

SPECIFIC FEATURES OF PSYCHOLOGICAL WARFARE ON FACEBOOK IN ROMANIA

Narrative Vectors, Manipulative Tactics, and Psychological Mechanisms Analyzed through Chaos Theory, Chinese Water Drop Torture, and the Butterfly Effect

The thesis examines the concrete ways in which psychological warfare can be observed and analyzed in posts circulating on Facebook in Romania. Manipulation and disinformation in the online environment are already well documented, so I sought to understand in detail how a manipulative post is constructed: what narrative it conveys, what techniques are used to set it in motion, and what psychological vulnerabilities it attempts to exploit.

This led to the three-dimensional $V \times T \times RP$ analytical framework. Narrative vectors (V) show what is being said, manipulative tactics (T) show how the message is delivered, and psychological mechanisms (RP) examine why it may work on the recipient. In this way, I sought to break each post down into its components and analyze, at the same time, its theme, construction, and targeted psychological mechanism.

The topic is also important because of Romania's particular situation. Romania has the longest border with Ukraine among NATO and EU member states and is located next to a hybrid conflict in which information and psychological operations are ongoing. The common language shared with the Republic of Moldova allows materials to circulate very easily between the two spaces, while these geopolitical factors overlap with internal vulnerabilities, including distrust in institutions, reliance on social media for information, low levels of media literacy, functional illiteracy, and social grievances and frustrations that can easily be incorporated into manipulative constructions.

I chose Facebook as the research platform because of its large number of users, its technical and social characteristics, and its electoral relevance in Romania. The platform facilitates community formation, the rapid distribution of content, and the emergence of echo chambers, while the algorithm amplifies messages that provoke strong reactions. In addition, Facebook is used predominantly by mature users, who are more active voters and have a greater capacity to influence the political opinions of those around them. For these reasons, I consider Facebook to remain the most important social media platform for socio-political and electoral analysis in Romania.

The theoretical section examines psychological and information warfare, propaganda, disinformation, and established mechanisms of influence. Theories such as Agenda-Setting, Framing, the Spiral of Silence, and Gatekeeping are useful for explaining processes that existed before social media, but in today's digital environment these processes are strongly influenced by algorithms and by the very high speed at which information circulates. However, I did not build the research around these classical theories. Instead, I used three conceptual and metaphorical reference points borrowed from other fields (torture/psychology, mathematics/physics, and meteorology) to analyze the data: Chinese Water Drop Torture, Chaos Theory, and the Butterfly Effect.

Chinese Water Drop Torture describes the mechanism of repetition. A single post may have little effect, but repeating the same ideas across multiple sources and over a longer period can produce familiarity and sedimentation. Chaos Theory is used to analyze the unpredictability of a post's evolution within a complex system shaped by the algorithm, the timing of publication, the audience's state, and many other factors. The Butterfly Effect examines the differences between posts that remain at an ordinary level of engagement and those that become viral.

The research was built around four questions. Q1 seeks to identify the dominant narrative vectors, tactics, and psychological mechanisms. Q2 examines whether they form recurring combinations and profiles. Q3 follows the evolution of narratives over time and their relationship with events during the period analyzed. Q4 examines the extent to which the three metaphors can explain the empirical results. In relation to these questions, I formulated three hypotheses. H1 states that, if the Chinese Water Drop Torture mechanism is at work, there should be narrative vectors that recur across multiple sources and throughout the entire period. H2 states that a post's complexity score cannot predict engagement well if virality depends to a large extent on the external environment. H3 states that viral posts should have a $V \times T \times RP$ construction that differs from that of posts that do not become viral.

Research Methodology

The corpus consists of 843 Facebook posts collected manually between January 1 and February 28, 2026. Of these, 788 come from 16 fixed sources monitored systematically, while 55 were identified in the feed of the account created and used for the research. The period covers 59 days, divided into nine weeks.

Each post was analyzed using a coding framework of 148 categories, consisting of 40 narrative vectors, 34 manipulative tactics, and 74 psychological mechanisms. Applying the framework to the entire corpus produced 17,707 individual codes: 2,612 vectors, 5,978 tactics, and 9,117 psychological mechanisms, or an average of almost 21 codes per post.

In addition to the main coding, I determined for each post the dominant targeted emotion, the complexity score, the engagement rate, its classification as a drop or a butterfly, the presence of interventions carried out using artificial intelligence, its connection to events during the period, and cross-source redundancy.

The complexity score, denoted SC, is the sum of the number of vectors, tactics, and psychological mechanisms identified in a post. I do not consider it a universal measure of manipulation, but rather an indicator constructed for the comparisons made in this research.

