

ISBN: 978-9934-619-55-7

Authors: Gundars Bergmanis-Korāts, Giorgio Bertolin, Adele Pužule, Yukai Zeng

Acknowledgements: Logically.ai

Project Manager: Gundars Bergmanis-Korāts

Content Editor: Merle Read

Design: Inga Ropša

Riga, January 2024

NATO STRATCOM COE

11b Kalnciema iela,

Riga, LV1048, Latvia

stratcomcoe.org

@stratcomcoe

This publication does not represent the opinions or policies of NATO or NATO StratCom COE.

© All rights reserved by the NATO StratCom COE. Reports may not be copied, reproduced, distributed or publicly displayed without reference to the NATO StratCom COE. The views expressed here do not represent the views of NATO.

AI in Support of StratCom Capabilities

Contents

Abstract	5
Understanding the role of StratCom in NATO	5
The social perspective: what is required	6
PsyOps requirements	6
Target audience analysis	7
Audience segmentation	8
Construct validity	8
Fundamental attribution error	9
Operationalising the variables	10
When we fail, why do we fail?	11
Introducing the behaviour model: COM-B as a framework for TAA	11
Capability	11
Opportunity	12
Motivation	12
Overview	12
Example: TAA in a marketing company's digital advertising campaign	15
Technical prerequisites of a successful campaign	16
Borrowing useful tools from IEA	16
Hashtags and entities	17
Challenges with sentiment analysis: from polar sentiment towards directional sentiment	19
From topics to narratives	23
Network analysis	24
Deriving insights from surveys, interviews and focus groups	25
Multimodal perspectives	26
Fusing everything together	28
Steering towards an AI toolbox for targeted communication	31
The legal perspective	37
The Artificial Intelligence Act	39
Ethics Guidelines for Trustworthy AI	40
Legal framework for military purposes	41
North Atlantic Treaty Organization	41
The European Union	41
Conclusions and recommendations	42
The legal considerations	44
Endnotes	45

Abstract

The report aims to guide information environment assessment (IEA) practitioners. This includes understanding the information environment and audiences, particularly in online campaigns, and covers necessary technical elements and legal factors.

Key questions addressed include: what artificial intelligence (AI) functions are essential for Strategic Communications? Which models

need improvement? What is the projected future of AI in this field?

The report offers current knowledge to enhance practitioners' ability to navigate an AI-driven information environment securely, efficiently and in line with legal requirements.

Understanding the role of StratCom in NATO

Strategic communications (StratCom), according to Allied Joint Doctrine for Strategic Communications (AJP-10 (2023), hereinafter **AJP-10**), is the command group function responsible for **understanding the information environment** (IE) across all relevant audiences and, based on that understanding, using all means of communication—comprising actions, images and words—to appropriately **inform** and **influence** an **audience's attitudes** and **behaviours** through a **narrative-led and behaviour-centric approach** in pursuit of the desired end state. In the NATO military context, StratCom handles the integration of communication capabilities and information staff function with other military activities, to understand and shape the IE in support of NATO strategic aims and objectives.

NATO's J10-Strategic Communications directorate (J10-StratCom) includes the information staff function of Information Operations (Info Ops) and the communication capabilities of Military Public Affairs (Mil PA) and Psychological Operations (PsyOps). In our research report, although the primary topic is the role of AI in StratCom, we have chosen to focus particularly on PsyOps because of the significant interplay between these areas.

NATO PsyOps is based on true and factual information attributable to NATO or a concurring Ally, partner nation or organisation. The PsyOps staff within J10-StratCom exist only at and below the operational level and contribute to the commander's decision-making process. They advise on feasible PsyOps communication activities and the psychological impact of planned actions, and they recommend information and communication activities to be performed by PsyOps and non-PsyOps personnel and units to create effects in the IE. Within J10-StratCom the PsyOps staff provide five distinct functions:

1. Audience analysis
2. Psychological operations planning
3. Counter-hostile information and disinformation
4. Web operations
5. Combined joint psychological operations task force headquarters.

In the next section of the report, an overview is provided on StratCom capabilities and functions focusing on specific PsyOps requirements and prerequisites.

The social perspective: what is required

AI technologies, with their capabilities for data analysis, pattern recognition and predictive modelling, have become increasingly integral to the implementation of PsyOps in the digital realm. The ability of AI to target and customise messages based on individual psychological

profiles can greatly enhance the impact of PsyOps. Furthermore, understanding how AI can be leveraged in PsyOps can provide crucial insights into AI's broader role in StratCom, making this focus essential for a comprehensive exploration of the subject matter.

PsyOps requirements

To understand how AI can best support the functionality of PsyOps it is necessary to start by identifying vulnerable points in the standard procedures of these functions and investigating whether potential gaps can be addressed by AI-enabled solutions. Within this part of the report, gap analysis will be performed on PsyOps processes and functions and types of PsyOps activities, and comparing them with similar functions used for civilian purposes.

Allied Joint Doctrine for Psychological Operations (AJP-3.10.1 (2015)), like AJP-10, outlines the requirement to define an audience that is to be targeted through PsyOps, which, in itself, highlights the need to understand the target and take it into account when crafting a PsyOps campaign. Research is carried out to:

- Conduct baseline audience analysis¹
- Identify targets and audiences
- Carry out target audience analysis (TAA)
- Assess the success of individual campaigns.

The purpose then for both information activities and PsyOps products most typically is a change in perception or behaviour.

The closest comparison to these functions for civilian purposes is social marketing, which is a distinct marketing discipline referring primarily to efforts focused on influencing behaviours to improve health, prevent injuries,

protect the environment and contribute to communities.² For functions such as finding targets, performing TAA and evaluating the success rate of specific campaigns, usage of AI systems is widely known and accepted. Best practices and off-the-shelf tools can migrate from civilian purpose to military purpose. However, for such adaptation to be successful, the main differences between possible uses of AI systems for PsyOps and social marketing must be considered.

- The development of AI systems utilised for civilian purposes, especially after enactment of the Artificial Intelligence Act (see The legal perspective for more details)³, has more transparent innovation processes, allowing the broader public to follow AI advances, and AI developers to get and incorporate feedback more rapidly. Given the numerous sensitivities that must be kept in mind during the evolution of PsyOps campaigns for military purposes, such a quick turnaround is rarely possible.
- Success metrics for AI utilised for social marketing are widely discussed and understood, but those for PsyOps, again taking into account the sensitivity of such campaigns, are not as publicly discussed. More work is required when integrating AI systems into the military context, and

AI companies without sector-specific knowledge are limited in what they can innovate with PsyOps in mind.

- Social marketing metrics are more straightforward—measurables for a social media campaign, for example, could be post engagements and how those translate to sales. With PsyOps the measurable variables can still be engagements with material; however, it is harder to quantify the action (or inaction) performed by the target in response to the campaign.

By understanding these differences and applying the necessary adjustments, AI systems that have proven themselves in a civilian social marketing context can be carried over to the military sector and used to improve the function of PsyOps campaigns.

To carry out gap analysis and understand the utility of AI systems within PsyOps, an overview of the various IEA tools that are or can be relevant in PsyOps will be provided in the following sections.

Target audience analysis

Table 1 presents a collection of definitions and explanations pertaining to the concept of TAA as defined by NATO doctrines. TAA is a critical component of StratCom, involving a detailed examination of intended audiences to achieve desired effects. The diversity in definitions reflects the nuances in interpreting TAA across different contexts. Each entry in this table comprises the source of the definition, the term or concept used by that source, and the respective definition or explanation. Allied Joint Doctrine for Joint Targeting

(AJP-3.9) states that special care should be taken in the targeting process to consider possible unintended consequences, while aiming to achieve the maximum desired effect. When applied to audiences, it defines the examination of potential targets for information activities, to determine importance, priority and the means required to obtain a desired effect. To achieve this, audience segmentation is necessary. The following section will look at this process more closely.

Source	Term/concept	Definition/explanation
AJP-10	Target	An area, infrastructure, object, audience or organisation against which activities can be directed to create desired effects.
AJP-10	Audience	An individual, group or entity whose interpretation of events and subsequent behaviour may affect the attainment of the end state. Note: the audience may consist of publics, stakeholders and actors.
AJP-10	Target audience analysis	The focused examination of targeted audiences to create desired effects.
AJP-3.9	Joint targeting requirements	Joint targeting involves taking actions in one or more of the operational domains, using all capabilities available, against a target, to create an effect in one or more of the physical, virtual, or cognitive dimensions. [...] The use of some capabilities that create nonlethal effects requires the same type of special considerations because, while they may reduce the potential for death and physical destruction, their improper or untimely use also may have unintended consequences that are detrimental to creating the desired effects. (AJP-3.9, p. 1-3)

TABLE 1. Definitions of target audience analysis

Audience segmentation

Segmenting an audience means dividing it according to different parameters, to identify smaller groups that present certain characteristics. The easiest and most common way to do so is by using demographics: distinguishing audience members by parameters such as gender, age, education and so on.

However, using demographics as a primary basis for segmentation has been described as ‘error-ridden to the degree that the correlation between demographics and the actual determinants of the behaviour are imperfect. [...] The argument for using demographics is fundamentally based on complex causal relationships: that shared education, race, and cultural factors equate to common life experiences that will tend to engender similar motivations and situational constraints.’⁴ Academic research generally agrees in asserting that demographic variables alone are of limited value in identifying audience segments that are homogeneous with respect to expected behaviour.⁵ With regard to PsyOps, audience segmentation based on demographics is useful. However, to cover potential vulnerabilities, additional segmentation methods must be introduced, such as: socio-economic parameters

- geography
- access to information channels
- authority/orientation
- values/triggers
- locus of control.⁶

In recent years the ability of technology to collect, process and provide actionable insights into audience data has revamped interest in audience segmentation. In addition to demographic indicators, psychosocial factors are increasingly being considered; examples include targeted advertisements and messages based on linking personality traits to consumer behaviour.⁷ Given these advances in audience segmentation, for further analysis two questions must be posed:

- Is it possible to gather a personality profile without delivering proper testing?
- How much can a personality profile (profile information, interests, engagement behaviour and other features) indicate a person’s expected behaviour?

We address both questions later in this report.

Construct validity

Construct validity pertains to the extent to which the inferences drawn from the measurements in research can be justifiably applied to the conceptual frameworks that underpin those measurements. The question is: are the variables measured the best choice to give the desired results?

Establishing construct validity requires three essential components: unidimensionality, reliability and convergent/discriminant validity.⁸ In short:

- Unidimensionality involves demonstrating that a group of observed measures is linked to a single

underlying construct and not to any other construct.

- Reliability concerns the extent to which a measurement produces consistent or stable results; this quality is negatively related to the extent to which a measure is contaminated by random error.
- Convergent validity evaluates how well different methods of measuring a particular variable generate similar results; discriminant validity assesses the extent to which measures of distinct latent variables are unique.

In other words, for the measure to be considered valid, the variance in the measure should reflect only the variance attributable to its intended latent variable (and not to other latent variables).

For effective audience engagement, it is essential to include specific considerations in the TAA process. While this evaluation remains a primary duty for analysts, AI tools are not meant to replace them but act as a ‘force multiplier’ to enhance their efforts. These AI-driven tools should be transparent, ensuring analysts can easily comprehend how a particular graph or data representation is generated. The case of sentiment analysis clearly illustrates this point. This feature is present in most tools supporting TAA, and

often consists of a feature that assesses sentiment on a bidimensional scale (negative–positive).⁹ Yet, emotion is a complex dimension. The lack of clarity that follows the use of a single score to represent multidimensional constructs has been identified as an invalidating factor in social science research in general, and in psychology in particular.¹⁰ An analyst’s job is to understand the possibilities offered by the tools available, and to take responsibility for the methodology that will be chosen. The positive aspect is that the ‘black box’ understanding is slowly evolving among communities, institutions, governments and policymakers, as they increasingly recognise the realistic expectations and limitations of AI-driven technologies, rather than being swayed by media hype and exaggerated promises.

Fundamental attribution error

The methodology¹¹ of ‘psychographic profiling’ can be boiled down to a five-factor model of personality traits.¹² The tacit assumption here is that personality traits are the most important predictor of human behaviour. Correctly recognising a person’s traits will therefore enable opinion shift, persuasion or behaviour change (in simpler terms, *manipulation*). There is reason to believe that such methodology is oversimplified. In the field of social psychology, fundamental attribution error describes a cognitive bias in which observers tend to overestimate dispositional and personality-based reasons for an actor’s behaviour, while underestimating the impact of situational and environmental factors.

Any *modus operandi* based on the sole observation of dispositional causes taps directly into this bias. There tends to be more comfort in attributing an actor’s behaviour to their character traits rather than to situational influences.¹³ When someone’s personal traits are overemphasised as the explanation for their behaviour, any forecasting is equally distorted.¹⁴ While the weight of this bias vis-à-vis the actual effects of personality traits is

contested,¹⁵ it is worth considering whether such an approach brings more benefit or harm in the pursuit of designing a campaign suited for a specific target. There is research indicating that marketing campaigns based on psychometrics are more effective than those that do not employ audience segmentation based on these principles.¹⁶ However, that does not mean that an approach based on the gathering of psychometric data is necessarily the superior method of audience segmentation, as such techniques are often portrayed to be.

A comprehensive TAA therefore extends beyond simply understanding demographics and psychographics. External factors can significantly shape the behaviour¹⁷ and perceptions of audiences, necessitating their inclusion in any thorough analysis.¹⁸ Such factors encompass a wide range of elements, including the political landscape, cultural norms, socio-economic conditions and significant global events. Behavioural models which consider the impact of external factors on audiences can be considered for TAA. These include the capability-opportunity-motivation-behaviour model (COM-B model) for

behavioural change,¹⁹ with its focus on capability, opportunity and motivation being able to capture these complexities, acknowledging that the surrounding environment (opportunity) plays a significant role alongside individual capabilities and motivations.

To this end, AI and automation offer substantial benefits for behavioural models for TAA. Machine learning algorithms can be trained to sort through vast amounts of data,

identifying patterns and trends that would be challenging and time-consuming for humans to discern.²⁰ Natural language processing (NLP) can help analyse sentiment and attitudes from social media posts or survey responses, giving insight into both individual motivations and the larger societal context.²¹ AI can help make sense of demographic and psychographic data, alongside crucial external factors,²² providing a more holistic understanding of the audience and enhancing the effectiveness of StratCom.

Operationalising the variables

Keeping in mind the caveat outlined above regarding the potential oversimplification of methodologies which overemphasise personality traits as the explanation for behaviour, it is, however, possible to consider how TAA tools can be made more effective. In short, this is achieved by operationalisation, that is, by turning abstract conceptual ideas into measurable observations. Through operationalisation, data can be systematically collected on processes and phenomena that are not directly observable. This is the case for attitudes that are routinely investigated through TAA. For example, while the popularity of a narrative cannot be gauged directly, metrics are available to keep track of how much a story spreading that narrative is popular in the IE.

