process helped the group reach a more effective path to a solution, while Group 1 in the second scenario seemed to have more trouble structuring their deliberations and combining the synergies of the cards. Group 2 reached near-consensus with sufficient time remaining for multiple model runs in the SEN to test which of two alternative strategies would lead to better results. Ultimately, choice of the strategy rejected by Group 2 through this process would have led to less successful model runs than Group 1, and potentially to a loss in scenario 2, so the time the group was able to invest in this aspect of the deliberation seems to have been well spent.

The group dynamics described highlight some of the skills and approaches which determined the winning group. In particular, leadership, level of organization and structure of the decision-making process, along with an effective use of the technical capabilities of the SEN to which the participants had access contributed to Group 1's better performance in the first scenario, and Group 2's in the second scenario. These conclusions about the group dynamics in DTEX provide important insights for the successful selection process of technological solutions with a high level of impact against disinformation. They may be used in future iterations of this simulation to increase the productivity and competitiveness of both teams, thus ensuring a better learning experience for the participants and a more careful re-assessment of the IoS cards, previously ranked by experts, based on their characteristics.

5.2 IoS Cards: Strengths and Synergies

As noted at the outset, the purpose of the SEN (SIR model) and wargame virtual simulation in this case was to evaluate proposed anti-disinformation technological tools submitted to NATO through an innovation challenge. This section discusses the results of that evaluation which is based upon the totality of the information collected including the actions and arguments made by wargame participants, expert rankings, and simulation results.

Prior to the DTEX wargame the IoS cards were ranked by experts for their ability to impact five different characteristics of disinformation spread in the SEN, and then evaluated by the competing teams to construct compelling and synergistic combinations of the cards. The characteristics were: A - Reduces Initial Outbreak Size, β - Reduces Disinformation Spread Chance, τ - Increases Fact Check Frequency, γ - Increases Recovery Chance, and ρ - Increases Gain Resistance Chance. The probability that a resistant agent will trigger a fact check by a connected infected agent (Ψ) was added after DTEX based on the simulation experience and so is not included in this section. Based upon the expert rankings and the results of the wargame, including qualitative analysis of participant discussion and arguments we have categorized each card in Table 2 in terms of the best card(s) for addressing each aspect.

Containing initial outbreak size is potentially very important, especially if once the outbreak is identified, effective tools are available to curtail the spread of the outbreak. Card #33 was rated as providing the best impact on initial outbreak size. This technology provides a dashboard for decision-makers that "monitors all aspects of the spread of information (about COVID-19) and predicts what and how other topics will spread." The key aspect of this platform for curtailing initial outbreak size is that ideally this platform will allow rapid identification of outbreaks of disinformation, allowing agile targeting responses to those outbreaks using various other tools before the outbreaks have time to become widespread.

Once an outbreak of disinformation has begun, a critical factor shaping its spread is the extent to which individuals or media *infected* with disinformation spread it to others. The three best-rated cards for curtailing the disinformation spread chance were implemented in different strategies, suggesting potential for fruitful combination between these cards for larger impact. IoS card #5 SGOOF uses data-mining, classification, and machine learning classification to develop a 'truth score' and classification for information. This could be fed into a dashboard similarly to #33, but it also could potentially be used in public-facing applications. IoS card #20 DeepDetector is a more specialized software application aimed at detecting and identifying deep-fakes in video

Table 2. Cards with largest impacts on each aspect of SEN based upon expert ratings.

A Reduces Initial Outbreak Size	β Reduces Disinformation Spread Chance	τ Increases Fact Check Frequency	γ Increases Recovery Chance	ρ Increases Gain Resistance Chance	Average Impact Z-score
Best: #33. Covid-19 MAP Media Analytics Platform. **Second Best**: A tie between #7, Combat Misinformation through Social Media, and #35 Profiling fake news spreaders on Social Media.	**Best**: A three way tie between #20 DeepDetector, #5 SGOOF, and #35 Profiling fake news spreaders on Social Media.	**Best**: #29 Intelligence Dashboard **Second Best**: #45 mLAi Analytics.	**Best**: #39 PULSE **Second Best**: #7 Combat Misinformation Through Social Media.	**Best**: A three way tie between #7 Combat Misinformation Through Social Media, #9 Zetane, and #22 Nunki.	**Best**: #7 Combat Misinformation Through Social Media.

footage. The current prototype is asserted to have a 95-98% accuracy and could provide an important tool both if fed into a dashboard and as a public-facing application to allow for rapid identification of likely faked video content in order to catalyze actions to limit its spread. Another IoS card - #35 Profiling fake news spreaders on Social Media takes a somewhat different tactic. Potentialize synergizing with #5 and #20, this machine learning application focuses on the profiles of fake news spreaders instead of on the news content itself. This could provide particularly valuable information in order to facilitate rapid response to the spread of fake news that targets accounts being used to spread disinformation.

Once disinformation has begun to spread widely, combatting it involves in part triggering fact checking that potentially leads individuals to believe they should not trust the disinformation. The best rated card for increasing fact check frequency was #29 Intelligence Dashboard. This dashboard proposal utilizes a combination of AI and human fact checking to identify and classify the most prevalent information. As with other dashboard proposals, the primary focus here is on enabling decisionmakers to take effective actions to increase fact check frequency or provide targeted individuals with fact checks of disinformation which they have been exposed to. Individuals who have come to believe disinformation may eventually *recover* by believing fact checks which disabuse them of belief in the false narratives provided by the disinformation source. The best rated card for increasing recovery chance was #39 PULSE. This proposal emphasizes the important counter-insurgency principle that all combatants are intelligence gatherers. It provides a framework for submissions from "front-line workers" to identify and cluster information on unaddressed issues and challenges. This could be an important component of any dashboard, helping decision-makers operate with better information concerning the current state of play in the spread of disinformation, and potentially facilitating the identification of unaddressed issues.

A key factor in ultimately containing a disinformation outbreak is the development of resistance to it in the form of individuals who are no longer susceptible to the disinformation. Three technology cards received the highest ratings for this element: #7, #9, and #22, and pursue two quite distinct strategies that would need to be synergized for the largest impact. IoS card #7 aims to achieve resistance through counter-spreading measures, a unique and very important aspect of this card compared to most of the other proposed technologies. In essence, the strategy behind using it is to achieve resistance to disinformation by identifying potential spreaders, and swamping the disinformation signal with alternative signals. This more active resistance by jamming disinformation signals moves beyond most other cards which emphasize identification of disinformation rather than active counter-information measures. Card #9 Zetane is a dashboard that aids in visualization of the geographic and regional trends in false information spread. #22 Nunki is another dashboard application which focuses on alerts concerning events and news spread, hopefully facilitating rapid response. Obviously, the dashboard applications would be most fruitfully combined with other measures, such as IoS card #7, since with dashboard strategies the resistance developed would involve societal level rapid-response to renewed spread of disinformation.

Fortunately, as discussed above, multiple technologies can be combined to address the challenges of disinformation. However, if only a single technology was to be used, the best overall technology in terms of impact relative to the others across the five categories is #7 Combat Information Through Social Media. What makes this strategy stand out is its emphasis on active measures. The high ratings given this card suggest that efforts to develop a suite of different active signal-jamming measures to combat disinformation would be well worth while. Combination of such measures with good dashboard and intelligence to identify threats would probably help to magnify the effectiveness of this technology.

6 Conclusions

The simulation involving a virtual wargame using SEN succeeded across several dimensions. The DTEX project, described in this paper, set forth multiple objectives – producing ideas, testing them in a realistic scenario and observing the visualized effects of these ideas, educating the participants about the harmful effects of disinformation and the strengths and weaknesses of possible solutions, and testing the use of an internet-based virtual wargame. The fact that DTEX was conducted in a fully-online environment was also a step forward toward making such simulations and wargames more accessible across nations and thus more inclusive, diverse, and valuable. Another benefit of DTEX was that it created a collaborative setting in which participants from different backgrounds can contribute, as disinformation is a multidisciplinary topic that is researched by scholars and practitioners from various fields. The DTEX model also outlined opportunities for development and testing of solutions that pertain not only to other similar-to-disinformation issues, such as propaganda, and recruitment by radical organizations, but also to a wide range of other security issues, important to the international community.

One of the key elements of the DTEX war game scenario design involves the opportunity for groups to deliberate and play out the interaction between multiple technologies, as no single technology is likely to solve all of the problems presented by the scenarios, but some technologies are more compatible with each other than others. Deliberations about the tradeoffs between technologies provide important data about the challenges associated with integrating diverse (and potentially overlapping or competing) technologies to solve a problem, and their potential synergies. Hence, the experiment succeeded in building knowledge about the potential of the technology choices and the ways in which they could be effectively combined.

Another of the key elements of this study involved the use of SENs to facilitate interaction and evaluation in the context of a virtual wargame. Because the wargame was played out virtually, participants could be physically located in multiple NATO countries on multiple continents. By applying an epidemic-spread model to depict the spread of disinformation about the COVID-19 pandemic, these environment help participants visualize, conceptualize, apply, and analyze the consequences of the potential technological solutions for disinformation spread. The simulation as a case study demonstrated the utility of the SIR simulation as SEN for the virtual wargame.

In the process of describing our study, we also modified the SIR model to better capture some dynamics of disinformation flow, and those modifications (e.g., the possibility that resistance itself may be 'catching') can be incorporated into subsequent models of disinformation.

There were none the less some important limitations of this experiment. While the diversity of backgrounds of participants was a significant asset to the experiment, it also revealed some inequality in terms of how to best respond to the given scenario. For instance, students from political science backgrounds generally demonstrate more awareness about the way NATO is structured and how the different member-states work together. At the same time, they may not be equipped to assess the various technologies that were presented to them in the form of IoS cards from a more technical perspective. Another issue pertains to the ability to operate the SEN in which the cards were tested. In a fully asynchronous environment, which has the ability to overcome limitations of different time-zones, facilitators may not be able to be as helpful as they were in the synchronous online version of DTEX which this paper describes.

Aside from these limitations, the goals for which DTEX was designed and intended – innovation, education and collaboration, were successfully fulfilled mainly because of the virtual environment that helped participants. With the input and efforts of specialists from various fields, the simulation will further evolve and attempt to solve more of the problems of the future.

Acknowledgements

The authors of this paper would like to thank and acknowledge the NATO Innovation Hub and its partners for making DTEX possible and to express their appreciation for the helpful feedback of the anonymous reviewers who helped us improve this work.

Bibliography

Bauckhage, C., Kersting, K., & Rastegarpanah, B. (2014). Collective Attention to Social Media Evolves According to Diffusion Models. WWW'14 Companion: Proceedings of the 23rd International Conference on World Wide Web. Seoul: ACM.

