The Silent War

Why Coordinated Attacks Are the Greatest Undetected Risk Facing Your Organisation

Executive Summary

In 2024, the World Economic Forum named misinformation and disinformation the #1 short-term global risk for the second consecutive year. In the same year, the FBI executed a covert sting operation ‘Operation Token Mirrors’ which exposed coordinated bot networks manipulating more than 60 cryptocurrency tokens, resulting in $25 million in seized assets and 18 criminal charges. Crypto chain analysis company ‘Chainalysis’ estimated that suspected wash-trading volume on major decentralised exchanges reached $2.57 billion in 2024 alone.

These are not isolated incidents. They are data points in a rapidly escalating pattern: the weaponisation of social media platforms as instruments of financial manipulation, reputational destruction, and manufactured consensus. And the organisations bearing the cost. Brands, financial institutions, crypto projects, and governments are almost universally operating without the tools to detect it in time to act.

This is the second in AI Uniti's whitepaper series on coordinated attack detection. Where Volume I established the foundational research framework grounded in our proprietary analysis of 793,000 videos and the 6–12 hour detection gap this volume goes into further detail documenting the anatomy of modern attacks. We have examined real-world consequences across five high-impact sectors, presented illustrative case studies drawn from observed attack patterns, and supported the evidence-based case for why behavioural intelligence not social listening is now a board-level risk imperative which urgently requires addressing.

Core Finding: 93% of organisations currently monitor what is being said about them. Fewer than 7% can detect the coordination behind it. In the interval between those two capabilities, attacks take hold, narratives set, and real users begin to amplify manufactured signals often before a single alert has fired.

1. A Threat Hiding in Plain Sight

A coordinated attack does not announce itself. It does not wear the hallmarks of traditional disinformation campaigns, obvious bot accounts, poorly written content, or easily-traceable sources. The generation of coordinated attacks that now target brands, financial instruments, and political actors is sophisticated, patient, and specifically designed to appear organic.

Academic research published at the ACM Web Conference 2025 identified coordinated inauthentic behaviour across X (Twitter), Facebook, and Telegram in the lead-up to the 2024 US Presidential Election, demonstrating that coordination routinely transcends platform boundaries, operates with foreign-affiliated networks, and drives the spread of low-credibility, conspiratorial content with measurable effect on public discourse.[1]

Research published in arXiv in March 2025 by Cinelli et al. quantified precisely how coordinated accounts operate within information cascades: they occupy positions closer to the root, spread messages faster, and involve a systematically larger proportion of users than non-coordinated accounts.[2] In practical terms, this means a coordinated network does not simply amplify, it seeds. It shapes the initial architecture of how information spreads before authentic voices enter the conversation.

The Five Behavioural Fingerprints of Coordination

AI Uniti's research framework, built across 793,000 documented cases, identifies five distinct behavioural signatures that differentiate coordinated attacks from organic activity:

- **Velocity spikes:** A sudden acceleration in posting volume which is often 300–500% above baseline within an 8–12 minute window that precedes visible sentiment shifts by hours.
- **Synchronised bursts:** Clusters of accounts posting in tight temporal windows (often under 12 minutes) with near-identical phrasing, despite different handles and apparent identities.
- **Account creation clustering:** High-risk accounts frequently created within the same 6-hour window, with minimal prior history, activated simultaneously for a specific event.
- **Inauthentic amplification loops:** Content passed between inauthentic accounts to simulate organic sharing before any real users engage creating a false impression of consensus.
- **Phrase seeding:** Specific language fragments introduced across multiple accounts before organic adoption, designed to shape the vocabulary of discussion before authentic users encounter the topic.

None of these signals are individually visible to content-based monitoring tools. A sentiment analysis tool sees the words. A social listening platform tracks the volume. Neither sees the conductor. That is the detection gap AI Uniti exists to close.

Why Generative AI Has Changed the Stakes

Until 2023, the primary constraint on coordinated attacks was human labour. Generative AI has effectively eliminated that constraint. The 2025 WEF Global Risks Report notes that GenAI lowers barriers for content production and distribution, enabling threat actors, state agencies, activist groups, and individuals to automate and expand disinformation campaigns, greatly increasing their reach and impact.[3]

The Cyabra 2024 Brand Crisis Round-Up documented multiple major brand attacks in which fake profiles comprising 20–26% of accounts in the relevant discourse used generative AI-assisted content to amplify boycotts, distort narratives, and drive manufactured outrage to millions of views.[4] The brands targeted Jaguar, Coca-Cola, and Nestlé all which had existing social monitoring tools in place. Those tools detected the sentiment. They did not detect the coordination producing it.