The process of operationalisation involves three steps: identifying the main concepts, choosing a variable to be assigned to each concept and selecting indicators for each of these variables. The risk, for PsyOps, is that such a process might be carried out only 'once and for all', crystallising a structure that needs continuous readjustment to be effective. In practice, this means measuring only certain indicators because they represent the most easily available ones, those that were used previously and that are better understood by analysts.

Ideally, operationalisation should inform the segmentation of audiences. It could be useful to consider best practices in marketing. For a market segmentation scheme to support

public engagement campaigns, it must have five qualities:

1. Segments should be different, with members within a segment sharing enough similarities to be targeted with the same strategy.
2. Segments should align with the campaign's goals.
3. Segments should be sizable enough to warrant the effort to target them.
4. It should be possible to identify individuals' segment status.
5. The campaign organisation must be able to target at least one of the segments, possibly requiring adjustments to its structure or systems.²³

What is described above should be an integral part of the audience targeting process. Better targeting means better execution of the PsyOps campaign. However, this process is also a potential source of mistakes, reverberating through the whole campaign. It is paramount to ensure that we are measuring what we set out to measure in the first place: in other words, it is necessary to ensure construct validity for the indicators we chose. We will now consider this issue in more detail.

When we fail, why do we fail?

When the targeting process fails, with regard to audiences, how does that happen, and why? It can be useful to consider at which level the failure takes place.

One reason to consider is the difference in speed of innovation and turnover when comparing military and civilian sectors –because of, for example, sensitive information, that is true not only for AI system innovation, but also in broader contexts. Considerable difference between sectors can also be found in the availability of resources.

Previous publicly discussed PsyOps campaigns have failed because the targeting process was not nuanced enough. However, better dashboards and more statistical data were not what were required—in fact, there were objections to the overreliance on these tools at the expense of human contact.²⁴ While in abstract terms it is hard to argue against more sophisticated TAA tools, a risk connected to their use in practice is that this will come at the expense of investing in other forms of research. Budget constraints might force a choice between investing in AI-powered TAA tools or carrying out focus groups and face-to-face interviews.

Introducing the behaviour model: COM-B as a framework for TAA

In the context of TAA in NATO StratCom, the COM-B model for behaviour change serves as an insightful framework for understanding key factors influencing the audience, moving beyond demographic and psychographic understanding.²⁵ Using this model, the audience's capability, opportunity and motivation, in the context of the desired behaviour change, serve as vital components guiding the design of more effective messages and campaigns. AI can further enhance this process by automating certain aspects of the analysis and providing data-driven insights.²⁶ For instance, AI-driven NLP and sentiment analysis tools can assess public

opinion, while machine learning algorithms can identify patterns and trends in large data sets.

By understanding the capabilities, opportunities and motivations of the audience, and leveraging recently developed AI capabilities, NATO communicators can more efficiently design StratCom campaigns that resonate with their audience and successfully influence behaviour to achieve the desired objectives. We will now introduce the definitions of the key components of the COM-B model—capability, opportunity and motivation—within the context of TAA.²⁷

Capability

Capability refers to the audience's psychological and physical capacity to engage in the desired behaviour. Understanding the audience's knowledge, skills and abilities is crucial for developing effective messages and interventions. Three separate processes can be done in TAA to identify the audience's capabilities: (a) identify the audience's current

knowledge, skills and abilities related to the desired behaviour change; (b) assess the gaps and barriers that may prevent the audience from adopting the desired behaviour; and (c) develop communication strategies to address these gaps and barriers, including educational materials, training and resources.

Opportunity

Opportunity involves the external factors that affect the audience's ability to engage in the desired behaviour, such as social influences and the environmental context. Three separate processes can be done in TAA to identify the opportunities of the audience: (a) identify the social, cultural and environmental factors that influence the audience's behaviour; (b) assess how these factors may promote or hinder the desired behaviour change; and (c) develop communication strategies that address these factors, such as creating supportive social networks or modifying the environment. Within the opportunities the communication channels must be identified that the audience trusts and uses regularly. Assessing the specific information consumption behaviour of an

audience is vital, particularly in diverse, multilingual and multicultural societies. Different audience segments may use different channels to consume information and may have distinct preferences and behaviours within those channels. Therefore, to ensure effective communication and increase the likelihood of the campaign message being received, it is crucial to understand these behaviours and channel preferences at a granular level. Moreover, identifying barriers to behaviour changes within the audience is also an essential step in formulating an effective campaign strategy. Once these barriers are identified, strategies can be developed to overcome them, thereby making the campaign more successful.

Motivation

Motivation refers to the audience's conscious and unconscious mental processes that drive their behaviour. Understanding the audience's beliefs, values and emotions is essential for crafting persuasive messages and interventions. Three separate processes can be done in TAA to identify the motivations of the audience: (a) identify the audience's

beliefs, values and emotions related to the desired behaviour change; (b) assess how these factors may promote or hinder the adoption of the desired behaviour; and (c) develop communication strategies that leverage or address these factors, such as appealing to shared values or using emotional appeals.

Overview

While the COM-B model can guide the identification of what needs to change for a behaviour to occur, it does not provide specific guidance on how to bring about these changes. The key is to use the audience behavioural insights gained from the COM-B analysis to inform the development of tailored strategies that address the specific barriers and facilitators needed to invoke a change in audience behaviour.

In the context of comprehensively understanding and applying the COM-B model to TAA, it is important to examine how data can be employed to inform each facet of this behaviour

model—capability, opportunity and motivation. Table 2 shows a summary of this process, interrelating various categories of data and their sources—whether online or offline—with different facets of audience behaviour. Here, we briefly touch on the role of AI in facilitating data collection, processing and interpretation to discern intricate patterns and trends. The linking of various data types and sources with the diverse behavioural aspects of the audience that each source seeks to study and measure offers a practical demonstration of the COM-B model in action within TAA, while introducing how AI could aid this process.

Data type*	Data sources	Online/offline	Usage of AI†	Audience behavioural aspect
Survey data	Online survey data	Online	AI can aid in the creation and distribution of online surveys, and in analysis of responses to assess audience attitudes, beliefs, goals and values. This includes clustering or segmenting audiences based on various demographic, socio-economic and psychographic characteristics.	Capability Opportunity Motivation
	In-person survey data	Offline		
	Census data			
	Publicly available statistics			
Usage data	App usage data	Online	AI can identify patterns in usage, such as the most commonly used features, and the times of day when the app is most used. This provides insights into users' skills and knowledge, and their preferences and habits. AI can also help in the real-time collection of data, triggering certain actions based on user behaviours or providing personalised recommendations.	Capability
	Website interaction data		AI can analyse how users navigate through a website, which pages they visit, how much time they spend on each page, etc. By identifying patterns in data, AI can help in understanding aspects of websites users find most useful or engaging. Anonymised web traffic metadata can be turned into quantifiable measures of attention to individual narratives, topics and themes.‡	
Educational data	Educational records	Offline	AI can be used to identify correlations and trends to indicate a population's capabilities.	Capability
	Competency assessments			
Health data	Medical records	Offline	AI can be used to analyse medical records to gain insights into an individual's or a population's physical and mental capabilities. AI can extract key information from structured and unstructured data, such as diagnosis codes or physician notes. AI can also automate the collection of data by interfacing with electronic health record systems and pulling relevant data for analysis. §	Capability
	Health assessments			
Observational data	Direct observation	Offline	NLP can be used to analyse observers' notes, or computer vision can be used to analyse photos or other visual data. For example, sentiment analysis could be applied to observers' notes to quantify emotional responses, or image recognition could be used to identify patterns or anomalies in visual data.	Capability Opportunity Motivation
	Video/audio recordings		Object detection, facial recognition, gesture recognition and emotion detection can be used to gain insights from video data. Emotion detection algorithms can be used to infer people's responses to different stimuli or situations, providing valuable insight into their capabilities. Recognition software can identify individuals in video footage. Object recognition can list the different elements present in a scene. Speech recognition can transcribe spoken words in audio or video recordings. Sentiment analysis can evaluate the emotional tone of those words.	
	Environmental studies		AI-enabled devices can be used to collect environmental data like temperature, noise levels or air quality. This data can provide insights into the physical aspects of the environments where the audience live or work, which may influence their behaviour.	
	Ethnographic studies		AI can analyse text, visual and auditory data collected.	

Data type*	Data sources	Online/ offline	Usage of AI†	Audience behavioural aspect
Qualitative data	Interview transcripts	Offline	AI can be used to analyse transcripts. Text analytics, sentiment analysis and topic modelling techniques can extract key themes, emotional tones and specific points of interest. For instance, sentiment analysis can identify positive or negative sentiments in interview transcripts, while topic modelling can identify common themes that emerge in focus group discussions. AI can help streamline this traditionally labour-intensive process and add objectivity to the analysis.	Capability Motivation
	Focus group discussions			
	Open-ended survey responses			
	Online media	Online	NLP can be used to understand the tone and emotional context of news articles, social media posts and other forms of media. Content analysis can be done, identifying trends and patterns and how topics or groups are represented in the media.	Opportunity Motivation
Cultural insights data	Academic research	Offline	NLP can be used for topic modelling, to aid understanding of cultural dynamics that affect the opportunities available to the audience.	Opportunity Motivation
	Ethnographic studies		NLP can be used to analyse interview transcripts, field notes and other text-based data to identify themes, sentiments and patterns. Computer vision can be used to analyse visual data like photographs or videos, to aid understanding of cultural factors.	
	Media analysis	Online or offline	See 'Online media' analysis above. This can reveal cultural biases and norms that may influence the opportunities available to the audience, and their motivation.	
Social media data	Publicly available posts and comments on platforms (X (Twitter), Facebook, Instagram, Telegram, VK, TikTok, etc.)	Online	NLP can be used to analyse sentiments and stances to gauge public opinion, attitudes and emotions related to topics.	Motivation

* All data types listed are relevant to IEA elements.

† Based on the complexity of the data presented above, AI can be used for text analysis, natural language understanding, data categorisation and classification, trend identification, insights generation and natural language generation. However, for non-textual data types like health data, usage data or educational records, large language models (LLMs) may not be directly applicable. These might need more specialised models or algorithms designed for numerical or categorical data analysis. Yet, LLMs can support interpreting or communicating the results of these analyses.

‡ Contributed by Predata.

§ The discussion of AI's potential use in analysing health data is hypothetical and not practised by the NATO StratCom COE, with all theoretical applications being considered in strict compliance with GDPR standards.

TABLE 2. Integrating COM-B and AI in TAA: a classification of data sources and their applications

Example: TAA in a marketing company's digital advertising campaign

In the rapidly evolving digital landscape, the marketing industry has consistently been at the forefront of harnessing data for TAA: its advanced methodologies provide invaluable insights into audience engagement, making it an exemplary sector to delve into for understanding the intricacies of TAA. For an applicable illustration, we explore the scenario in which a typical marketing company conducts TTA while orchestrating a digital advertising campaign.²⁸ The goal is to underline the relevance of data-driven decision-making within diverse business ecosystems, and the objective is to promote comprehension and influence attitudes about its suite of services across various sectors. A thorough understanding of the audience's capability, opportunity and motivation is vital for effective execution. AI is central to TAA, serving as a key component from data collection to the analysis of various stages in the process.

In appraising the capability and motivation of the audience, AI tools can parse through both online and offline data sources: online data includes survey responses, social media interactions and website engagement metrics, while offline data consists of interview transcripts, feedback from focus groups and field research data. These sources provide an understanding of public awareness of the company's services and the audience's beliefs, values and emotional responses to data-driven decision-making. NLP is applied to analyse qualitative and textual data for nuanced insights. The opportunity for the audience is scrutinised to identify and analyse social, cultural and environmental factors that could influence audience engagement. This again involves examining both online data—which includes public statistics, market reports and industry-specific digital content, allowing for audience segmentation based on demographic and socio-economic characteristics—and offline data sources, consisting of ethnographic studies, market surveys and field studies conducted using AI-enabled devices to collect environmental data. These

are collectively enriching the understanding of cultural dynamics and potential opportunities for engagement.

The knowledge derived from this AI-aided TAA directly informs the design of a StratCom campaign. The campaign produces customised messages to engage specific audience segments, with the intent to close understanding gaps, leverage cultural norms, and appeal to values and emotions linked to data-driven decision-making. Post-campaign evaluations use both online and offline data sources. Online feedback is gathered through surveys and social media sentiment analysis to find out if there is a positive shift in business perception regarding the company's services in various industry sectors. Offline data, including focus group discussions and one-on-one interviews, is used to affirm the online sentiment analysis results. Increased engagement with content related to data-driven decision-making on the company's digital platforms is then used as a measurement of the effectiveness of the digital advertising campaign.

Transitioning from the commercial context of a marketing campaign to the more specific domain of PsyOps, there are underlying commonalities between these two domains. The above exploration of a marketing company's AI-enabled digital advertising campaign serves as a practical demonstration of how elements of TAA can be employed to shape public perception and attitudes. An effective PsyOps strategy requires comprehensive insight into the audience's knowledge, abilities, environmental factors and internal drive to behave in certain ways. It is through this understanding that messages and campaigns can be formulated to resonate, evoke desired reactions and ultimately influence behaviour in audiences.

For the process to be effective, the use of AI is crucial. AI offers the ability to gather meaningful information quickly and enhance the overall results. To enhance the efficacy of

influence campaigns, understanding the digital environment, especially social media platforms, is essential. Alongside a deep understanding of the audience from the COM-B analysis, it is equally important to have a technical grasp of how these platforms function. The combination of audience insights and technical platform knowledge can significantly increase the impact of an influence campaign.

Before initiating a campaign, it is beneficial to have a clear understanding of the workings of social media platforms, their revenue generation methods and their algorithms. Such knowledge ensures improved content delivery, increased audience engagement and effective navigation of the digital realm, all of which are vital for strategic information operations.

Technical prerequisites of a successful campaign

Understanding communication platforms in campaign design with the focus on web operations, grasping the platforms' functionality, revenue models and algorithms, allows for the optimisation of content and messaging, ensuring greater reach and engagement with the intended audience. This comprehensive knowledge empowers campaign planners to navigate the digital landscape effectively, using platform-specific dynamics to achieve desired goals in the realm of information operations.²⁹ For example, highly redacted open-source Twitter source code reveals a fraction of algorithm details and ranking parameters, which could potentially be exploited by both malicious actors and government/military campaigners to manipulate public discourse and launch targeted influence campaigns (in much the same way as advertising marketers).³⁰ In addition to thorough audience research, a deep understanding of social media algorithms is essential for crafting the most effective online campaigns. Open-source

code enables a deeper comprehension of these algorithms, allowing for more strategic content creation. Overall, deep research of individual platform-targeting mechanisms is highly necessary to reach the desired targets in the most cost-effective way, whether through recommendation mechanisms used by ad frameworks or automated channel messaging mechanisms like Telegram bots.