Bauckhage, C., Hadiji, F., & Kersting, K. (2015). How Viral Are Viral Videos? *Proceedings of the 9th Conference on Web and Social Media*. Oxford. PDF

BBC. (July 30, 2020). *Hackers Post Fake Stories on Real News Sites 'to Discredit NATO'*. HTML

Beskow, D. M, Kumar, S., & Carley, K. M. (2020). The Evolution of Political Memes: Detecting and Characterizing Internet Memes with Multi-Modal Deep Learning. *Information Processing and Management*, 57(2), 102170. doi

Chipman, J. A. (November 16, 2018). New Geopolitical Challenge to the Rules-Based Order. *The International Institute for Strategic Studies*. HTML

Daley, D. J. & Kendall, D. G. (1964). Epidemics and Rumours. *Nature*, 204, 1118. doi

Daw, A. J. (2005). On the Use of Synthetic Environments for the Through Life Delivery of Capability. In *Analytical Support to Defence Transformation* (pp. 9-1 – 9-16). Meeting Proceedings RTO-MP-SAS-055, Paper 9. Neuilly-sur-Seine: RTO. HTML

Drake, J. M., Bakach, I., Just, M. R., O'Regan, S. M., Gambhir, M., & Fung, I. C. (2015). Transmission Models of Historical Ebola Outbreaks. *Emerging Infectious Diseases*, 21(8), 1447-1450. doi

Goffman, W. & Newill, V. A. (1964). Generalization of Epidemic Theory: An Application to the Transmission of Ideas. *Nature*, 204, 225-228. doi

Harko, T., Lobo, F., & Mak, M. K. (2014). Exact Analytical Solutions of the Susceptible-Infected-Recovered (SIR) Epidemic Model and of the SIR Model with Equal Death and Birth Rates. *Applied Mathematics and Computation*, 236, 184–194. doi

Harper, C. (2020). 'Game out' decision making. *Atlantic Council*. HTML

Ji, S., Lü, L., Yeung, C. H., & Hu, Y. (2017). Effective Spreading from Multiple Leaders Identified by Percolation in the Susceptible-Infected-Recovered (SIR) Model. *New Journal of Physics*, 19(7), 073020. doi

Kermack, W. O., & McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceediings of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 115(772), 700-721. doi

NATO ACT. (2010). *Disruptive Technology Assessment Game 'DTAG'*, Handbook v0.1. PDF

Schiller, R. J. (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton, NJ: Princeton University Press. doi

Spitzack, C. A. (2018). *Gray Is the New Black: Great Power Competition in the Gray Zone* (Publication No. 0000-0003-1707-0956) [Master's Thesis, The University of Texas at Austin], TexasScholar Works. HTML

Stonedahl, F. & Wilensky, U. (2008). *NetLogo Virus on a Network model*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. HTML

Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among Memes in a World with Limited Attention. *Scientific Reports*, 2(1), 335. doi

Weng, L., Menczer, F., & Ahn, Y. (2013). Virality Prediction and Community Structure in Social Networks. *Scientific Reports*, 3(1), 2522. doi

Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. HTML

Zhao, L., Cui, H., Qiu, X., Wang, X., & Wang, J. (2013). SIR Rumor Spreading Model in the New Media Age. *Physica A: Statistical Mechanics and its Applications*, 392(4), 995–1003. doi

Copyright Information

Appendix 1. SIR model code

```
turtles-own
[
  infected?           ;; if true, the turtle is infectious
  resistant?          ;; if true, the turtle can't be infected
  fact-check-timer    ;; number of ticks since this turtle's last fact-check
]

to setup
  clear-all
  setup-nodes
  setup-spatially-clustered-network
  ask n-of initial-outbreak-size turtles
    [ become-infected ]
  ask links [ set color white ]
  reset-ticks
end

to setup-nodes
  set-default-shape turtles "circle"
  create-turtles number-of-nodes
  [
    ; for visual reasons, we don't put any nodes *too* close to the edges
    setxy (random-xcor * 0.95) (random-ycor * 0.95)
    become-susceptible
    set fact-check-timer random fact-check-frequency
  ]
end

to setup-spatially-clustered-network
  let num-links (average-node-degree * number-of-nodes) / 2
  while [count links < num-links ]
  [
    ask one-of turtles
    [
      let choice (min-one-of (other turtles with [not link-neighbor? myself])
                   [distance myself])
      if choice != nobody [ create-link-with choice ]
    ]
  ]
  ; make the network look a little prettier
  repeat 10
  [
    layout-spring turtles links 0.3 (world-width / (sqrt number-of-nodes)) 1
  ]
end

to go
  if all? turtles [not infected?]
```

```
    [ stop ]
  ask turtles
  [
     set fact-check-timer fact-check-timer + 1
     if fact-check-timer >= fact-check-frequency
       [ set fact-check-timer 0 ]
  ]
  spread-disinformation
  do-fact-checks
  tick
end

to become-infected  ;; turtle procedure
  set infected? true
  set resistant? false
  set color red
end

to become-susceptible  ;; turtle procedure
  set infected? false
  set resistant? false
  set color blue
end

to become-resistant  ;; turtle procedure
  set infected? false
  set resistant? true
  set color gray
  ask my-links [ set color gray - 2 ]
end

to spread-disinformation
  ask turtles with [infected?]
    [ ask link-neighbors with [not resistant?]
        [ if random-float 100 < disinformation-spread-chance
            [ become-infected ] ] ]
end

to do-fact-checks
  ask turtles with [infected? and fact-check-timer = 0]
  [
    if random 100 < recovery-chance
    [
      ifelse random 100 < gain-resistance-chance
        [ become-resistant ]
        [ become-susceptible ]
    ]
  ]

  ask turtles with [infected? and any? link-neighbors with [resistant?] ]
  [
    if random 100 < resistance-fact-check-probability
    [
      if random 100 < recovery-chance
      [
        ifelse random 100 < gain-resistance-chance
          [ become-resistant ]
          [ become-susceptible ]
      ]
    ]
  ]
end
```

NOTE: This model is a modified version of the NetLogo Virus on a Network model (Stonedahl & Wilensky, 2008), copyright 2008 Uri Wilensky. This work is licensed under the Creative Commons Attribution-NonCommercial-Share-Alike 3.0 License.

Appendix 2. DTEX scenarios

Scenario 1

Background

The Supreme Allied Commander Transformation (SACT) has handpicked you for a small task force that will assist an Allied Command Operations (ACO) team in the ongoing fight against disinformation. You have been asked to pick 5 technologies (IoS cards) that you believe will help solve the problems described in the following scenario.

Note: Please stick to the details given in the IoS cards. The only creative license you can take is during the prediction and explanation of the outcomes in the future (where the technologies you've chosen will be implemented). Feel free to ask questions about the scenario, operating environment, and IoS cards. Your facilitator will be your main point of contact and will be available in your zoom room at all times.

Description

1. In the midst of increased fear about new waves of COVID-19, there has been a barrage of fake posts across several social media platforms in multiple languages claiming that there has been large outbreaks of COVID-19 within NATO forces that are part of the Enhanced Forward Presence - a NATO-allied forward deployed defense and deterrence military posture in Central Europe through Poland and Northern Europe through Estonia, Latvia, and Lithuania.
2. NATO analysts have noticed that the dissemination of disinformation is happening largely through numerous small-scale 'influencers' - whose accounts are getting hacked or imitated. These accounts are spreading different messages depending on the populations they're targeting.
3. Highly graphic visuals and deep-fake videos are being used to depict highly dramatized scenes that are far from reality yet convincingly real. Videos with fake information - in the form of text alongside images - are the primary vectors. These videos seem to be designed to elicit strong emotional responses that seem to have the ultimate goal of creating a rift within NATO.
4. These social media posts are also well crafted. The language and cultural contexts are too good for AI to differentiate easily. Human-AI partnerships may be necessary. The type of fake personalities delivering these fake news reports also seem to be very effective in making the message look authentic. Forensic psychologists at NATO claim that they will be able to solve part of the disinformation problématique if more information about these 'talking heads' were made available to them.
5. The populations that were targeted by these disinformation attempts need to be identified in order to target mitigation efforts towards the same population. Managing such efforts also require dashboards that aggregate and visualize data using maps and other tools.

You can use details from the following reports/articles to guide and support your choices of IoS cards:

- Canadian-led NATO battlegroup in Latvia targeted by pandemic disinformation campaign
- Hackers Broke Into Real News Sites to Plant Fake Stories
- Pillars of Russia's Disinformation and Propaganda Ecosystem (Infographics on pages 8, 10)

Expectations

1. **Pick five IoS cards** and explain why you think they are the best choices to address the issues.
2. **Develop a plan** that leverages the five IoS cards you chose - both their individual strengths and their combined synergies. This plan should counter or mitigate the effects of disinformation campaigns. Explain how your IoS cards can combine their strengths.
3. **Present your plan to the jury**, during the 'confrontation session' with the other team and convince them that your plan is the better one. Focus on explaining (a) how you plan to use the IoS cards and how you plan to combine their strengths, and (b) what effects you intend to achieve through your plan. Below is the full list of desired effects:
    1. **Identification** of malicious communication material online
    2. **Categorization** of information (real v. fake)
    3. **Attribution**: Finding sources of fake information
    4. **Additional Analyses**: Processing and analysis of collected information to fulfill other objectives
    5. **Visualization** of analyses
    6. **Mitigation of Effects**: Countering disinformation and their effects by shielding the audience being targeted, disseminating counternarratives, etc.

Scenario 2

Background

The Supreme Allied Commander Transformation (SACT) has once again handpicked you for a small task force that will assist an Allied Command Operations

(ACO) team in the ongoing fight against disinformation. You have been asked to pick 5 technologies (IoS cards) that you believe will help solve the problems described in the following scenario.

Note: Please stick to the details given in the IoS cards. The only creative license you can take is during the prediction and explanation of the outcomes in the future (where the technologies you've chosen will be implemented). Feel free to ask questions about the scenario, operating environment, and IoS cards. Your facilitator will be your main point of contact and will be available in your zoom room at all times.

Description

1. NATO teams have been monitoring COVID-19-related disinformation efforts for a while but are still not able to efficiently sort disinformation. Both bots and humans have been actively spreading disinformation but the teams are not able to differentiate the sources. These efforts seem to be targeting civilian populations across NATO nations. These disinformation campaigns are somehow able to target populations that seem to have low levels of awareness of the real nature of the pandemic and of the best practices to prevent spread. Experts suggest that such targeting is meant to spread anxiety about the future.
2. Troves of data have been collected by NATO teams which have been analyzing these bots. However, analysts are no longer able to extract actionable insights from these datasets. Team leaders have been affected by sensory overload caused by ineffective tools that are not able to aggregate and analyze such datasets.
3. Analysts have been manually aggregating and visualizing data points to present the big picture to their leaders and other decision makers. This has been drastically slowing down reaction times, allowing disinformation campaigns to spread virally in the meantime. Team leaders are skeptical of tools that oversimplify analyses because they believe they can lead to serious oversights. Analysts are not able to find tools that strike the right balance between sensory overload and potentially irresponsible reductionism.
4. NATO's sociologists and other interdisciplinary researchers are also not able to extract useful insights from these large datasets. Their goal is to connect bits and pieces, highlight similar narratives, and craft better counter-narratives and responses. These experts are also unable to obtain real time feedback on the spread of disinformation.
5. NATO is interested in using these large datasets to forecast future trends. Team leaders and policy makers currently lack such tools in their planning and decision-making processes.

