EU DISINFO LAB

vera.ai

Visual assessment of CIB in disinformation campaigns

JANUARY 2025

TABLE OF CONTENTS

Author: **Ana Romero-Vicente**

Reviewer: **Maria Giovanna Sessa**

This publication has been made in the framework of the Vera AI project. Vera.ai is co-funded by the European Commission under grant agreement ID 101070093, and the UK and Swiss authorities. This website reflects the views of the vera.ai consortium and respective contributors. The EU cannot be held responsible for any use which may be made of the information contained herein.

INTRODUCTION

This study examines three disinformation, manipulation and interference campaigns, using visual representations to illustrate how they align with key indicators of Coordinated Inauthentic Behaviour (CIB) campaigns. This assessment includes a range of indicators to spot coordination, (in)authenticity, the source, as well as the impact and distribution in order to make the CIB phenomenon more easily identifiable and understandable to a wider audience.

METHODOLOGY

- To determine if a case study exhibits CIB, a set of 50 CIB generic indicators has been defined. These indicators can be used to analyse any disinformation, manipulation and interference campaign and are based on EU DisinfoLab's previous work, Revisit the Coordinated Inauthentic Behaviour Detection Tree.
- The next step is to evaluate whether these indicators are present. If they are, the likelihood of CIB activity increases. Each indicator is marked as "Y" (Yes) if present and "N" (No) if not —whether due to lack of detection, occurrence, or inclusion in the research. For this reason, the absence of an indicator does not decrease the overall likelihood of CIB.
- Each indicator present in the analysis carries equal weight, contributing proportionally to a score between 0 and 100. This score represents the probability of CIB, with 0 indicating no indicators detected (0% probability of CIB) and 100 indicating all indicators present (100% probability). For instance, if a case study contains 40 out of 50 indicators, the score results in an 80% probability of CIB.
- To simplify interpretation, 5 gauges are used to visually present the results: Coordination, Authenticity, Source, Impact and Final Assessment. The gauges are colour-coded as follows: red for scores lower than 25 (low likelihood of CIB), yellow for scores between 25 and 75 (medium likelihood of CIB) and green for scores above 75 (high likelihood of CIB).

