Research on Exploiting Cognitive Hacks (REACH)

Marco Marsili

Cà Foscari University of Venice

https://www.marcomarsili.it - info@marcomarsili.it

Ca' Foscari University of Venice

Department of Philosophy and Cultural Heritage

This poster was supported by the Ministry of University and Research (MUR), Italy, through the Young Researchers-Seal of Excellence (SOE) grant funded by NextGenerationEU (NGEU) under the National Recovery and Resilience Plan (NRRP)

Introduction

Are you afraid of the Metaverse, its enabling technologies, and their misleading use in the information environment? Maybe you should be. Today's world is characterized by global interconnectivity. Digital innovations are increasingly touching all aspects of life and have a profound impact on defense, security and civilian society and disrupt or overturn traditional methods and practices, and may revolutionize governmental structures, economies, and international security (Kosal & Regnault 2020).

Information is "the foundation of all human interaction". The intersection of the information, physical and cognitive/social domains, empowered by the digital ecosystem–Internet, social media, and communication applications–creates the conditions for cognitive hacking, a cyberattack that seeks to manipulate the perception of people by exploiting their psychological vulnerabilities and is considered a threat from disinformation.

Nothing new in its single components, the novelty in cognitive attacks is the speed and power of dissemination of beliefs–false or true–instilled deeply in the consciousness of targets. One of the main challenges in the digital age is the dissemination of false information with the consequences of influencing public opinions, affecting political decisions, and even the psychological well-being of individuals (Guess et al. 2020).

In the near future, artificial intelligence-based technology will take an increasing role in the digital information environment, where the speed of machine-driven decision-making process leaves little or no time for the human to intervene–to maintain any meaningful oversight, not least control–and therefore trigger concerns and pose serious challenges associated with information accountability and assessment (discerning between intentional mal/mis/disinformation and valid counterhypotheses/arguments/evidence). AI-assisted operations are expected to have an immense impact in the information environment with effects and influence in virtual, physical, and cognitive dimensions.

Emotions play a significant role in how people interpret and react to online information (Pessoa 2009), influencing their susceptibility to manipulation (Diemer et al. 2015). Previous studies have highlighted the role of emotions, anger, and personality traits in influencing susceptibility to fake news and decision-making (Lee & Nass 2003). Additionally, cognitive biases, personality tendencies, and individual decision-making processes can shape people's propensity to believe and spread unverified news contents (Kahneman & Tversky 1979).

Therefore, as human cognition is highly susceptible to manipulation and deception, a cognitive strategy aims to influence thinking processes, such as perceptions, decision making and behavior. Cognitive attacks affect perceptions, beliefs, interests, aims, decisions, and behavior by deliberately targeting the human mind. This weaponized use of information serves to build and reinforce biased or false narratives to alter the perception and the behavior of individuals and finally of society by undermining social cohesion (Bovet & Makse 2019). Indeed, cognitive operations target influential individuals, specific groups, and large numbers of citizens selectively and serially in society, with the potential to fracture and fragment an entire society or disrupt alliances (Kosal & Regnault 2020).

The human-machine interaction, accelerated and expanded by technologies with a tempo and scale previously unimaginable, is a fundamental component of cognitive operations, and plays a central and crucial role due to the way our perception and judgment are affected, thus making it an unprecedented challenge to contemporary society (Allcott & Gentzkow 2017).

Cognitive activities are a component of modern warfare and do not necessarily carry a kinetic component or directly tangible outcomes, such as territorial or resource acquisition–as with other hybrid threats (Marsili 2023). These activities vary greatly and may encompass supporting or conflicting cultural or personalized components–social psychology, game theory, and ethics are all contributing factors.

While fast technological change makes the future of warfare uncertain and unpredictable, the Metaverse, with its growing popularity and immersive nature, provides a unique context for exposure to this distorted information, and seems to be the "natural" environment to conduct information and cognitive attacks. The term "Metaverse" was coined in 1992 by visionary author Neal Stephenson in his dystopian sci-fi thriller *Snow Crash*, which predicted the Metaverse as a convergence between the real and the virtual world; a universe beyond the physical where physical reality is merged and interacts with digital virtuality facilitated by emerging and disruptive technologies (EDTs).

Metaverse is the next disruptive technology, a transformative or revolutionary technology that, also because of its dual use nature, is poised to have a significant effect (positive and negative) on societies and decision-makers over the next 20 years. Still in the early stage of its development, the Metaverse is expected to be mature by 2030, in an order of magnitude beyond what is available today with an estimated 1 billion users by then (McKinsley & Co. 2022; Tucci & Needle 2023; Crawford & Aulanier 2023), with a huge impact on society in the coming decades. For the Metaverse may revolutionize aspects of our societies, the implications of culture, concepts, risk-tolerance, organizational structure, policies, treaties, human capital, morals and ethics must be fully appreciated.

Objective

The overall objective of the project aims to investigate the impact of cognitive hacks on individuals and society–especially in relevant groups and sub-groups of key-actors–in the context of the Metaverse, and to provide recommendations to mitigate it. This way, the research will contribute to promote a resilient society and reinforce social cohesion.

