TOWARDS THE USE OF THE DISARM FRAMEWORK AS A LEVER FOR RAISING AWARENESS AND PROTECTING AUDIENCES TARGETED BY DISINFORMATION OPERATIONS.

Dr. Jean-Philippe RIANT – DATASK – PEC – Reserve Officer

Ugo PEYRE – DATASK – Etudiant M2 Sciences Po Rennes

jp@riant.fr -
ugopeyre42@gmail.com

C&ESAR 2024 by DGA

31st Computer & Electronics Security Application Rendezvous

21 novembre 2024

CONTEXT AND IMPORTANCE

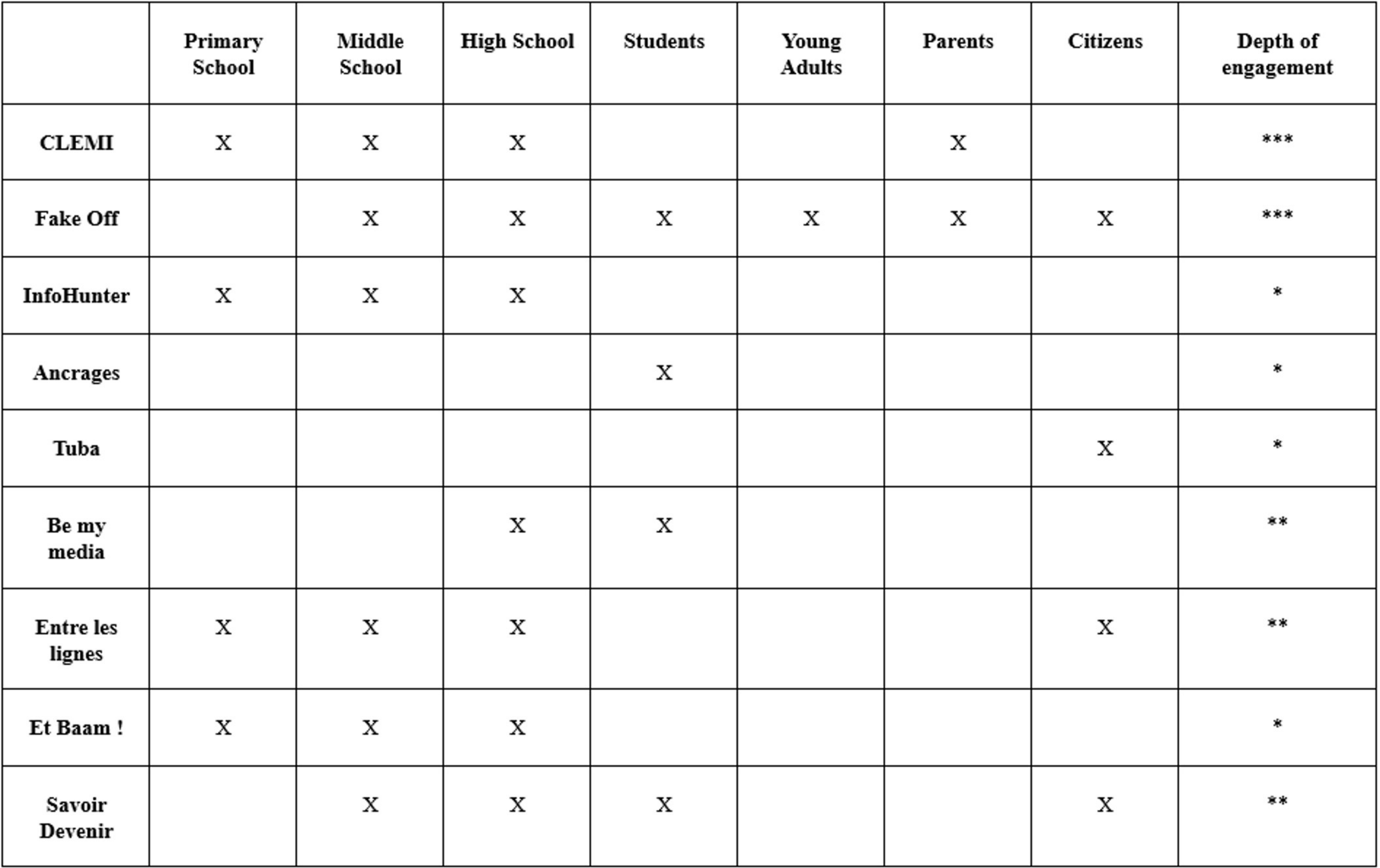
- Disinformation manipulates information to influence decisions and create instability.
- It is increasingly used in geopolitics and hybrid warfare strategies.
- Our article aims to explore countermeasures and build resilience at the individual scale level using DISARM Framework as a scenario engine.

THE ROLE OF THE INDIVIDUAL

- Disinformation targets individual biases and emotions to spread.
- Collective individual behaviors amplify its societal impact.
- Each person acts as a node in networks of information and disinformation.

PROTECTION MEASURES, FR & EU

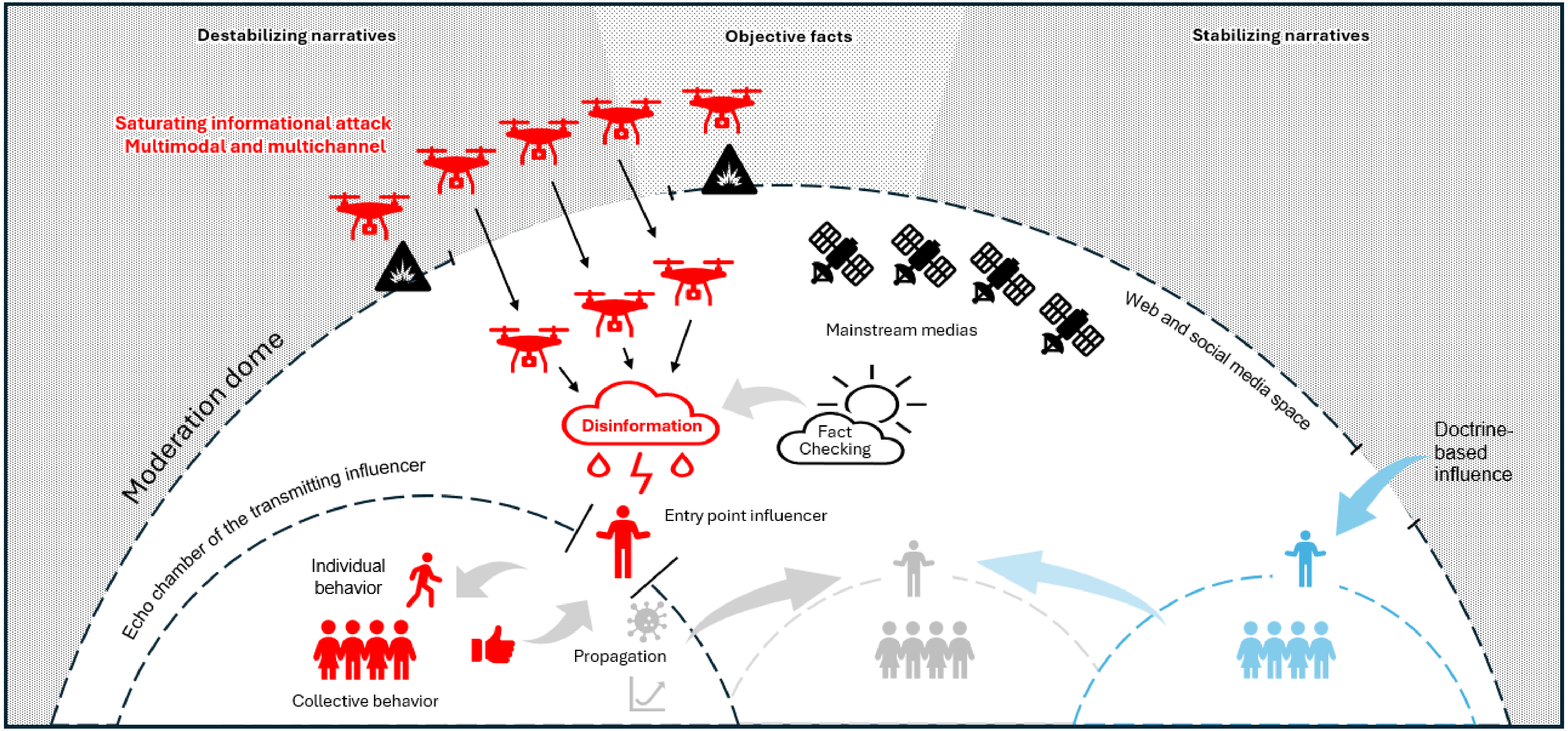
- Media education is the primary tool against disinformation in France.
- Implementation varies across regions, leading to geographic disparities.
- Finland offers a model of systematic media literacy for other nations.



