






Unmasking Digital Deception: Defending Your Organization Against AI-Powered Misinformation


 US panic-buying feeds into fuel shortage, sending gasoline prices higher

 Iran's Military Chief Says Coalition Needed to 'Resolve Palestinian Issue' Amid Raging Volence

 'Crimes that must be prosecuted': Germany vows 'zero tolerance' for 'anti-Semitic' attacks amid Israeli-Palestinian tensions

 Israeli airstrikes on Gaza resume as Tel Aviv thanks Biden administration for blocking UN statement calling for ceasefire

 Iran's Military Chief Says Coalition Needed to 'Resolve Palestinian Issue' Amid Raging Volence

 'Crimes that must be prosecuted': Germany vows 'zero tolerance' for 'anti-Semitic' attacks amid Israeli-Palestinian tensions

 'Crimes that must be prosecuted': Germany vows 'zero tolerance' for 'anti-Semitic' attacks amid Israeli-Palestinian tensions



Welcome!

I'm Nick Loui, Co-Founder & CEO of PeakMetrics

LinkedIn: <https://www.linkedin.com/in/nloui/>

X (Twitter): @nloui



**“A lie can travel halfway
around the world while the
truth is still putting on its
shoes.”**

Not Just a Government Problem Anymore



A forged DoD memo stated that Broadcom's acquisition of CA prompted national security concerns, causing the stocks of both companies to fall.



Users of 4Chan spread a rumor that the coffee giant was giving free drinks to undocumented immigrants, forcing the company to respond.



When Lapsus\$ hacked Okta, they were not inside Okta's network as they claimed. They had limited access and created a social media campaign to exaggerate the attack.



A Chinese influence operation tried to mobilize U.S. protests against an Australian rare earths mining company planning an expansion in Texas



A deepfake of an explosion at the Pentagon briefly rippled through the stock market after getting spread by RT (Russia Today) and a fake "verified" Bloomberg News account

Misinformation vs. Disinformation

Disinformation

Content that is intentionally false and designed to cause harm. It is motivated by three factors: to make money; to gain political influence, either foreign or domestic; or to cause trouble for the sake of it.

Misinformation

Misinformation also describes false content, but the person sharing doesn't realize that it is false or misleading. Often a piece of disinformation is picked up by someone who doesn't realize it's false and that person shares it with their networks, believing that they are helping.

Disinformation for Hire: The Dark Side of the Information Economy

Fake News Stories

Recontextualized Stories

Deep Fakes

Troll Farms

Bots

Fake News Sites

Influencers

Brand / Reputation Damage

Loss of Trust

Financial Loss

Operational Disruption

Cybersecurity

Societal Disruption

Legal / Regulatory

▶ TR / 01 ▶ 03
▶ TR / 01 ▶ 03

▶ SEARCH ▶ 01 ▶ 03
▶ SEARCH ▶ TR / 01 ▶ 03

▶ RS / 011

Disinformation as a Product



Cabin Crew Take Secret Pictures, You Won't Believe The Results
(Time To Break)



20 Celebrities Who Have Beaten Cancer
(Celebstars)



The Must-See Technology That Is Changing the Way You Listen...
(Sound Online by Sony)



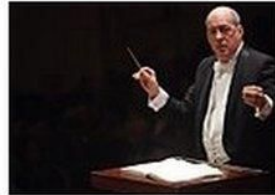
15 Reasons Why You Should Never Support Horse Racing
(Sports Mozo)