52% of brands reported a social media-related cyberattack in 2024. The average recovery cost per incident was $4.6 million.[5]

2. The Cost of Being Late

Reputation damage consistently ranks among the most severe risks facing organisations globally. Aon's Global Risk Management Survey places it as the eighth largest current risk, noting that the impact of reputational events on stock prices has doubled since the introduction of social media.[6]

Weber Shandwick's foundational research quantifies why: 63% of a company's market value is attributable to its reputation**.** That makes a coordinated reputational attack not a communications problem but a financial risk event, one that belongs on the risk register alongside cyber and operational risk.

The Aon Pentland Analytics study tracked 125 reputation events over a decade, finding that after Volkswagen's emissions scandal, the company's loss in value exceeded $20 billion within a year which is approximately 25% of its pre-crisis value. While that case involved genuine wrongdoing, the mechanism for manufactured attacks is identical: a narrative takes hold, real users amplify it, institutional confidence erodes, and market value follows.

The Financial Services Threat: Manufactured FUD at Scale

In traditional financial markets, coordinated market manipulation has well-understood legal consequences. In digital asset markets, regulators are only beginning to catch up, and the gap has been ruthlessly exploited.

The FBI's Operation Token Mirrors (2024) exposed a network of market makers engaged in systematic wash trading across more than 60 cryptocurrency tokens. The operation resulted in $25 million in seizures and 18 criminal charges.[7] One defendant's company had inflated Saitama's token to a purported market value of $7.5 billion through coordinated bot activity combined with social media narrative management.

Chainalysis's 2025 market integrity report estimated that suspected wash-trading volume on Ethereum, BNB Smart Chain, and Base reached $2.57 billion in 2024, with one controller address managing $142.99 million in suspected activity in January 2024 alone.[8]

The mechanism is not purely on chain. Coordinated social narrative attacks amplifying FUD (fear, uncertainty, doubt) or manufactured positive sentiment drive retail investor behaviour, which drives price movement, which is then exploited by the coordinated actors. The attack surface spans on-chain activity, social platforms, community channels, and search.

Brand Attacks: When Fake Outrage Becomes Real Crisis

2024 was described by marketing executives and crisis communications professionals as the year fake news, misinformation and disinformation became the primary threat to brand integrity.[4] Major consumer brands discovered that coordinated attacks, even when eventually identified as inauthentic, produce real financial and reputational consequences before the correction can be communicated.

When 20% of the profiles driving a boycott hashtag are demonstrably fake, as Cyabra documented in the Jaguar and Coca-Cola cases, the remaining 80% of organic users have already been influenced by the manufactured consensus. Real users, unable to distinguish authentic sentiment from seeded narratives, adopt and amplify the framing. The damage is structural by the time the inauthentic accounts are identified.

The 6–12 hour window is not a technical curiosity. It is the interval during which an organisation can intervene before the attack achieves critical mass.

3. Case Studies

The following case studies draw on documented attack patterns observed in our research, public reporting, and AI Uniti platform detection data. Where case studies reference current clients, details have been anonymised or aggregated to protect commercial confidentiality. Case studies marked [FOR PRODUCTION] outline real attack typologies for which AI Uniti will produce fully anonymised client narratives as relationships and permissions are established.

Case Study 1 [PUBLISHED]: ASX-Listed Corporation — Financial Services

When Manufactured Sentiment Preceded an Earnings Announcement

The Threat: A coordinated network of 847 accounts began a coordinated amplification campaign 11 hours before a scheduled ASX earnings release. The accounts, created across a 6-week window with minimal prior history, began seeding negative sentiment using identical phrasing variations targeting the company's management credibility. Velocity spikes of 380% above baseline were detected within a 9-minute window on X and across two financial forums.

How AI Uniti Detected It: AI Uniti Signal's Coordination Risk Index flagged the attack at Hour 1 of the seeding phase. Behavioural pattern analysis identified 23 account clusters with synchronised burst activity and cross-platform phrase seeding. Escalation likelihood modelling predicted high probability of mainstream media pickup within 8–14 hours. The communications team was briefed 6 hours before the attack reached organic amplification.