	Primary School	Middle School	High School	Students	Young Adults	Parents	Citizens	Depth of engagement
CLEMI	X	X	X			X		***
Fake Off		X	X	X	X	X	X	***
InfoHunter	X	X	X					*
Ancrages				X				*
Tuba							X	*
Be my media			X	X				**
Entre les lignes	X	X	X				X	**
Et Baam !	X	X	X					*
Savoir Devenir		X	X	X			X	**

INFORMATION ATTACKS: FROM GLOBAL TO INDIVIDUAL

- Disinformation attacks occur globally, systemically, and individually.
- Platforms struggle with moderating harmful content effectively.
- Layered defenses, like the Iron Dome analogy, are essential for resilience.



Destabilizing narratives

Objective facts

Stabilizing narratives

Saturating informational attack
Multimodal and multichannel

Moderation dome

Mainstream medias

Web and social media space

Disinformation

Fact Checking

Doctrine-based influence

Echo chamber of the transmitting influencer

Entry point influencer

Individual behavior

Collective behavior

Propagation

Community sensitive to disinformation

Other communities

Stabilizing communities

Scale of global actors' strategies

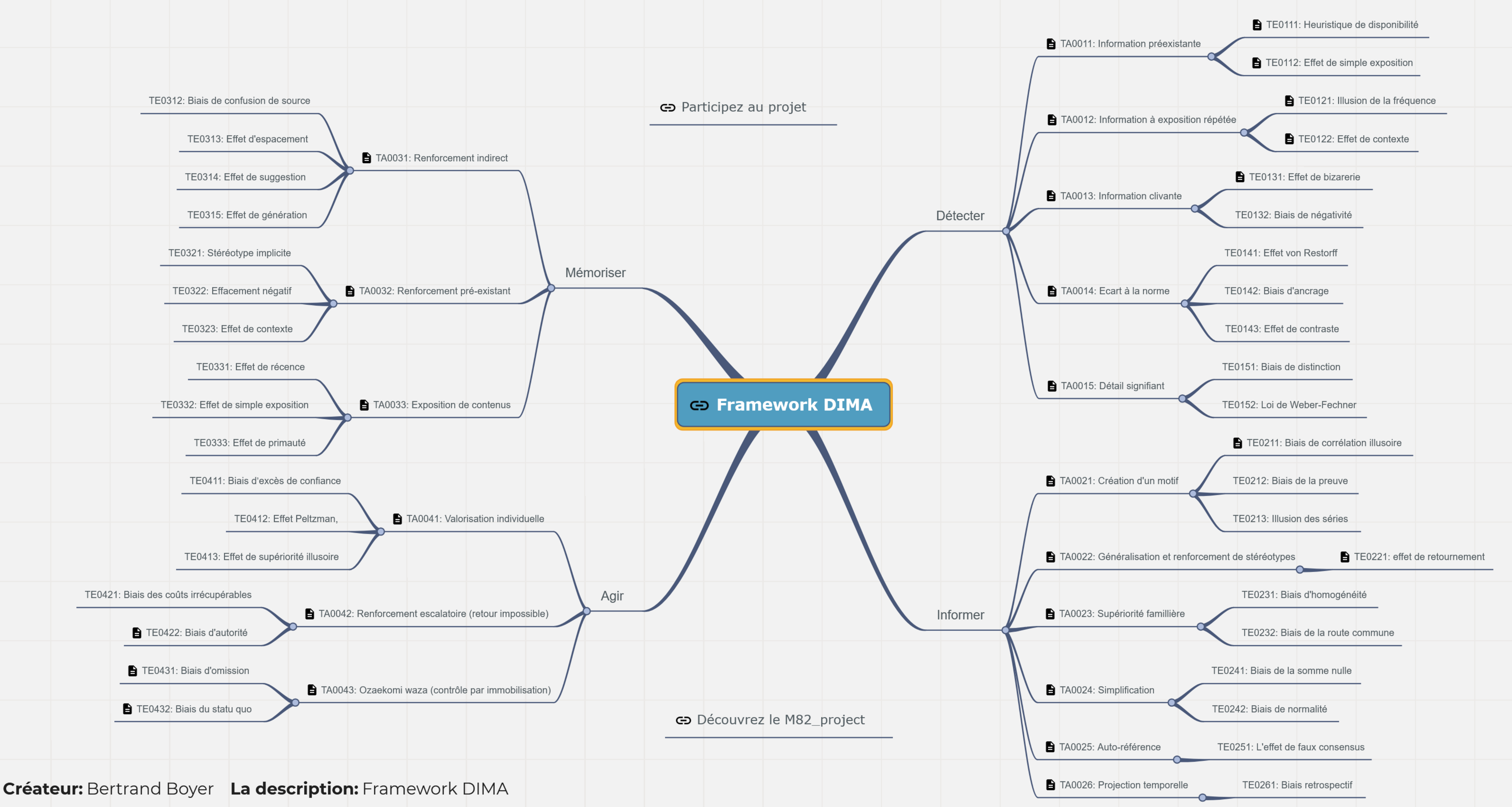
Scale of systemic defense and broadcasters

Scale of influence

Scale of the individual

COGNITIVE MECHANISMS

- Confirmation bias and emotional responses drive disinformation's spread.
- Sensational falsehoods are more memorable than bland truths.
- Repeated exposure to false information makes it seem credible.

