# AI-Driven Disinformation Campaigns on Twitter (X) in the Russia-Ukraine War

## Description

*Editorâ??s Note: this article is being republished with the permission of Small Wars Journal as part of a republishing arrangement between IWI and SWJ. The original article was published on 10.02.2025 and is available here.*

image

## Abstract

In the Russia-Ukraine war, sectors like the digital and psychological ones were added to the conflict, and AI-driven fake news on Twitter (X) played a major part. Propaganda was disseminated to large numbers by operatives who used automated tools powered by artificial intelligence. They applied data mining and AI techniques to read through large data sets, find out what people felt about certain situations, and then send messages that mattered to specific groups. People were exposed to AI-created content, which included articles, images, and deepfakes, all aimed at copying the look of trustworthy sources, tricking users, and harming what people think about Ukraineâ??s leadership and help from the West.

## Introduction

The Russia-Ukraine conflict, which reached its peak in February 2022, has fought a massive struggle of information dissemination in addition to its existing military hostilities. Current media organizations work to provide reports about the situation, but people turn primarily to Twitter (X) platforms to share news and opinions with disinformation. Twitter (X)â??s network features, including instant posting and tagging patterns, enable users to distribute content widely, thus creating an effective means of influencing public opinion. Artificial intelligence (AI) stands as an essential tool for people who wish to control and shape information during this particular time. Artificial intelligence enables the production and dissemination of false information on Twitter (X), which produces major changes in how audiences think about the conflict. The analysis of disinformation campaign mechanisms on Twitter (X) enables people to understand the effects these techniques have on their perceptions and perspectives. During the Russia-Ukraine war, Twitter (X) and AI have enabled multiple disinformation methods that altered global public perception and war narrative formation. AI bots, together with AI-backed accounts, generate artificial propaganda that opposes alternate perspectives and advocates opposing viewpoints.

The level of complexity with which AI handles disinformation campaigns on Twitter (X) continues to advance. It utilizes machine learning algorithms to evaluate big datasets from the platform, which helps them detect patterns and user attitudes alongside population statistics. AI analytics produce targeted disinformation programs that speak directly to various audiences through content that appeals to them. Comprehensively detailed messages take advantage of peopleâ??s existing viewpoints, therefore becoming more convincing in their delivery and difficult to detect AI-generated content.

## The Role of Twitter (X) in Information Warfare

Twitter (X) functions as a major platform for delivering news from the Russia-Ukraine conflict, to presents live information worldwide to its users. Through its message-sharing system, Twitter (X) allows people to transmit brief updates together with their opinions and situational analyses rapidly. Fast sharing of information creates opportunities to transmit fake information because of Twitter (X)â??s real-time transmission capabilities. Twitter (X) has been structured to promote shareable content through its features that include hashtags along with retweet capabilities, enabling fast dissemination to many users. Information exchanged at fast speeds on Twitter (X) enables people to mistake unverified details as authentic, which results in misinterpretations of ground realities.

Twitter (X) turned into a leading source of news updates after the conflict began, as both official accounts and common users transmitted real-time developments through their posts. The hashtags *#UkraineUnderAttack* and *#RussiaTerroristState* gained wide usage to mobilize international support and political focus on Ukraine. The genuine expressions of solidarity, as well as accurate news reporting, were accompanied by fabricated hashtags. The pro-Russian accounts invented *#FakeNewsUkraine* to attack the Ukrainian narrative with massive hashtags per minute, and Twitter (X) users quickly reacted to instantaneous information, thus becoming more likely to accept information they saw as facts, even if they were untrue. The accumulation of deceptive narratives caused noticeable changes in public understanding of the conflict, which subsequently guided both international reactions and public discussions about the matter.