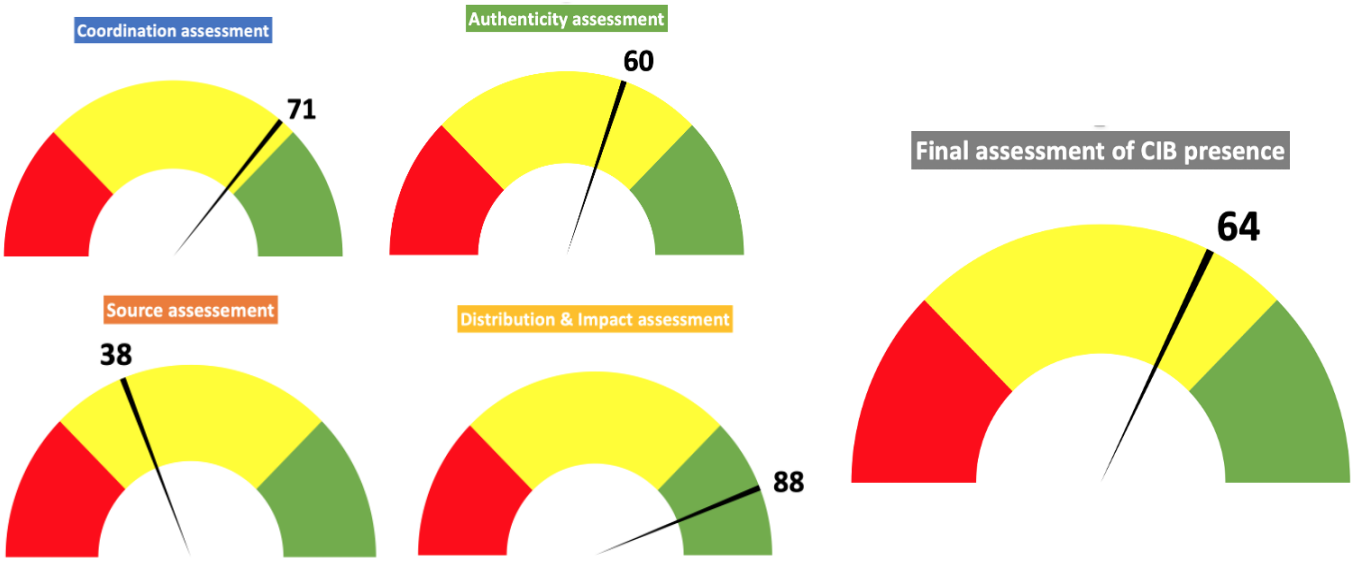
#CASE STUDY 1: OPERATION OVERLOAD

"Operation Overload" exposes a large-scale cross-country, multi-platform disinformation campaign, revealed by CheckFirst, aimed at manipulating media and fact-checkers in Europe. This pro-Russian operation, with clear indicators of foreign interference and information manipulation (FIMI), targets organisations with fake content through emails and social media, urging them to verify misleading claims, diverting their efforts from genuine reporting and fact-checking. The perpetrators employ sophisticated techniques to create a false sense of widespread acceptance and viral traction for their fabricated content. The investigation was updated in September, 2024

Check the EU DisinfoLab Webinar "Operation Overload: 'Please check' – how pro-Russian propagandists' try to manipulate newsrooms", with Guillaume Kuster, Check First.

CIB ASSESSMENT

Medium-high likelihood of CIB, especially in the distribution, coordination and authenticity assessment.



INDICATORS	COORDINATION	Y/N	AUTHENTICITY	Y/N	SOURCE	Y/N	DISTRIBUTION & IMPACT
CONTENT	1. Accounts copy-pasting and sharing the same or similar textual content.	Y	15. Poor translation, misspelling, and typos.	N	35. Mentions of specific names, organisations, locations, or keywords that hint at the involvement of known malign actors or behind the campaign.	N	43. The campaign content aims at a specific target.
	2. Use of evocative images, memes or collages to simplify complex issues and trigger emotional reactions.	Y	16. Sense of unnaturalness, with a repetitive tone (e.g., frequent use of specific phrases, hashtags, slogans).	N			44. The campaign leverages high-profile events for maximum impact.
	3. Translating and posting the same or similar content in different languages.	Y	17. Extreme content polarisation (amplifying specific narratives) or whimsical and conspiratorial plots.	Y			
	4. Single-topic accounts.	Y	18. Manipulated, forged or fabricated texts, including using fake endorsements by public figures.	Y			
METADATA	5. Multiple accounts using the same IP address, device and configurations.	N	19. Activity from IP ranges associated with VPNs or proxies.	N	36. Metadata analysis of content, visuals or links provides insights into the identity of the creator.	Y	/
			20. Sudden spikes in API requests that far exceed normal patterns.	N	37. Monitoring IP addresses, timestamps, and request details uncovers the actors behind the campaign.	N	
			21. Activity from IP geographic locations that are known for hosting bot farms.	N	38. Geographic tracking of API requests reveals physical locations linked to the actors behind the campaign.	N	
IDENTITY	6. Multiple accounts share identical or similar profile name or bio.	N	22. Identity theft (name, location and any other private detail from a person or entity).	Y	39. Account registration details (e.g., emails, domains, registered companies, phone numbers, or physical addresses) share a common source.	Y	/
			23. Lack of personalisation: generic profiles with minimal personal information.	Y			
			24. Profile names with random numbers or letters.	Y			
VISUALS	7. Multiple accounts share identical or similar profile picture or cover photos.	N	25. Visual theft of images or logos impersonating a legitimate person or entity.	Y	40. Visual elements in shared images reveal identifiable clues about the campaign's origin or source.	N	/
			26. Simultaneous profile pictures updates on multiple accounts.	N			