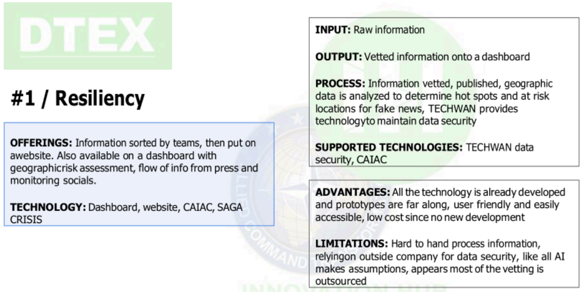
You can use details from the following reports/articles to guide and support your choices of IoS cards:

- NATO's approach to countering disinformation: a focus on COVID-19
- 'Ghostwriter' Influence Campaign: Unknown Actors Leverage Website Compromises and Fabricated Content to Push Narratives Aligned with Russian Security Interests
- NATO Chief Rebukes China Over Coronavirus Disinformation

Expectations

1. **Pick five IoS cards** and explain why you think they are the best choices to address the issues.
2. **Develop a plan** that leverages the five IoS cards you chose - both their individual strengths and their combined synergies. This plan should counter or mitigate the effects of disinformation campaigns. Explain how your IoS cards can combine their strengths.
3. **Present your plan to the jury**, during the 'confrontation session' with the other team and convince them that your plan is the better one. Focus on explaining (a) how you plan to use the IoS cards and how you plan to combine their strengths, and (b) what effects you intend to achieve through your plan. Below is the full list of desired effects:
   1. **Identification** of malicious communication material online
   2. **Categorization** of information (real v. fake)
   3. **Attribution**: Finding sources of fake information
   4. **Additional Analyses**: Processing and analysis of collected information to fulfill other objectives
   5. **Visualization** of analyses
   6. **Mitigation of Effects**: Countering disinformation and their effects by shielding the audience being targeted, disseminating counternarratives, etc.

Appendix 3. IoS Cards used in DTEX



#1 / Resiliency

OFFERINGS: Information sorted by teams, then put on awebsite. Also available on a dashboard with geographicrisk assessment, flow of info from press and monitoring socials.

TECHNOLOGY: Dashboard, website, CAIAC, SAGA CRISIS

INPUT: Raw information

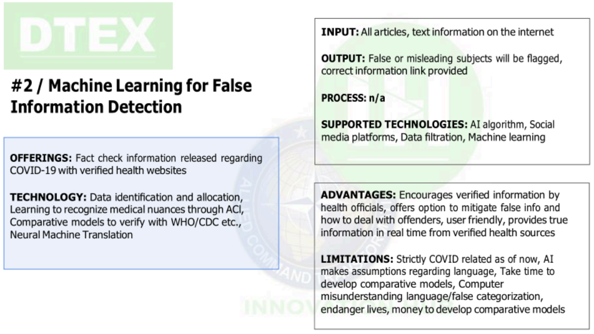
OUTPUT: Vetted information onto a dashboard

PROCESS: Information vetted, published, geographic data is analyzed to determine hot spots and at risk locations for fake news, TECHWAN provides technologyto maintain data security

SUPPORTED TECHNOLOGIES: TECHWAN data security, CAIAC

ADVANTAGES: All the technology is already developed and prototypes are far along, user friendly and easily accessible, low cost since no new development

LIMITATIONS: Hard to hand process information, relyingon outside company for data security, like all AI makes assumptions, appears most of the vetting is outsourced



#2 / Machine Learning for False Information Detection

OFFERINGS: Fact check information released regarding COVID-19 with verified health websites

TECHNOLOGY: Data identification and allocation, Learning to recognize medical nuances through ACI, Comparative models to verify with WHO/CDC etc., Neural Machine Translation

INPUT: All articles, text information on the internet

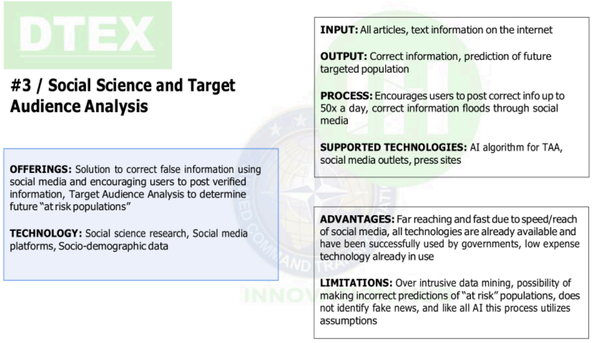
OUTPUT: False or misleading subjects will be flagged, correct information link provided

PROCESS: n/a

SUPPORTED TECHNOLOGIES: AI algorithm, Social media platforms, Data filtration, Machine learning

ADVANTAGES: Encourages verified information by health officials, offers option to mitigate false info and how to deal with offenders, user friendly, provides true information in real time from verified health sources

LIMITATIONS: Strictly COVID related as of now, AI makes assumptions regarding language, Take time to develop comparative models, Computer misunderstanding language/false categorization, endanger lives, money to develop comparative models



#3 / Social Science and Target Audience Analysis

OFFERINGS: Solution to correct false information using social media and encouraging users to post verified information, Target Audience Analysis to determine future "at risk populations"

TECHNOLOGY: Social science research, Social media platforms, Socio-demographic data

INPUT: All articles, text information on the internet

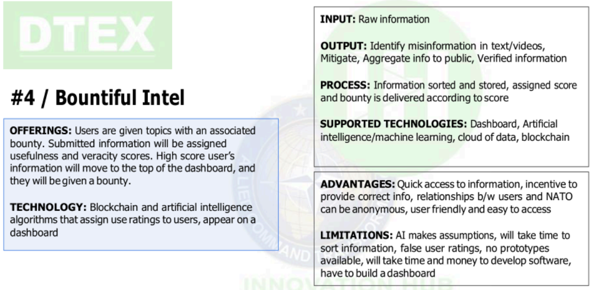
OUTPUT: Correct information, prediction of future targeted population

PROCESS: Encourages users to post correct info up to 50x a day, correct information floods through social media

SUPPORTED TECHNOLOGIES: AI algorithm for TAA, social media outlets, press sites

ADVANTAGES: Far reaching and fast due to speed/reach of social media, all technologies are already available and have been successfully used by governments, low expense technology already in use

LIMITATIONS: Over intrusive data mining, possibility of making incorrect predictions of "at risk" populations, does not identify fake news, and like all AI this process utilizes assumptions



#4 / Bountiful Intel

OFFERINGS: Users are given topics with an associated bounty. Submitted information will be assigned usefulness and veracity scores. High score user's information will move to the top of the dashboard, and they will be given a bounty.

TECHNOLOGY: Blockchain and artificial intelligence algorithms that assign use ratings to users, appear on a dashboard

INPUT: Raw information

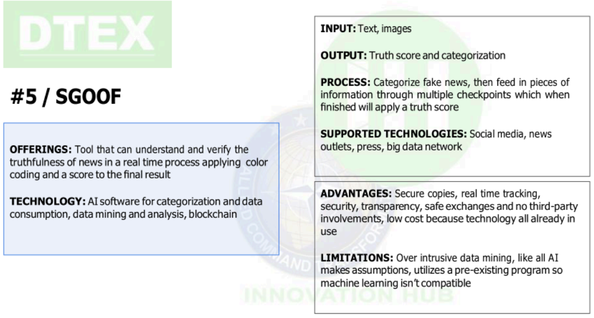
OUTPUT: Identify misinformation in text/videos, Mitigate, Aggregate info to public, Verified information

PROCESS: Information sorted and stored, assigned score and bounty is delivered according to score

SUPPORTED TECHNOLOGIES: Dashboard, Artificial intelligence/machine learning, cloud of data, blockchain

ADVANTAGES: Quick access to information, incentive to provide correct info, relationships b/w users and NATO can be anonymous, user friendly and easy to access

LIMITATIONS: AI makes assumptions, will take time to sort information, false user ratings, no prototypes available, will take time and money to develop software, have to build a dashboard



#5 / SGOOF

OFFERINGS: Tool that can understand and verify the truthfulness of news in a real time process applying color coding and a score to the final result

TECHNOLOGY: AI software for categorization and data consumption, data mining and analysis, blockchain

INPUT: Text, images

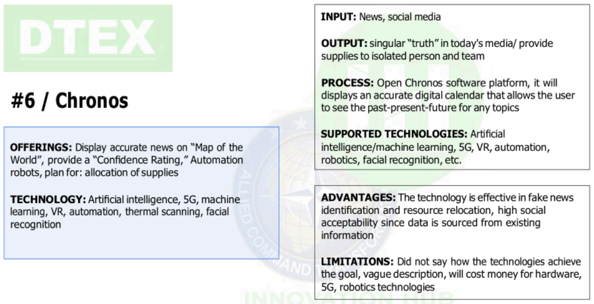
OUTPUT: Truth score and categorization

PROCESS: Categorize fake news, then feed in pieces of information through multiple checkpoints which when finished will apply a truth score

SUPPORTED TECHNOLOGIES: Social media, news outlets, press, big data network

ADVANTAGES: Secure copies, real time tracking, security, transparency, safe exchanges and no third-party involvements, low cost because technology all already in use

LIMITATIONS: Over intrusive data mining, like all AI makes assumptions, utilizes a pre-existing program so machine learning isn't compatible



#6 / Chronos

OFFERINGS: Display accurate news on "Map of the World", provide a "Confidence Rating," Automation robots, plan for: allocation of supplies

TECHNOLOGY: Artificial intelligence, 5G, machine learning, VR, automation, thermal scanning, facial recognition

INPUT: News, social media

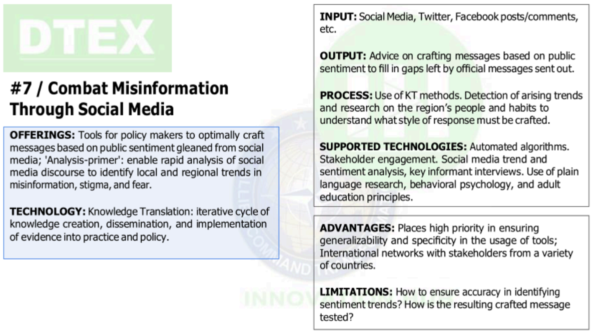
OUTPUT: singular "truth" in today's media/ provide supplies to isolated person and team

PROCESS: Open Chronos software platform, it will displays an accurate digital calendar that allows the user to see the past-present-future for any topics

SUPPORTED TECHNOLOGIES: Artificial intelligence/machine learning, 5G, VR, automation, robotics, facial recognition, etc.

ADVANTAGES: The technology is effective in fake news identification and resource relocation, high social acceptability since data is sourced from existing information

LIMITATIONS: Did not say how the technologies achieve the goal, vague description, will cost money for hardware, 5G, robotics technologies



#7 / Combat Misinformation Through Social Media

OFFERINGS: Tools for policy makers to optimally craft messages based on public sentiment gleaned from social media; 'Analysis-primer': enable rapid analysis of social media discourse to identify local and regional trends in misinformation, stigma, and fear.

TECHNOLOGY: Knowledge Translation: iterative cycle of knowledge creation, dissemination, and implementation of evidence into practice and policy.