In the following sections, we will delve into the technical prerequisites necessary for orchestrating a successful campaign. We will expound on various facets of IEA, illustrating their roles and importance in the broader scheme. Further, we will reveal how the latest advances in AI can not only support these elements but also significantly enhance their efficacy. This integration of AI with IEA forms a powerful synergy that can be further harnessed in PsyOps.

Borrowing useful tools from IEA

Choosing the most effective current AI models and using innovative AI systems is a crucial aspect of influence campaigns. However, we must not overlook the unprecedented pace at which AI systems are evolving. It is imperative to focus not just on technology, and in this context we advocate for a process-oriented approach designed with

flexibility at its core. By building processes that are model-agnostic, we can create a 'plug-and-play' framework that allows for easy integration of new and improved models as they become available. This approach not only ensures our influence campaigns remain at the forefront of technological advancements but also allows for continual refinement and

optimisation of the processes involved, creating a sustainable and adaptable system.

For a clearer understanding, we will walk through different technical tools used in IEA. We will show how to combine these tools

Hashtags and entities

Hashtags are key markers in social media platforms used to categorise and link related content. They play a vital role in the IEA by enabling the tracking of discussions, sentiments and trending topics related to specific themes or issues. By analysing hashtags, analysts can gain insights into narratives, identify influencers and understand the dynamics of discussions across different platforms. Thus, hashtags serve as crucial navigation tools in the vast and complex information landscape, helping to assess and interpret the pulse of the digital environment.

Many social media listening tools use donut charts to visualise the share of top trending hashtags on different social media platforms. For a practical example we looked at data from X (formerly Twitter) between 1 January and 1 February 2022 (Figure 1) and between 24 February and 24 March 2022 (Figure 2), preceding and following the Russian invasion of Ukraine. Figure 1 shows that the top three hashtags refer to cryptocurrency (#bitcoin), the

to efficiently support operations. To make our examples relevant, we have drawn from the digital IE during the early stages of Russia's invasion of Ukraine.

English Premier League (#mufc) and COVID-19 (#covid). Notice there is already some discussion about the Russia–Ukraine situation during this period, as related hashtags show up in the top ten list. Compare this to the chart in Figure 2, corresponding to the weeks after the start of the invasion on 24 February, which shows a huge bump in activity for these hashtags.

The hashtags #ukraine, #russia and #nato accounted for 25 per cent of online activity during this period. Therefore, public interest shifted away from COVID-19 following the start of the war.

Another aspect when considering hashtags is that these **conversation handles** are prone to hijacking, which is the practice of co-opting a popular or trending hashtag to divert attention to an unrelated topic or agenda. This disruptive tactic can dilute the original message, cause confusion and even damage reputations. Large language models (LLMs)

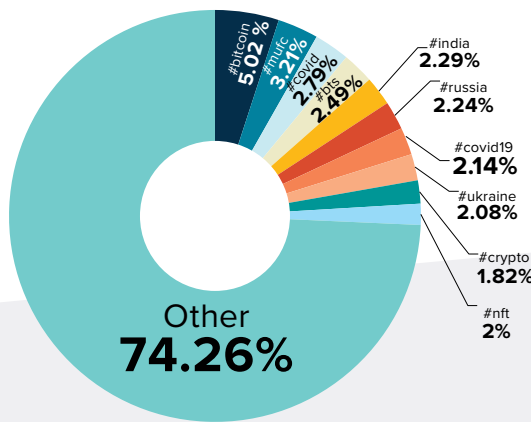


FIGURE 1. Top hashtags, 1 January – 1 February 2022 (sample size 5.5m)

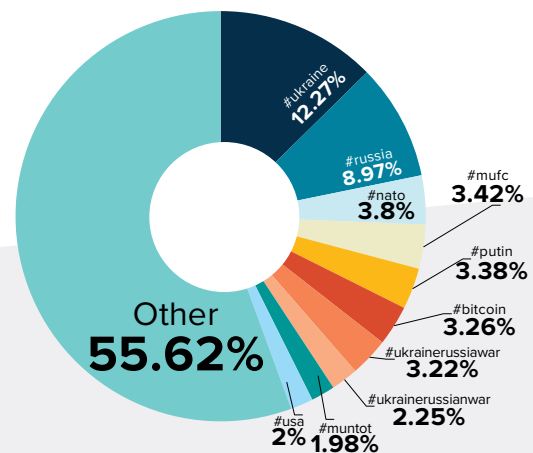


FIGURE 2. Top hashtags, 24 February – 24 March 2022 (sample size 6m)

provide a sophisticated solution to identify and combat hashtag hijacking. By employing historical analysis, these models can also recognise common hijacking patterns, aiding in the development of proactive strategies to protect the integrity of online conversations.

In addition to usage pattern analysis (seeing a sudden spike in usage of a hashtag), here is how an LLM might help identify instances of narrative/hashtag hijacking:

- **Content analysis:** LLMs can aid in detecting shifts in topics and narratives behind a hashtag. In other words, the reasoning capabilities of a model help analyse the content of posts using a hashtag, allowing identification of unrelated topics when they diverge from the original subject. Similarly, comparing recent posts with the historical usage of the hashtag could further confirm that the original purpose had been hijacked.

- **Sentiment/stance/intent/support analysis:** more sophisticated stance and sentiment analysis has the potential to detect a shift from positive to neutral or even negative sentiments when analysing communications on social media.

By flagging these inconsistencies, LLM-powered tools can alert moderators or automated systems to take action, such as removing the unrelated posts or even banning the offending accounts. The same goes for politically or issue-related hashtags, as may be used by militaries and governments.

This is important in the context of this report because analysts now have the capacity to detect changes in public interest and the surrounding topic categories of interest, and to assess clusters of related social media accounts and actors. Analysts are then able to focus efforts on those posts corresponding to the hashtag of interest and further apply NLP models such as named entity recognition (NER) and sentiment classification.

NER is a component of information extraction that detects and categorises named entities in text into set groups like persons, organisations, locations, times and quantities. NER is a crucial aspect in information retrieval, question answering and content summarisation. By pinpointing these specific data points in large volumes of text, NER provides structured understanding and context, thereby enhancing the efficiency of data analysis. In the context of an online campaign, NER enables more effective tailoring of messaging strategies, for instance by recognising key persons, organisations or locations within specific topic discussions that resonate among groups of interest, thus improving the impact and reach of the campaign.

However, the approach is not without its limitations. Factors such as language specifics and biases in training data, including the underrepresentation of certain geographic locations, individuals or other entities, can affect the model's effectiveness. Nonetheless, the advent of LLMs has promised (at least theoretically) to mitigate part of these challenges. Even though LLMs are powerful and, in many cases, multilingual, employing techniques like few-shot learning (using the AI model's reasoning capabilities for prediction/classification tasks etc. by showing a small sample of examples), as well as carefully defining the entities being sought, can help improve results.

Another important aspect is that LLMs, like GPT-4, are trained on a vast array of sources, incorporating a more comprehensive understanding of languages and contexts. Manual entity linking (e.g. capital of Estonia = Tallin) can be learned by the model. Careful and responsible use of these models in tools can be an invaluable asset for international research and analysis, where consistency across languages is crucial. Experiments show that this technology, though progress is as yet incomplete, continues to accelerate in performance.³¹

Challenges with sentiment analysis: from polar sentiment towards directional sentiment

A sentiment classification model takes a document as input and produces a label that indicates the author's sentiment towards the content, based on the language used. More advanced modelling also allows for a more nuanced analysis, to include degrees of sentiment such as hostility, support and hate (and many more) towards extractable entities.

For simple experiments we used a transformer model³² designed specifically for the task of classifying text into positive, negative or neutral according to the author's sentiment towards the content. Figure 3 shows the breakdown by sentiment for the top three hashtags on X/Twitter during the first four weeks after the start of the full-scale Russian invasion.

Content expressing negative sentiment is the most frequent for all three hashtags. Of the three, #ukraine has a slightly higher (6 per cent) share of content with positive sentiment compared to #russia (5 per cent) and #nato (4.5 per cent). We could further track attributes of the various kinds of content such as source domain (the platform the content was scraped from), source location (the 'stated'/defined

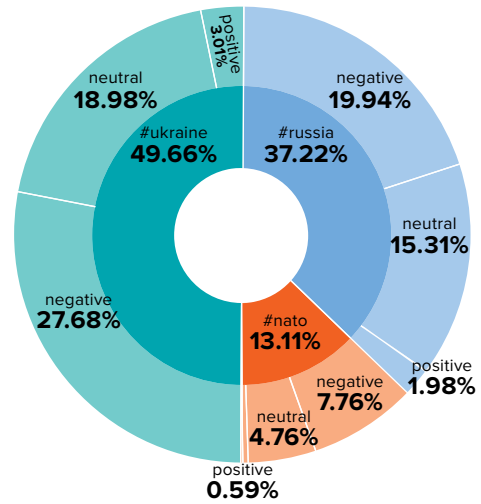


FIGURE 3. Sentiment of content, 24 February – 24 March 2022 (sample size 1.5m)

location of the account posting the content) and source language (the language of the content detected using a language detection model).

There is a problem in that positive/negative/neutral classification by itself is not informative. It gives only the broad perspective of the overall sentiment. However, especially when

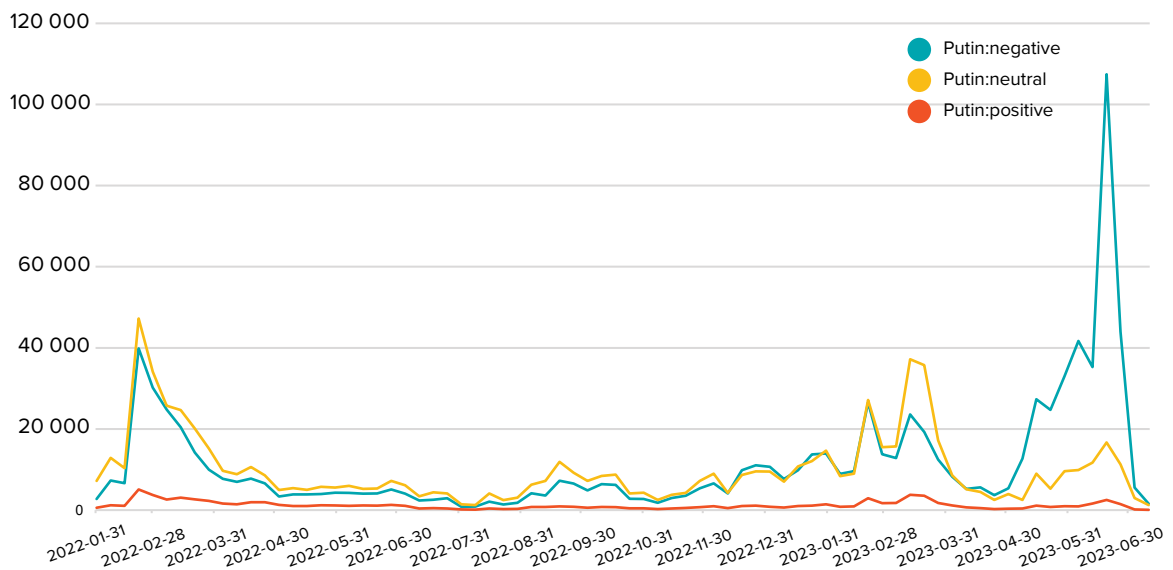


FIGURE 4. Sentiment towards Putin (sample size 1m)

dealing with understanding specific conversations of various audience segments, it is important to estimate and thus understand the audience's attitude or opinion towards a specific topic or named entity. This is done by combining entity detection and sentiment estimation by training a model that does both (Google Cloud NLP provides entity sentiment analysis³³ capability for a limited number of languages) or by doing sentiment estimation on specific text segments where the named entity is mentioned. Figure 4 shows an example of how such a model can be used to track sentiment towards an entity of interest (Putin) along the time dimension.

As a practical example, the line plots of a data sample consisting of over 1 million documents (tweets, blogs, posts) drawn from various sources between February 2022 and March 2023 present a graphic visualisation of the number of documents expressing positive, negative or neutral sentiment towards Putin. Negative sentiment towards Putin was the highest at the start of the war

(i.e. February–March 2022). It tapered down steeply and started plateauing around May 2022. There are some noticeable spikes thereafter, such as around September–October 2022, February–March 2023 and in July–August 2023. Positive sentiment is low ($n = 73K$ out of 1m documents) throughout this period. Neutral sentiment had a similar trajectory to negative sentiment, indicating that in any given sub-interval of time, public opinion was split at a similar proportion between negative and neutral sentiment. These trends can be correlated with major events in the war to gain further insights.

In this way on social media we can detect accounts and groups of accounts that post negatively or positively towards a specific named entity, such as Putin, Zelenskyy, NATO or US President Joe Biden's visit to Ukraine. Such an approach can help IEA teams to detect coordinated behaviours online.

Now, to gain deeper insights, we can use LLMs to estimate whether posts, news articles,

Defining the task	You are an advanced text processing model. Your task is to determine the sentiment towards the following entities: [NATO; Latvia and the Latvian military; Ukraine and the Ukrainian military; Russia and the Russian military].
Describing the reasoning	The European Union (EU), the United States of America (US) are the West. Give results on a scale of -1 to 1, where -1 is extremely negative, 0 is neutral and 1 is extremely positive. In the event the named entities are not explicitly mentioned and cannot be implied, return 99 for NA. Additionally, write down who is the villain, aggressor in the narrative if applicable.
Ensuring that the model will reason based on the data	Carefully consider the examples as well as the narrative titles—sometimes the examples may posit the narrative of the opposite side and refute it.
Defining the output	Finally, indicate scores from 0 to 1 to show how relevant the topic is to assessing the political and security implications in Europe, NATO, Russia, the Baltics and the West in general. Scores are always from 0 to 1, where 0 is not relevant and 1 is extremely relevant. E.g., a narrative about the weather in Ukraine would be 0, a narrative about the war in Ukraine would be 1. Do not print any introduction, simply return your answer in the specified format: Latvia&Baltics: FLOAT\nNATO: FLOAT\nUA: FLOAT\nRU: FLOAT\nWest: FLOAT\nRelevance: FLOAT\nAggressor: STR Repeat for each narrative (if more than 1).

TABLE 3. GPT-4 prompt to extract the relevance of a news article or social media post

transcriptions of TV/radio broadcasts, podcasts or videos support or oppose specific topics and narratives. For this example we designed the GPT prompt shown in Table 3.

We executed this prompt on every news article in our example dataset so that each news article for the respective news domain would have a vector with five float numbers. Once finished, we computed an average value for each news domain and visualised the results using the coloured heatmap as shown in Figure 5. The example dataset consists of 1.5K news articles mentioning 'Latvia' and shows AI estimated support for or disagreement with news domains vs a specific topic/entity as described in the prompt in Table 3. We sorted the data by the *Latvia&Baltics* column to show news domains that strongly

disagree with the statements related to Latvia and the Baltic States. Notice the neutrality and support for Russia-related topics and negativity towards the *West*.