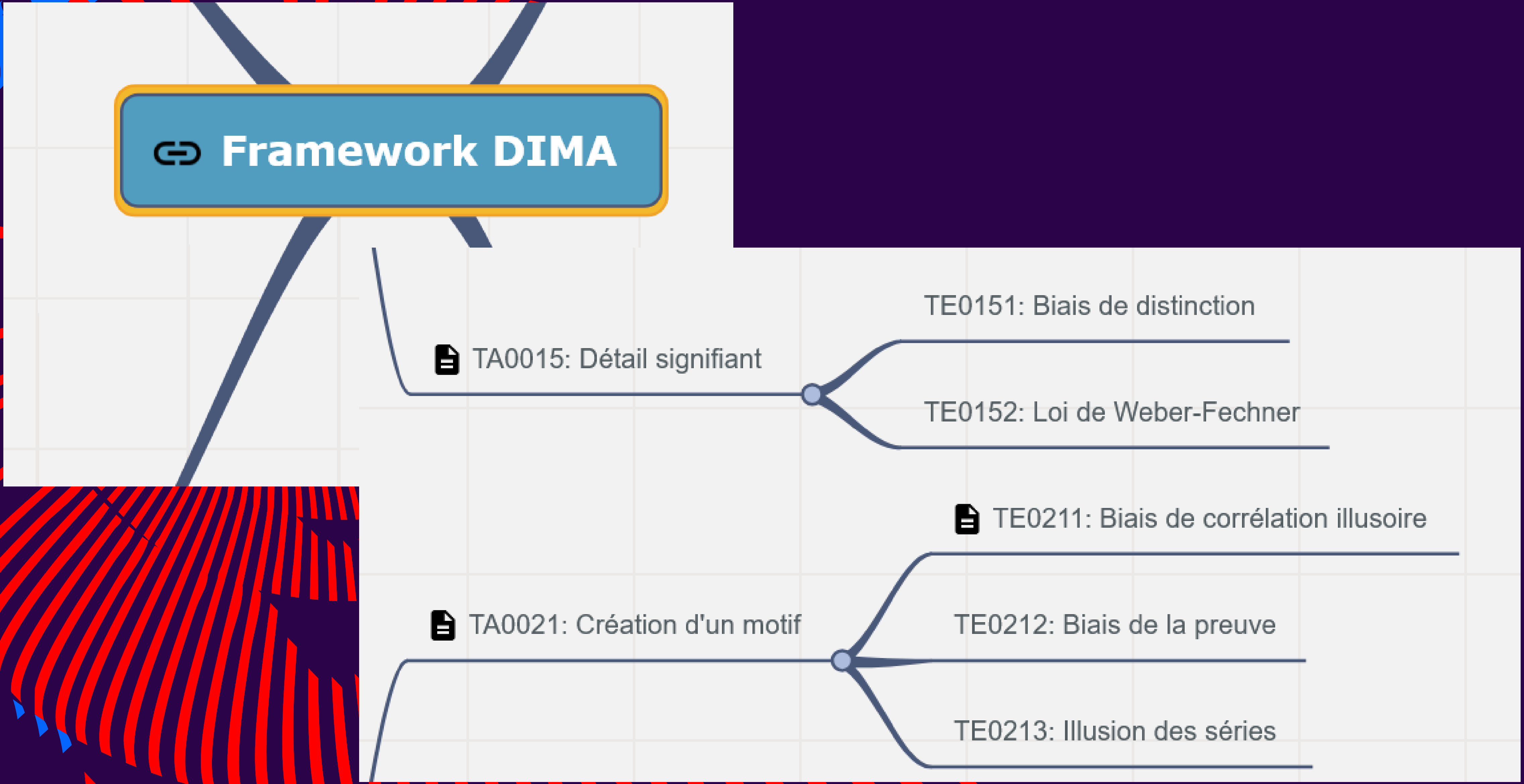


Participez au projet

Découvrez le M82_project

- Framework DIMA
  - Détecter
    - TA0011: Information préexistante
      - TE0111: Heuristique de disponibilité
      - TE0112: Effet de simple exposition
    - TA0012: Information à exposition répétée
      - TE0121: Illusion de la fréquence
      - TE0122: Effet de contexte
    - TA0013: Information clivante
      - TE0131: Effet de bizarerie
      - TE0132: Biais de négativité
    - TA0014: Ecart à la norme
      - TE0141: Effet von Restorff
      - TE0142: Biais d'ancrage
      - TE0143: Effet de contraste
    - TA0015: Détail signifiant
      - TE0151: Biais de distinction
      - TE0152: Loi de Weber-Fechner
  - Informer
    - TA0021: Création d'un motif
      - TE0211: Biais de corrélation illusoire
      - TE0212: Biais de la preuve
      - TE0213: Illusion des séries
    - TA0022: Généralisation et renforcement de stéréotypes
      - TE0221: effet de retournement
    - TA0023: Supériorité familière
      - TE0231: Biais d'homogénéité
      - TE0232: Biais de la route commune
    - TA0024: Simplification
      - TE0241: Biais de la somme nulle
      - TE0242: Biais de normalité
    - TA0025: Auto-référence
      - TE0251: L'effet de faux consensus
    - TA0026: Projection temporelle
      - TE0261: Biais retrospectif
  - Mémoriser
    - TA0031: Renforcement indirect
      - TE0312: Biais de confusion de source
      - TE0313: Effet d'espacement
      - TE0314: Effet de suggestion
      - TE0315: Effet de génération
    - TA0032: Renforcement pré-existant
      - TE0321: Stéréotype implicite
      - TE0322: Effacement négatif
      - TE0323: Effet de contexte
    - TA0033: Exposition de contenus
      - TE0331: Effet de récence
      - TE0332: Effet de simple exposition
      - TE0333: Effet de primauté
  - Agir
    - TA0041: Valorisation individuelle
      - TE0411: Biais d'excès de confiance
      - TE0412: Effet Peltzman,
      - TE0413: Effet de supériorité illusoire
    - TA0042: Renforcement escalatoire (retour impossible)
      - TE0421: Biais des coûts irrécupérables
      - TE0422: Biais d'autorité
    - TA0043: Ozaekomi waza (contrôle par immobilisation)
      - TE0431: Biais d'omission
      - TE0432: Biais du statu quo