44 Stunning Images Of Things You Had No Idea Existed (Pics)



Life Insurance Companies Hate This New Trick

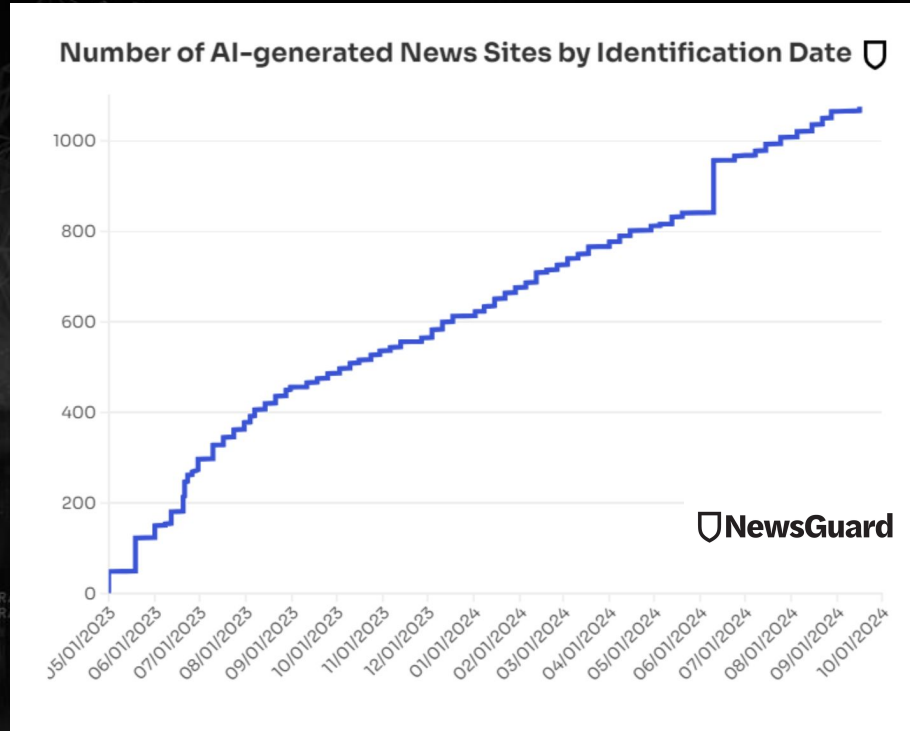


New Leader for Atlanta Symphony
(ArtsBeat)



Fiji send England Rugby World Cup warning ahead of...

The AI Effect



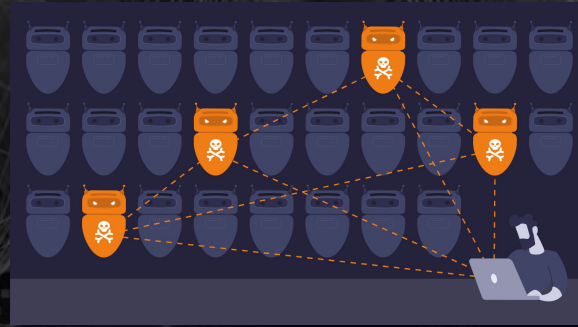
Disinformation has become a critical aspect of many cyber-attacks

1. Disinformation is less expensive than traditional cyberattacks
2. It's costly for companies
3. Anyone can spread it
4. It spreads quickly

1. Misinformation & Disinformation Are Cheap

Disinformation is an inexpensive way to launch an attack on a business.

Anyone who wants to can actually purchase a disinformation campaign against your company, complete with fake news and misinformation that can be quickly spread.



\$78 Billion.

2. It's Costly For Companies

With generative AI advancing, disinformation is now more sophisticated and widespread, making it crucial to detect and combat AI-driven threats before they cause damage to your bottom line.



Eli Lilly and Company 
@EliLillyandCo

We are excited to announce insulin is free now.

1:36 PM · 11/10/22 · [Twitter for iPhone](#)

554 Retweets **171** Quote Tweets **3,324** Likes

Eli Lilly: A tweet sent by a Twitter account impersonating Eli Lilly & Co. said, "insulin is free now," causing the company's stock to drop over 4% and leading the company to suspend all activity and advertising on Twitter.

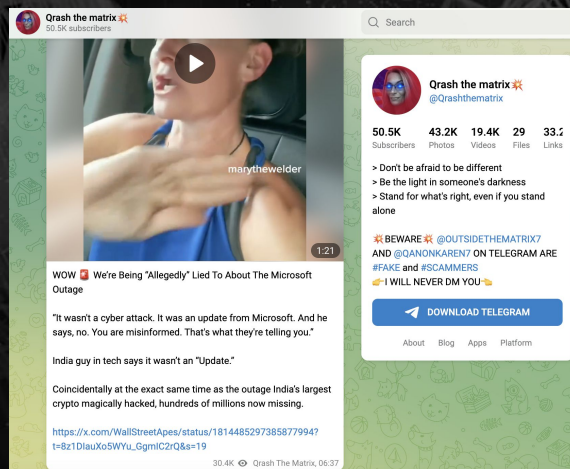
RS / 011
RS / 011

RS / 0211R / DN
RS / 0211R / DN

3. Anyone Can Spread It

One of the biggest risks of disinformation is that it can come from anyone, anywhere.