			27. Spoofed, fabricated AI-generated visuals, including low quality ones.	Y			
BEHAVIOURAL	8. Multiple accounts created around the same time.	Y	28. Account activity suddenly resumed after a long period (dormant accounts).	Y	/		45. Peripheral accounts amplify the content of core accounts.
	9. Similar posting timestamps across different accounts.	Y	29. Proof that account has been hijacked for the campaign.	Y			46. Evidence of shifts in user behaviour, such as increased support or opposition to a campaign-targeted issue.
	10. Sudden spikes in messaging around a certain narrative or event (this may also suggest automated bot traffic).	Y					
	11. Significant activity from locations or time zones that do not align with the alleged user location (if any).	N					
	12. Strategically and repeatedly expose viewers to the same false stories exploiting the illusory truth effect.	Y					
NETWORK	13. Accounts engaging with each other in a synchronised manner, often from disproportionately interconnected clusters of accounts that follow each.	Y	30. Accounts with little activity that rarely interact with users outside of their network.	Y	41. Early-interacting accounts reveal the primary source or key participants.	N	47. Accounts in the network engage with each other across platforms, amplifying the campaign's impact.
	14. Cross-platform coordination: Similar posting patterns across different social media platforms.	Y	31. High interaction posts from low-activity accounts with incomplete profiles or minimal content history.	Y	42. Analysis of cross-platform activity reveals recurring patterns and coordinated interactions that trace back to key accounts or nodes central to the campaign's origin.	Y	48. Content is amplified by state media.
			32. Unusual high levels of likes, shares, or comments.	N			49. Content is amplified by public figures known to spread disinformation.
			33. Abnormal changes in follower count over short periods of time.	N			50. Content is amplified by fringe or junk outlets.
AUTOMATION	/		34. Bot behaviour: unusually rapid engagement with content, such as liking, sharing, following, or commenting.	Y	/		/

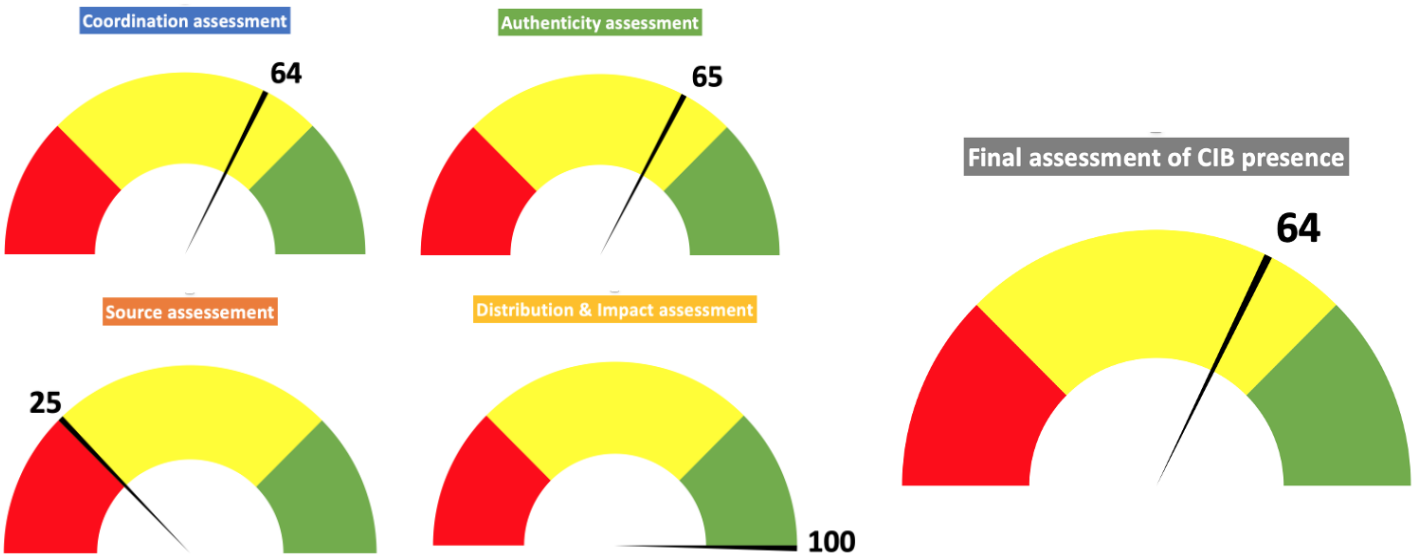
#CASE STUDY 2: MASSIVE RUSSIAN INFLUENCE OPERATION TARGETED FORMER UKRAINIAN DEFENCE MINISTER ON TIKTOK

A joint investigation by DFRLab and BBC Verify uncovered a massive Russian influence operation targeting former Ukrainian Defence Minister, Oleksii Reznikov, on TikTok. This campaign, described as the largest on the platform, involved over 12,800 accounts spreading false corruption allegations using AI-generated audio in multiple languages. These videos, viewed millions of times, aimed to damage Ukraine's image and erode Western support by portraying Ukrainian officials as corrupt, thus undermining their credibility during the ongoing conflict with Russia.