Créateur: Bertrand Boyer **La description:** Framework DIMA



Framework DIMA

TA0015: Détail signifiant

TE0151: Biais de distinction

TE0152: Loi de Weber-Fechner

TA0021: Création d'un motif

TE0211: Biais de corrélation illusoire

TE0212: Biais de la preuve

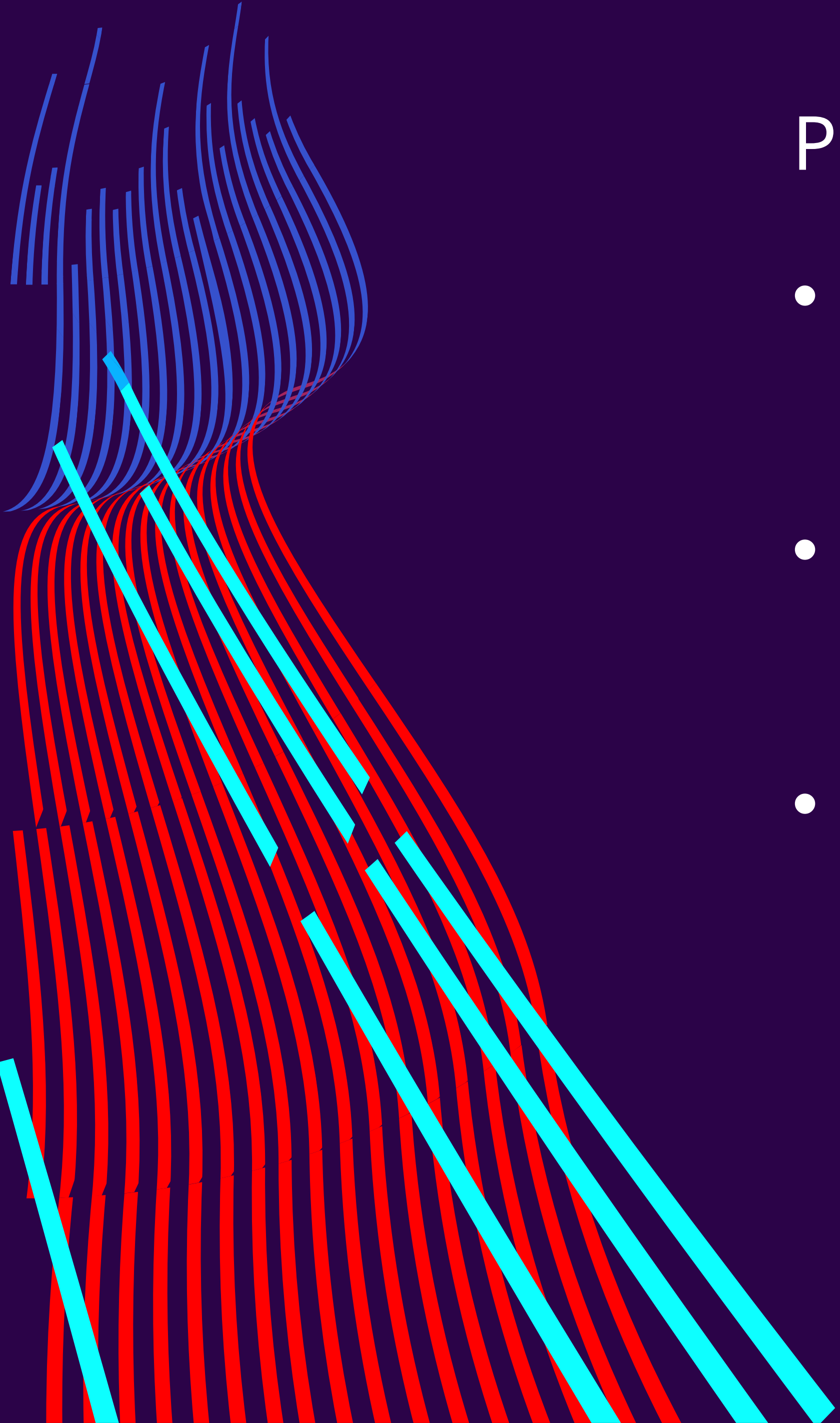
TE0213: Illusion des séries

SOCIAL MEDIA DYNAMICS

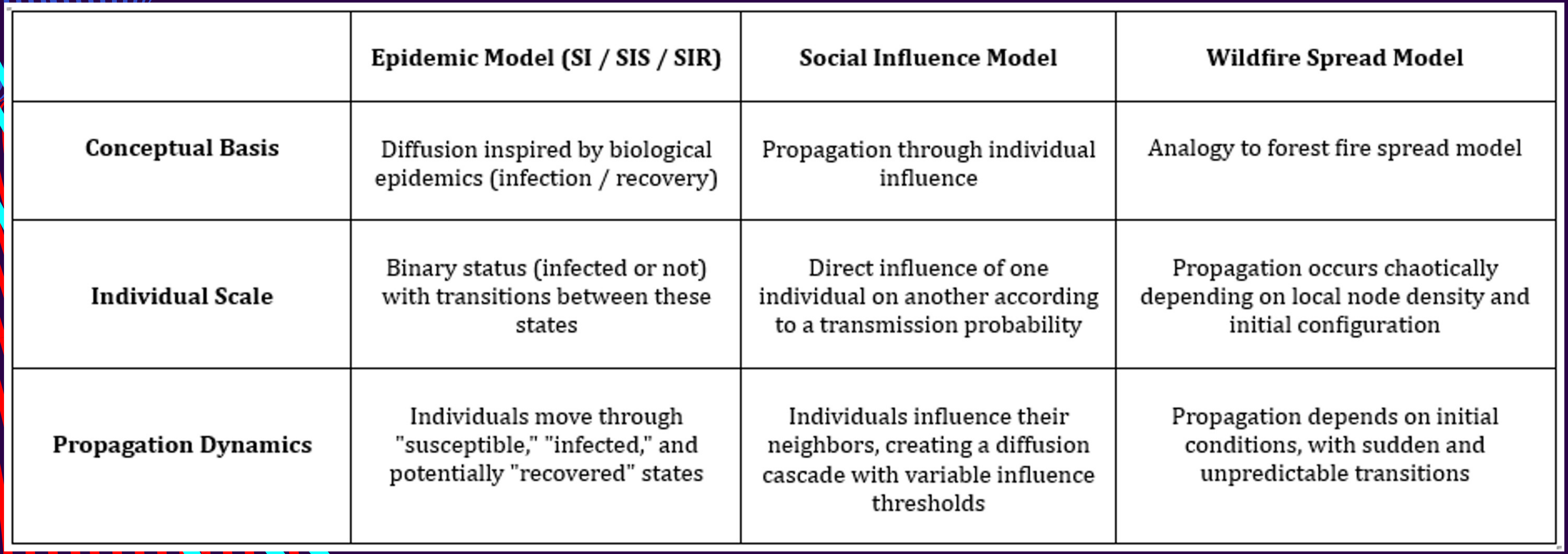
- Social media creates echo chambers that reinforce existing beliefs.
- Influencers act as opinion leaders, sharers, or activists in disinformation campaigns.
- Platforms' dopamine-driven reward systems encourage rapid sharing.

PROPAGATION MECHANISMS

- Epidemic models categorize users as susceptible, infected, or immune.
- Wildfire models show how disinformation spreads explosively through networks.
- The speed and reach of disinformation depend on individual behaviors.