The experiment indicates the presence of two distinct information spheres, each with differing narratives and values. This phenomenon can be analysed through the simple structuring of prompts, revealing underlying biases and the complex interaction of information in the region. The contrast between these two spheres reflects regional tensions and may also be indicative of broader international dynamics that must be considered while designing a campaign.

In an analogous way the communication of various audience segments can

source	Latvia&Baltics	NATO	RU	UA	West
russian.rt.com	-0,3	0	-0,5	0	0
altapress.ru	-0,3	0	-0,6	0	0
radiosputnik.ria.ru	-0,333	0	0,133	0	0
finance.rambler.ru	-0,35	0	-0,5	0	0
aif.ru	-0,373	0	-0,1	0,171	-0,04
vm.ru	-0,375	0	-0,367	0,133	-0,45
mk.ru	-0,375	0	-0,025	0	-0,3
rusonline.org	-0,4	0	0,4	-0,8	-0,5
ryb.ru	-0,456	0	0,189	-0,5	-0,4
nakanune.ru	-0,5	0	-0,5	0	-0,5
putin-today.ru	-0,5	0	0,5	0	-0,5
theins.ru	-0,5	0	1	-1	-0,5
fishki.net	-0,5	0	-0,633	0	0
rg.ru	-0,567	0	0,467	0	-0,2
belta.by	-0,583	-1	0,667	-0,125	-0,675
vz.ru	-0,6	0	0,233	0,2	-0,2
mpsh.ru	-0,75	0	-0,025	0,033	-0,563
pnp.ru	-0,8	0	0,2	0	0
segodnia.ru	-0,8	0	0,2	0	-0,5
russiapost.su	-0,8	0	0,35	-0,7	-0,9

FIGURE 5. Heatmap of the average estimated support/opposition scores by GPT-4 where the estimated relevance to the topic is above 0.5. Zero filled missing values indicate that within the chosen sample there were no relevant topics mentioned in the text of news articles for a specific news domain

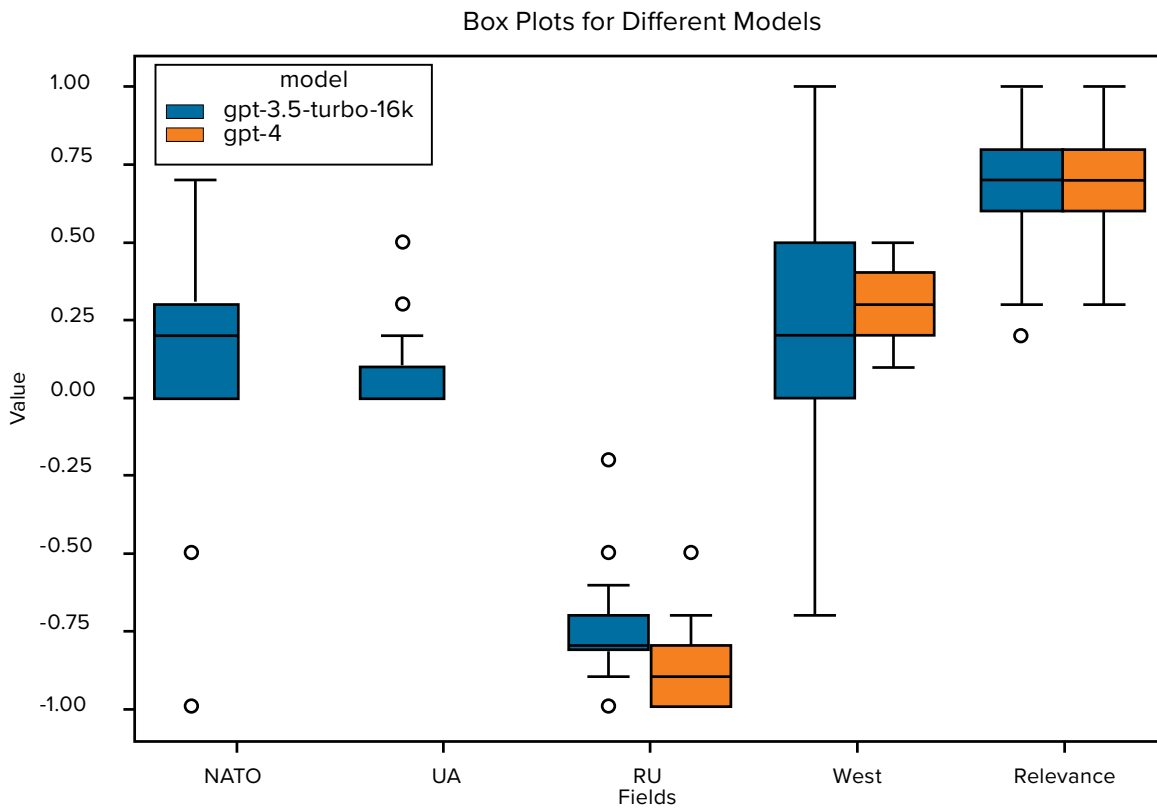


FIGURE 6. Box plots of estimated overall support for Russia in the Latvian language

be analysed. If social media post timelines of identified individuals are available, LLM-enabled tools can help to find the ones that support or oppose a specific campaign, for example. It is expected that the better the reasoning capabilities of the model are, the greater will be the increase in *zero-shot* and *few-shot* learning capabilities.

Finally, an important aspect to mention is the challenge of controlling the LLM outputs (output consistency and therefore reliability). We can demonstrate this by executing the same query from Table 3 on a single Latvian news article³⁴ in the Latvian language. We use GPT-4 and GPT-3.5-turbo-16k (hereinafter, GPT-4 and GPT-3.5, respectively) models without any modifications, 100 times per model. Figure 6 shows box plots of the estimated support variables in defined topics and overall relevance. The experiment shows high variation for both models; however, the GPT-3.5 model with very high variation introduces scores for topics related to NATO and UA, whereas GPT-4 does not. Also, GPT-4 seems to

be more consistent and confident that this article is negative towards Russia and almost neutral or positive towards West. However, the GPT-3.5 model struggles with estimating stance towards the West and gives a very wide range of values. When estimating article relevance to the overall topic of our interest, both perform almost equally. High variation questions the reliability and consistency of the results and raises the question of the difficulty of controlling current LLMs.

Overall, GPT-3.5 is still an appealing model due to (1) its large data context, as well as (2) its speed and (3) its cost. But it requires much more careful data cleaning and result validation when compared with the more expensive GPT-4 if data is processed using prompt engineering only, without fine-tuning the model for a specific use case.

But can we at this point fully rely on the current generative AI capabilities? The short answer is no! However, there is immense potential in using them to filter the data, to

generate summaries and as zero/few-shot classifiers. Large-scale open-source and commercial models have limited effectiveness, particularly in languages with minimal support. Therefore, when military organisations and governments consider employing LLMs to comprehend non-English information environments, they must tackle not only performance challenges but also ensure data privacy and security concerns are properly managed. This is essential to thoroughly recognize and address the capability gaps in multilingual and multimodal contexts.

It is vital to understand that the drawbacks presented arise from simple prompting and can be mitigated or significantly reduced

From topics to narratives

The rise of social media and blogs makes it challenging to filter important data in the IE. Identifying the main messages or narratives³⁹ is key for analysts who study online discussions that include multiple languages and types of data. When narratives are evaluated alongside fact-checked information, it becomes possible to identify and study disinformation campaigns. Understanding narratives is also crucial for assessing the effectiveness of communication campaigns, especially when integrated with technical methods like trend analysis, NER and sentiment analysis. Approaches to identify and analyse narratives vary. A possible example of AI-driven automated narrative extraction is explained by Sadler.⁴⁰

The general design requires filtering out irrelevant or outdated content. The process begins by gathering data and generating contextual text/image embeddings (each document, text chunk or even image is processed using a specific machine learning model and turned into a numerical feature vector called an embedding vector) using cutting-edge embedding vector generators.⁴¹ These embeddings help group relevant documents together using an agglomerative (hierarchical) clustering algorithm to form clusters that represent possible

in two steps: by incorporating representative examples in the prompt, and by fine-tuning a large model for a narrow task.³⁵ Also, these solutions tend towards the more commercial side in promising to respect the company's data. We may speculate that more research organisations and social media analytics companies will be able to use such models on more sensitive data.³⁶ Although the OpenAI API interface promises not to use any data for human-in-the-loop operations, sending sensitive data on remote servers is prohibited. Therefore, especially for military purposes using open-source LLMs such as Llama 2,³⁷ Bloom Big Science³⁸ and others must be considered as these and similar models play an increasing role in today's data processing pipelines.

raw narratives⁴² related to specific stories. Then a filter should be subsequently applied to these clusters to retain only documents containing argumentative or factual information and meeting readability standards. This filtering step removes content with relevant hashtags or keywords but lacking meaningful context, as well as content that doesn't contribute to the narrative. Finally, clusters exceeding a specific size threshold can be presented to the end user as suggested emerging narratives. Users can then explore these narratives based on their individual needs and interests, providing a structured view of the initial information space. However, LLMs can be used here also to summarise the contents of each cluster and extract the higher-level messages as narratives using generative AI capabilities. To achieve this, there are multiple options to choose from OpenAI models with a 16K or even 32K token context window (sometimes referred to as a data context window),⁴³ or Anthropic with a 100K data context window. Meta's recently released Llama and Llama 2 models, together with many other open-source LLMs, open new possibilities to develop privacy-respecting tools to achieve performance like that of commercial ones maintained by big tech companies.

Network analysis

In the context of this report, we also want to emphasise the network analysis that has been widely used to understand and categorise the spread of evolving content on social media. By modelling relationships between entities such as individual accounts, hashtags or domains as nodes and their interactions as connections, various networks such as retweet and co-hashtag networks can be analysed, revealing low-density and high-sparsity networks with complex community structures. Overall, metrics like network density and edge weight, along with centrality metrics like in-degree and PageRank centrality, are widely applied to understand the properties of the social network. Furthermore, these structures allow us to identify the relationships of social media accounts and behaviours, and, more importantly, coordinated behaviours. In studying the Russia-Ukraine war retweet data (Figure 7), communities and influential users were identified using the

Louvain method and k-core decomposition algorithm. The visualisation of these communities underscores the significance of micro-targeting, given the distinct beliefs, behaviours and information patterns of individuals within each community.

Network analysis is crucial for comprehending dynamics within social media online communities. By identifying accounts and communities, generative AI models can distinguish content-based patterns and detect hostile narratives like those from pro-Kremlin or Chinese media agencies. This goes beyond content filtering; it entails analysing connections, influencers and information diffusion. Using algorithms that evaluate network structure and content, hidden affiliations or disinformation campaigns can be revealed. Such insights aid in formulating strategies against disinformation, fostering transparency and reinforcing democratic values in owned communications.

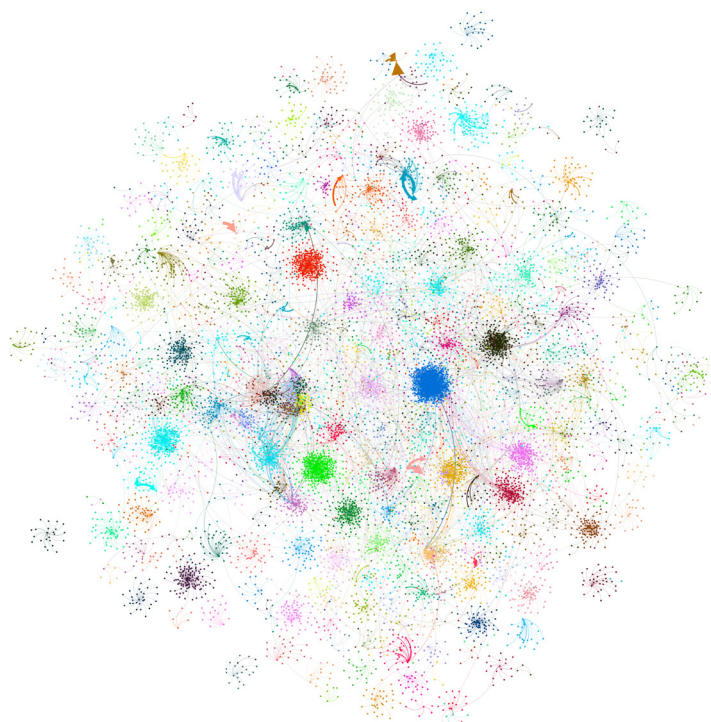


FIGURE 7. Retweet network (1 January 2022 – 1 May 2023), with colour coding indicating Louvain communities

Deriving insights from surveys, interviews and focus groups

The IEA elements outlined in AJP-10 provide a structured framework to understand the broader information environment. The combined strength of insights from surveys, interviews and focus groups, paired with online and social media data, can be directly mapped to these IEA elements, offering a holistic understanding of the audiences' perceptions, beliefs, motivation and behaviour. Interviews delve deeper into the qualitative aspects of IEA, shedding light on the underlying motivations, cultural nuances and cognitive biases that influence information consumption and dissemination. **Deploying surveys** to a cross-section of target demographics can provide quantitative metrics that gauge the effectiveness of communication campaigns. For example, post-campaign surveys can assess recall rates, sentiment towards the message and even the propensity for the audience to act on the communicated information. In a digital age, these surveys can also track the source of the audience's information—for example, if they gained awareness of the campaign from an X post, a YouTube ad or an Instagram story. This allows communication strategists to allocate resources more effectively to platforms yielding the highest impact.⁴⁴ StratCom often aims to evoke emotions or inspire actions. **Individual interviews** help decipher the 'why' behind observed behaviours from online data. For instance, if analytics from a Facebook campaign reveal high engagement but low conversion, interviews might unearth that, while the content was engaging, the call to action was not compelling enough, or there were barriers to the next step.⁴⁵ **Focus groups** can be particularly insightful when launching new communication

campaigns or evaluating existing ones. By observing the dynamics of group discussions, dominant narratives, shared viewpoints or contentious issues can be identified. For example, if a teaser for a new campaign is released to a focus group, the collective feedback, both positive and negative, can offer invaluable insights before a full-scale roll-out. The group setting might also mirror discussions happening organically in online communities or social media platforms.⁴⁶

In tandem, LLMs can be instrumental in various stages of survey research. In designing the survey instrument, LLMs can aid in crafting well-phrased and consistent questions and suggesting optimal response options to gauge respondents' opinions. They can revolutionise sampling by selecting representative samples and proposing suitable techniques for recruiting participants, even employing conversational AI for intelligent interviewing. When it comes to data cleaning and management, LLMs can ensure more accurate and consistent survey data by spotting inconsistent selections, low-quality entries and errors. Innovations like ChatGPT's Code Interpreter plugin have even been utilised to automate data analysis, employing statistical and qualitative methods to uncover insights. Finally, LLMs can streamline reporting and dissemination by summarising findings, creating visualisations and crafting presentations or written reports. Their ability to implement data science code can further enhance the accessibility and comprehension of the results, illustrating the potential of LLMs to add efficiency, accuracy and sophistication to the entire survey process.⁴⁷

Multimodal perspectives

In exploring another aspect of LLMs, we examined the multimodal functions of ChatGPT and Google Bard by requesting descriptions of a set of images. Unfortunately the current versions cannot process images with humans or human faces due to implemented security guardrails (safety mechanisms) that would greatly benefit visual content processing incentives. The tested models also refused to describe images of burning military vehicles and graphic content. Two example descriptions are shown in Figure 8.