The coding was carried out through a partially innovative hybrid method based on the triangulation and convergence of three different AI systems—Claude, ChatGPT, and Gemini—in their highest-performance versions available at the time, followed by human moderation and adjudication. Three different prompts were used for each post, one for each V, T, and RP dimension, and each prompt was run separately with the three AI assistants. This resulted in nine AI analysis/coding sessions, with human moderation, for each post.

I developed this method because relying on a single AI system created too high a risk that model-specific biases would influence the result. In the coding process, differences in sensitivity and

perspective emerged between the systems: Gemini tended to code more broadly, Claude was more restrictive, and ChatGPT generally fell between the two. The differences were mainly differences of threshold and only rarely reflected completely different interpretations. Human moderation remained essential, especially for local context, irony, caricatures, subtleties, political meaning, and similar elements.

Main Results

The overall map shows a fairly strong concentration of content around a limited number of themes. The most frequent vectors are V34 - Cult of the Savior and Messianic Leader, V03 - Systemic Incompetence, V38 - Corruption, V02 - Betrayal of the Public Interest, V01 - Informal Decision-Making Networks/Parallel State, V21 - Autonomy and Sovereignty, and V06 - Economic Problems. V34 appears in almost one third of the posts, while the anti-system cluster, formed mainly by V03, V38, V02, and V01, occupies a very large share of the corpus.

This produces a construction in which the current political and institutional order is repeatedly described in terms of incompetence, corruption, betrayal, or hidden control, while, in parallel, the image of a leader capable of offering an alternative is constructed. At the tactical level, the most frequent mechanisms include manipulation of visual material, deliberate provocation, construction of an alternative narrative, misleading headlines, imposition of a single version of the truth, and semantic manipulation. At the psychological level, moral indignation, admiration and idealization, attribution of ill intent, contempt, the need for simplicity, cohesion through a common enemy, frustration, and attraction to shocking content appear very frequently.

The analysis of $V \times T \times RP$ combinations led to the identification of five manipulation recipes, which together cover 90.3% of the corpus. This result is important because the 40 vectors, 34 tactics, and 74 psychological mechanisms can form 100,640 $V \times T \times RP$ triplets. Of these, 47,995 actually appear in the corpus, while the ten most frequent triplets account for only 0.6% of all occurrences. If the analysis stopped here, manipulation might appear extremely diverse and lacking in structure. However, when the combinations are analyzed at the level of broader profiles, five families explain approximately nine out of ten posts.

The first recipe, Indignant Moral Denunciation, includes 305 posts, or 36.2% of the corpus. It attacks institutions, politicians, or opponents through themes of incompetence, corruption, and betrayal, and relies psychologically above all on indignation, contempt, frustration, and the attribution of ill intent. Anger is the dominant emotion.

The second recipe, Cult of the Savior Leader, appears in 214 posts, or 25.4%. Here, the emphasis is on constructing a positive figure presented as a protector, moral leader, and solution to the problems described by the rest of the content. Hope is the dominant emotion in 47% of the posts in this recipe, while Pride is dominant in 44%. Together, the first two recipes account for 61.5% of the corpus, and the relationship between them is important because one delegitimizes the opponent and the existing system, while the other constructs the alternative.

The third recipe is Existential Fear, with 120 posts, or 14.2%, in which themes such as war, collapse, economic problems, and decadence are delivered predominantly through fear. Globalist Conspiracy, present in 78 posts, or 9.3%, is the most complex of the five and explains different

events through the idea that hidden forces are acting behind the scenes. Restorative Nostalgia, with 44 posts, or 5.2%, has the simplest structure and draws on an idealized past, with Sadness as the dominant emotion in 59% of cases.

The temporal analysis shows two main modes of operation. One group of vectors runs relatively constantly regardless of events, while others react much more strongly to the immediate context. Of the 25 vectors frequent enough for stable analysis, five were classified as constant, 13 as reactive, and seven fell between the two categories. The constant layer appears in 73.1% of posts, while the reactive layer appears in 65.8%. In 43.2% of posts, elements from both categories appear simultaneously.

This difference can be seen clearly in the evolution of certain themes. Economic problems increase during weeks associated with taxes, Western decadence reacts to economic or geopolitical events, and systemic collapse rises during the week in which inflation becomes a major topic. At the same time, themes such as the savior leader, corruption, or sovereignty remain continuously present.

Hypothesis Testing

H1, related to Chinese Water Drop Torture, is confirmed. Of the 40 vectors, 37 meet the recurrence criteria established in the research. More importantly, seven central vectors—V34, V03, V38, V02, V01, V21, and V06—appear across all 16 fixed sources plus the feed and in all nine weeks analyzed. Together, they account for 49.7% of all vector occurrences. The result shows that there is a narrative background that repeats with very high regularity. The research demonstrates this recurrence, but not the existence of conscious coordination among all sources, which would require other methods, particularly network analysis.