PROPAGATION MECHANISMS

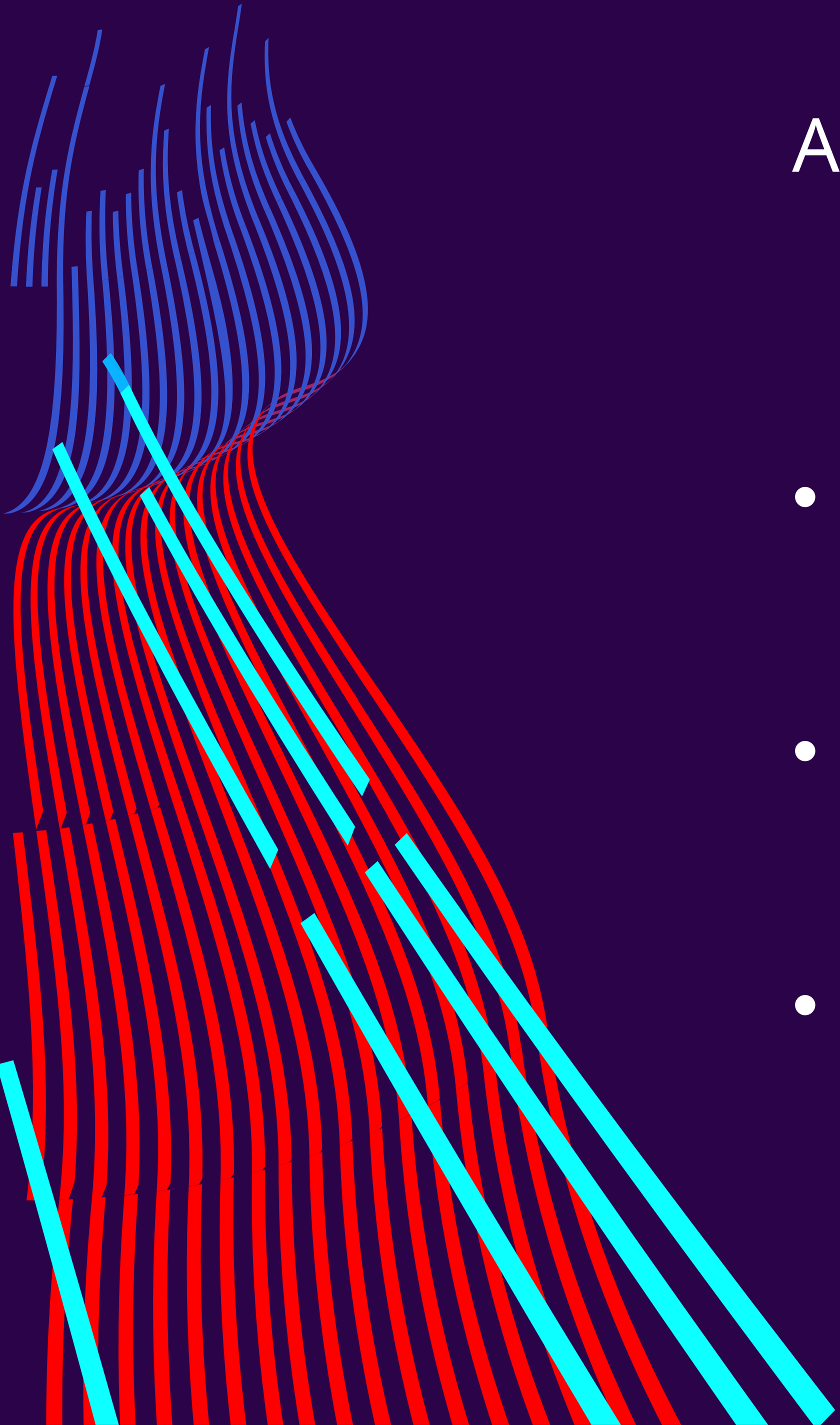


	Epidemic Model (SI / SIS / SIR)	Social Influence Model	Wildfire Spread Model
Conceptual Basis	Diffusion inspired by biological epidemics (infection / recovery)	Propagation through individual influence	Analogy to forest fire spread model
Individual Scale	Binary status (infected or not) with transitions between these states	Direct influence of one individual on another according to a transmission probability	Propagation occurs chaotically depending on local node density and initial configuration
Propagation Dynamics	Individuals move through "susceptible," "infected," and potentially "recovered" states	Individuals influence their neighbors, creating a diffusion cascade with variable influence thresholds	Propagation depends on initial conditions, with sudden and unpredictable transitions

THE INDIVIDUAL AS A NODE

- Individuals serve as nodes that spread or disrupt disinformation.
- Social validation mechanisms amplify false narratives.
- Active users play a critical role in network-wide propagation dynamics.

AWARENESS LEVERS



- Traditional media education alone is insufficient to combat disinformation.
- Innovative tools like simulations and games enhance learning outcomes.
- Digital environments offer opportunities for immersive awareness-building.

DISARM FRAMEWORK: OVERVIEW

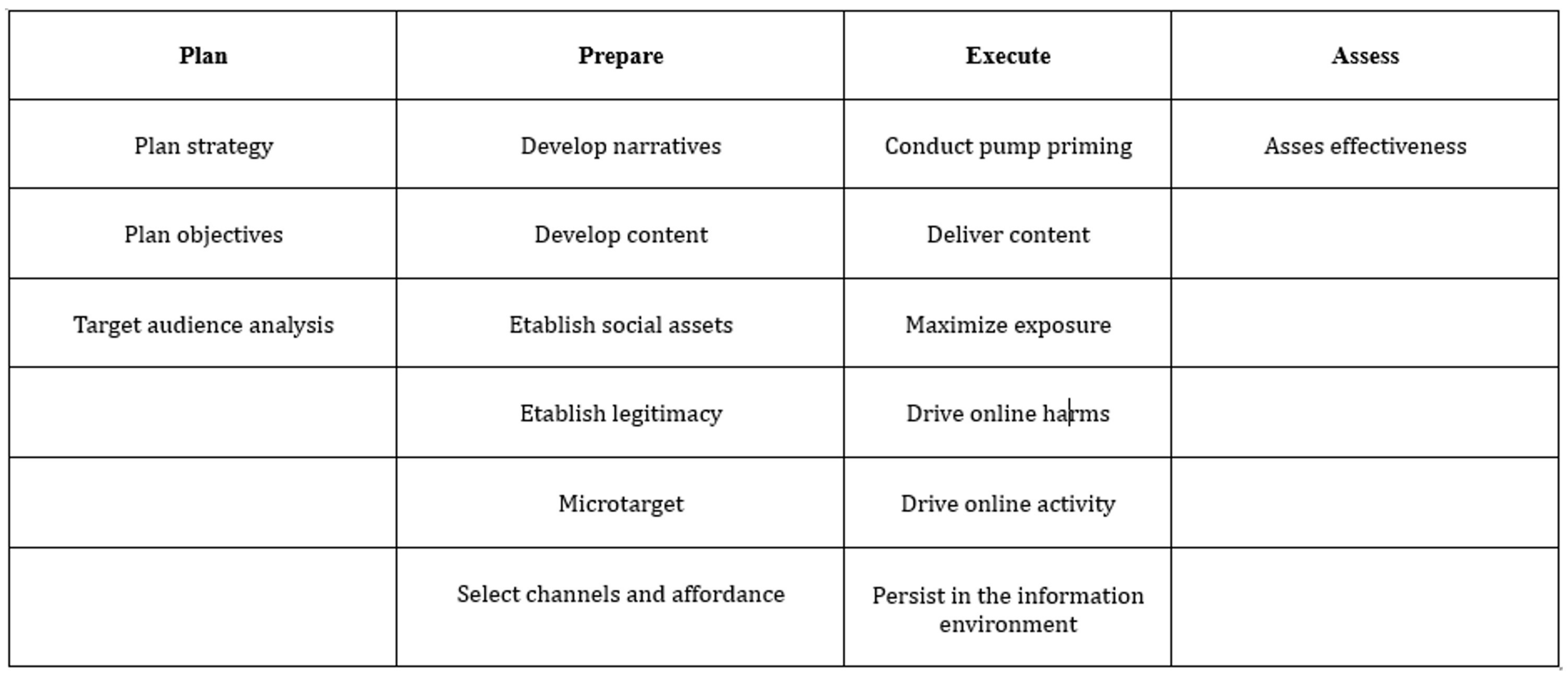
- DISARM’s red matrix maps attacker strategies; the blue matrix suggests defenses.