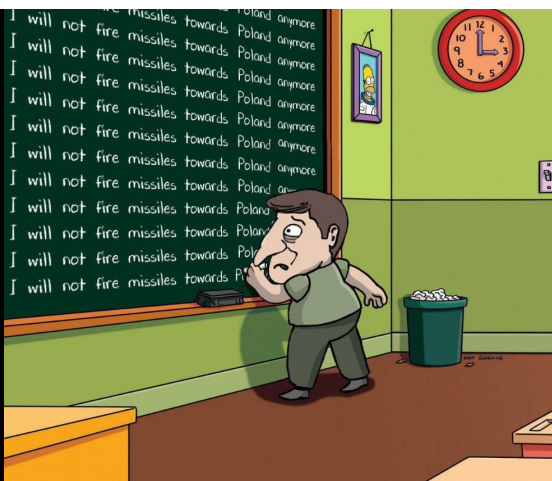
Images with complex symbols, less known geographic locations and so on may

result in confident yet incorrect results (hallucinations), so human review is essential. Despite being in their initial stages, multimodal capabilities show promise and are anticipated to be integrated into social media analytics platforms. This progression allows for broader analyses, though the technology is still evolving. As it advances, these capabilities may provide a fresh approach to targeted communication, aligning with NATO's goals and underscoring the importance of continuous improvement.



ChatGPT:

The image is a black and red emblem. At its centre, there's a menacing skull with a crosshair target encompassing its forehead. Around the skull, written in a circular manner, are the words 'PMC Wagner Group' at the top and the Cyrillic text 'Группа Вагнера' (translated as 'Group Wagner') at the bottom. There are also two white stars on either side of the circular text. The emblem conveys a militant or paramilitary symbolism.



ChatGPT:

The image is a satirical illustration styled similarly to 'The Simpsons' animated series. It features a caricatured man, presumably a representation of a political figure, writing on a chalkboard. The repeated phrase on the chalkboard is 'I will not fire missiles towards Poland anymore.' The repetition mimics a common trope from 'The Simpsons' where Bart Simpson writes lines on a chalkboard as a form of punishment. On the wall, there's a framed image of a cartoon character, and a clock showing the time. A trash bin filled with crumpled paper can be seen nearby. The caption at the bottom reads 'image used for illustration purposes only', suggesting that the image is making a humorous or critical commentary on a political or current event.

Google Bard (with incorrect description underlined in red):

The image you sent shows a skull with a cross in the center. The skull is wearing a helmet with the word 'Wagner' on it. The Wagner Group is a Russian private military company that has been accused of committing war crimes in Ukraine.

The image can be interpreted in a number of ways. It could be seen as a symbol of the death and destruction that the Wagner Group has caused in Ukraine. It could also be seen as a warning to the Wagner Group that their actions will not be tolerated.

Google Bard:

The image you sent shows a cartoon of a man writing on a blackboard. The man is wearing a military uniform and has a swastika armband on his arm. The blackboard is covered in writing, but the only words that are clear are 'I will not fire missiles at Poland anymore.'

The image is likely a reference to the recent accusations that Russia has been preparing to fire missiles at Poland. The image is a warning to Russia that such an attack would be met with severe consequences.

The image is also likely to evoke a range of emotions in viewers, including anger, sadness, and disgust. It is a powerful reminder of the human cost of war and the need to hold those responsible for atrocities accountable.

Here is a more creative interpretation of the image:

The image is a memento mori, a reminder of our own mortality. The skull represents death, the cross represents hope. The helmet with the word "Wagner" on it represents the forces of darkness.

The image is a warning to those who would seek to do harm to others. It is a reminder that we are all connected, and that our actions have consequences. It is a call to action to stand up against evil and create a better world.

The image is also likely to evoke a range of emotions in viewers, including anger, disgust, and fear. It is a powerful reminder of the dangers of war and the need to stand up against aggression.

Here is a more creative interpretation of the image:

The image is a metaphor for the power of words to change the world. The man with the swastika armband represents the forces of darkness and hatred. The blackboard represents the power of truth and justice.

The image is a reminder that we all have the power to speak out against evil and create a better world. It is a call to action to stand up for what is right, even in the face of danger.

Figure 8. ChatGPT and Google Bard descriptions of two images

Fusing everything together

In this section we will explore the integration of various elements to form a coherent strategy. The use of AI in targeted StratCom campaigns is no longer a novel concept but an essential tool. The focus will be on how AI can enhance pre-campaign planning, streamline campaign execution and provide comprehensive post-campaign analysis, all within the bounds of responsible and ethical use. This toolbox aims to elevate communication effectiveness and adapt to the dynamic landscape of information dissemination.

Building upon the relationship between AI technologies and StratCom, particularly in PsyOps, a systematic approach is essential for enhancing the effectiveness of communication campaigns.⁴⁸ This approach is detailed in Figure 9 (columns titled ‘Campaign Stages’ and ‘Campaign Phases’). The pre-campaign phase starts with Strategic Campaign Planning, where the foundation for the entire campaign is set by outlining clear objectives and strategies. This

is closely followed by Target Audience Analysis, which ensures that the campaign’s messages are tailored to resonate with the intended audience. Concluding the pre-campaign phase, the Campaign Intervention Strategy is crafted, providing a detailed roadmap for the forthcoming interventions.

Transitioning to the during-campaign phase, the Intervention step actualises the strategies and ensures alignment with the insights gathered about the audience. Once the campaign has run its course, the post-campaign phase commences with the Measure of Effectiveness (MOE). This phase is pivotal for evaluating the campaign’s impact, analysing its outcomes and measuring its overall success. In dynamic campaigns that adapt in real time, both the Intervention and MOE steps can be conducted simultaneously during the campaign’s progression. This structured approach, encompassing pre-campaign, during-campaign, and post-campaign phases, ensures

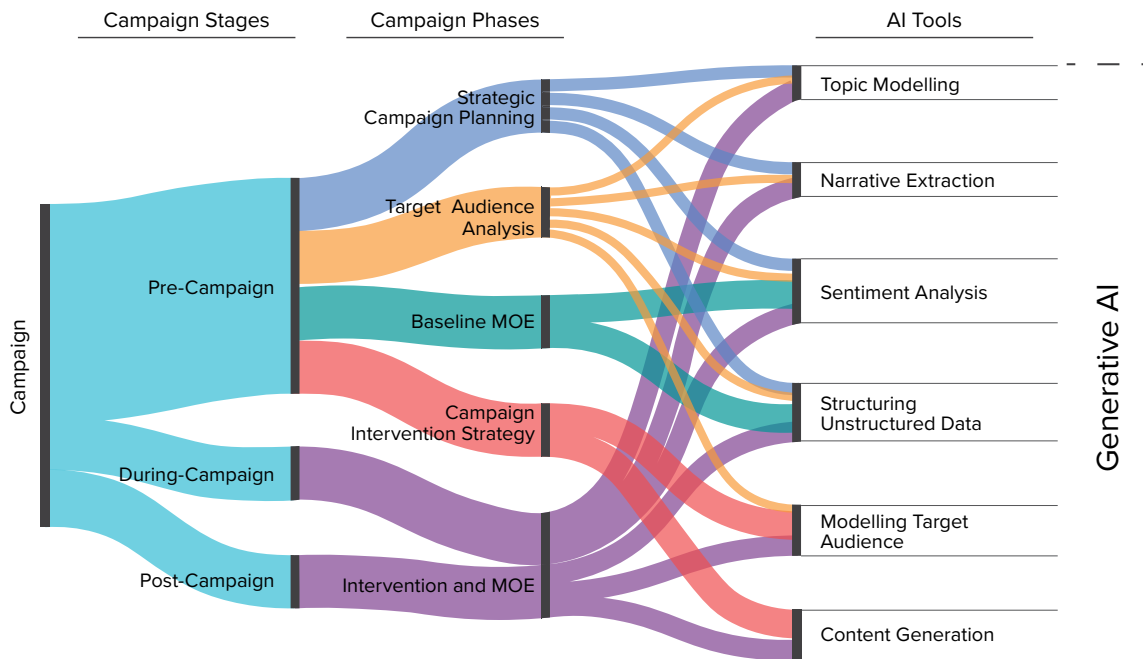


FIGURE 9. Summary of usage of AI tools in various stages of a StratCom campaign

that communication interventions, especially those leveraging AI and PsyOps insights, are both impactful and adaptive to the dynamic IE.

The utilisation of specific AI tools can significantly enhance the effectiveness of communication campaigns across various phases (see Figure 9). In the Strategic Campaign Planning phase, Topic Modelling aids in identifying prevalent themes and subjects within vast data sets, ensuring that the campaign's focus aligns with current trends and interests. Narrative Extraction complements this by pulling out dominant stories or narratives from the data. Sentiment Analysis gauges the emotional tone behind the data, helping strategists understand public sentiment towards particular topics, which is crucial for shaping the campaign's messaging. Lastly, Structuring Unstructured Data organises the vast amounts of information, making it more accessible and actionable for campaign planners.

As the campaign moves to the Target Audience Analysis phase, Modelling Target Audiences would help create detailed profiles of the intended audience segments. This tool, combined with the insights from Topic Modelling, Narrative Extraction and Sentiment Analysis, ensures that the campaign's messages are tailored to resonate with the intended audience. For the Campaign Intervention Strategy, the focus shifts to crafting actionable plans: Modelling Target Audiences provides insights into how different audience segments might react to various interventions, guiding the strategy formulation. Content Generation, being a subset of Generative AI, plays an important role here, helping strategists devise content that is not only relevant but also engaging.

During the Intervention phase, Content Generation can be used to create real-time content that aligns with the campaign's objectives. Tools like Topic Modelling and Narrative Extraction continuously monitor the IE, ensuring that the campaign remains relevant. Sentiment Analysis provides real-time feedback on public sentiment, allowing for timely adjustments. Lastly, in the MOE phase, the campaign's impact is assessed. Topic

Modelling and Narrative Extraction help in understanding the post-campaign discourse, while Sentiment Analysis gauges the emotional impact of the campaign. Structuring Unstructured Data organises post-campaign data for analysis, and Modelling Target Audiences assesses how different audience segments reacted to the campaign.

Building upon the comprehensive capabilities of AI tools in campaign management, it is essential to delve deeper into the transformative potential of Generative AI, and particularly LLMs. At its core, Generative AI is designed to create and generate content, patterns or data that was not previously in its training set, making it a powerful tool in StratCom. As shown in Figure 9, when considering the functions of the AI tools listed, Generative AI's capabilities can be harnessed to perform and even enhance many of these tasks. For instance, in Topic Modelling, not only can Generative AI identify prevalent themes but it can also predict emerging topics based on current trends. With Narrative Extraction it can generate potential narratives or storylines that might resonate with a target audience, based on the data it has been trained on. For Sentiment Analysis, beyond just gauging the emotional tone, Generative AI can predict how certain content might be received emotionally by different audience segments. When it comes to Structuring Unstructured Data, Generative AI can organise and even fill in gaps within data sets, making predictions based on patterns it identifies. In Modelling Target Audiences, Generative AI can create detailed and dynamic audience profiles, adapting them as more data becomes available. Lastly, with Content Generation, the possibilities are vast. Generative AI can craft articles, social media posts or even video content tailored to specific audiences.

Recognising the uses of generative AI in StratCom campaigns, it is imperative to integrate the IEA into our strategic toolkit to further refine and focus our communication strategies, ensuring they are both data-driven and contextually relevant. Figure 10 shows how the IEA elements are linked with a subset of

generative AI tools to ensure a comprehensive and effective approach. For instance, Baseline Analysis is closely tied with Topic Modelling, which aids in understanding the prevailing themes and topics within the IE: this ensures that communication strategies are aligned with the existing narratives. Similarly, Human Factor Analysis benefits from Narrative Extraction, allowing for a deeper dive into narratives that resonate with human emotions, beliefs and perceptions. These integrations ensure adaptive, data-driven campaign strategies.

The synergy between IEA elements and generative AI and LLMs emphasises the evolving landscape of StratCom: transitioning to AI-augmented strategies not only refines campaign precision, but also introduces innovative methodologies previously beyond reach.

IEA elements with different data sources pose challenges in terms of fusing the information together. Therefore, the cornerstone of

data-driven and AI-powered communication lies in the establishment of standardised data practices. This entails ensuring that data is collected, stored and managed in a consistent and coherent manner across all communication channels and platforms. By adopting standardised data practices, communicators can break down silos, foster collaboration, and gain a comprehensive understanding of their audience preferences and behaviours. Seamless exchange of information between diverse communication systems, eliminating data silos and enabling the aggregation of data from multiple sources provide a holistic view of the communication landscape, allowing communicators to identify trends, patterns and insights that were previously invisible. Therefore, along with AI capabilities, the data standardisation, storing and sharing must be prioritised to ensure a long-term robust data infrastructure that fuels innovation and enhances the effectiveness of communication strategies.