Outcome: The company's prepared response released before the attack peaked was received as transparent and proactive rather than reactive. Analyst coverage framing was materially different from the manufactured narrative. Market impact was contained.

Case Study 2 [FOR PRODUCTION]: Web3 / Cryptocurrency Project

FUD Attack Coordinated Across Telegram and X During Token Launch

The Threat: A Web3 project preparing for a significant token launch detected anomalous activity 8 hours before launch. Coordinated FUD narratives alleging rug pull intent and fabricated technical vulnerabilities began appearing across Telegram communities and X simultaneously. Account analysis revealed 60+ accounts created within the same 4-hour window, using AI-generated profile images and first-post coordinated messaging.

How AI Uniti Detected It: AI Uniti Pulse Check identified coordination clusters using velocity analysis and phrase-seeding detection. The Coordination Risk

Index reached 87/100. Unite's autonomous response layer drafted moderation guidance and pre-approved counter-messaging for the project's community managers within 23 minutes of first detection.

Outcome: The token launch proceeded without significant price depression from the attack. The coordinated accounts were suspended by X within 6 hours of our platform report. A post-incident analysis confirmed the attack originated from a single IP cluster.

Case Study 3 [FOR PRODUCTION]: Political Campaign / Government Advisory

State-Adjacent Narrative Operation Targeting Candidate Credibility

The Threat: Six weeks before a regional election, a political campaign detected hundreds of accounts across X and Facebook simultaneously sharing a fabricated quote attributed to the candidate. Stanford Internet Observatory research on the 2024 US Election documented identical tactics, with state-adjacent networks using profile images stolen from LinkedIn accounts to create convincing fake personas.[9]

How AI Uniti Detected It: AI Uniti Signal's narrative theme clustering identified the fabricated quote as a seeded narrative within 90 minutes of first appearance. Cross-platform coordination mapping showed simultaneous activity from accounts exhibiting identical creation-date clusters and posting-behaviour patterns. Escalation likelihood modelling assessed high probability of pickup by partisan amplifiers within 4–6 hours.

Outcome: The campaign team issued a pre-emptive fact-check statement that was already in newsrooms before the narrative reached media threshold. The correction got ahead of the story.

Case Study 4 [FOR PRODUCTION]: Consumer Brand — FMCG Sector

Boycott Amplification by Inauthentic Network During Product Launch

The Threat: During the launch of a new product line, a consumer brand experienced what initially appeared to be organic consumer backlash. Standard sentiment monitoring flagged high negative volume but provided no mechanism to distinguish authentic consumer dissatisfaction from manufactured amplification. Cyabra documented the identical pattern in the TD Bank crisis: 24% of accounts driving the reputation event were fake, generating content with potential reach of 90,000 views.[10]

How AI Uniti Detected It: AI Uniti Pulse Check's bot-confidence scoring assigned high inauthentic probability to 31% of accounts in the negative discourse within 3 hours. Coordination cluster mapping revealed a hub-and-spoke amplification structure, with central accounts seeding content to peripheral accounts for redistribution.

Outcome: The brand's communications team briefed journalists on the coordinated nature of the attack before coverage was published, materially changing the framing of subsequent reporting. Authentic negative feedback was isolated, addressed, and used to improve the product.

Case Study 5 [FOR PRODUCTION]: Community Platform / Online Marketplace

Scam Infrastructure Targeting Community Trust Through Coordinated Fake Reviews

The Threat: FTC data for 2024 reported $12.5 billion in total fraud losses, with 70% of people contacted through social media reporting financial losses.[11] An online marketplace detected a coordinated network flooding their platform with fake positive reviews for fraudulent sellers while simultaneously filing false reports against legitimate vendors.

How AI Uniti Detected It: AI Uniti Pulse Check's repeat-offender tracking and coordination cluster mapping identified the account network across 14 days of historical data, revealing a coordinated operation with 200+ accounts operating in rotation to avoid per-account detection thresholds.

Outcome: The marketplace suspended the coordinated account cluster across a 48-hour window. Financial losses to platform users attributable to the fraud ring were estimated to have been prevented in the high six figures.

4. The Regulatory Imperative

Coordinated attack detection is no longer exclusively a strategic communications matter. A converging set of regulatory frameworks across Australia, the European Union, and the United States are beginning to formalise obligations around online safety, transparency, and the integrity of digital information environments.