## Automated Bots and AI-generated content on Twitter (X)

AI-based automated bots deployed through Twitter (X) disinformation campaigns represent the most widespread AI implementation for spreading misinformation. These bots use programmed AI systems that produce content rapidly while attempting to imitate human dialogue behavior. Artificial intelligence bots operate in an automated style to both post tweets and retweets from other profiles while they initiate social interactions with users to build fake natural dialogue, duplicating human tone. AI systems produce fake human-shaped texts quickly, which introduces false information into the public discussion. Twitter (X) proved the existence of Russian disinformation through its removal of thousands of fraudulent accounts that disseminated fabricated information using bots. Researchers at the University of Adelaideâ??s School of Mathematical Sciences analyzed 5,203,764 Twitter (X) posts, including tweets, retweets, quote tweets, and replies that carried hashtags *#StandWithPutin*, *#StandWithRussia*, *#SupportRussia*, *#StandWithUkraine*, *#StandWithZelenskyy*, and *#SupportUkraine* from 23 February through 8 March 2022. Research revealed that AI bot accounts comprised 60 to 80 percent of the total tweets bearing the studied hashtags during this specific timeframe.

> A recent study discovered that about 1,000 fake American AI bots had generated accounts to spread pro-Russian propaganda, which targeted Ukrainian backlash and shaped geopolitical stories to benefit Ukrainian interests by polluting public opinion.

The extent of automation is worth mentioning, as highlighted by the Washington Post, where researchers examined 1.3 million accounts that regularly tweeted about Russian politics, underscoring that 45% or 585,000 of these accounts were bots, also known as â??political botsâ?•. Thousands of social media bots and fake AI-supported accounts spread false information about the Russia-Ukraine conflict, severely affecting public opinion during the conflict. The bot networks employ artificial intelligence to behave similarly to real people so they can connect with users as they distribute state-approved Russian self-defense messages. A recent study discovered that about 1,000 fake American AI bots had generated accounts to spread pro-Russian propaganda, which targeted Ukrainian backlash and shaped geopolitical stories to benefit Ukrainian interests by polluting public opinion.

Deepfakes serve as a main method in the Russia-Ukraine war to distribute misinformation through the Twitter (X) social media platform. Artificial intelligence systems create fake videos and audio files that alter real-world public figures, such that their images, as well as recorded statements present false information that quickly captures widespread attention. In March 2022, the release of a deepfake video depicted Ukrainian President Volodymyr Zelenskyy supposedly asking his troops to lay down their weapons and surrender. Even though the video proved to be fake, it received thousands of retweets on Twitter (X) before people discovered its fraud. According to research from the Digital Forensic Research Lab, deepfakes and other misinformation about the conflict spread to more than 70 million Twitter (X) users during the first few weeks of the Russian invasion. A manipulated video falsely showing Putin making a declared martial law statement successfully spread through Twitter (X),using hashtags including #PutinSurrender to deceive people about his actions. Deepfakes are distributed at an alarming rate because they create significant problems for readers trying to spot factual information, since manipulated media warps true events and can shape how people think about reality.

These artificial intelligence bots use algorithms to identify trends, along with generating messages that represent prevailing public sentiments. The bots leverage this ability to expand messages that represent Russian military ventures as suitable reactions to outside dangers. During 2024, the Pravda network released around three and a half million inaccurate AI-produced articles because they wanted to test AI chatbots and confuse their responses. Artificial intelligence bots produced fake tweets that presented false statements about the Ukrainian military committing violent actions toward civilians so Russia could legitimize its military actions under humanitarian pretenses. A study showed that elite Twitter (X)bots spread Russian falsified information successfully through their platforms with a success rate of approximately 33%.

Although Twitter (X) is officially blocked in Russia, the platform continues to host accounts and users that are backed by AI Twitter (X)bots, which are strategically employed to counter Ukrainian narratives and promote pro-Russian viewpoints. The artificial intelligence-based bots operate to spread Russian-supporting content as a method to generate trends that back up Kremlin positions. The bots enhance Russian control of Twitter (X) through their continuous production of pro-Russian content and counteraction of opposing viewpoints, thus determining how people perceive the Ukrainian conflict. These AI bots execute three elements: post pro-Russia tweets that link to active trends and hashtags, follow users to enhance certain messaging, and create the impression of backing for Russian activity. The coordinated campaign works to discredit Ukrainian perspectives between nations as it leverages AI technology to transform modern information warfare between countries.