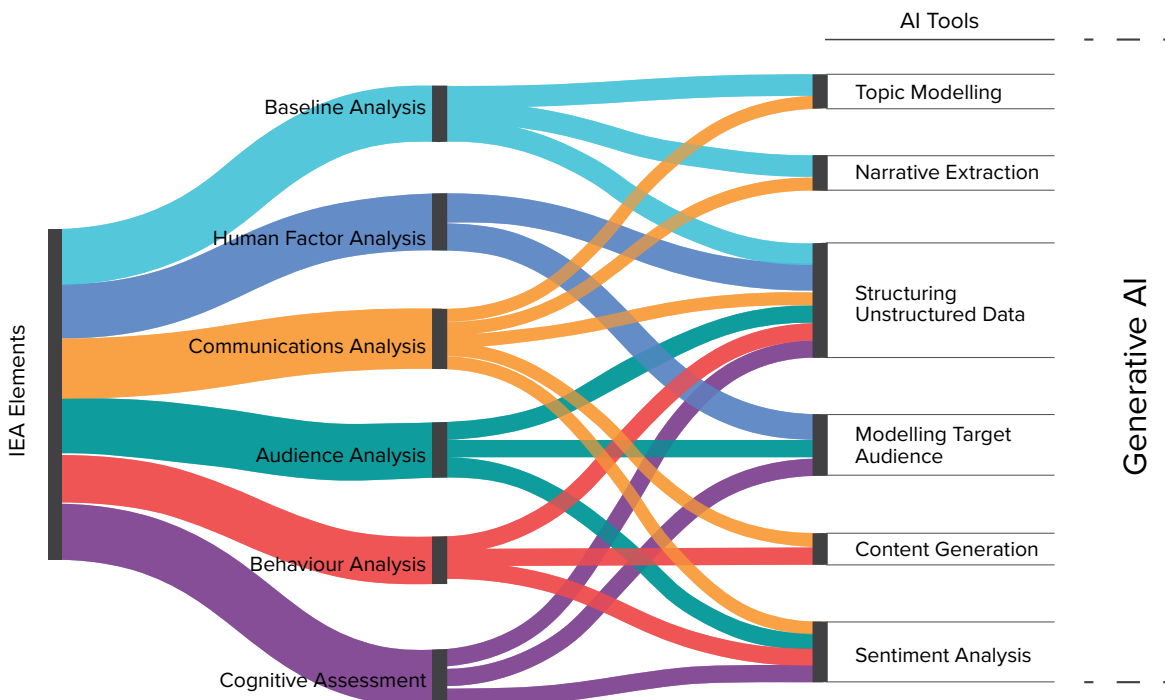


FIGURE 10. Summary of how IEA features are linked to AI tools

Steering towards an AI toolbox for targeted communication

In earlier sections, we examined IEA tools and the role and importance of LLM capabilities. This section aims to start a discussion on the evolving role of generative AI in communications. One LLM strength lies in its ability to structure unstructured data, which enables the building of software processing pipelines for extracting data insights. For this example, we can use the GPT-4 API to structure textual description into the structured JSON format which allows the target descriptions to be stored:

Structuring unstructured data is also helpful when performing human factor analysis where generating products such as PMESII/ASCOPE⁵⁰ tables is just one of many use cases where LLMs play a crucial role.

Using outputs from the hypothetically *completed*⁵¹ IEA cycle, we now delve into the design of a basic hypothetical campaign, showing in simple terms how AI can be used in shaping strategies for targets (T). Let's define a hypothetical problem:

TA⁴⁹ = "Young, Russian-speaking sports enthusiasts, 18–25 years old, likes Russian rap music, fast cars, hates NATO and LGBTQ+. Supports traditional values, favours Russia's war in Ukraine. Lives in large EU cities, in closed Russian communities. Uses VK, TikTok, Telegram."

Generate a JSON document based on the description provided. Description: {TAA}

```
{
  "name": "Young Russian Sports Enthusiasts",
  "demographics": {
    "age": {
      "min": 18,
      "max": 25
    },
    "gender": "unspecified",
    "education": "unspecified",
    "location": "large EU cities, in closed Russian communities"
  },
  "psychographics": {
    "attitudes": [
      "Likes Russian rap music",
      "Likes fast cars",
      "Hates NATO",
      "Hates LGBTQ+"
    ],
    ...
  }
}
```

The Baltic States, alarmed by the threat from the Wagner Group, a Russian private military company, have announced joint international military exercises in Latvia for autumn 2023. Involving NATO and partner countries like Sweden and Finland, the exercises will focus on countering the Wagner Group and improving defence against Russian aggression. Held in various locations in Latvia, the exercises will include activities like live-fire and urban warfare training. This significant step is aimed at enhancing interoperability with NATO allies and sending a clear message to Russia about the Baltic States' readiness to defend themselves.

The IEA's findings suggest two primary targets⁵² that should be specifically addressed prior to conducting military exercises:

T1 Young, Russian-speaking sports enthusiasts, 18–25 years old, likes Russian rap music, fast cars, hates NATO and LGBTQ+. Supports traditional values, favours Russia's war in Ukraine. Lives in large EU cities, in closed Russian communities. Uses VK, TikTok, Telegram.

T2 Middle-aged, Latvian-speaking males, 35–50 years old. Believes that vaccines are bad. Thinks NATO is behind the war in Ukraine and that LGBTQ+ is a disease. Supports only traditional values. Lives in the border areas of the Baltic States close to Belarus and Russia. Mostly gets information from Telegram, newspapers, Russian broadcast media.

Based on the results of the TAA, our aim is to investigate how rapidly evolving generative AI can support the campaign design.

Furthermore, when TAA is done on a regular basis, JSON documents of individual TAAs can be stored in the database, allowing AI agent systems⁵³ to be built that can reason over TAA data and connect structured, unstructured and semi-structured data sources to further extend the TAA.

Now let's examine the possible campaign structures with the help of LLMs (ChatGPT in this example). We start by prompting them as follows:

You are an advanced Audience Response Simulator that acts as the target audience for my campaign. I will provide you with a description of my target audience, and you will describe how my target audience might react to a planned military exercise that I will explain further. In the end answer three main questions:

1) What and where are the possible threats?

2) How should the pre-exercise deterrence campaign be designed?

3) What are the implications of demonstrating zero tolerance and force?

Then describe the COM-B model for this target audience to change their opinion on upcoming military exercises.

Target audience: {}

The generated outputs are shown in Table 4, with the most important parts highlighted in green. As expected, the outputs for T1 and T2 are generic as both share strong anti-NATO sentiments, have similar beliefs and attitudes, and rely on the same media platforms (Telegram, newspapers, Russian broadcast media, etc.).

According to GPT-4, the suggested campaign strategies for both targets would

T1	T2
----	----

Capabilities

<p>Physical: This audience is likely physically active, given their interest in sports. They might have a predisposition to valuing strength and stamina.</p> <p>Psychological: Their support for traditional values and Russia’s war in Ukraine, combined with their antipathy towards NATO and LGBTQ+, suggests strong nationalist sentiments and a conservative worldview.</p>	<p>Physical: Given the age range, they may have experienced the Soviet era and its aftermath, influencing their views and beliefs.</p> <p>Psychological: Their beliefs about vaccines, NATO, and LGBTQ+ suggest a deep-rooted mistrust in Western narratives and a preference for traditional or alternative perspectives.</p>
---	--

Opportunities

<p>Physical: Living in large EU cities means they have access to global perspectives and are exposed to diverse viewpoints. However, their closed communities might limit this exposure.</p> <p>Social: Their main social networks are VK, TikTok, and Telegram, platforms that can sometimes echo their pre-existing beliefs if they’re not exposed to a diverse range of content.</p>	<p>Physical: Living in the border areas of the Baltic states near Belarus and Russia, they might be influenced by both Baltic and Russian narratives, given their geographical location.</p> <p>Social: Their primary sources of information are Telegram, newspapers, and Russian broadcast media. This indicates a mix of modern and traditional media consumption, with a significant influence from Russian narratives.</p>
---	---

Motivation

<p>Reflective: Their interests in Russian rap music, fast cars, and traditional values give insight into their cultural and moral values.</p> <p>Automatic: Emotional ties to Russian identity, possibly driven by music, sports, and their closed communities.</p>	<p>Reflective: Their mistrust in vaccines, belief in NATO’s involvement in Ukraine, and views on LGBTQ+ reflect a strong inclination towards conspiracy theories and traditionalism.</p> <p>Automatic: Their emotional responses might be rooted in historical, cultural, and possibly personal experiences that align with their current beliefs.</p>
---	--

What and where are possible threats?

<p>Online Platforms: VK, TikTok, and Telegram are their primary sources of information. Track narratives and potential misinformation or radicalisation campaigns on these platforms.</p> <p>Closed Communities: These communities might have influential figures or groups that can sway opinions. Monitoring sentiments and identifying key influencers can help gauge potential threats.</p>	<p>Media Channels: Monitor narratives on Telegram, Russian broadcast media, and local newspapers for misinformation or antagonistic views.</p> <p>Border Areas: Given their geographical location, there might be a higher chance of direct influence from Belarus and Russia, both in terms of information and potential physical actors or agitators.</p>
--	---

How should the pre-exercise deterrence campaign be designed?

<p>Cultural Connection: Incorporate elements of Russian rap music or sports into the campaign to establish a cultural connection. This can make the message more relatable.</p> <p>Transparent Communication: Clearly communicate the goals and intentions behind the military exercises to dispel myths and reduce fear.</p> <p>Engage Influencers: Identify and engage with influencers within their communities or on their favourite platforms to spread positive or neutral narratives about the exercises.</p>	<p>Trust-building: Focus on building trust by presenting facts clearly, possibly using trusted local figures or influencers to convey the message.</p> <p>Local Media: Leverage local newspapers and possibly local Telegram channels to provide balanced views on the military exercises.</p> <p>Cultural Sensitivity: Ensure that the campaign respects and understands their attachment to traditional values. Addressing their concerns without attacking their beliefs is crucial.</p>
--	---

What are the implications of demonstrating zero tolerance and force?

<p>Backlash: Given their strong nationalist sentiments and conservative worldview, a forceful approach might be met with severe resistance and could further alienate this group.</p> <p>Strengthening of Pre-existing Beliefs: Demonstrating force can validate their negative beliefs about NATO or the EU and further solidify their support for traditional values and Russia's stance.</p> <p>Potential for Radicalisation: Extreme measures could push some individuals towards radical actions or beliefs, especially if they feel their identity or community is under threat.</p>	<p>Intensified Mistrust: Forceful actions or zero-tolerance policies can validate their existing beliefs about NATO or Western intentions, further alienating them.</p> <p>Strengthened Alternative Narratives: Using force can bolster the narratives they already believe in, making them more susceptible to misinformation and anti-Western sentiments.</p> <p>Risk of Radicalisation: Feeling threatened or cornered can push some individuals towards more radical actions or beliefs, especially if they feel their traditional values are under direct assault.</p>
--	---

TABLE 4. Generated outputs for T1 and T2

need to engage with these shared characteristics, possibly through community engagement and careful balance between showing force and tolerance. Threatening and showing force might alienate both targets and confirm their cognitive biases. The main differences between the targets lie in their age and language specifics, while they have traditional values in common, as well as potential increased exposure to Russian media influence. Overall, the model suggests an inclusive approach that allows both targets to associate their values

with it. However, it should be stressed that we have poorly informed the model for it to reason accordingly.

Following this reasoning, GPT-4 suggested campaign slogans for each target using the following prompt:



Now, as an Audience Response Simulator, for both target audiences generate Midjourney prompts for

each narrative. Midjourney is an AI image generator where each image must be explained in detail. Midjourney is not able to generate text on the images, nor logos.

Figure 11 shows the combined output with generated Midjourney prompts requested for each narrative and the respective Midjourney image generated using the GPT-4 prompt.

The results appear generic, and regardless of the model outputs it is necessary to approach current generative AI capabilities with caution due to their inherent limitations.

Still, AI opens up new horizons for tailoring content that aligns with strategic objectives. Despite challenges in terms of controlling the outputs of current generative AI models, they offer novel means to engage audiences with AI-generated customised messages. While AI outputs in this instance might not be directly applicable, for innovative idea generation informed by IEA outcomes such AI proves invaluable. If appropriately harnessed with due respect to data privacy and cybersecurity, StratCom practitioners might experiment with reaching varied demographics, always keeping core values in mind. The potential to amplify NATO's communication is vast, and while much is still to be discerned, the promise is considerable.

T1	T2
<p>Slogan: 'Strength in Unity: NATO's Promise for a Peaceful Future.'</p>	<p>Slogan: 'NATO: Protecting Homes, Preserving Traditions.'</p>
<p>Messaging narratives:</p> <p><i>Common Goals, Common Future: Just like in sports, teamwork is key. NATO exercises promote unity and collaboration to build a world where everyone can enjoy their passions in peace.</i></p> <p>GPT-4 generated Midjourney image prompt:</p> <p>A football pitch at twilight with players from different nationalities locking hands in a circle. The sunset in the background paints a picturesque scene of unity and solidarity. The focus should be on the diverse facial expressions of players showing camaraderie.</p>	<p>Messaging narratives:</p> <p><i>Guardians of Peace: From the Baltic borders to the Atlantic shores, NATO ensures that every home is safe, and every tradition is respected.</i></p> <p>GPT-4 generated Midjourney image prompt:</p> <p>A serene image of the Baltic landscape at dawn with a watchtower in the distance. In the foreground, a father and child are seen flying a kite together, while the shadows of NATO soldiers subtly line the horizon, guarding the peace.</p>
	

Bridge of Cultures: NATO values diversity. While respecting all, we focus on common goals to ensure peace and harmony across nations.

GPT-4 generated Midjourney image prompt:

An urban setting with a mural being painted on a brick wall. The mural showcases a fusion of Russian rap symbols, like a microphone, with international musical notes floating around. Young people from diverse backgrounds should be seen collaborating on the artwork.



Informed Decisions: In a world of fast information, NATO believes in the truth. We encourage everyone to seek out facts, understand different perspectives, and make informed choices about what they believe.

GPT-4 generated Midjourney image prompt:

A man is seen from the back, sitting at a rustic wooden table, intently reading a newspaper with a vintage radio beside him. On the table lies an open Telegram app on a smartphone, displaying a globe icon, signifying diverse sources of information.



Freedom of Choice: Just as everyone has the right to enjoy the music, sports, or lifestyle they love, NATO stands for every nation's right to choose its own path.

GPT-4 generated Midjourney image prompt:

A juxtaposed image of a bustling urban market street on one side and a serene countryside on the other. The urban side showcases stalls selling vinyl records and headphones, while the countryside flaunts a vast landscape of choices, signifying freedom.



Strength in Tradition: Just as traditional values have their strength and importance, NATO's enduring commitment is to protect every nation's right to its heritage and future.

GPT-4 generated Midjourney image prompt:

A traditional Latvian festival with people dancing around a midsummer bonfire. Along the fringes, there are subtle hints of NATO's influence like a peace dove flying overhead or a distant flag at half-mast, signifying respect for traditions while emphasizing unity.



FIGURE 11. Combined GPT-4 output with generated Midjourney prompts and images

LLMs can help develop varied training situations and change the way educational content is designed. Using their language reasoning skills, LLMs can assist in analysing campaign results. They can generate new or edit existing content that is relevant and tailored, simplifying military training. With LLMs trainers can make learning more interactive and personalised, leading to better training outcomes. This adaptability from LLMs improves the overall training process.⁵⁴

Current AI features, including multimodality and generative attributes, although promising, bring up complexities, especially when adopted by governmental and military organisations. Legally it is vital to examine the rules governing this technology to ensure its ethical and responsible use within NATO's scope. In the following section, we will investigate various regulations that communicators should be aware of, and we will outline the military's position in this context.

The legal perspective

†NB: At the time of drafting this report (June, 2023), the Artificial Intelligence Act (Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (Brussels, 21 April 2021, COM(2021) 206 final, 2021/0106 (COD)) is draft law, having passed the European Parliament vote, and awaiting the vote of the Council of the European Union. The legal analysis provided is based on draft law (with European Parliament amendments published on 14 June 2023) that is still subject to amendments. At any point where the term 'AI Act' is mentioned, it should be considered to include the Artificial Intelligence Act as it stands with the amendments adopted by the European Parliament on 14 June 2023.