- Its comprehensive taxonomy simplifies complex disinformation operations, with a timeline.

- DISARM supports collaboration across fields and adapts to emerging threats.

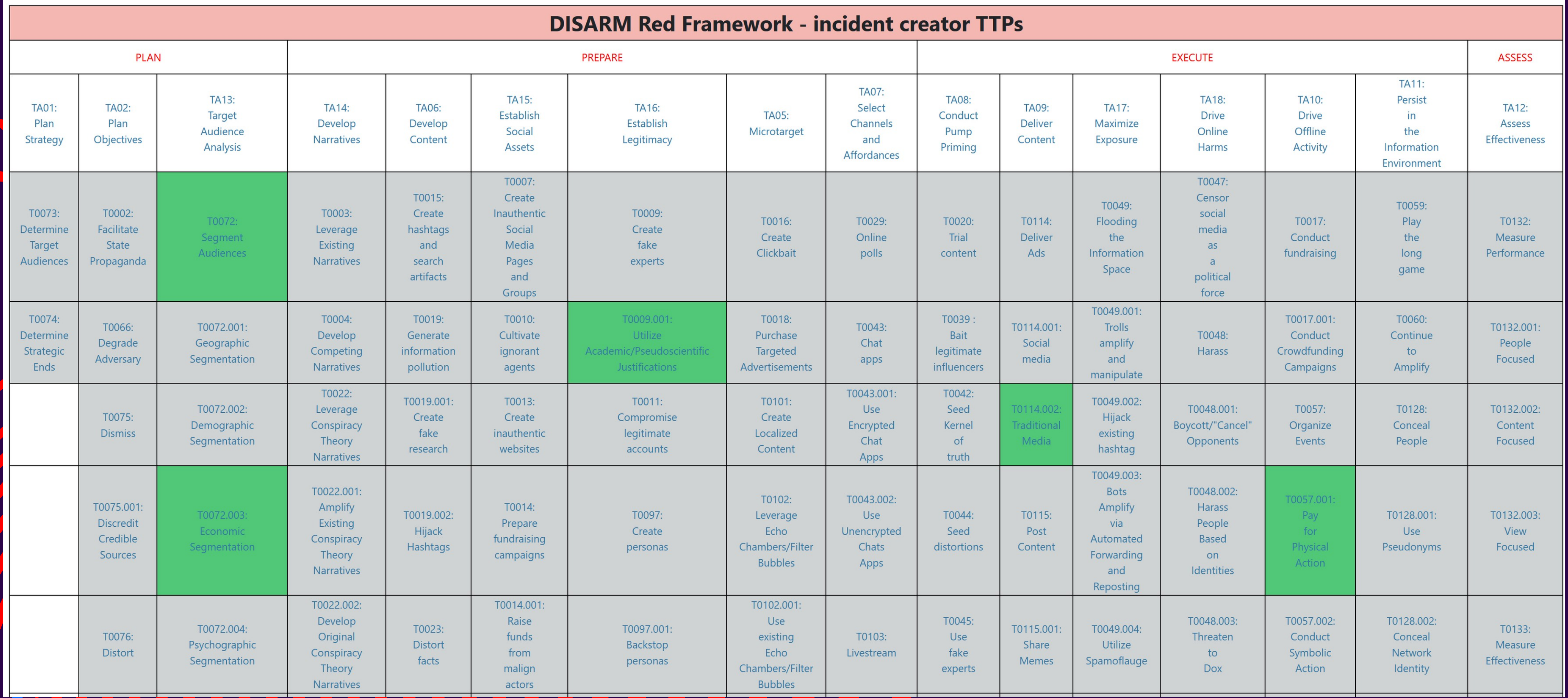
DISARM FRAMEWORK RED

TIME



Plan	Prepare	Execute	Assess
Plan strategy	Develop narratives	Conduct pump priming	Asses effectiveness
Plan objectives	Develop content	Deliver content	
Target audience analysis	Etablish social assets	Maximize exposure	
	Etablish legitimacy	Drive online harms	
	Microtarget	Drive online activity	
	Select channels and affordance	Persist in the information environment	