Epic Games: The group Mogilevich on Telegram claimed to have hacked Epic Games and stolen 189GB of data, which Epic quickly denied. Mogilevich eventually admitted the claims were false, calling themselves “professional fraudsters” looking for quick cash.



Microsoft: Last week, a conspiracy theory about Microsoft began circulating on Telegram before spreading to other platforms, like X. The narrative claimed, we're being lied to about a Microsoft update, suggesting that the outage was not due to an update, as officially stated, but a cyberattack. The claim is from a very “good friend” in India.

4. It Spreads Quickly

Once a disinformation campaign has started, it's hard to put a stop to it.



SEC: On January 9, The Securities and Exchange Commission's official X account was hacked, falsely announcing the approval of bitcoin ETFs.

Bitcoin briefly surged to nearly \$48,000 before SEC chair Gary Gensler clarified the post was unauthorized, stating that no bitcoin ETFs had been approved. The SEC later confirmed its account had been compromised.

Is Your Organization a Target?

High Visibility

**Taking a Public
Stance**

**Undergoing a
Major Deal**

**A Strong Social
Media Presence**

New Products

**Broader
Industry Issues**

**Sensitive Data /
Regulated
Industry**

»RS/ 0211TR / DN
»RS/ 0211TR / DN

DISARM: The Foundation for Cognitive Security

DISARM Red Framework - incident creator TTPs															
PLAN			PREPARE					EXECUTE							ASSESS
TA01: Plan Strategy	TA02: Plan Objectives	TA13: Target Audience Analysis	TA14: Develop Narratives	TA06: Develop Content	TA15: Establish Social Assets	TA16: Establish Legitimacy	TA05: Microtarget	TA07: Select Channels	TA08: Conduct	TA09: Deliver	TA17: Maximize	TA18: Drive	TA10: Drive	TA11: Persist in	TA12: Assess

DISARM Blue Framework - responder TTPs															
TA01: Plan Strategy	TA02: Plan Objectives	TA05: Microtarget	TA06: Develop Content	TA07: Select Channels and Affordances	TA08: Conduct Pump Priming	TA09: Deliver Content	TA11: Persist in the Information Environment	TA12: Assess Effectiveness	TA15: Establish Social Assets						
T0073: Determine Target Audiences	T0002: Facilitate State Propaganda	T0072: Segment Audiences	T0003: Leverage Existing Narratives	T0016: Create hashtags and search artifacts	T0007: Create authentic Social Media Pages and Groups	T0009: Create fake experts	T001: Create Click								
T0074: Determine Strategic Ends	T0066: Degrade Adversary	T0072.001: Geographic Segmentation	T0004: Develop Competing Narratives	T0019: Generate information pollution	T0010: Cultivate ignorant agents	T0009.001: Utilize Academic/Pseudoscientific Justifications	T001: Purch Target Advertis								
T0075: Dismiss	T0072.002: Demographic Segmentation	T0022: Leverage Conspiracy Theory Narratives	T0019.001: Create fake research	T0013: Create inauthentic websites	T0011: Compromise legitimate accounts	T001: Create Local Cont									
T0075.001: Discredit Credible Sources	T0072.003: Economic Segmentation	T0022.001: Amplify Existing Conspiracy Theory Narratives	T0019.002: Hijack Hashtags	T0014: Prepare fundraising campaigns	T0097: Create personas	T010: Level: Ech Chamber Bubb									
T0076: Distort	T0072.004: Psychographic Segmentation	T0022.002: Develop Original Conspiracy Theory Narratives	T0023: Distort facts	T0014.001: Raise funds from malign actors	T0097.001: Backstop personas	T0102: Use exist Ech Chamber Bubb									
T0077:	T0072.005: Political	T0040: Demand	T0023.001: Reframe	T0014.002: Raise funds	T0098: Establish inauthentic	T0102: Create F									
C00016: Censorship	C00207: Run a competing disinformation campaign - not recommended	C00065: Reduce political targeting	C00085: Mute content	C00195: Redirect searches away from disinformation or extremist content	C00117: Downgrade / de-amplify so message is seen by fewer people	C00147: Make amplification of social media posts expire (e.g. can't like/retweet after n days)	C00138: Spam domestic actors with lawsuits	C00140: "Bomb" link shorteners with calls	C00040: third party verification for people						
C00017: Repair broken social connections	C00164: compatriot policy	C00066: Co-opt a hashtag and drown it out (hijack it back)	C00014: Real-time updates to fact-checking database	C00098: Revocation of allowed/ "verified" status	C00119: Engage payload and debunk.	C00128: Create friction by marking content with ridicule or other "decelerants"	C00139: Weaponise youtube content matrices	C00148: Add random links to network graphs	C00059: Verification of project before posting fund requests						
C00019: Reduce effect of division-enablers	C00092: Establish a truth teller reputation score for influencers	C00178: Fill information voids with non-disinformation content	C00032: Hijack content and link to truth- based info	C00105: Buy more advertising than misinformation creators	C00120: Open dialogue about design of platforms to produce different outcomes	C00129: Use banking to cut off access	C00131: Seize and analyse botnet servers	C00149: Poison the monitoring & evaluation data	C00058: Report crowdfunder as violator						
C00021: Encourage in-person communication	C00222: Tabletop simulations	C00216: Use advertiser controls to stem flow of funds to bad actors	C00071: Block source of pollution	C00103: Create a bot that engages / distract trolls	C00121: Tool transparency and literacy for channels people follow.	C00182: Redirection / malware detection/ remediation	C00143: (botnet) DMCA takedown requests to waste group time	C00172: social media source removal							
C00022: Inoculate. Positive campaign to promote feeling of safety	C00070: Block access to disinformation resources	C00130: Mentorship: elders, youth, credit. Learn vicariously.	C00072: Remove non-relevant content from special interest groups - not recommended	C00101: Create friction by rate-limiting engagement	C00112: "Prove they are not an op!"	C00200: Respected figure (influencer) disavows misinfo		C00056: Encourage people to leave social media							
C00066: Charge for social media	C00169: develop a creative content hub	C00074: Identify and delete or rate limit identical content	C00097: Require use of verified identities to contribute to poll or comment	C00100: Hashtag jacking	C00109: Dampen Emotional Reaction			C00053: Delete old accounts / Remove unused social media accounts							
C00024: Promote healthy narratives	C00060: Legal action against for-profit engagement factories	C00075: normalise language	C00099: Strengthen verification methods	C00154: Ask media not to report false information	C00211: Use humorous counter-narratives			C00052: Infiltrate platforms							