With the rapid development of AI and its steady integration into the capabilities and functions that contribute to IEA, this section of the report will examine the legal structure governing IEA operations⁵⁵ and noteworthy legislative steps.

The most notable development of the AI regulatory framework has been led by the European Commission and resulted in the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on AI (Artificial Intelligence Act) cited above, hereinafter the **AI Act**.

One of the main contributions the AI Act makes within both the legal and technical sectors is the introduction of a framework that covers the whole process of creating and using AI systems, hereinafter **AI systems**. Comparative legal acts have not been focused on the life cycle of AI systems, but on the consequences of AI system vulnerabilities/risks, such as discrimination⁵⁶ or crashes of self-driving cars.⁵⁷ The AI Act focuses on systems themselves, requiring accountability from developers and users.

Additionally, the AI Act will be binding on all European Union member states, which comprise a majority of NATO member states.

The AI Act and the research leading up to the current version of the draft law form the basis for the legal analysis provided within this section of the report.

In the process of drafting the AI Act, three main documents were created: (a) guidelines regarding the creation of trustworthy AI systems,⁵⁸ hereinafter **Ethics Guidelines**; (b) policy recommendations;⁵⁹ and (c) White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, hereinafter **White Paper** (Figure 12).⁶⁰ The White Paper, building upon the Ethics Guidelines and policy recommendations, outlines the necessity for regulating AI systems. Emphasising the necessity to regulate systems that are involved in almost every aspect of civilians' lives,⁶¹ the White Paper discusses the need for cooperation with

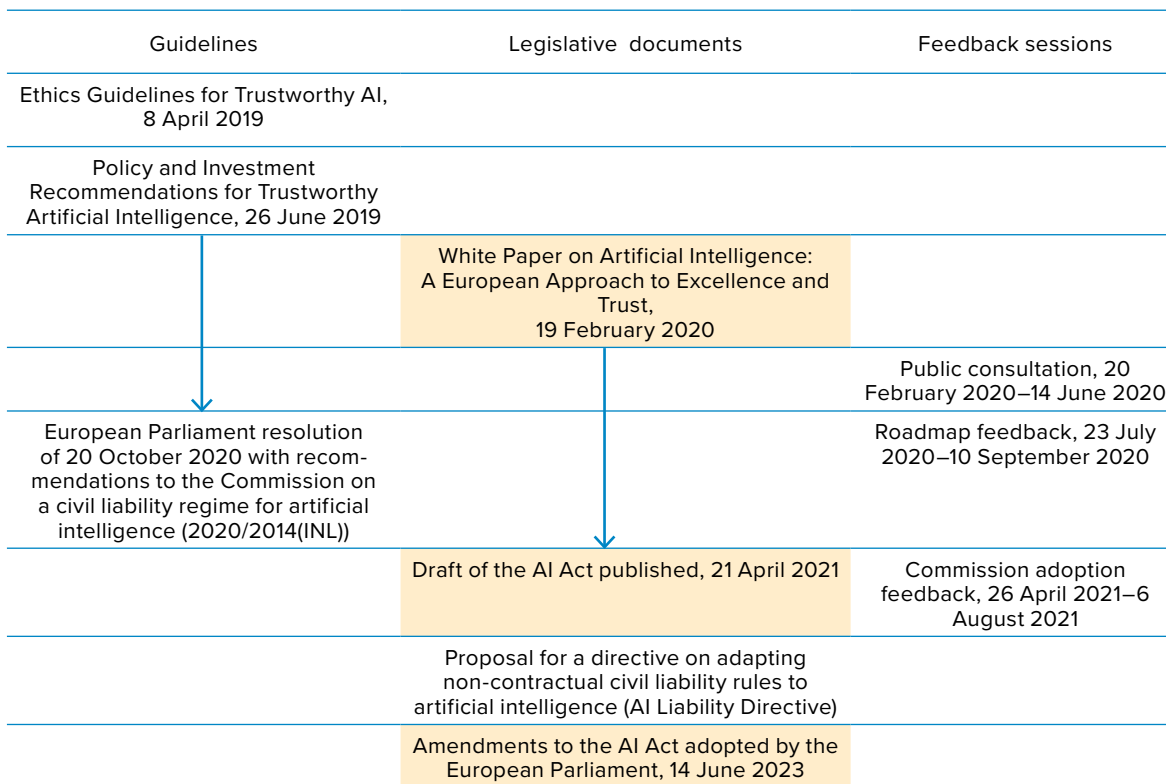


FIGURE 12. Development of the AI Act

member states of the European Union, partnership with the private sector and focus on small and medium enterprises, as well as the requirement to promote the adoption of AI by the public sector.⁶²

Significantly, the Ethics Guidelines that served as the basis for the White Paper have been the foundation for most drafts of and amendments to the AI Act: that is, while specific articles and principles of the AI Act have been amended during the writing of this

report, at no point have legislators abandoned the conclusions reached within the Ethics Guidelines, and all amendments recorded have strived to only increase the influence the Ethics Guidelines have on the final piece of the legal act.⁶³ This is important to note, as it indicates a specific approach in regulating AI systems—**even in instances when there is no set legal framework, the existing Ethics Guidelines can serve as an indicator in responsibly creating and using AI systems.**⁶⁴

The Artificial Intelligence Act

The draft AI Act is composed of a memorandum and twelve sections, discussing scope and definitions, prohibited AI practices, high-risk AI systems, transparency obligations for certain AI systems, measures in support of innovation, governance and implementation, codes of conduct and final provisions.

For the purposes of this report, the following elements of the AI Act should be taken into account:

- (a)** Article 2 Section 3 of the AI Act states that the AI Act does not apply to AI systems that are developed or used exclusively for military purposes.

The AI Act clearly states that whenever a system has been created for the usage of military necessities, the regulations laid out within the AI Act do not have to be considered. The same can be applied to systems that might have been developed for commercial use but end up being used for military purposes.

Within the European Union and its binding agreements, defence and military operations are national considerations, which are not legislated on a European Union level. Because of that, the underlying recommendation to apply the existing Ethics Guidelines persists in situations where the legal framework is lacking.

- (b)** Subsection 5.2.3 of the explanatory memorandum added to the AI Act gives an overview of Title III, High-Risk AI Systems, that is, systems that create a high risk to the health and safety or fundamental rights of natural persons.

PsyOps requires data analysis and segmentation, which, in turn, means working with the data of natural and legal persons and can affect both their safety and fundamental rights (see e.g. Articles 7 and 8 of the Charter of Fundamental Rights of the European Union⁶⁵). This is why, in order to conduct PsyOps, it is crucial to understand the principles according to which high-risk AI systems are regulated.

Apart from horizontal obligations on providers of high-risk AI systems (Chapter 3 of the AI Act), outlined conformity assessment procedures (Chapters 4 and 5 of the AI Act) and a description of internal control checks, the explanatory memorandum notes that principles of safe usage of high-risk AI systems are based on the Ethics Guidelines for Trustworthy AI.

- (c)** Amendment 633 of the AI Act (Article 69) states that codes of conduct can be drawn up for all AI systems to encourage the development and usage of trustworthy and transparent systems.

Considerations outlined within the amendments (AI literacy, the danger of advancing existing biases and discrimination, etc.) further develop considerations outlined within the Ethics Guidelines. That, in turn, gives all entities involved in the AI system life cycle the opportunity to build upon the required transparency and accountability, even for systems that have not been classified as high risk.

Ethics Guidelines for Trustworthy AI

To better understand the content of the AI Act, the main principles described within the Ethics Guidelines are outlined below.

1. Trustworthy AI throughout the system's life cycle should be lawful, ethical and robust.

While compliance with laws and regulations, as well as adherence to ethical principles and values, might be straightforward, the third component—robustness from both a technical and a social perspective—might be the more interesting component to analyse and take into account.⁶⁶

As AI learns from existing data, transparency and human oversight are important to make sure that whatever the AI system is used for does not end up infringing upon the fundamental human rights and safety of individuals. Not only does this continue to justify additional codes of conduct (see Article 69 of the AI Act), but it also highlights the necessity to innovate and the requirement to safeguard fundamental human rights.

2. In assessing trustworthy AI components, seven requirements are suggested:

- (1)** human agency and oversight,
- (2)** technical robustness and safety,
- (3)** privacy and data governance,
- (4)** transparency,
- (5)** diversity, non-discrimination and fairness,
- (6)** environmental and societal well-being, and
- (7)** accountability.⁶⁷

While some of the requirements have been widely discussed both in this report and in public discourse, there are additional points worth mentioning.

An added consideration with data governance is the quality and integrity of data, as also highlighted in the AI Act amendments. Among other aspects, the amendments discuss the importance of the quality of the data fed into systems. This not only puts an additional responsibility onto the developers and users, but also opens a wider discussion on the transparency of data used in training AI systems and models. The requirement of privacy and data governance within the Ethics Guidelines also includes the necessity of proper communication regarding said AI systems.

Another consideration, included within the requirement of diversity, is stakeholder participation. The AI Act on some level ensures stakeholder participation within the requirement of user-completed reports; however, the Ethics Guidelines introduce the idea of stakeholder participation within the development stage of an AI system life cycle. That, in turn, would ensure the creation of a system that does not advance biases or discrimination.⁶⁸

Pursuant to the explanatory memorandum, all restrictions related to high-risk AI systems are derived from the aforementioned principles and requirements. While the Ethics Guidelines outline the ideal requirements of an AI system (not all of which will be translated into law), in order to conform to good practice such principles and requirements should be taken into account.

Legal framework for military purposes

As previously mentioned, the AI Act is not required to be used when developing or using AI systems for military purposes. Below a brief overview of the legal framework relevant for functions performed for military purposes

is provided. This is not directly linked to the use of AI systems, but provides an insight into the regulatory frameworks relevant to the military sector.

North Atlantic Treaty Organization

NATO is an international organisation, established under the Washington Treaty, hereinafter the **Treaty**, which provides the foundation for its functioning. International organisations are governed on the basis of their founding documents (for NATO—the Treaty) and any treaties or agreements they sign.

The Treaty includes the following legal remarks:

- The preamble states that NATO is built upon reaffirmation of faith in the principles of the Charter of the United Nations, hereinafter the **UN Charter**, and is founded on the principle of the rule of law. Article 1 states that any international dispute must be settled in such a manner that justice is not endangered and references the UN Charter again.
- Article 5 references the UN Charter and its Article 51 (the right to self-defence).

- Article 7 states that the Treaty shall not be considered as contradictory to the UN Charter.
- Article 12 invites the future reassessment of the Treaty, considering other regional arrangements and the UN Charter.

The Treaty does not include an obligation to govern NATO within the framework of any one specific legal document, apart from the Treaty itself, excluding specific references to the principles of the UN Charter. Therefore, NATO, as an international organisation, is not required to abide by any pieces of legislation mentioned in this report. That, however, is not true for NATO member states when acting as autonomous states. Autonomous states (including all NATO member states) must abide by any piece of legislation they sign and ratify.

This means that the AI Act, when enforced, will be binding on separate member states of the European Union, but will not be binding on NATO.

The European Union

For European Union member states' military considerations, insofar as EU law goes, Common Foreign and Security Policy is regulated under Title V of the Treaty on European Union.

The main requirements for EU member states' military considerations as outlined in this document are:

- Cooperation and solidarity: the member states are expected to work together to enhance and develop their mutual political solidarity (Article 24). They are also

expected to engage in systematic cooperation in the conduct of the common foreign and security policy (Article 25).

- Compliance with Union policy: the member states must refrain from any action contrary to the interests of the Union or likely to impair its effectiveness in international relations (Article 24). In addition, they are required to uphold the Union's positions in international organisations and conferences (Article 34).
- Coordination in international settings: member states are required to coordinate their actions in international organisations and conferences (Article 34). Those states that are members of the United Nations Security Council must keep other member states and the High Representative of the Union for Foreign Affairs and Security Policy informed and defend the positions and interests of the Union (Article 34).
- Implementation of Union decisions: diplomatic and consular missions of the

member states and the Union delegations must cooperate to ensure decisions defining Union positions and actions are complied with and implemented (Article 35).

- Respect for data privacy: when processing personal data as part of activities under this chapter, member states are required to follow rules laid down by the Council for the protection of individuals' data and the free movement of such data (Article 39).
- Upholding treaty provisions: The implementation of the common foreign and security policy should not affect the application of the procedures and the extent of the powers of the institutions as laid down by the Treaties for the exercise of the Union competences (Article 40).

The above indicates that European Union member states have considerable autonomy in this area.

Conclusions and recommendations

In this report we have discussed data-driven techniques from the IEA, shedding light on the pros and cons of AI-enabled data processing pipelines. Key insights from the IEA can aid information campaigns, as explained below.

Generative AI has recently emerged as a significant transformative force, enhancing the capabilities of communicators, programmers, data analysts and a host of other professionals. Despite this progressive trajectory, it must be acknowledged that LLMs exhibit inherent limitations in semantic grounding, a shortcoming underscored by the occurrence of hallucinations and inconsistencies in AI-generated content. Such fluctuations in reliability are not uniform but vary according to the specific characteristics of each model and are further

compounded by the linguistic nuances and origin of the underlying data. It is therefore imperative that users and developers of generative AI systems maintain a critical awareness of these constraints as they harness the considerable powers of these advanced computational tools. The recent introduction of GPT-4-vision models, alongside a growing list of open-source alternatives, provides analysts with new capabilities such as understanding political memes or proactively filtering graphic content. These advanced multimodal capabilities facilitate a more comprehensive analysis by interpreting and synthesising information across various forms of data, including textual and visual inputs.

Moving beyond simple summarisation is vital, aiming to leverage the advanced reasoning abilities of LLMs. Detailed analysis of identified hostile narratives and assessment of message intent and their relevance to predetermined topics represent initial steps in this direction. The inherent reasoning capabilities of LLMs can be instrumental in verifying facts (when a fact-checking database is available), provided the information source is deemed credible.

An emergent concern is the potential misuse of LLMs as instruments of disinformation by adversaries. The combination of creative content generation with high-speed content delivery presents both opportunities and challenges for defenders. While it is vital to adopt innovative techniques to identify and counteract hostile information activities, this very technology can be employed to enhance owned communication efforts. With adversaries likely adopting such technological advances for malign activities, it becomes imperative to harness tools like narrative extraction, clustering and summarisation. AI-powered tools facilitate the identification of dominant narratives on social networks and allow the analysis to move beyond traditional sentiment analysis.⁶⁹ Narrative tracking offers insights into the evolution and traction of both our own campaigns as well as adversarial narrative evolution. Also, advances in multimodal generative AI models, namely text-to-image applications such as Stable Diffusion or DALL-E, have increased the magnitude of challenges related to identifying and analysing disinformation, as well as providing tools to assess complex data types such as political memes. Both ChatGPT and Google Bard can describe complex images, but both tools are subject to numerous restrictions before they can be fully integrated into analysts' data processing pipelines. However, the technology is maturing fast, and it is expected that the first multimodal social media analytics tools will be on the market soon.