The research will scrutinize if, how and why the targets of cognitive hacks: 1) (inter)act in the Metaverse; 2) process, store, and respond to biased information or narratives (true or false) spread in the Metaverse; 3) are likely to change their mind, perceptions, opinions, beliefs, interests, aims, decisions, and behaviors 4) the durations of the effects of such actions.

The study will analyze the structure of relations between the comprehensive set of values and the range of behaviors, and aims to unveil the motivational underpinnings of value-behavior relation. Values are important for understanding various social-psychological phenomena (Schwartz & Bardi 2001).

Methodology

A qualitative oriented mix of content-analytical methods fits this project: explorative or generative (inductive category development formulating new categories out of the material); descriptive or explanatory (working through the texts registering occurrences/frequencies); relational (cross-tabulation of categories with person variables); causal (content-analytical variable). Qualitative research focuses on capturing subjective insights and aims to understand the underlying reasons, motivations, and behaviors of individuals. Quantitative research, on the other hand, involves collecting and analyzing numerical data to identify patterns, trends, and significance. It aims to quantify user behaviors, preferences, and attitudes, allowing for generalizations and statistical insights.

Numerical data derived from qualitative content analysis (QCA), a research tool used to determine the presence of certain words, themes, or concepts within some given qualitative data, will give them meaning. The QCA will be used as a mixed methods approach: assignment of categories to text as qualitative step, working through many text passages, and analysis of frequencies of categories as quantitative step. The qualitative data will be categorized quantitatively and subjected to statistical analysis. Researchers will then make inferences about the messages within the texts, the writer(s), the audience, and even the culture and time of surrounding the text. Conceptual analysis will determine the existence and frequency of concepts in a text, while relational analysis will develop the conceptual analysis further by examining the relationships among concepts in a text. Semantic analysis, will support the understanding of natural language (text) by extracting insightful information such as context, emotions, and sentiments from unstructured data–semantics is about universally coded meaning, and pragmatics, the meaning encoded in words that is then interpreted by an audience (Goddard 2013).

Likert scale, one of the most reliable ways to measure opinions, perceptions, and behaviors, will be used to measure attitudes and opinions with a great degree of nuance. This method will uncover degrees of opinion that could make a real difference in understanding the feedback of the survey.

The research will be conducted applying the concept of triangulation, which combines different methods–exploratory, descriptive, and analytical–and fits to qualitative studies, and will conduct multivariate statistical analyses, including analysis of variance (ANOVA) and regression analysis, to examine the relationships between the psychological variables examined and the exposure to false contents (Tabachnick & Fidell 2013).

Impact

This ambitious project will disclose the impact of cognitive hacks conducted through the Metaverse on European targets, i.e., individuals–the broad audience, political and military leaders, policy and decision-makers, media, and communication professionals–and society, and how international relations and may be affected by such actions and will contribute to strengthening the European identity and promote a resilient society.

Bibliography

1. Kosal M.E. & Regnault H. (2020). Introduction, in: Kosal M., ed., Disruptive and Game Changing Technologies in Modern Warfare. Advanced Sciences and Technologies for Security Applications (Cham: Springer), pp. 1-11. DOI: https://doi.org/10.1007/978-3-030-28342-1_1.
2. Guess, A., Nyhan, B. & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour,* 4(5), pp. 472-480.
3. Pessoa, L. (2009). How do emotion and motivation direct executive control?. *Trends in cognitive sciences*, 13(4), pp. 160-166.
4. Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y. & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, 6, pp. 1-9. DOI: 10.3389/fpsyg.2015.00026.
5. Lee, K. M. & Nass, C. (2003). Experimental tests of normative group influence and representation effects in computer-mediated communication: Evidence for the social identity model of deindividuation effects. *Communication Research*, 30(1), pp. 36-52.
6. Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), pp. 263-291. Reprinted in: L.C. MacLean & W.T Ziemba, eds. (2023). *Handbook of the Fundamentals of Financial Decision Making*, World Scientific Handbook in Financial Economics Series: Vol. 4. Singapore: World Scientific Publishing, pp. 99-127. DOI: https://doi.org/10.1142/9789814417358_0006.
7. Bovet, A. & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), pp. 1-10.
8. Allcott, H. & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), pp. 211-236.
9. Marsili, M. (2023), Guerre à la Carte: Cyber, Information, Cognitive Warfare and the Metaverse, Applied Cybersecurity & Internet Governance (ACIG), 2(1), pp. 1-11. DOI: 10.60097/ACIG/162861.
10. Schwartz, S. H., & Bardi, A. (2001). Value hierarchies across cultures: Taking a similarities perspective. Journal of Cross Cultural Psychology, 32, pp. 268-290.
11. Goddard, C. (2013). *Semantic Analysis: An Introduction.* 2nd ed. New York: Oxford Univ. Press.
12. Tabachnick, B.G. & Fidell, L.S. (2013). *Using Multivariate Statistics.* 6th ed. Boston, MA: Pearson.

DOI: https://doi.org/10.5281/zenodo.11507818