The false tweets generated by AI bots and large language models (LLM) generated false public consensus about narratives, including the widespread defense justification for Russian actions and the narrative about sovereignty protection. Automated bots succeed in manipulation by producing substantial content quantity rapidly.  Users who received the same message from different accounts three times were inclined to believe these narratives were true. The misinformation spread by these bots altered public opinion, thus creating wrong perceptions about the conflict while promoting Russian propaganda.

## Viral Hashtag Manipulation

Through the use of #*StopUkrainianAggression*, pro-Russian accounts wanted to change the publicâ??s perspective by making Russia appear defensive. Several untrue viral tweet contents spread quickly because they contained frequently used hashtags, making them accessible to a wider listener base. A generating campaign alleging Ukrainian military attacks against civilians would gain rapid spread across social media, employing strategic AI-created hashtag promotion. The hashtag mechanism formed a propagation pattern that allowed false information to spread rapidly throughout different groups of audience members. AI bots use automated content creation to support hashtag propagation and distribute disinformation across trending topics. Due to their data-oriented algorithms, AI bots execute their work, which enables them to spread false information regardless of the truth. These bots maintain pro-Russian content creation alongside user interaction to build artificial consensus about fake messages until it becomes hard for users to identify authentic information from fake content. The discussions about Ukraineâ??s conflict evolved through misinformed content that spread through popular hashtag trending topics. The artificial intelligence algorithms significantly contributed to this misinformation spread by causing bots to interact with the hashtag, thus producing a continuous stream of false content that deepened its acceptance in public discussion. Research showed how millions of people received messages through the #UkrainianNazis hashtag, and one specific post became viral within only a few hours after being posted. By controlling hashtags, these operators disrupted public discussions, so information accuracy became nearly impossible to find since the noise drowned out all else. The false narratives reached wide public visibility, which altered how people viewed the conflict and how they understood both parties.

## Targeted Disinformation Campaigns to gain public support

Many of the AI tools operated by Russian operatives analyzed the listener groups that reacted best to specific informational messages during the conflict. The Russian strategists engineered propaganda through campaigns to depict Western Ukraineâ??s backing as a violation of Russiaâ??s regional control area. Twitter (X) messages that were specifically tailored presented alleged cases of Ukrainian governmental corruption while spreading the claims that Western countries are backing a â??failedâ?• state. Russian operatives during the Russia-Ukraine war implemented artificial intelligence for creating targeted disinformation, which used enhanced propaganda through precise audience targeting capabilities. The campaigns used machine learning algorithms to analyze user data, which helped propagandists deliver messages specific to their target groupsâ?? political and demographic characteristics. Artificial intelligence bots spread propaganda by increasing message efficacy, so false information seemed to have broad support from the public. Deepfake videos alongside media further added credibility to fabricated narratives.

Public opinion has seen a considerable transformation because of deliberately crafted misinformation. The campaigns achieved their objectives by targeting specific audiences, strengthening pre-existing beliefs, and making false information more convincing to counter. These refined strategies helped to increase societal polarization because people started creating self-contained information bubbles that eliminated dissenting opinions. These narratives grew more pervasive in worldwide dialogue, which made diplomatic solutions more challenging for the crisis.

## Conclusion

Twitter (X) served extensively for spreading disinformation throughout the Russia-Ukraine conflict to show how information warfare is becoming more sophisticated. Modern AI systems have extensively boosted the production and diffusion of false content, specifically identified for target audiences, allowing them to define public opinions regarding the ongoing conflict. Twitter (X) provides a case study of how social media platforms can become tools for weaponization in modern times through mechanisms ranging from false narrative automation by AI bots to orchestrating disinformation operations that confirm existing biases.

*Tayyaba Rehan is a student of the National Defence University, Islamabad, Pakistan. She is currently pursuing her degree in Defence and Strategic Studies. She has worked with multiple governmental and non-governmental organizations. Her numerous articles have been published on national and international platforms. Her area of interest includes national security, terrorism, and peacekeeping.*

*If you value reading the Irregular Warfare Initiative, please consider supporting our work. And for the best gear, check out the IWI store for mugs, coasters, apparel, and other items.*

**Date Created**
2025/11/12