INPUT: Social Media, Twitter, Facebook posts/comments, etc.

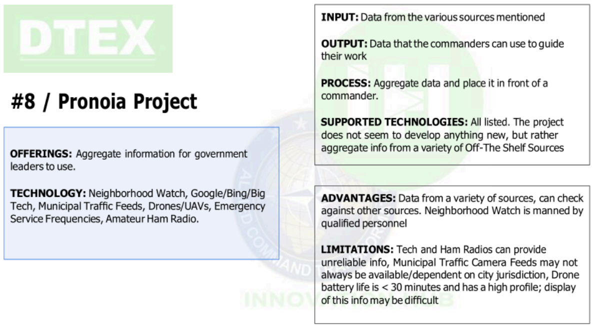
OUTPUT: Advice on crafting messages based on public sentiment to fill in gaps left by official messages sent out.

PROCESS: Use of KT methods. Detection of arising trends and research on the region's people and habits to understand what style of response must be crafted.

SUPPORTED TECHNOLOGIES: Automated algorithms. Stakeholder engagement. Social media trend and sentiment analysis, key informant interviews. Use of plain language research, behavioral psychology, and adult education principles.

ADVANTAGES: Places high priority in ensuring generalizability and specificity in the usage of tools; International networks with stakeholders from a variety of countries.

LIMITATIONS: How to ensure accuracy in identifying sentiment trends? How is the resulting crafted message tested?



#8 / Pronoia Project

OFFERINGS: Aggregate information for government leaders to use.

TECHNOLOGY: Neighborhood Watch, Google/Bing/Big Tech, Municipal Traffic Feeds, Drones/UAVs, Emergency Service Frequencies, Amateur Ham Radio.

INPUT: Data from the various sources mentioned

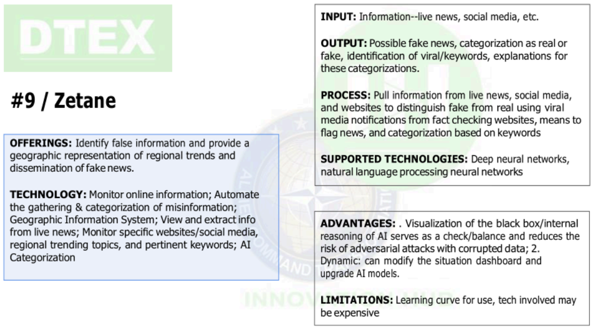
OUTPUT: Data that the commanders can use to guide their work

PROCESS: Aggregate data and place it in front of a commander.

SUPPORTED TECHNOLOGIES: All listed. The project does not seem to develop anything new, but rather aggregate info from a variety of Off-The Shelf Sources

ADVANTAGES: Data from a variety of sources, can check against other sources. Neighborhood Watch is manned by qualified personnel

LIMITATIONS: Tech and Ham Radios can provide unreliable info, Municipal Traffic Camera Feeds may not always be available/dependent on city jurisdiction, Drone battery life is < 30 minutes and has a high profile; display of this info may be difficult



DTEX

#9 / Zetane

OFFERINGS: Identify false information and provide a geographic representation of regional trends and dissemination of fake news.

TECHNOLOGY: Monitor online information; Automate the gathering & categorization of misinformation; Geographic Information System; View and extract info from live news; Monitor specific websites/social media, regional trending topics, and pertinent keywords; AI Categorization

INPUT: Information--live news, social media, etc.

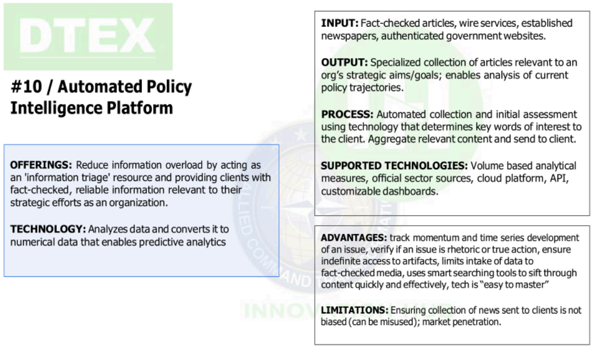
OUTPUT: Possible fake news, categorization as real or fake, identification of viral/keywords, explanations for these categorizations.

PROCESS: Pull information from live news, social media, and websites to distinguish fake from real using viral media notifications from fact checking websites, means to flag news, and categorization based on keywords

SUPPORTED TECHNOLOGIES: Deep neural networks, natural language processing neural networks

ADVANTAGES: . Visualization of the black box/internal reasoning of AI serves as a check/balance and reduces the risk of adversarial attacks with corrupted data; 2. Dynamic: can modify the situation dashboard and upgrade AI models.

LIMITATIONS: Learning curve for use, tech involved may be expensive



DTEX

#10 / Automated Policy Intelligence Platform

OFFERINGS: Reduce information overload by acting as an 'information triage' resource and providing clients with fact-checked, reliable information relevant to their strategic efforts as an organization.

TECHNOLOGY: Analyzes data and converts it to numerical data that enables predictive analytics

INPUT: Fact-checked articles, wire services, established newspapers, authenticated government websites.

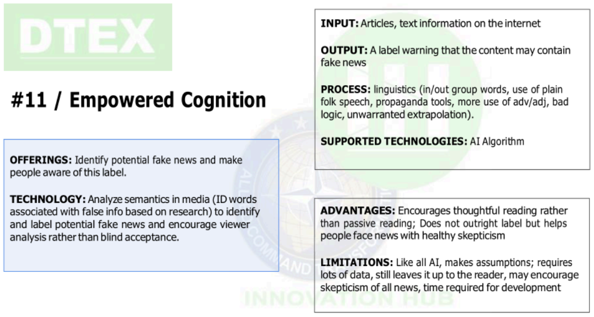
OUTPUT: Specialized collection of articles relevant to an org's strategic aims/goals; enables analysis of current policy trajectories.

PROCESS: Automated collection and initial assessment using technology that determines key words of interest to the client. Aggregate relevant content and send to client.

SUPPORTED TECHNOLOGIES: Volume based analytical measures, official sector sources, cloud platform, API, customizable dashboards.

ADVANTAGES: track momentum and time series development of an issue, verify if an issue is rhetoric or true action, ensure indefinite access to artifacts, limits intake of data to fact-checked media, uses smart searching tools to sift through content quickly and effectively, tech is "easy to master"

LIMITATIONS: Ensuring collection of news sent to clients is not biased (can be misused); market penetration.



DTEX

#11 / Empowered Cognition

OFFERINGS: Identify potential fake news and make people aware of this label.

TECHNOLOGY: Analyze semantics in media (ID words associated with false info based on research) to identify and label potential fake news and encourage viewer analysis rather than blind acceptance.

INPUT: Articles, text information on the internet

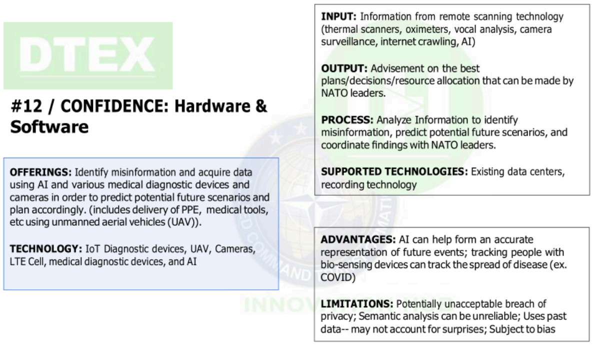
OUTPUT: A label warning that the content may contain fake news

PROCESS: linguistics (in/out group words, use of plain folk speech, propaganda tools, more use of adv/adj, bad logic, unwarranted extrapolation).

SUPPORTED TECHNOLOGIES: AI Algorithm

ADVANTAGES: Encourages thoughtful reading rather than passive reading; Does not outright label but helps people face news with healthy skepticism

LIMITATIONS: Like all AI, makes assumptions; requires lots of data, still leaves it up to the reader, may encourage skepticism of all news, time required for development



DTEX

#12 / CONFIDENCE: Hardware & Software

OFFERINGS: Identify misinformation and acquire data using AI and various medical diagnostic devices and cameras in order to predict potential future scenarios and plan accordingly. (includes delivery of PPE, medical tools, etc using unmanned aerial vehicles (UAV)).

TECHNOLOGY: IoT Diagnostic devices, UAV, Cameras, LTE Cell, medical diagnostic devices, and AI

INPUT: Information from remote scanning technology (thermal scanners, oximeters, vocal analysis, camera surveillance, internet crawling, AI)

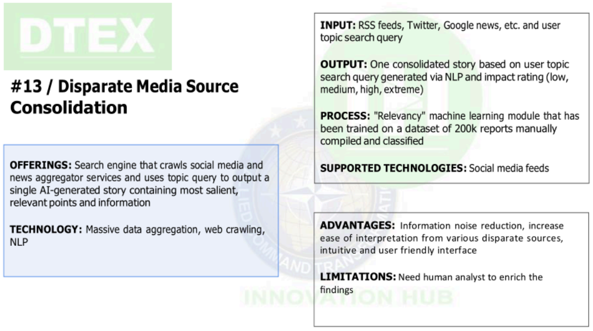
OUTPUT: Advisement on the best plans/decisions/resource allocation that can be made by NATO leaders.

PROCESS: Analyze Information to identify misinformation, predict potential future scenarios, and coordinate findings with NATO leaders.

SUPPORTED TECHNOLOGIES: Existing data centers, recording technology

ADVANTAGES: AI can help form an accurate representation of future events; tracking people with bio-sensing devices can track the spread of disease (ex. COVID)

LIMITATIONS: Potentially unacceptable breach of privacy; Semantic analysis can be unreliable; Uses past data-- may not account for surprises; Subject to bias



DTEX

#13 / Disparate Media Source Consolidation

OFFERINGS: Search engine that crawls social media and news aggregator services and uses topic query to output a single AI-generated story containing most salient, relevant points and information

TECHNOLOGY: Massive data aggregation, web crawling, NLP

INPUT: RSS feeds, Twitter, Google news, etc. and user topic search query

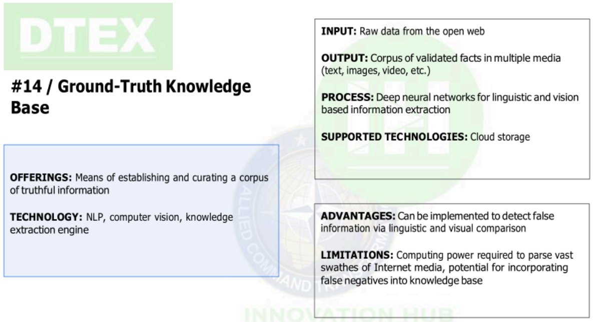
OUTPUT: One consolidated story based on user topic search query generated via NLP and impact rating (low, medium, high, extreme)

PROCESS: "Relevancy" machine learning module that has been trained on a dataset of 200k reports manually compiled and classified

SUPPORTED TECHNOLOGIES: Social media feeds

ADVANTAGES: Information noise reduction, increase ease of interpretation from various disparate sources, intuitive and user friendly interface

LIMITATIONS: Need human analyst to enrich the findings



DTEX

#14 / Ground-Truth Knowledge Base

OFFERINGS: Means of establishing and curating a corpus of truthful information

TECHNOLOGY: NLP, computer vision, knowledge extraction engine

INPUT: Raw data from the open web

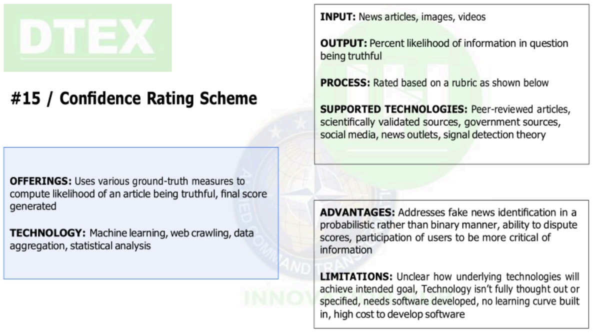
OUTPUT: Corpus of validated facts in multiple media (text, images, video, etc.)