What is Narrative Intelligence?

Narrative Intelligence

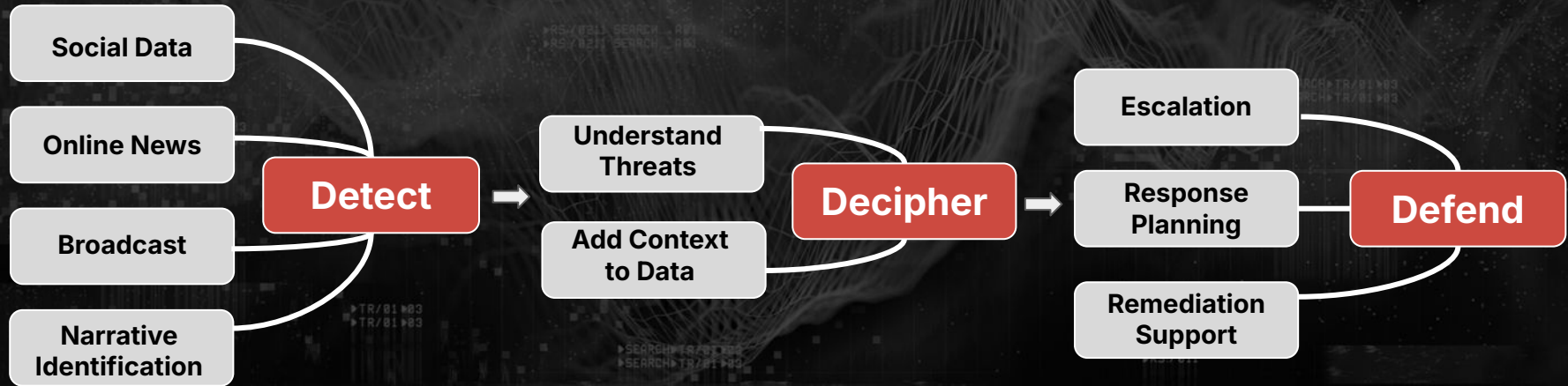
The strategy governments and organizations use to detect, decipher, and defend their reputation from AI-powered narrative threats, like misinformation, disinformation, deepfakes, and bots.

Narrative Attacks

Target a brand's reputation by spreading false or misleading narratives across social media, news outlets, and other platforms, like the deep web.