H2, related to Chaos Theory, is confirmed with qualifications. The mean SC score is 21.01, and its relationship with the engagement rate is weak, explaining only a small part of the variation in engagement. The rest depends both on external factors—the algorithm, publication timing, events, audience behavior, and so on—and on content elements that the SC score does not capture.

The group analysis, however, revealed a very interesting threshold around SC=20. Below this level, the virality rate is around 23%, while above it the rate rises to approximately 42%. Median engagement increases from 0.35–0.58 to 1.75–2.29. The threshold is not a universal rule and does not guarantee virality, but within the corpus it marks a clear change in probability.

H3, related to the Butterfly Effect, is also confirmed with qualifications. Of the 843 posts, 275 were classified as butterflies, with engagement above 3%, while 568 were classified as drops. Clear differences appear between the two groups in the distribution of emotions. Pride increases from 10.6% among drops to 24.0% among butterflies, Hope from 13.2% to 23.6%, and Sadness from 4.2% to 10.5%. Fear decreases from 15.5% to 3.6%, while Anger decreases from 33.1% to 26.2%. These differences show that high-engagement posts have a distinct emotional profile compared with low-engagement posts.

V34 - Cult of the Savior Leader appears in 52.7% of viral posts, compared with 23.2% of drops. RP54 - Admiration/Idealization increases from 24.8% to 58.5%, while tactics involving the construction of a protector figure, the aestheticization of authority, and the legitimization of authoritarianism are also much more prevalent in viral posts. The mean SC score is 22.44 for

butterflies and 20.32 for drops, showing that a complexity difference exists, but it is much smaller than the differences in composition. Therefore, what matters is not only how many elements are used, but especially which elements are used.

The interpretation must, however, be made carefully. The combination of V34 with Hope or Pride represents 40.7% of the butterflies in the corpus and is concentrated mainly around Călin Georgescu and, occasionally, George Simion. These posts have greater cross-source overlap and higher mean engagement than the other butterflies, which makes the hypothesis of coordinated amplification plausible, without the method used being able to demonstrate its existence directly.

Original Contributions of the Research

I consider this section one of the most important in the thesis because it shows what the research adds to what already existed in the literature. The contributions are theoretical, methodological, and empirical, and the three are interconnected.

The first theoretical contribution is the integration of Chinese Water Drop Torture, Chaos Theory, and the Butterfly Effect into a single analytical framework. The metaphors were not left at the level of explanatory comparisons, but were linked to concrete variables and tested on the corpus. The Drop was tested through the recurrence of vectors over time and across sources, Chaos through the relationship between the complexity score and engagement, and the Butterfly through comparison of the $V \times T \times RP$ profile of viral posts with that of posts that remain drops. The three explanations are not mutually exclusive; rather, they describe different parts of the phenomenon and complement one another.

Combining impact with the complexity score also produces a classification into four types of posts: simple drop, dense drop, simple viral, and dense viral. This classification separates construction complexity from the outcome achieved and shows that a post can be very dense without becoming viral, or can become viral even with a lower score. The model was constructed and tested under the conditions of this corpus, and I do not present it as a general rule, but I consider its logic sound and potentially applicable to other networks or types of content, with the necessary calibration.

The second important contribution is the three-dimensional $V \times T \times RP$ framework, with its 148 categories. It allows the same post to be analyzed simultaneously in terms of content, delivery technique, and psychological mechanism. The level of granularity is high, but this is precisely what made it possible to identify relationships that would have been difficult to observe with simplified coding. The framework does not determine whether information is true or false, nor does it directly measure its effect on people; these questions require other levels of analysis. It shows how the message is constructed and what it attempts to activate.

The methodological contribution is complemented by a coding triangulation and convergence procedure carried out using three AI systems in nine separate sessions for each post, with human moderation at every stage. I used AI not to make the final decision, but to produce parallel analyses that could be compared. I consider this hybrid method a useful model for research in which the volume of data is too large for exclusively manual coding by a single researcher, while the content is too complex to be left entirely to automated processing. The procedure is described in detail precisely so that it can be verified, criticized, replicated, and improved in other studies.

The empirical contribution consists of the map constructed from the 843 posts and the 17,707 codes applied to them. This analysis produced six findings that I consider important.

The first is the existence of a core of seven main vectors: V34, V03, V38, V02, V01, V21, and V06. At least one of these appears in 699 of the 843 posts, or approximately 83% of the corpus, and all are present throughout the entire period.