PROCESS: Deep neural networks for linguistic and vision based information extraction

SUPPORTED TECHNOLOGIES: Cloud storage

ADVANTAGES: Can be implemented to detect false information via linguistic and visual comparison

LIMITATIONS: Computing power required to parse vast swathes of Internet media, potential for incorporating false negatives into knowledge base



DTEX

#15 / Confidence Rating Scheme

OFFERINGS: Uses various ground-truth measures to compute likelihood of an article being truthful, final score generated

TECHNOLOGY: Machine learning, web crawling, data aggregation, statistical analysis

INPUT: News articles, images, videos

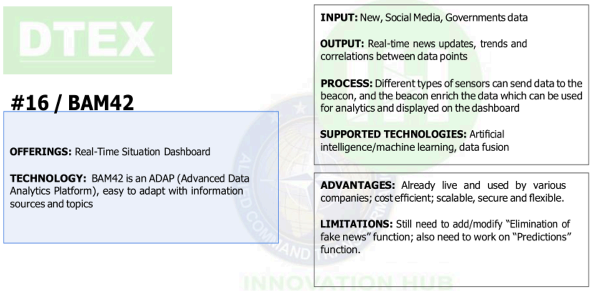
OUTPUT: Percent likelihood of information in question being truthful

PROCESS: Rated based on a rubric as shown below

SUPPORTED TECHNOLOGIES: Peer-reviewed articles, scientifically validated sources, government sources, social media, news outlets, signal detection theory

ADVANTAGES: Addresses fake news identification in a probabilistic rather than binary manner, ability to dispute scores, participation of users to be more critical of information

LIMITATIONS: Unclear how underlying technologies will achieve intended goal, Technology isn't fully thought out or specified, needs software developed, no learning curve built in, high cost to develop software



DTEX

#16 / BAM42

OFFERINGS: Real-Time Situation Dashboard

TECHNOLOGY: BAM42 is an ADAP (Advanced Data Analytics Platform), easy to adapt with information sources and topics

INPUT: New, Social Media, Governments data

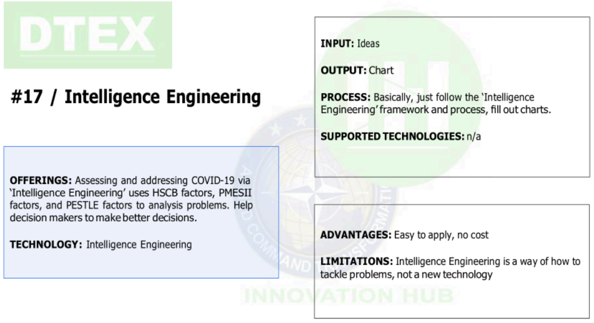
OUTPUT: Real-time news updates, trends and correlations between data points

PROCESS: Different types of sensors can send data to the beacon, and the beacon enrich the data which can be used for analytics and displayed on the dashboard

SUPPORTED TECHNOLOGIES: Artificial intelligence/machine learning, data fusion

ADVANTAGES: Already live and used by various companies; cost efficient; scalable, secure and flexible.

LIMITATIONS: Still need to add/modify "Elimination of fake news" function; also need to work on "Predictions" function.



DTEX

#17 / Intelligence Engineering

OFFERINGS: Assessing and addressing COVID-19 via 'Intelligence Engineering' uses HSCB factors, PMESII factors, and PESTLE factors to analysis problems. Help decision makers to make better decisions.

TECHNOLOGY: Intelligence Engineering

INPUT: Ideas

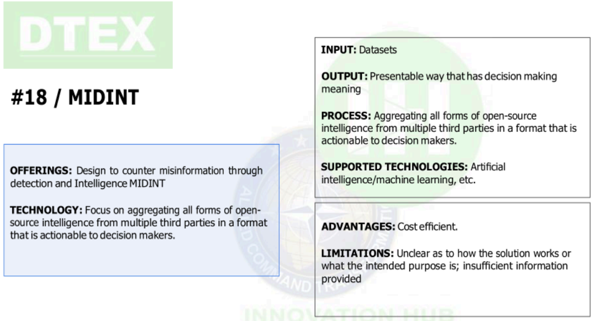
OUTPUT: Chart

PROCESS: Basically, just follow the 'Intelligence Engineering' framework and process, fill out charts.

SUPPORTED TECHNOLOGIES: n/a

ADVANTAGES: Easy to apply, no cost

LIMITATIONS: Intelligence Engineering is a way of how to tackle problems, not a new technology



DTEX

#18 / MIDINT

OFFERINGS: Design to counter misinformation through detection and Intelligence MIDINT

TECHNOLOGY: Focus on aggregating all forms of open-source intelligence from multiple third parties in a format that is actionable to decision makers.

INPUT: Datasets

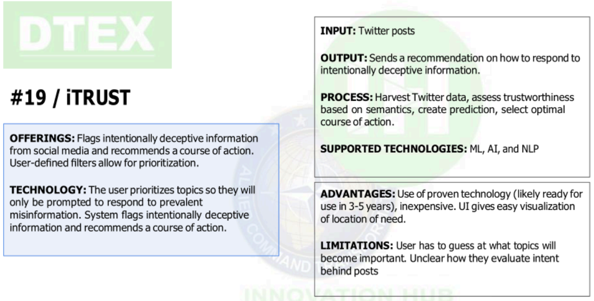
OUTPUT: Presentable way that has decision making meaning

PROCESS: Aggregating all forms of open-source intelligence from multiple third parties in a format that is actionable to decision makers.

SUPPORTED TECHNOLOGIES: Artificial intelligence/machine learning, etc.

ADVANTAGES: Cost efficient.

LIMITATIONS: Unclear as to how the solution works or what the intended purpose is; insufficient information provided



DTEX

#19 / iTRUST

OFFERINGS: Flags intentionally deceptive information from social media and recommends a course of action. User-defined filters allow for prioritization.

TECHNOLOGY: The user prioritizes topics so they will only be prompted to respond to prevalent misinformation. System flags intentionally deceptive information and recommends a course of action.

INPUT: Twitter posts

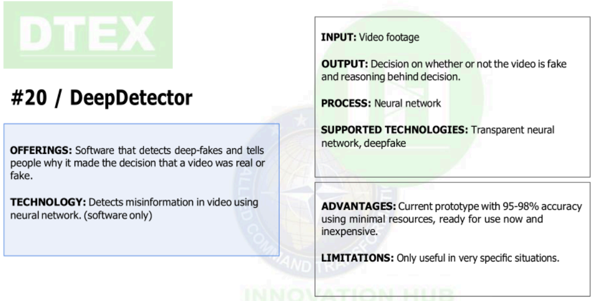
OUTPUT: Sends a recommendation on how to respond to intentionally deceptive information.

PROCESS: Harvest Twitter data, assess trustworthiness based on semantics, create prediction, select optimal course of action.

SUPPORTED TECHNOLOGIES: ML, AI, and NLP

ADVANTAGES: Use of proven technology (likely ready for use in 3-5 years), inexpensive. UI gives easy visualization of location of need.

LIMITATIONS: User has to guess at what topics will become important. Unclear how they evaluate intent behind posts



DTEX

#20 / DeepDetector

OFFERINGS: Software that detects deep-fakes and tells people why it made the decision that a video was real or fake.

TECHNOLOGY: Detects misinformation in video using neural network. (software only)

INPUT: Video footage

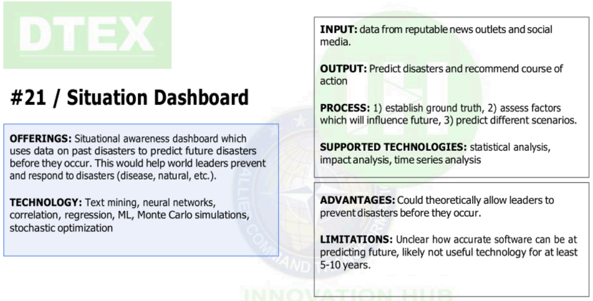
OUTPUT: Decision on whether or not the video is fake and reasoning behind decision.

PROCESS: Neural network

SUPPORTED TECHNOLOGIES: Transparent neural network, deepfake

ADVANTAGES: Current prototype with 95-98% accuracy using minimal resources, ready for use now and inexpensive.

LIMITATIONS: Only useful in very specific situations.



DTEX

#21 / Situation Dashboard

OFFERINGS: Situational awareness dashboard which uses data on past disasters to predict future disasters before they occur. This would help world leaders prevent and respond to disasters (disease, natural, etc.).

TECHNOLOGY: Text mining, neural networks, correlation, regression, ML, Monte Carlo simulations, stochastic optimization

INPUT: data from reputable news outlets and social media.

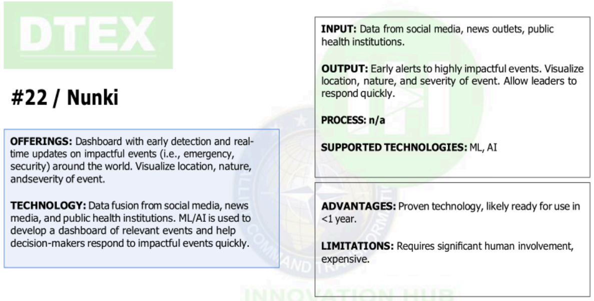
OUTPUT: Predict disasters and recommend course of action

PROCESS: 1) establish ground truth, 2) assess factors which will influence future, 3) predict different scenarios.

SUPPORTED TECHNOLOGIES: statistical analysis, impact analysis, time series analysis

ADVANTAGES: Could theoretically allow leaders to prevent disasters before they occur.

LIMITATIONS: Unclear how accurate software can be at predicting future, likely not useful technology for at least 5-10 years.



DTEX

#22 / Nunki

OFFERINGS: Dashboard with early detection and real-time updates on impactful events (i.e., emergency, security) around the world. Visualize location, nature, andseverity of event.

TECHNOLOGY: Data fusion from social media, news media, and public health institutions. ML/AI is used to develop a dashboard of relevant events and help decision-makers respond to impactful events quickly.

INPUT: Data from social media, news outlets, public health institutions.