Detect. Decipher. Defend. Framework

A customizable framework for cyber security teams to proactively understand and defend against emerging online narrative threats like misinformation, disinformation and deepfakes.

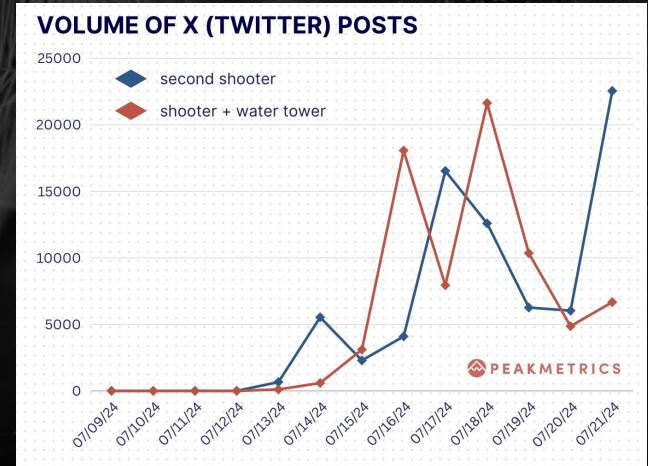


Detect

Timing is critical for cyber security teams. Proactive detection allows teams to get ahead of issues through early identification and helps to understand if digital deception is at play.

- Always on intelligence feed
- Monitor fringe platforms
- Team or tool to identify narratives taking shape

Politics: The day after the assassination attempt on Trump, YouTuber and musician Ryan Upchurch posted on TikTok, questioning why the authorities were "covering up the water tower." His post sparked others to share Google Earth images of the Butler County fairgrounds. These images were later picked up by conspiracy accounts like SGT News on Telegram and John Cullen on X, who alleged the involvement of a second shooter.



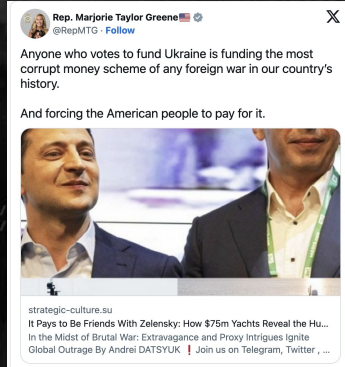
Decipher

Add context to gain a complete understanding and assess threat levels.

- Check the source credibility.
- Uncover the origin, key authors and domains driving the conversation.
- Monitor languages to understand the conversation across regions.
- Develop a list of top factors to quantify the potential impact of a narrative threat.
- Look at social media follower count as a way to discover bot-like-activity.

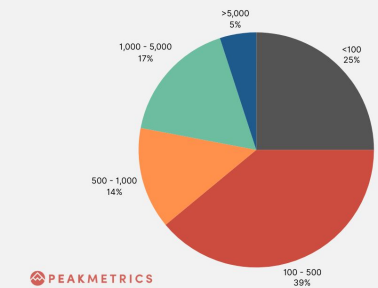
All these factors help to determine if it's a threat you need to defend against, or a passing fad.

Politics: Last November, Rep Marjorie Taylor Greene sparked backlash after sharing Russian Propaganda claiming Ukrainian President Zelensky's closest associates bought two yachts. Knowing the origin of a post is critical to understanding the credibility.



Olympics: Ahead of the Paris Olympics, Iran was online amplifying calls to ban Israel from the games. PeakMetrics examined indicators of bot-like activity in the 50,000+ mentions of the #BanIsrael narrative. 25% of these posts were from accounts with less than 100 followers.

**BAN ISRAEL FROM THE OLYMPICS:
FOLLOWERS PER ACCOUNT SHARING NARRATIVE**



Defend

Cybersecurity professionals are the first line of defense, detecting risks and assessing threat levels. Once a narrative is identified as a threat, the next step is your defense strategy.

- Set up an internal alert process
- Communicate the determined threat level and perceived risks.
- Know when to loop in your Communications team with a repeatable response plan to counter threats.
- Provide clarity on the truth.
- Issue takedown notices when necessary

Future Proofing: Why This Matters

It's the age of digital transformation. And it's accelerating.

Customers reward security champions. As new threat vectors, like misinformation and disinformation become more sophisticated, having a plan in place to identify and combat risks is critical.



Thank You!



LinkedIn:

<https://www.linkedin.com/in/nloui/>