Check the EU DisinfoLab webinar "Massive Russian influence operation on TikTok and beyond", with Roman Osadchuk, DFRLab.

CIB ASSESSMENT

Medium-high likelihood of CIB, especially in distribution, coordination and authenticity assessment.



INDICATORS	COORDINATION	/N	AUTHENTICITY	Y/N	SOURCE	Y/N	DISTRIBUTION & IMPACT	Y/N
CONTENT	1.Accounts copy-pasting and sharing the same or similar textual content.	Y	15. Poor translation, misspelling, and typos.	N	35. Mentions of specific names, organisations, locations, or keywords that hint at the involvement of known malign actors or behind the campaign.	Y	43. The campaign content aims at a specific target.	Y
	2.. Use of evocative images, memes or collages to simplify complex issues and trigger emotional reactions.	Y	16. Sense of unnaturalness, with a repetitive tone (e.g., frequent use of specific phrases, hashtags, slogans).	Y			44.The campaign leverages high-profile events for maximum impact.	Y
	3. Translating and posting the same or similar content in different languages.	Y	17. Extreme content polarisation (amplifying specific narratives) or whimsical and conspiratorial plots.	Y				
	4. Single-topic accounts.	Y	18. Manipulated, forged or fabricated texts, including using fake endorsements by public figures.	Y				
METADATA	5. Multiple accounts using the same IP address, device and configurations.	N	19. Activity from IP ranges associated with VPNs or proxies.	N	36. Metadata analysis of content, visuals or links provides insights into the identity of the creator.	N	/	
			20. Sudden spikes in API requests that far exceed normal patterns.	N	37. Monitoring IP addresses, timestamps, and request details uncovers the actors behind the campaign.	N		
			21. Activity from IP geographic locations that are known for hosting bot farms.	N	38. Geographic tracking of API requests reveals physical locations linked to the actors behind the campaign.	N		
IDENTITY	6. Multiple accounts share identical or similar profile name or bio.	N	22. Identity theft (name, location and any other private detail from a person or entity).	Y	39. Account registration details (e.g., emails, domains, registered companies, phone numbers, or physical addresses) share a common source.	N	/	
			23. Lack of personalisation: generic profiles with minimal personal information.	Y				

			24. Profile names with random, numbers or letters.	Y				
VISUALS	7. Multiple accounts share identical or similar profile picture or cover photos.	N	25. Visual theft of images or logos impersonating a legitimate person or entity.	Y	40. Visual elements in shared images reveal identifiable clues about the campaign's origin or source.	N	/	
			26. Simultaneous profile pictures updates on multiple accounts.	N				
			27. Spoofed, fabricated AI-generated visuals, including low quality ones.	Y				
BEHAVIOURAL	8. Multiple accounts created around the same time.	N	28. Account activity suddenly resumed after a long period (dormant accounts).	Y	/		45. Peripheral accounts amplifying the content of the core accounts.	Y
	9. Similar posting timestamps across different accounts.	Y	29. Proof that account has been hijacked for the campaign.	N			46. Evidence of shifts in user behaviour, such as increased support or opposition to a campaign-targeted issue.	Y
	10. Sudden spikes in messaging around a certain narrative or event (this may also suggest automated bot traffic).	Y						
	11. Significant activity from locations or time zones that do not align with the alleged user location (if any).	N						
	12. Strategically and repeatedly expose viewers to the same false stories exploiting the illusory truth effect.	Y						
ETWORK	13. Accounts engaging with each other in a synchronised manner, often from disproportionately interconnected clusters of accounts that follow each.	Y	30. Accounts with little activity that rarely interact with users outside of their network.	Y	41. Early-interacting accounts reveal the primary source or key participants.	N	47. Accounts in the network engage with each other across platforms, amplifying the campaign's impact.	Y