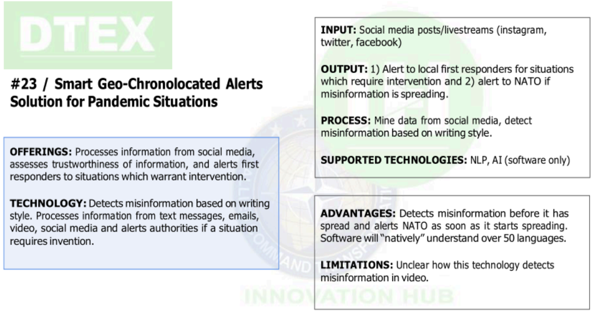
OUTPUT: Early alerts to highly impactful events. Visualize location, nature, and severity of event. Allow leaders to respond quickly.

PROCESS: n/a

SUPPORTED TECHNOLOGIES: ML, AI

ADVANTAGES: Proven technology, likely ready for use in <1 year.

LIMITATIONS: Requires significant human involvement, expensive.



DTEX

#23 / Smart Geo-Chronolocated Alerts Solution for Pandemic Situations

OFFERINGS: Processes information from social media, assesses trustworthiness of information, and alerts first responders to situations which warrant intervention.

TECHNOLOGY: Detects misinformation based on writing style. Processes information from text messages, emails, video, social media and alerts authorities if a situation requires invention.

INPUT: Social media posts/livestreams (instagram, twitter, facebook)

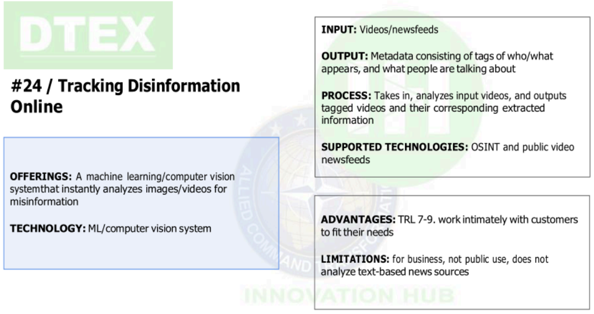
OUTPUT: 1) Alert to local first responders for situations which require intervention and 2) alert to NATO if misinformation is spreading.

PROCESS: Mine data from social media, detect misinformation based on writing style.

SUPPORTED TECHNOLOGIES: NLP, AI (software only)

ADVANTAGES: Detects misinformation before it has spread and alerts NATO as soon as it starts spreading. Software will "natively" understand over 50 languages.

LIMITATIONS: Unclear how this technology detects misinformation in video.



DTEX

#24 / Tracking Disinformation Online

OFFERINGS: A machine learning/computer vision systemthat instantly analyzes images/videos for misinformation

TECHNOLOGY: ML/computer vision system

INPUT: Videos/newsfeeds

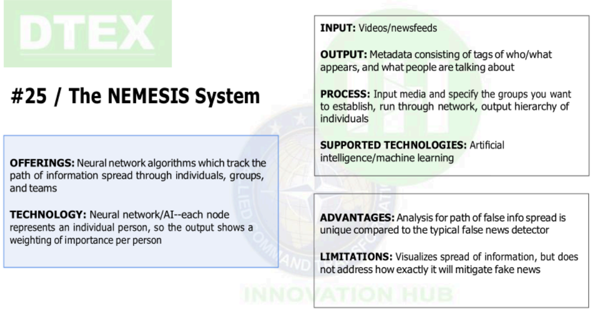
OUTPUT: Metadata consisting of tags of who/what appears, and what people are talking about

PROCESS: Takes in, analyzes input videos, and outputs tagged videos and their corresponding extracted information

SUPPORTED TECHNOLOGIES: OSINT and public video newsfeeds

ADVANTAGES: TRL 7-9. work intimately with customers to fit their needs

LIMITATIONS: for business, not public use, does not analyze text-based news sources



DTEX

#25 / The NEMESIS System

OFFERINGS: Neural network algorithms which track the path of information spread through individuals, groups, and teams

TECHNOLOGY: Neural network/AI--each node represents an individual person, so the output shows a weighting of importance per person

INPUT: Videos/newsfeeds

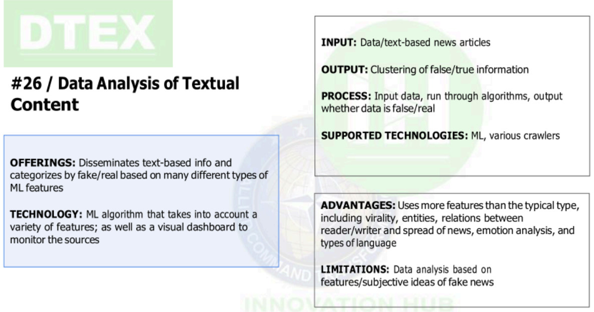
OUTPUT: Metadata consisting of tags of who/what appears, and what people are talking about

PROCESS: Input media and specify the groups you want to establish, run through network, output hierarchy of individuals

SUPPORTED TECHNOLOGIES: Artificial intelligence/machine learning

ADVANTAGES: Analysis for path of false info spread is unique compared to the typical false news detector

LIMITATIONS: Visualizes spread of information, but does not address how exactly it will mitigate fake news



DTEX

#26 / Data Analysis of Textual Content

OFFERINGS: Disseminates text-based info and categorizes by fake/real based on many different types of ML features

TECHNOLOGY: ML algorithm that takes into account a variety of features; as well as a visual dashboard to monitor the sources

INPUT: Data/text-based news articles

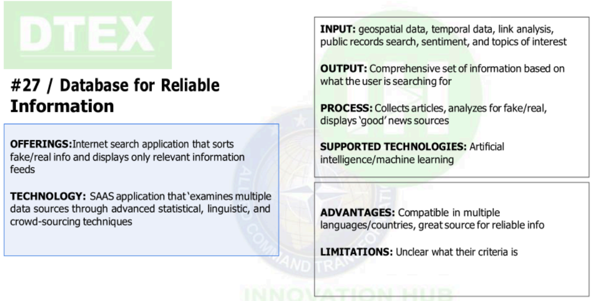
OUTPUT: Clustering of false/true information

PROCESS: Input data, run through algorithms, output whether data is false/real

SUPPORTED TECHNOLOGIES: ML, various crawlers

ADVANTAGES: Uses more features than the typical type, including virality, entities, relations between reader/writer and spread of news, emotion analysis, and types of language

LIMITATIONS: Data analysis based on features/subjective ideas of fake news



DTEX

#27 / Database for Reliable Information

OFFERINGS: Internet search application that sorts fake/real info and displays only relevant information feeds

TECHNOLOGY: SAAS application that 'examines multiple data sources through advanced statistical, linguistic, and crowd-sourcing techniques

INPUT: geospatial data, temporal data, link analysis, public records search, sentiment, and topics of interest

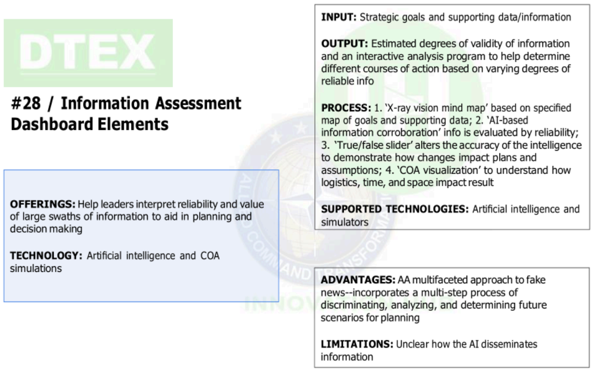
OUTPUT: Comprehensive set of information based on what the user is searching for

PROCESS: Collects articles, analyzes for fake/real, displays 'good' news sources

SUPPORTED TECHNOLOGIES: Artificial intelligence/machine learning

ADVANTAGES: Compatible in multiple languages/countries, great source for reliable info

LIMITATIONS: Unclear what their criteria is



DTEX

#28 / Information Assessment Dashboard Elements

OFFERINGS: Help leaders interpret reliability and value of large swaths of information to aid in planning and decision making

TECHNOLOGY: Artificial intelligence and COA simulations

INPUT: Strategic goals and supporting data/information

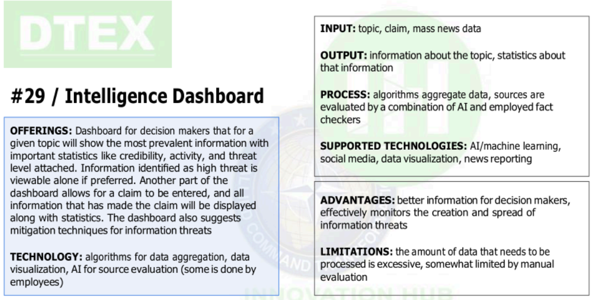
OUTPUT: Estimated degrees of validity of information and an interactive analysis program to help determine different courses of action based on varying degrees of reliable info

PROCESS: 1. 'X-ray vision mind map' based on specified map of goals and supporting data; 2. 'AI-based information corroboration' info is evaluated by reliability; 3. 'True/false slider' alters the accuracy of the intelligence to demonstrate how changes impact plans and assumptions; 4. 'COA visualization' to understand how logistics, time, and space impact result

SUPPORTED TECHNOLOGIES: Artificial intelligence and simulators

ADVANTAGES: AA multifaceted approach to fake news--incorporates a multi-step process of discriminating, analyzing, and determining future scenarios for planning

LIMITATIONS: Unclear how the AI disseminates information



DTEX

#29 / Intelligence Dashboard

OFFERINGS: Dashboard for decision makers that for a given topic will show the most prevalent information with important statistics like credibility, activity, and threat level attached. Information identified as high threat is viewable alone if preferred. Another part of the dashboard allows for a claim to be entered, and all information that has made the claim will be displayed along with statistics. The dashboard also suggests mitigation techniques for information threats

TECHNOLOGY: algorithms for data aggregation, data visualization, AI for source evaluation (some is done by employees)

INPUT: topic, claim, mass news data

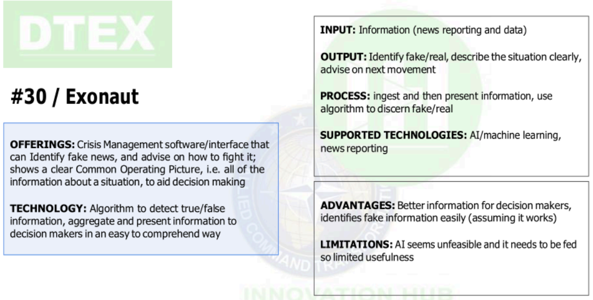
OUTPUT: information about the topic, statistics about that information

PROCESS: algorithms aggregate data, sources are evaluated by a combination of AI and employed fact checkers

SUPPORTED TECHNOLOGIES: AI/machine learning, social media, data visualization, news reporting

ADVANTAGES: better information for decision makers, effectively monitors the creation and spread of information threats

LIMITATIONS: the amount of data that needs to be processed is excessive, somewhat limited by manual evaluation



DTEX

#30 / Exonaut

OFFERINGS: Crisis Management software/interface that can Identify fake news, and advise on how to fight it; shows a clear Common Operating Picture, i.e. all of the information about a situation, to aid decision making