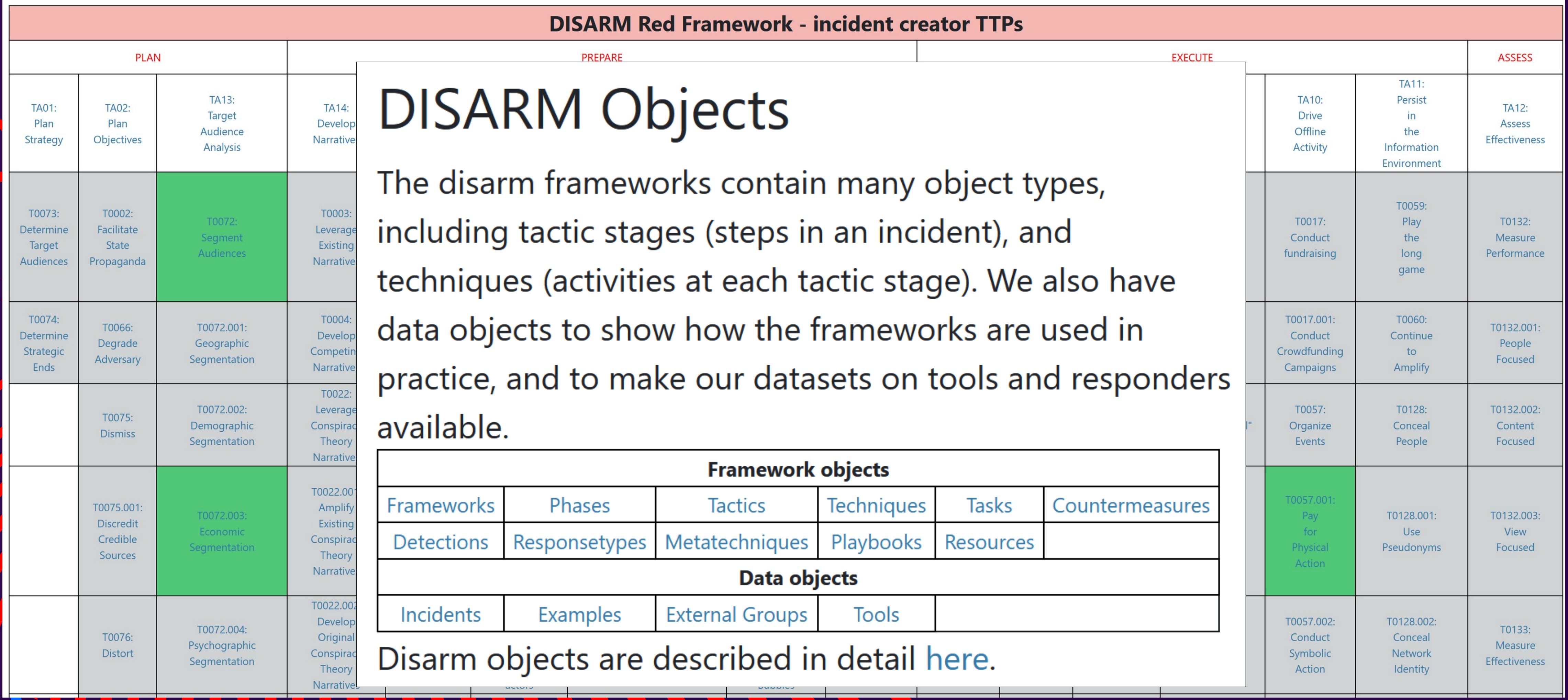
DISARM FRAMEWORK RED



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 and Affordances	TA08: Conduct Pump Priming	TA09: Deliver Content	TA17: Maximize Exposure	TA18: Drive Online Harms	TA10: Drive Offline Activity	TA11: Persist in the Information Environment	TA12: Assess Effectiveness
T0073: Determine Target Audiences	T0002: Facilitate State Propaganda	T0072: Segment Audiences	T0003: Leverage Existing Narratives	T0015: Create hashtags and search artifacts	T0007: Create Inauthentic Social Media Pages and Groups	T0009: Create fake experts	T0016: Create Clickbait	T0029: Online polls	T0020: Trial content	T0114: Deliver Ads	T0049: Flooding the Information Space	T0047: Censor social media as a political force	T0017: Conduct fundraising	T0059: Play the long game	T0132: Measure Performance
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	T0018: Purchase Targeted Advertisements	T0043: Chat apps	T0039 : Bait legitimate influencers	T0114.001: Social media	T0049.001: Trolls amplify and manipulate	T0048: Harass	T0017.001: Conduct Crowdfunding Campaigns	T0060: Continue to Amplify	T0132.001: People Focused
	T0075: Dismiss	T0072.002: Demographic Segmentation	T0022: Leverage Conspiracy Theory Narratives	T0019.001: Create fake research	T0013: Create inauthentic websites	T0011: Compromise legitimate accounts	T0101: Create Localized Content	T0043.001: Use Encrypted Chat Apps	T0042: Seed Kernel of truth	T0114.002: Traditional Media	T0049.002: Hijack existing hashtag	T0048.001: Boycott/"Cancel" Opponents	T0057: Organize Events	T0128: Conceal People	T0132.002: Content Focused
	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	T0102: Leverage Echo Chambers/Filter Bubbles	T0043.002: Use Unencrypted Chats Apps	T0044: Seed distortions	T0115: Post Content	T0049.003: Bots Amplify via Automated Forwarding and Reposting	T0048.002: Harass People Based on Identities	T0057.001: Pay for Physical Action	T0128.001: Use Pseudonyms	T0132.003: View Focused
	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	T0102.001: Use existing Echo Chambers/Filter Bubbles	T0103: Livestream	T0045: Use fake experts	T0115.001: Share Memes	T0049.004: Utilize Spamoflauge	T0048.003: Threaten to Dox	T0057.002: Conduct Symbolic Action	T0128.002: Conceal Network Identity	T0133: Measure Effectiveness

DISARM FRAMEWORK RED



DISARM Red Framework - incident creator TTPs

PLAN			PREPARE	EXECUTE		ASSESS
TA01: Plan Strategy	TA02: Plan Objectives	TA13: Target Audience Analysis	TA14: Develop Narrative	TA10: Drive Offline Activity	TA11: Persist in the Information Environment	TA12: Assess Effectiveness
T0073: Determine Target Audiences	T0002: Facilitate State Propaganda	T0072: Segment Audiences	T0003: Leverage Existing Narrative	T0017: Conduct fundraising	T0059: Play the long game	T0132: Measure Performance
T0074: Determine Strategic Ends	T0066: Degrade Adversary	T0072.001: Geographic Segmentation	T0004: Develop Competin Narrative	T0017.001: Conduct Crowdfunding Campaigns	T0060: Continue to Amplify	T0132.001: People Focused
	T0075: Dismiss	T0072.002: Demographic Segmentation	T0022: Leverage Conspirac Theory Narrative	T0057: Organize Events	T0128: Conceal People	T0132.002: Content Focused
	T0075.001: Discredit Credible Sources	T0072.003: Economic Segmentation	T0022.001 Amplify Existing Conspirac Theory Narrative	T0057.001: Pay for Physical Action	T0128.001: Use Pseudonyms	T0132.003: View Focused
	T0076: Distort	T0072.004: Psychographic Segmentation	T0022.002 Develop Original Conspirac Theory Narrative	T0057.002: Conduct Symbolic Action	T0128.002: Conceal Network Identity	T0133: Measure Effectiveness