			31. High interaction posts from low-activity accounts with incomplete profiles or minimal content history.	Y				
	14. Cross-platform coordination: Similar posting patterns across different social media platforms.	Y	32. Unusual high levels of likes, shares, or comments.	Y	42. Analysis of cross-platform activity reveals recurring patterns and coordinated interactions that trace back to key accounts or nodes central to the campaign's origin.	Y	48. Content is amplified by state media.	Y
			33. Abnormal changes in follower count over short periods of time.	N			49. Content is amplified by public figures known to spread disinformation.	Y
							55. Content is amplified by fringe or junk outlets.	Y
AUTOMATION	/		34. Bot behaviour: unusually rapid engagement with content, such as liking, sharing, following, or commenting.	Y	/		/	

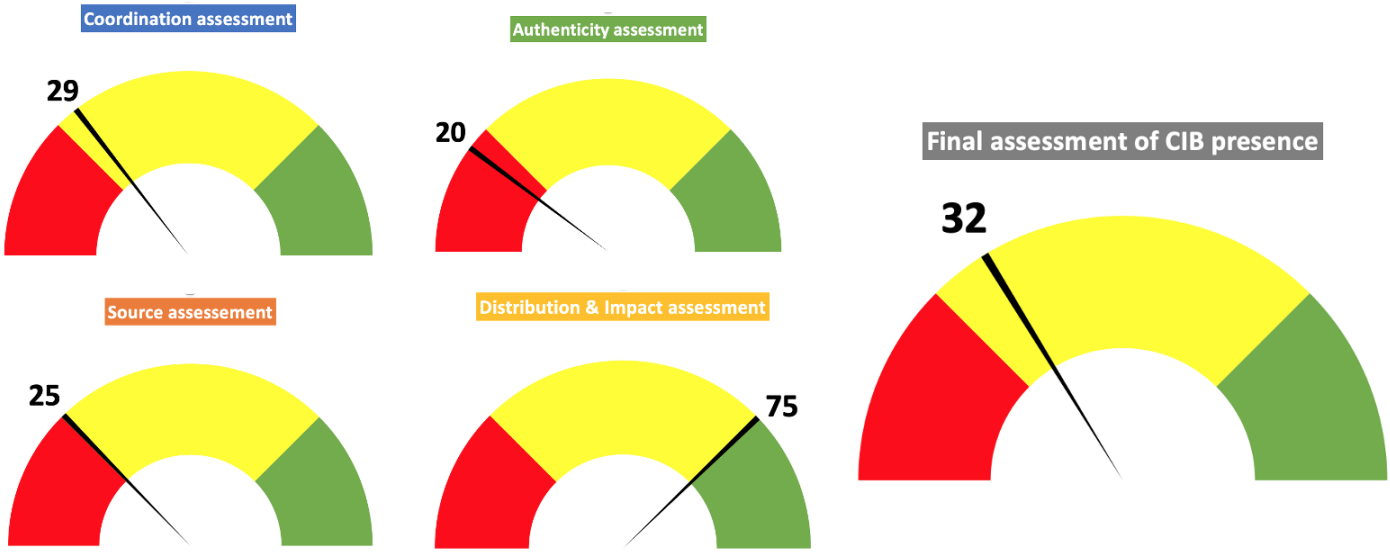
#CASE STUDY 3 - QANON'S 'SAVE THE CHILDREN' CAMPAIGN

In the summer of 2020, the QAnon movement hijacked the #SaveTheChildren hashtag, associated with a reputable charity, using it to introduce their conspiracy theories to a broader audience. Malicious actors spread falsehoods, disguising them within an allegedly legitimate cause solely to amplify their narrative (i.e., elites harvest a chemical called adrenochrome from children for longevity). They leveraged the increased public concern about child trafficking to promote their agenda, creating a false sense of widespread support to manipulate public opinion.

Multiple stakeholders such as Vox, AP, the New York Times, and Save the Children reported the case, as well as researchers from the Queensland University of Technology and from the University of Maryland and the Georgia State University.

CIB ASSESSMENT

Medium low likelihood of CIB despite high distribution and impact assessment.