TECHNOLOGY: Algorithm to detect true/false information, aggregate and present information to decision makers in an easy to comprehend way

INPUT: Information (news reporting and data)

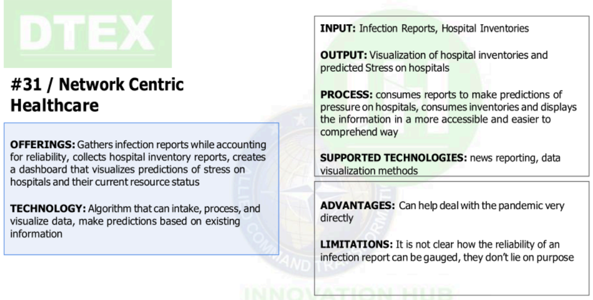
OUTPUT: Identify fake/real, describe the situation clearly, advise on next movement

PROCESS: ingest and then present information, use algorithm to discern fake/real

SUPPORTED TECHNOLOGIES: AI/machine learning, news reporting

ADVANTAGES: Better information for decision makers, identifies fake information easily (assuming it works)

LIMITATIONS: AI seems unfeasible and it needs to be fed so limited usefulness



DTEX

#31 / Network Centric Healthcare

OFFERINGS: Gathers infection reports while accounting for reliability, collects hospital inventory reports, creates a dashboard that visualizes predictions of stress on hospitals and their current resource status

TECHNOLOGY: Algorithm that can intake, process, and visualize data, make predictions based on existing information

INPUT: Infection Reports, Hospital Inventories

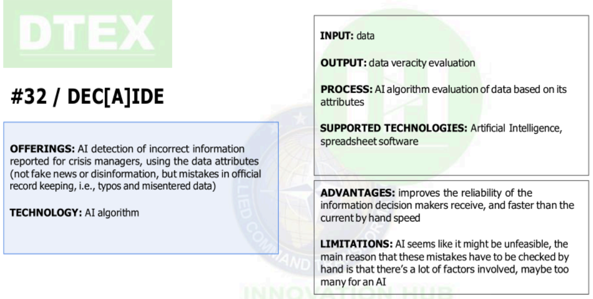
OUTPUT: Visualization of hospital inventories and predicted Stress on hospitals

PROCESS: consumes reports to make predictions of pressure on hospitals, consumes inventories and displays the information in a more accessible and easier to comprehend way

SUPPORTED TECHNOLOGIES: news reporting, data visualization methods

ADVANTAGES: Can help deal with the pandemic very directly

LIMITATIONS: It is not clear how the reliability of an infection report can be gauged, they don't lie on purpose



DTEX

#32 / DEC[A]IDE

OFFERINGS: AI detection of incorrect information reported for crisis managers, using the data attributes (not fake news or disinformation, but mistakes in official record keeping, i.e., typos and misentered data)

TECHNOLOGY: AI algorithm

INPUT: data

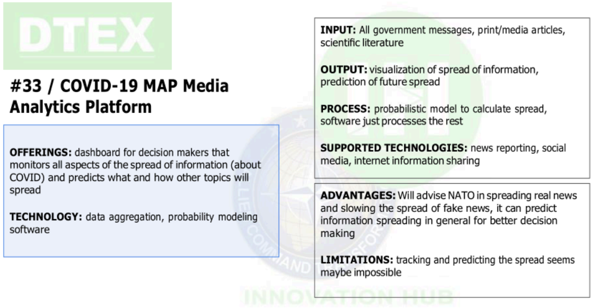
OUTPUT: data veracity evaluation

PROCESS: AI algorithm evaluation of data based on its attributes

SUPPORTED TECHNOLOGIES: Artificial Intelligence, spreadsheet software

ADVANTAGES: improves the reliability of the information decision makers receive, and faster than the current by hand speed

LIMITATIONS: AI seems like it might be unfeasible, the main reason that these mistakes have to be checked by hand is that there's a lot of factors involved, maybe too many for an AI



#33 / COVID-19 MAP Media Analytics Platform

OFFERINGS: dashboard for decision makers that monitors all aspects of the spread of information (about COVID) and predicts what and how other topics will spread

TECHNOLOGY: data aggregation, probability modeling software

INPUT: All government messages, print/media articles, scientific literature

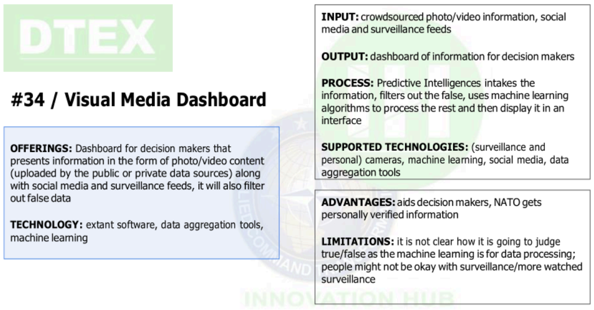
OUTPUT: visualization of spread of information, prediction of future spread

PROCESS: probabilistic model to calculate spread, software just processes the rest

SUPPORTED TECHNOLOGIES: news reporting, social media, internet information sharing

ADVANTAGES: Will advise NATO in spreading real news and slowing the spread of fake news, it can predict information spreading in general for better decision making

LIMITATIONS: tracking and predicting the spread seems maybe impossible



#34 / Visual Media Dashboard

OFFERINGS: Dashboard for decision makers that presents information in the form of photo/video content (uploaded by the public or private data sources) along with social media and surveillance feeds, it will also filter out false data

TECHNOLOGY: extant software, data aggregation tools, machine learning

INPUT: crowdsourced photo/video information, social media and surveillance feeds

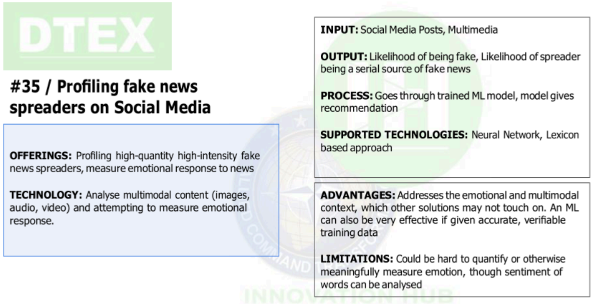
OUTPUT: dashboard of information for decision makers

PROCESS: Predictive Intelligences intakes the information, filters out the false, uses machine learning algorithms to process the rest and then display it in an interface

SUPPORTED TECHNOLOGIES: (surveillance and personal) cameras, machine learning, social media, data aggregation tools

ADVANTAGES: aids decision makers, NATO gets personally verified information

LIMITATIONS: it is not clear how it is going to judge true/false as the machine learning is for data processing; people might not be okay with surveillance/more watched surveillance



#35 / Profiling fake news spreaders on Social Media

OFFERINGS: Profiling high-quantity high-intensity fake news spreaders, measure emotional response to news

TECHNOLOGY: Analyse multimodal content (images, audio, video) and attempting to measure emotional response.

INPUT: Social Media Posts, Multimedia

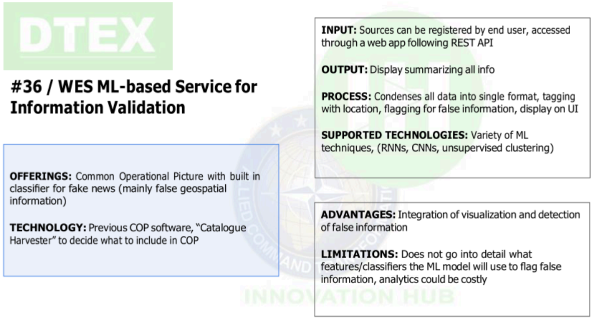
OUTPUT: Likelihood of being fake, Likelihood of spreader being a serial source of fake news

PROCESS: Goes through trained ML model, model gives recommendation

SUPPORTED TECHNOLOGIES: Neural Network, Lexicon based approach

ADVANTAGES: Addresses the emotional and multimodal context, which other solutions may not touch on. An ML can also be very effective if given accurate, verifiable training data

LIMITATIONS: Could be hard to quantify or otherwise meaningfully measure emotion, though sentiment of words can be analysed



#36 / WES ML-based Service for Information Validation

OFFERINGS: Common Operational Picture with built in classifier for fake news (mainly false geospatial information)

TECHNOLOGY: Previous COP software, "Catalogue Harvester" to decide what to include in COP

INPUT: Sources can be registered by end user, accessed through a web app following REST API

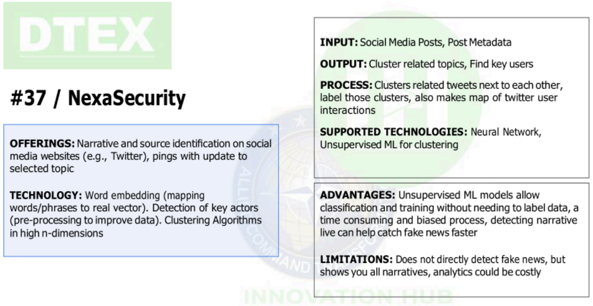
OUTPUT: Display summarizing all info

PROCESS: Condenses all data into single format, tagging with location, flagging for false information, display on UI

SUPPORTED TECHNOLOGIES: Variety of ML techniques, (RNNs, CNNs, unsupervised clustering)

ADVANTAGES: Integration of visualization and detection of false information

LIMITATIONS: Does not go into detail what features/classifiers the ML model will use to flag false information, analytics could be costly



#37 / NexaSecurity

OFFERINGS: Narrative and source identification on social media websites (e.g., Twitter), pings with update to selected topic

TECHNOLOGY: Word embedding (mapping words/phrases to real vector). Detection of key actors (pre-processing to improve data). Clustering Algorithms in high n-dimensions

INPUT: Social Media Posts, Post Metadata

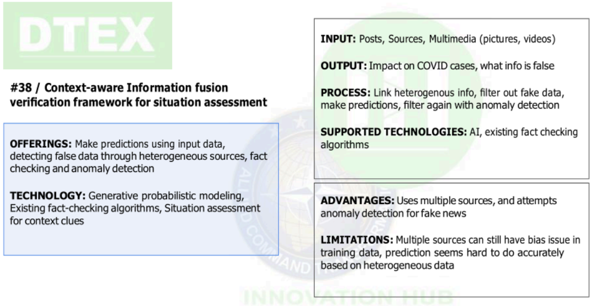
OUTPUT: Cluster related topics, Find key users

PROCESS: Clusters related tweets next to each other, label those clusters, also makes map of twitter user interactions

SUPPORTED TECHNOLOGIES: Neural Network, Unsupervised ML for clustering

ADVANTAGES: Unsupervised ML models allow classification and training without needing to label data, a time consuming and biased process, detecting narrative live can help catch fake news faster

LIMITATIONS: Does not directly detect fake news, but shows you all narratives, analytics could be costly



#38 / Context-aware Information fusion verification framework for situation assessment

OFFERINGS: Make predictions using input data, detecting false data through heterogeneous sources, fact checking and anomaly detection

TECHNOLOGY: Generative probabilistic modeling, Existing fact-checking algorithms, Situation assessment for context clues