The second is the difference between drops and butterflies. 67.4% of the posts are drops and 32.6% are butterflies. The typical drop post has engagement of approximately 0.28%, while the typical butterfly post reaches 8.94%, a ratio of about 32 to 1. The high percentage of viral posts should not be generalized to Romanian Facebook as a whole, because the monitored sources have highly active political audiences and there is also the possibility that part of the content was artificially amplified.

The third finding is the threshold around $SC=20$, above which the virality rate almost doubles. This result is important precisely because the relationship between complexity and engagement does not emerge clearly through a linear correlation. Denser content does not automatically produce better results, but two different performance regimes appear below and above this threshold.

The fourth finding is the existence of a distinct profile of viral posts. They use V34, RP54, construction of the protector figure, aestheticization of authority, Hope, and Pride much more frequently. At the same time, Fear, suspicion, and some attack tactics are more strongly associated with drop posts. The differences show that the virality observed during the research period is not distributed evenly across all forms of manipulation.

The fifth finding concerns what I called signature pairs between vectors and psychological mechanisms. Some narratives are very consistently linked to particular vulnerabilities. V34 - Savior Leader is associated with RP54 - Admiration/Idealization in approximately 88% of the cases in which it appears. V33 - Globalist Conspiracy is linked to RP51 - Generalized Orchestration in approximately 89%, V06 - Economic Problems to RP48 - Loss Aversion in approximately 88%, and V16 - Threat of War to RP13 - Fear in approximately 84%. V25 - Socioeconomic Nostalgia is, in turn, very strongly linked to RP67 - Nostalgia. These results suggest that the relationship between a narrative and a psychological mechanism is sometimes more stable than the relationship between a narrative and the tactic used to deliver it.

The sixth finding is the two-layer dynamic, with one constant layer and one reactive layer. The first keeps the core ideological themes in circulation, while the second exploits current events. The two do not operate separately, but appear simultaneously in a substantial part of the corpus. Thus, the same underlying themes are maintained over time but are rapidly connected to taxes, inflation, war, political statements, or other topics entering the public agenda.

Implications, Limitations, and Future Directions

The results also have implications for information security, because manipulation cannot be countered solely through technical or institutional measures. Media literacy remains important, but I believe it should be complemented by basic knowledge of psychology, because manipulation exploits natural human reactions, and people should understand at least the mechanisms through

which repetition, emotion, cognitive biases, or confirmation of their existing beliefs can influence judgment.

At the same time, institutions need to communicate faster and more clearly. Manipulative narratives are short, visual, emotional, and adapted to the channel, whereas institutional responses often arrive late and use unattractive language. Continuous monitoring of the digital space is also necessary, especially when the volume of content is very large. AI tools can be useful for filtering and analysis, but the final decision must remain with humans and must be open to verification. At the same time, the state must draw a clear distinction between opinions and actions. Criticism, even when harsh, should not be treated as manipulation. Intervention may be justified when the issue is no longer one of opinion, but involves the use of fake accounts, coordinated amplification, hidden financing, fraud, or incitement.

The research also has limitations. The corpus covers two months, one platform, and a single ideological area, and the sources were selected according to predetermined criteria rather than through statistical sampling. Engagement data were recorded 2–7 days after publication, which limits exact comparability. The 148-category framework provides a large amount of information, but it also increases the risk of overlap between codes. AI-assisted coding reduces part of the workload and enables triangulation, but it does not eliminate model errors or the subjectivity of the researcher who makes the final decision. In addition, the framework identifies the psychological mechanisms that the message attempts to activate, but does not demonstrate their actual effect on each recipient.

Future research would benefit from extending the time period, including other platforms, and comparing multiple ideological areas, while network analysis could show more clearly how the sources are connected and whether there are genuine indications of coordination. Experimental studies with participants could also measure the extent to which the mechanisms identified through coding produce the hypothesized psychological effects.

Conclusions

The four research questions were answered, and the three hypotheses were supported by the data to different degrees. Overall, the results show that the manipulation analyzed operates simultaneously through the repetition of stable themes, rapid adaptation to current events, and the use of combinations of narratives, tactics, and psychological mechanisms that can produce very different levels of engagement.

The research also shows that manipulation cannot be reduced to the mere circulation of false information. Many of the posts analyzed start from real problems and grievances, which are then selected, exaggerated, and connected within a broader political interpretation. For this reason, countering manipulation requires not only fact-checking, but also an understanding of the mechanisms through which messages are constructed and psychologically exploited.

I consider psychological warfare and digital manipulation to require continuous analysis, not only during electoral periods or after crises emerge. As technology evolves, manipulation tactics also change. The research provides a detailed picture of the period and ecosystem analyzed, as well as a framework that can be continued, extended, and adapted in future studies.