INDICATORS	COORDINATION	Y/N	AUTHENTICITY	Y/N	SOURCE	Y/N	DISTRIBUTION & IMPACT	Y/N
CONTENT	1. Accounts copy-pasting and sharing the same or similar textual content.	Y	15. Poor translation, misspelling, and typos.	N	35. Mentions of specific names, organisations, locations, or keywords that hint at the involvement of known malign actors or behind the campaign.	Y	43. The campaign content aims at a specific target.	Y
	2.. Use of evocative images, memes or collages to simplify complex issues and trigger emotional reactions.	Y	16. Sense of unnaturalness, with a repetitive tone (e.g., frequent use of specific phrases, hashtags, slogans).	Y			44. The campaign leverages high-profile events for maximum impact.	Y
	3. Translating and posting the same or similar content in different languages.	N	17. Extreme content polarisation (amplifying specific narratives) or whimsical and conspiratorial plots.	Y				
	4. Single-topic accounts.	N	18. Manipulated, forged or fabricated texts, including using fake endorsements by public figures.	N				
METADATA	5. Multiple accounts using the same IP address, device and configurations.	N	19. Activity from IP ranges associated with VPNs or proxies.	N	36. Metadata analysis of content, visuals or links provides insights into the identity of the creator.	N	/	
			20. Sudden spikes in API requests that far exceed normal patterns.	N	37. Monitoring IP addresses, timestamps, and request details uncovers the actors behind the campaign.	N		
			21. Activity from IP geographic locations that are known for hosting bot farms.	N	38. Geographic tracking of API requests reveals physical locations linked to the actors behind the campaign.	N		

IDENTITY	6. Multiple accounts share identical or similar profile name or bio.	N	22. Identity theft (name, location and any other private detail from a person or entity).	N	39. Account registration details (e.g., emails, domains, registered companies, phone numbers, or physical addresses) share a common source.	N	/	
			23. Lack of personalisation: generic profiles with minimal personal information.	N				
			24. Profile names with random, numbers or letters.	N				
VISUALS	7. Multiple accounts share identical or similar profile picture or cover photos.	N	25. Visual theft of images or logos impersonating a legitimate person or entity.	Y	40. Visual elements in shared images reveal identifiable clues about the campaign's origin or source.	N	/	
			26. Simultaneous profile pictures updates on multiple accounts.	N				
			27. Spoofed, fabricated AI-generated visuals, including low quality ones.	N				
BEHAVIOURAL	8. Multiple accounts created around the same time.	N	28. Account activity suddenly resumed after a long period (dormant accounts).	N	/		45. Peripheral accounts amplifying the content of the core accounts.	Y
	9. Similar posting timestamps across different accounts.	N	29. Proof that account has been hijacked for the campaign.	N			46. Evidence of shifts in user behaviour, such as increased support or opposition to a campaign-targeted issue.	Y
	10. Sudden spikes in messaging around a certain narrative or event (this may also suggest automated bot traffic).	Y	.					
	11. Significant activity from locations or time zones that do not align with the alleged user location (if any).	N						

	12. Strategically and repeatedly expose viewers to the same false stories exploiting the illusory truth effect.	Y						
NETWORK	13. Accounts engaging with each other in a synchronised manner, often from disproportionately interconnected clusters of accounts that follow each.	N	30. Accounts with little activity that rarely interact with users outside of their network.	N	41. Early-interacting accounts reveal the primary source or key participants.	N	47. Accounts in the network engage with each other across platforms, amplifying the campaign's impact.	Y
	14. Cross-platform coordination: Similar posting patterns across different social media platforms.	N	31. High interaction posts from low-activity accounts with incomplete profiles or minimal content history.	Y	42.Analysis of cross-platform activity reveals recurring patterns and coordinated interactions that trace back to key accounts or nodes central to the campaign's origin.	Y	48. Content is amplified by state media.	N
			32. Unusual high levels of likes, shares, or comments.	N			49. Content is amplified by public figures known to spread disinformation.	Y
			33. Abnormal changes in follower count over short periods of time.	N			55. Content is amplified by fringe or junk outlets.	N
AUTOMATION	/		34. Bot behaviour: unusually rapid engagement with content, such as liking, sharing, following, or commenting.	N	/		/	