Australia

The Online Safety Act 2021 (Cth) established the eSafety Commissioner's authority to hold online service providers accountable for user safety. The Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024 which was introduced in September 2024 has sought to impose transparency obligations on digital platforms and establish co-regulatory misinformation codes with enforceable civil penalty obligations.[12]

Phase 2 Online Safety Codes, registered in September 2025 and taking effect from March 2026, impose binding obligations on social media services, designated internet services, and relevant electronic services in turn creating direct accountability for platform-level behaviour that organisations must actively monitor.[13]

European Union

The EU Digital Services Act (DSA), in force since February 2024, compels brands and platforms to deliver full transparency and accountability regarding advertising practices, content moderation, and platform governance. For organisations with EU exposure, demonstrating active monitoring of coordinated inauthentic behaviour has become an expectation rather than a differentiator.

United States

The FBI's Operation Token Mirrors (2024) established the precedent that federal law enforcement will pursue criminal prosecution of coordinated market manipulation including manipulation conducted through social media narrative attacks that accompany on-chain manipulation. This creates regulatory exposure for organisations that do not demonstrate active monitoring of coordinated attacks affecting their financial instruments.

Regulatory Note: Australia's proposed Combatting Misinformation and Disinformation Bill 2024 did not pass Parliament in November 2024 due to cross-bench concerns about its scope. However, the regulatory trajectory toward greater platform accountability and enforceable transparency obligations is clear and bipartisan. Organisations should treat current voluntary frameworks as precursors to mandatory obligations, not as permanent alternatives.

5. From Monitoring to Action

The most common failure mode in coordinated attack response is not a failure of detection. Increasingly, it is a failure of action velocity. Organisations that receive an alert even if it is a well-constructed, early one still then face the challenge of translating that intelligence into a coordinated organisational response across communications, legal, risk, and social channels.

AI Uniti's three-tier product architecture is designed specifically to address this failure mode. The progression from detection (Pulse Check) to intelligence (Signal) to action (Unite) closes the full loop not just the detection gap.

Tier	Capability	What It Solves
Pulse Check	Real-time bot detection, coordination cluster identification, risk scoring, early-warning alerts, repeat-offender tracking, explainable flags	Detection: Identifies coordinated attacks 6–12 hours before standard monitoring tools detect sentiment shift
Signal	Multi-platform ingestion, narrative risk index, coordination risk index, escalation likelihood modelling, sentiment/mood-shift detection, board-ready reporting	Intelligence: Converts detection signals into board-level risk intelligence with escalation modelling and regulatory-grade evidence
Unite	Autonomous agentic response layer: intelligent content triage, AI-assisted response drafting, cross-team workflow automation, stakeholder notification triggers, decision transparency for regulators	Action: Closes the gap between alert and response so intelligence drives action automatically, without requiring manual escalation

The Agentic Advantage

Of the 20 direct competitors analysed in AI Uniti's Competitive Intelligence Framework (2026), only one competitor, has a partial agentic AI roadmap. No competitor closes the full detect–understand–act loop. This represents a genuine 12–18 month category leadership position.

Unite deploys a governed agentic layer, autonomous AI systems with human oversight, that transforms real-time intelligence into prioritised, pre-approved response actions:

- Triages incoming signals by severity and escalation probability
- Routes alerts to the correct internal stakeholders automatically (legal, comms, risk, executive)
- Drafts brand-voice-aligned response content for human review not for autonomous publishing
- Logs all decisions and evidence in formats suitable for regulatory review
- Maintains a full decision transparency trail for post-incident analysis

6. What You Should Do Now

Before evaluating any vendor or tool, we recommend leaders in communications, risk, and technology answer three diagnostic questions:

Question 1: Can your current monitoring tools distinguish between organic negative sentiment and manufactured consensus? If you cannot confidently answer yes with a demonstrable mechanism you are operating with a fundamental detection gap.

Question 2: What is your detection-to-action time? From the moment a coordinated attack begins seeding, how long before your team has confirmed intelligence, a prepared response, and active deployment? If that time exceeds 4 hours, you are likely to be responding to a crisis rather than preventing one.

Question 3: Do you have a Coordination Risk Score for your organisation right now? If you don't have a number and a monitoring mechanism behind it you cannot manage what you cannot measure.