INPUT: Posts, Sources, Multimedia (pictures, videos)

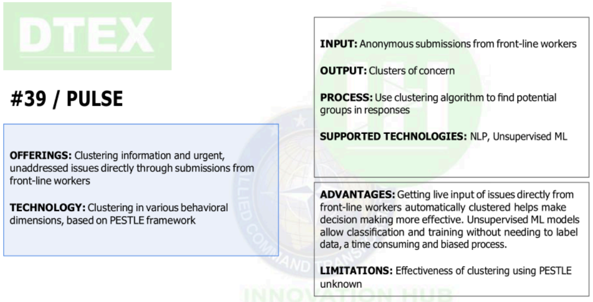
OUTPUT: Impact on COVID cases, what info is false

PROCESS: Link heterogenous info, filter out fake data, make predictions, filter again with anomaly detection

SUPPORTED TECHNOLOGIES: AI, existing fact checking algorithms

ADVANTAGES: Uses multiple sources, and attempts anomaly detection for fake news

LIMITATIONS: Multiple sources can still have bias issue in training data, prediction seems hard to do accurately based on heterogeneous data



#39 / PULSE

OFFERINGS: Clustering information and urgent, unaddressed issues directly through submissions from front-line workers

TECHNOLOGY: Clustering in various behavioral dimensions, based on PESTLE framework

INPUT: Anonymous submissions from front-line workers

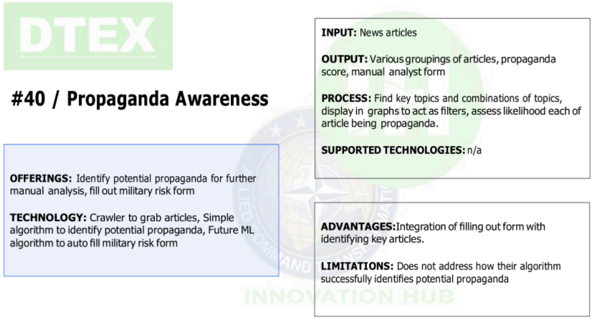
OUTPUT: Clusters of concern

PROCESS: Use clustering algorithm to find potential groups in responses

SUPPORTED TECHNOLOGIES: NLP, Unsupervised ML

ADVANTAGES: Getting live input of issues directly from front-line workers automatically clustered helps make decision making more effective. Unsupervised ML models allow classification and training without needing to label data, a time consuming and biased process.

LIMITATIONS: Effectiveness of clustering using PESTLE unknown



#40 / Propaganda Awareness

OFFERINGS: Identify potential propaganda for further manual analysis, fill out military risk form

TECHNOLOGY: Crawler to grab articles, Simple algorithm to identify potential propaganda, Future ML algorithm to auto fill military risk form

INPUT: News articles

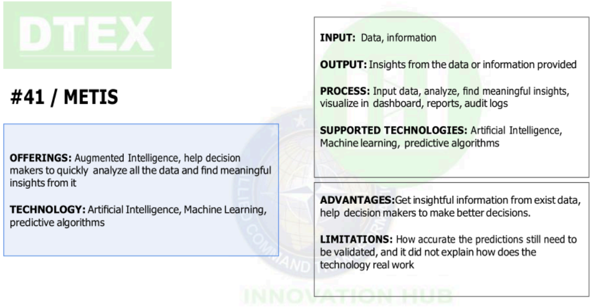
OUTPUT: Various groupings of articles, propaganda score, manual analyst form

PROCESS: Find key topics and combinations of topics, display in graphs to act as filters, assess likelihood each of article being propaganda.

SUPPORTED TECHNOLOGIES: n/a

ADVANTAGES: Integration of filling out form with identifying key articles.

LIMITATIONS: Does not address how their algorithm successfully identifies potential propaganda



#41 / METIS

OFFERINGS: Augmented Intelligence, help decision makers to quickly analyze all the data and find meaningful insights from it

TECHNOLOGY: Artificial Intelligence, Machine Learning, predictive algorithms

INPUT: Data, information

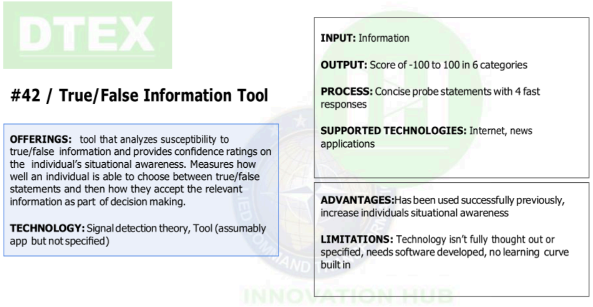
OUTPUT: Insights from the data or information provided

PROCESS: Input data, analyze, find meaningful insights, visualize in dashboard, reports, audit logs

SUPPORTED TECHNOLOGIES: Artificial Intelligence, Machine learning, predictive algorithms

ADVANTAGES: Get insightful information from exist data, help decision makers to make better decisions.

LIMITATIONS: How accurate the predictions still need to be validated, and it did not explain how does the technology real work



#42 / True/False Information Tool

OFFERINGS: tool that analyzes susceptibility to true/false information and provides confidence ratings on the individual's situational awareness. Measures how well an individual is able to choose between true/false statements and then how they accept the relevant information as part of decision making.

TECHNOLOGY: Signal detection theory, Tool (assumably app but not specified)

INPUT: Information

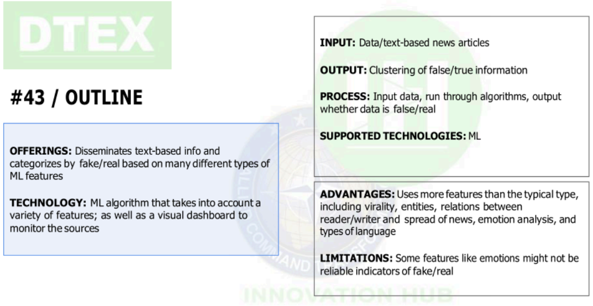
OUTPUT: Score of -100 to 100 in 6 categories

PROCESS: Concise probe statements with 4 fast responses

SUPPORTED TECHNOLOGIES: Internet, news applications

ADVANTAGES: Has been used successfully previously, increase individuals situational awareness

LIMITATIONS: Technology isn't fully thought out or specified, needs software developed, no learning curve built in



#43 / OUTLINE

OFFERINGS: Disseminates text-based info and categorizes by fake/real based on many different types of ML features

TECHNOLOGY: ML algorithm that takes into account a variety of features; as well as a visual dashboard to monitor the sources

INPUT: Data/text-based news articles

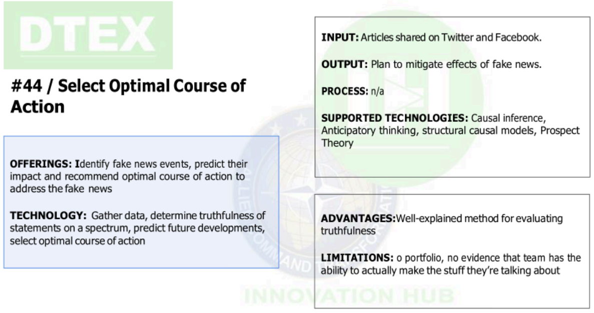
OUTPUT: Clustering of false/true information

PROCESS: Input data, run through algorithms, output whether data is false/real

SUPPORTED TECHNOLOGIES: ML

ADVANTAGES: Uses more features than the typical type, including virality, entities, relations between reader/writer and spread of news, emotion analysis, and types of language

LIMITATIONS: Some features like emotions might not be reliable indicators of fake/real



#44 / Select Optimal Course of Action

OFFERINGS: Identify fake news events, predict their impact and recommend optimal course of action to address the fake news

TECHNOLOGY: Gather data, determine truthfulness of statements on a spectrum, predict future developments, select optimal course of action

INPUT: Articles shared on Twitter and Facebook.

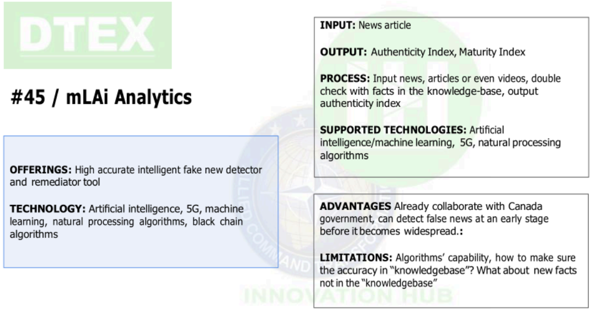
OUTPUT: Plan to mitigate effects of fake news.

PROCESS: n/a

SUPPORTED TECHNOLOGIES: Causal inference, Anticipatory thinking, structural causal models, Prospect Theory

ADVANTAGES: Well-explained method for evaluating truthfulness

LIMITATIONS: o portfolio, no evidence that team has the ability to actually make the stuff they're talking about



#45 / mLAi Analytics

OFFERINGS: High accurate intelligent fake new detector and remediator tool

TECHNOLOGY: Artificial intelligence, 5G, machine learning, natural processing algorithms, black chain algorithms

INPUT: News article

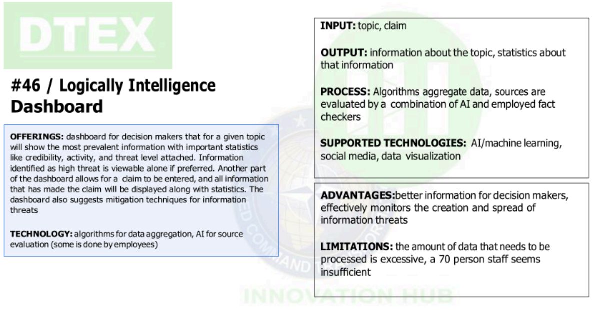
OUTPUT: Authenticity Index, Maturity Index

PROCESS: Input news, articles or even videos, double check with facts in the knowledge-base, output authenticity index

SUPPORTED TECHNOLOGIES: Artificial intelligence/machine learning, 5G, natural processing algorithms

ADVANTAGES Already collaborate with Canada government, can detect false news at an early stage before it becomes widespread.**:**

LIMITATIONS: Algorithms' capability, how to make sure the accuracy in "knowledgebase"? What about new facts not in the "knowledgebase"



#46 / Logically Intelligence Dashboard

OFFERINGS: dashboard for decision makers that for a given topic will show the most prevalent information with important statistics like credibility, activity, and threat level attached. Information identified as high threat is viewable alone if preferred. Another part of the dashboard allows for a claim to be entered, and all information that has made the claim will be displayed along with statistics. The dashboard also suggests mitigation techniques for information threats

TECHNOLOGY: algorithms for data aggregation, AI for source evaluation (some is done by employees)

INPUT: topic, claim

OUTPUT: information about the topic, statistics about that information

PROCESS: Algorithms aggregate data, sources are evaluated by a combination of AI and employed fact checkers

SUPPORTED TECHNOLOGIES: AI/machine learning, social media, data visualization

ADVANTAGES:better information for decision makers, effectively monitors the creation and spread of information threats

LIMITATIONS: the amount of data that needs to be processed is excessive, a 70 person staff seems insufficient