Through the incorporation of AI, specifically LLMs, the domain of network analysis has been significantly enhanced. By harnessing the sophisticated reasoning afforded by prompt engineering and the refinement provided by fine-tuning LLMs, these models are adept

at discerning content patterns, including those with nuanced sentiment stances. This advanced capability is instrumental in the identification of adversarial narratives and the pinpointing of social media accounts propagating pro-Kremlin or Chinese media narratives. The utilisation of AI in examining network structures sheds light on the pathways of information dissemination, revealing covert campaigns and affiliations. The insights garnered from AI-driven network analysis are invaluable; they inform the development of strategic communication initiatives designed to counter disinformation and advocate for transparency.

From the viewpoint of communication campaign planners, generative AI (in particular, in the context of LLMs) can emulate potential targets within a controlled environment, facilitating the execution of wargaming scenarios prior to the actual campaign. This sets the stage for forward-looking experiments, leading to a refined AI communication toolkit designed for precise micro-targeting. Campaigners can benefit from generative AI that helps design visual content based on TAA outputs and therefore offer various text prompts, resulting in artistic and, depending on the prompt, semi-realistic images tailored to specific target descriptions. Additionally, the newly announced custom assistant creation by OpenAI⁷⁰ enables the design of chatbots tuned to a specific communication style. NATO StratCom Centre of Excellence continues to investigate risks and opportunities enabled by newly introduced capabilities in the context of targeted communication.

AI starts with data. The successful implementation of creative cutting-edge AI solutions hinges on the adoption of standardised data storage and sharing practices across governments and organisations. Diverse modalities, file types, access restrictions, sharing capabilities and user habits all significantly influence the future of AI implementation. Moreover, the long-term integration and adoption of AI in large-scale government and institutional systems are directly tied to the underlying data infrastructure. This holds true for military applications, communication systems and all other domains.

The legal considerations

First, it is vital to recognise the most important lesson evident when tracking the legislation development process—most of the drafted regulations rely upon ethics guidelines. That has to be a consideration when applying AI to any necessary discipline—while the specific legal requirements are still to be enforced, many pre-emptive regulations and considerations can be put in place by taking into account ethics and the components of trustworthy AI systems derived henceforth. The better the understanding of legal requirements applicable to AI, the better the chance of transparent and responsible use of different systems.

Second, when using AI systems attention must be paid to the different legal systems at play. Regarding the usage of AI systems for civilian purposes (which includes, but is not limited to, political campaigns, marketing etc.), the most notable legal development is the AI Act (which is binding on EU member states). The AI Act, in turn, specifies that it is not meant for military purposes; therefore, in respect of AI usage in the military, national military institutions must abide by the relevant domestic laws. NATO, as per the Washington Treaty, is not obliged to consider any of the legislation.

Third, with regard to the development of AI regulation in general, it is essential to look at the whole life cycle of AI systems (including developers, users and stakeholders). Considering only isolated risks—or even worse, proposing regulations based on instances where a system has failed—will delay any innovation or successful implementation of AI.

Fourth, all of this points to one main consideration—while the continuum of competition reaches across different stages, some of which are more challenging to clearly differentiate, legal frameworks clearly indicate the rules applicable to specific stages of conflict. Therefore, one must exercise caution when moving between the concepts of ‘civilian’ and ‘military’ operations, and, where those concepts change, when moving between cooperation, rivalry and confrontation.

Endnotes

- 1 AJP-10: Allied Joint Doctrine for Strategic Communications.
- 2 P. Kotler and N. Lee (2008). *Social Marketing: Influencing Behaviors for Good*. SAGE.
- 3 European Parliament (2023). *Artificial Intelligence Act: Deal on Comprehensive Rules for Trustworthy AI*, 9 December.
- 4 M.D. Slater (1995). 'Choosing Audience Segmentation Strategies', in E.W. Maibach and R.L. Parrott (eds), *Designing Health Messages: Approaches from Communication Theory and Public Health Practice*. SAGE.
- 5 S.E. Boslaugh, M.W. Kreuter, R.A. Nicholson and K. Naleid (2005). 'Comparing Demographic, Health Status and Psychosocial Strategies of Audience Segmentation to Promote Physical Activity', *Health Education Research* 20(4), 430–38.
- 6 That is, a fundamental individual difference variable that reflects individuals' beliefs about the degree of control they have over events in their lives.
- 7 M. Rosenberg, N. Confessore and C. Cadwalladr (2018). '*How Trump Consultants Exploited the Facebook Data of Millions*', *New York Times*, 17 March.
- 8 S.W. O'Leary-Kelly and R.J. Vokurka (1998). 'The Empirical Assessment of Construct Validity', *Journal of Operations Management* 16(4), 387–405.
- 9 Ibid.
- 10 M.E. Strauss and G.T. Smith (2009). 'Construct Validity: Advances in Theory and Methodology', *Annual Review of Clinical Psychology* 5, 1–25.
- 11 M. Kosinski, D. Stillwell and T. Graepel (2013). 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior', *Proceedings of the National Academy of Sciences* 110(15), 5802–05.
- 12 OMR (2017). '*Alexander Nix: From Mad Men to Math Men*', OMR Festival 2017, Hamburg, Germany, #OMR17, YouTube video.
- 13 T. Gilovich, D. Keltner, S. Chen and R. Nisbett (2013). *Social Psychology*, 3rd edn. New York, NY: W.W. Norton.
- 14 Z. Berry and J. Frederickson (2015). 'Explanations and Implications of the Fundamental Attribution Error: A Review and Proposal', *Journal of Integrated Social Sciences* 5(1), 44–57.
- 15 See S. Epstein and L. Teraspulsy (1986). 'Perception of Cross-situational Consistency', *Journal of Personality and Social Psychology* 50(6), 1152–60.
- 16 J.P. Peter (1979). 'Reliability: A Review of Psychometric Basics and Recent Marketing Practices', *Journal of Marketing Research* 16(1), 6–17.
- 17 I. Ajzen (1991). 'The Theory of Planned Behavior', *Organizational Behavior and Human Decision Processes* 50(2), 179–211.
- 18 A. Bandura (2009). 'Social Cognitive Theory of Mass Communication'. *Media Effects*. Routledge, 110–40.
- 19 S. Michie, M.M. Van Stralen and R. West (2011). 'The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions', *Implementation Science* 6(1), 1–12.
- 20 Y. Duan, J.S. Edwards and Y.K. Dwivedi (2019). 'Artificial Intelligence for Decision Making in the Era of Big Data—Evolution, Challenges and Research Agenda', *International Journal of Information Management* 48, 63–71.
- 21 B. Liu (2012). 'Sentiment Analysis and Opinion Mining', *Synthesis Lectures on Human Language Technologies* 5(1), 1–167.
- 22 K. Sporrel et al. (2021). 'The Design and Development of a Personalized Leisure Time Physical Activity Application Based on Behavior Change Theories, End-User Perceptions, and Principles from Empirical Data Mining', *Frontiers in Public Health* 8, 528472.
- 23 E.W. Maibach, A. Leiserowitz, C. Roser-Renouf and C.K. Mertz (2011). 'Identifying Like-Minded Audiences for Global Warming

- Public Engagement Campaigns: An Audience Segmentation Analysis and Tool Development', *PLOS One* 6(3), e17571.
- 24 B. Boudreau (2016). 'We Have Met the Enemy and He Is Us'—An Analysis of NATO Strategic Communications: The International Security Assistance Force (ISAF) in Afghanistan, 2003–2014. NATO Strategic Communications Centre of Excellence.
 - 25 Michie et al. (2011).
 - 26 M.H. Huang and R.T. Rust (2021). 'A Strategic Framework for Artificial Intelligence in Marketing', *Journal of the Academy of Marketing Science* 49, 30–50.
 - 27 K. Klonowska and F. Bekkers (2021). 'Maneuvering in a Behavior-Oriented Approach', in *Behavior-Oriented Operations in the Military Context: Enhancing Capabilities to Understand and Anticipate Human Behavior*. Hague Centre for Strategic Studies, 57–75.
 - 28 D. Chaffey and F. Ellis-Chadwick (2019). *Digital Marketing*. Pearson UK.
 - 29 Arvind Narayanan (2023). *Understanding Social Media Recommendation Algorithms*. Knight Columbia Knight First Amendment Institute, 9 March. Constanza M. Vidal Bustamante et al. (2022). *Technology Primer on Social Media Recommendation Algorithms*. Belfer Center for Science and International Affairs, 25 August.
 - 30 Github, *Twitter Source Code Recap*. Ryan Mac and Kate Conger (2023). 'Twitter Says Parts of Its Source Code Were Leaked Online', *New York Times*, 26 March [Accessed 3 April 2023].
 - 31 The GDEL Project, *LLM Infinite Loops In LLM Entity Extraction: When Temperature & Basic Prompt Engineering Can't Fix Things; LLM Entity Extraction vs LLM Embeddings: Discontinuous Failure Points vs Smooth Responsiveness; LLM Infinite Loops & Failure Modes: The Current State of LLM Entity Extraction*, all 9 August 2023.
 - 32 R. Merritt (2022). 'What Is a Transformer Model?', *NVIDIA*, 25 March.
 - 33 *Google Cloud* project page.
 - 34 'Krieviju sasniedzis "dolāru un eiro" bads', TVNET | FINANCENET, 10 August 2023.
 - 35 A. Peng et al. (2023). 'GPT-3.5 Turbo, Fine-Tuning, and API Updates', *OpenAI*, 22 August.
 - 36 OpenAI (2023). 'OpenAI Partners with Scale to Provide Support for Enterprises' Fine-Tuning Models', *OpenAI*, 24 August.
 - 37 <https://ai.meta.com/llama/>.
 - 38 <https://bigscience.huggingface.co/blog/bloom>.
 - 39 Within the military context, AJP-10 defines narrative as 'a spoken or written account of events and information arranged in a logical sequence to influence the behaviour of a target audience'.
 - 40 For more details on how Twitter (X) can be viewed in terms of narratives and its usefulness, see Neil Sadler (2018). 'Narrative and Interpretation on Twitter: Reading Tweets by Telling Stories', *New Media & Society* 20(9), 3266–82.
 - 41 Open-source pretrained transformer models such as sentence transformers (sbert.net), OpenAI and Google Cloud NLP text embeddings, and Cohere offer diverse functionalities for natural language processing. Sentence transformers excel in creating sentence-level embeddings, enhancing semantic similarity and text classification tasks. OpenAI and Google Cloud NLP are effective for a range of applications, including sentiment analysis and language translation. Cohere and similar platforms provide advanced language understanding and generation capabilities. Each model has unique strengths, suitable for different NLP and information analysis tasks.
 - 42 Extracted but not yet post-processed short summaries or descriptions of larger text chunks, images etc. Usually they are direct, unedited accounts of events, identified using machine learning (clustering, NLP), without external interpretations.
 - 43 The data context refers to the quantity of information that an AI model can assimilate and process simultaneously. It is usually measured in 'tokens' that represent word-level text chunks.
 - 44 A. Bryman (2016). *Social Research Methods*. Oxford University Press.
 - 45 N.K. Denzin and Y.S. Lincoln (eds). (2011). *The SAGE Handbook of Qualitative Research*. SAGE.
 - 46 R.A. Krueger (2014). *Focus Groups: A Practical Guide for Applied Research*. SAGE.

- 47 Bernard J. Jansen, Soon-gyo Jung and Joni Salminen (2023). 'Employing Large Language Models in Survey Research', *Natural Language Processing Journal* 4, 100020.
- 48 Behavioural Dynamic Institute Methodology Team (2015). *BDI Methodology for Strategic Communications Campaigns*. BDI Institute.
- 49 TA: target audience. According to AJP-10, TA should be simply Targets (T) but for better model reasoning as we do not explain the model what simply "Targets" mean, we have stuck to TA.
- 50 *PMESII/ASCOPE tables* used in human factor analysis.
- 51 For the simplicity of this report, we assume we have the outputs from the IEA.
- 52 All mentioned and described targets in this report are completely hypothetical.
- 53 Llama Index AI: <https://www.llamaindex.ai>.
- 54 K. Manson (2023). 'The US Military Is Taking Generative AI out for a Spin', *Bloomberg Newsletter Cybersecurity*, 5 July.
- 55 AJP-10, 54, subsection d(1), Audience analysis. The PsyOps staff provide audience research and TAA for the IEA. They are an integral part of the Info Ops IEA function, but it is the decision of the director of communications whether they are teamed with or subordinated to that function.
- 56 The White House, Office of Science and Technology Policy, *Blueprint for an AI Bill of Rights*.
- 57 Federal Automated Vehicles Policy (USDOT). Occupant Protection for Vehicles with Automated Driving Systems (NHTSA). The Automated and Electric Vehicles Act 2018.
- 58 *Ethics Guidelines for Trustworthy AI*, 8 April 2019.
- 59 Policy and Investment Recommendations for Trustworthy Artificial Intelligence, 26 June 2019.
- 60 European Commission, White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, COM(2020) 65 Final, Brussels, 19 February 2020.
- 61 *Ibid.*, 3–4.
- 62 European Parliament, Artificial Intelligence Act, P9_TA(2023)0236, 14 June 2023. Texts adopted by the European Parliament on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))1, Ordinary legislative procedure—First reading.
- 63 White Paper, 13.
- 64 Legal developments have taken place outside the European Union as well, with multiple common law countries setting court precedent on the use of AI systems and copyright (*Commissioner of Patents v. Thaler*, Federal Court of Australia, Full Court, 13 April 2022, 289 FCR 45; *Nduka v. International Patent Office*, Chancery Division, 11 February 2016, [2016] 2 WLUK 300) and adopting similar legislation to the *General Data Protection Regulation*, which includes discussion on the use of AI and data privacy. There are also two notable pieces of legislation in the United States, namely, *the Electric Vehicles Act 2018* and the Algorithmic Accountability Act 2022 (not yet in force at the time of drafting this report), which respectively address liability that has risen from the use of AI and the danger of discrimination through the means of AI systems.
- 65 Charter of Fundamental Rights of the European Union, 2000/C 364/01, 18 December 2000, *Official Journal of the European Communities* C 364/1, Articles 7–8.
- 66 *Ethics Guidelines for Trustworthy AI*, 2.
- 67 *Ibid.*, 2.
- 68 *Ibid.*, 13.
- 69 Sentiment analysis that is able to estimate only general sentiment categories: 'positive', 'negative' and 'neutral'.
- 70 OpenAI, *GPTs vs Assistants* [Accessed 6 December 2023].