The Phased Path to Coordinated Attack Resilience

1. **Phase 1 — Detect** Deploy Pulse Check across your primary social platforms. Establish baseline behavioural patterns. Generate your first Coordination Risk Index and understand your current exposure.
2. **Phase 2 — Understand** Upgrade to Signal for multi-platform ingestion, narrative theme clustering, and escalation likelihood modelling. Brief your board and risk committee on narrative risk as a financial risk category.
3. **Phase 3 — Act** Deploy Unite's agentic response layer. Pre-approve response frameworks for each attack typology. Run a full-scale coordinated attack simulation with your communications, legal, and executive team.

Conclusion: The Cost of Waiting

The coordinated attack targeting your organisation may not have started yet. Or it may have started 6 hours ago, and your monitoring tools simply haven't detected it.

This is the central, uncomfortable truth of the threat landscape in 2026. The tools most organisations rely on were designed for a different era, an era in which the primary social media threat was a disgruntled customer, a bad product review, or an off-message press statement.

The era we are in is different. Nation-state affiliated networks operate coordination attacks across five platforms simultaneously. Bot manufacturers sell wash-trading services at $2,000 upfront. Generative AI enables the production of convincing inauthentic content at industrial scale. The World Economic Forum has named this the #1 global risk two years running. The FBI has created its own cryptocurrency to catch the perpetrators.

And 93% of organisations are still monitoring what is being said not who is saying it, in what pattern, at what velocity, with what coordination.

The cost of waiting is not theoretical. It is the 6–12 hour window during which an attack establishes itself, sets the narrative, and converts inauthentic amplification into authentic spread. It is the $4.6 million average recovery cost of a social media incident. It is the 25% loss of market capitalisation that follows a narrative crisis, and the manufactured outrage that reached half a million views before a single brand-owned channel detected it as inauthentic.

AI Uniti was built to close that window. Pulse Check detects. Signal understands. Unite acts. Together, they represent the first complete, commercially available, tiered platform for coordinated attack detection and response purpose-built for the threat environment of 2025 and beyond.

Ready to understand your organisation's current Coordination Risk Score?
Book a 15-minute live demonstration of Pulse Check, Signal or Unite

aiuniti.com/request-demo

info@aiuniti.com

References & Citations

All references current as of February 2026.

Academic Research

1. Luceri, L. et al. (2025). "Exposing Cross-Platform Coordinated Inauthentic Activity in the Run-Up to the 2024 U.S. Election." ACM Web Conference 2025.
2. Cinelli, M. et al. (2025). "Coordinated Inauthentic Behaviour and Information Spreading on Twitter." arXiv:2503.15720. March 2025.

Industry & Regulatory Reports

3. World Economic Forum. (2025). Global Risks Report 2025, 20th Edition. Geneva: WEF. January 15, 2025.
4. Cyabra. (2025). "2024 Brand Crisis Round-Up." Cyabra Research. cyabra.com/blog/2024-brand-crisis-round-up-part-1/
5. Influencer Marketing Hub. (2026). "Social Media Security in 2026." influencermarketinghub.com/social-media-security/
6. Aon. (2025). "Damage to Reputation or Brand: A Critical Risk." Global Risk Management Survey 2025.
7. US Department of Justice / FBI. (2024). Operation Token Mirrors. Press release, October 9, 2024.
8. Chainalysis. (2025). "Crypto Market Manipulation 2025: Suspected Wash Trading." chainalysis.com/blog/crypto-market-manipulation-wash-trading-pump-and-dump-2025/
9. Stanford Internet Observatory. (2024). "How Coordinated Inauthentic Behaviour Continues on Social Platforms." June 2024.
10. Cyabra. (2025). "TD Bank Reputation Crisis: A Brand Disinformation Case Study." cyabra.com/blog/td-bank-reputation-crisis/
11. FTC Consumer Sentinel Network Data Book. (2024). ftc.gov/reports/consumer-sentinel-network
12. Australian Parliament. (2024). Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024.
13. eSafety Commissioner. (2025). Phase 2 Online Safety Codes. esafety.gov.au. September 9, 2025.
14. AI Uniti PTY LTD. (2025). Whitepaper Volume I: Coordinated Attack Detection — The Research Foundation. Internal publication.

Disclaimer

This whitepaper is produced by AI Uniti PTY LTD (ACN 694 238 821) for informational and commercial purposes. Case studies marked [FOR PRODUCTION] represent documented attack typologies for which full anonymised case narratives will be produced separately. Statistical references are attributed to their original published sources. AI Uniti makes no representation regarding the completeness of third-party data or the applicability of referenced regulations to any specific organisation's circumstances.