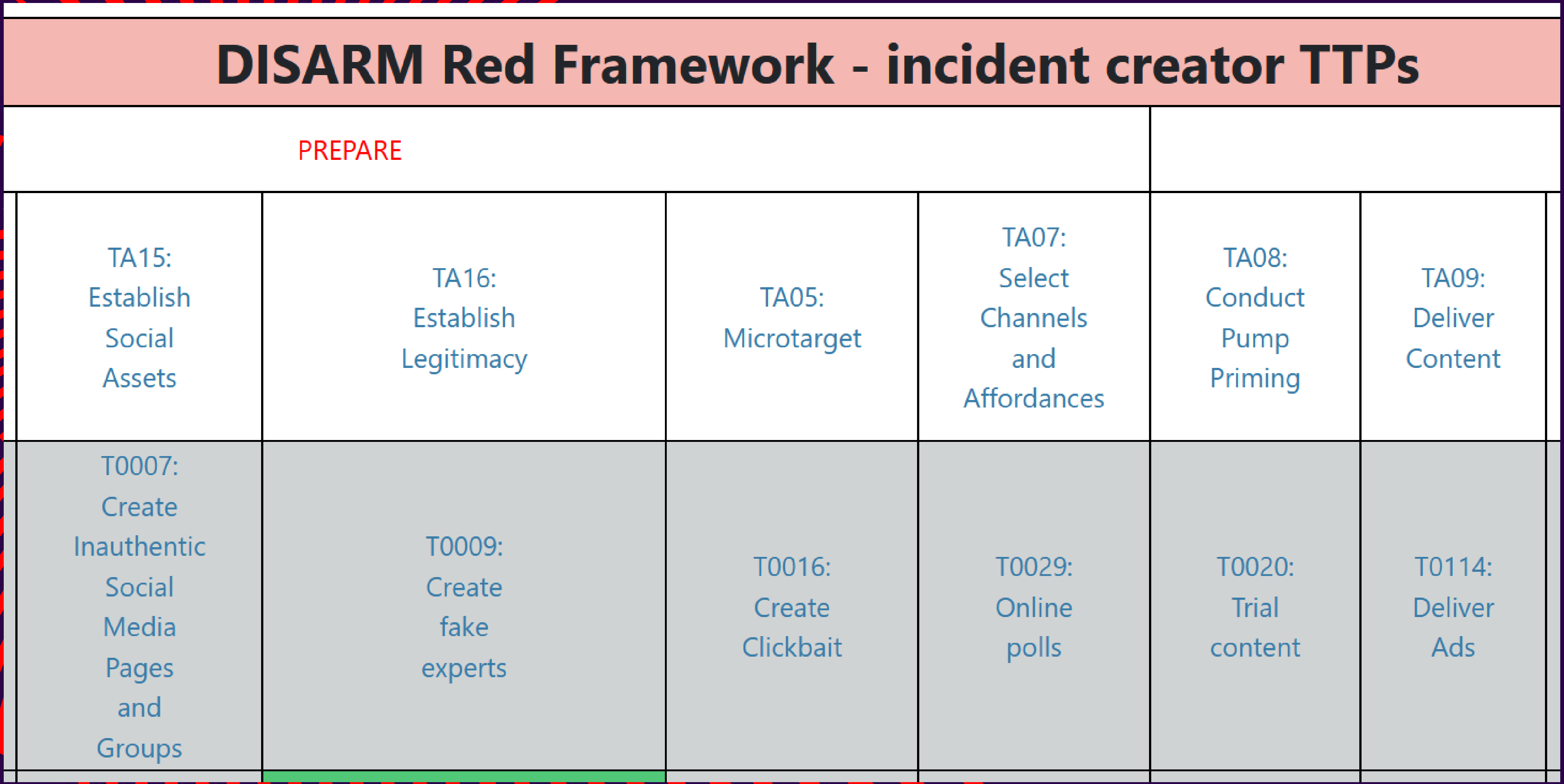
DISARM Objects

The disarm frameworks contain many object types, including tactic stages (steps in an incident), and techniques (activities at each tactic stage). We also have data objects to show how the frameworks are used in practice, and to make our datasets on tools and responders available.

Framework objects					
Frameworks	Phases	Tactics	Techniques	Tasks	Countermeasures
Detections	Responsetypes	Metatechniques	Playbooks	Resources	
Data objects					
Incidents	Examples	External Groups	Tools		

Disarm objects are described in detail here.

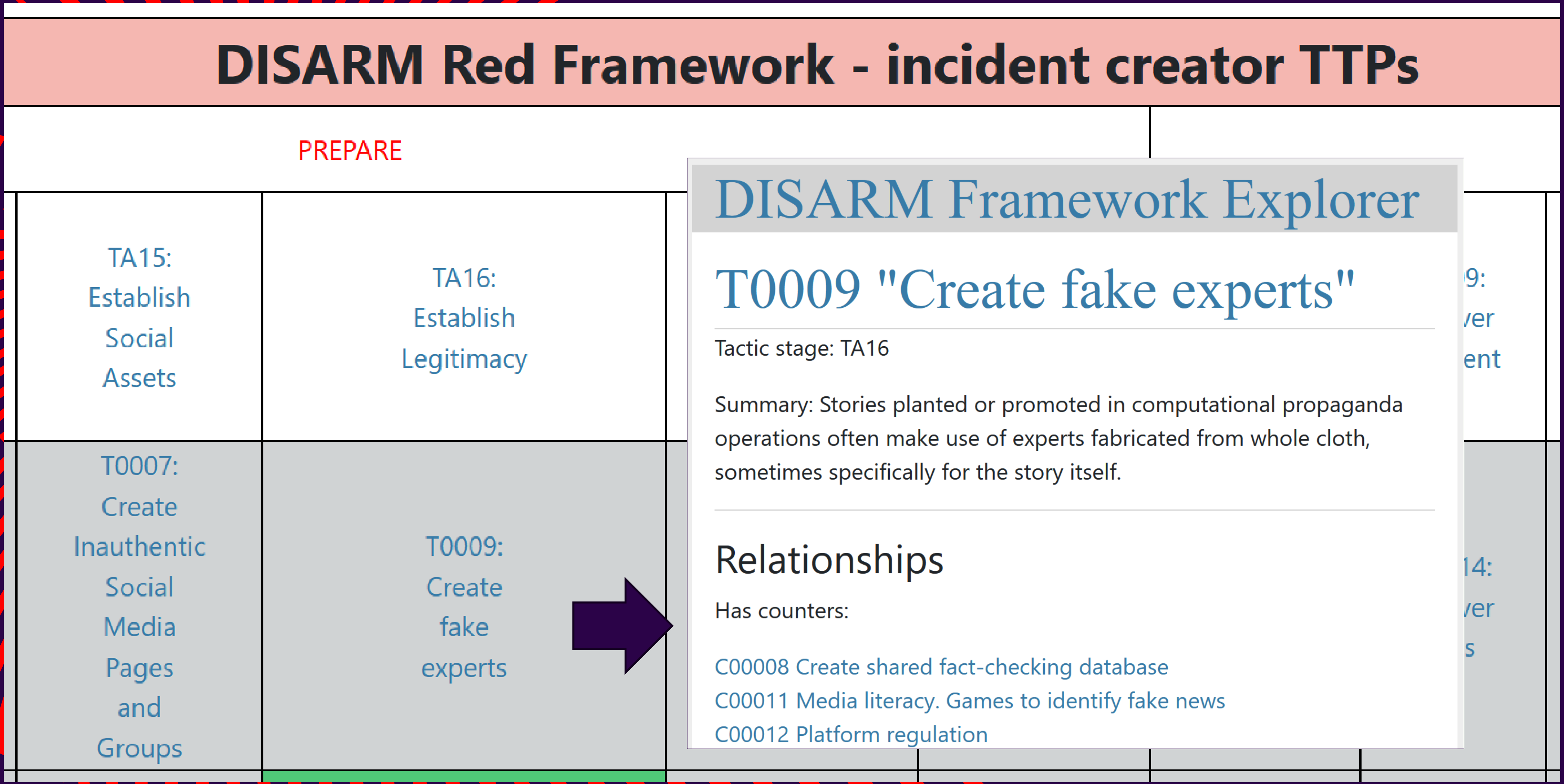
DISARM FRAMEWORK RED



DISARM Red Framework - incident creator TTPs

PREPARE					
TA15: Establish Social Assets	TA16: Establish Legitimacy	TA05: Microtarget	TA07: Select Channels and Affordances	TA08: Conduct Pump Priming	TA09: Deliver Content
T0007: Create Inauthentic Social Media Pages and Groups	T0009: Create fake experts	T0016: Create Clickbait	T0029: Online polls	T0020: Trial content	T0114: Deliver Ads

DISARM FRAMEWORK RED



DISARM Red Framework - incident creator TTPs

PREPARE		
TA15: Establish Social Assets	TA16: Establish Legitimacy	9: ver ent
T0007: Create Inauthentic Social Media Pages and Groups	T0009: Create fake experts	14: ver s

DISARM Framework Explorer

T0009 "Create fake experts"

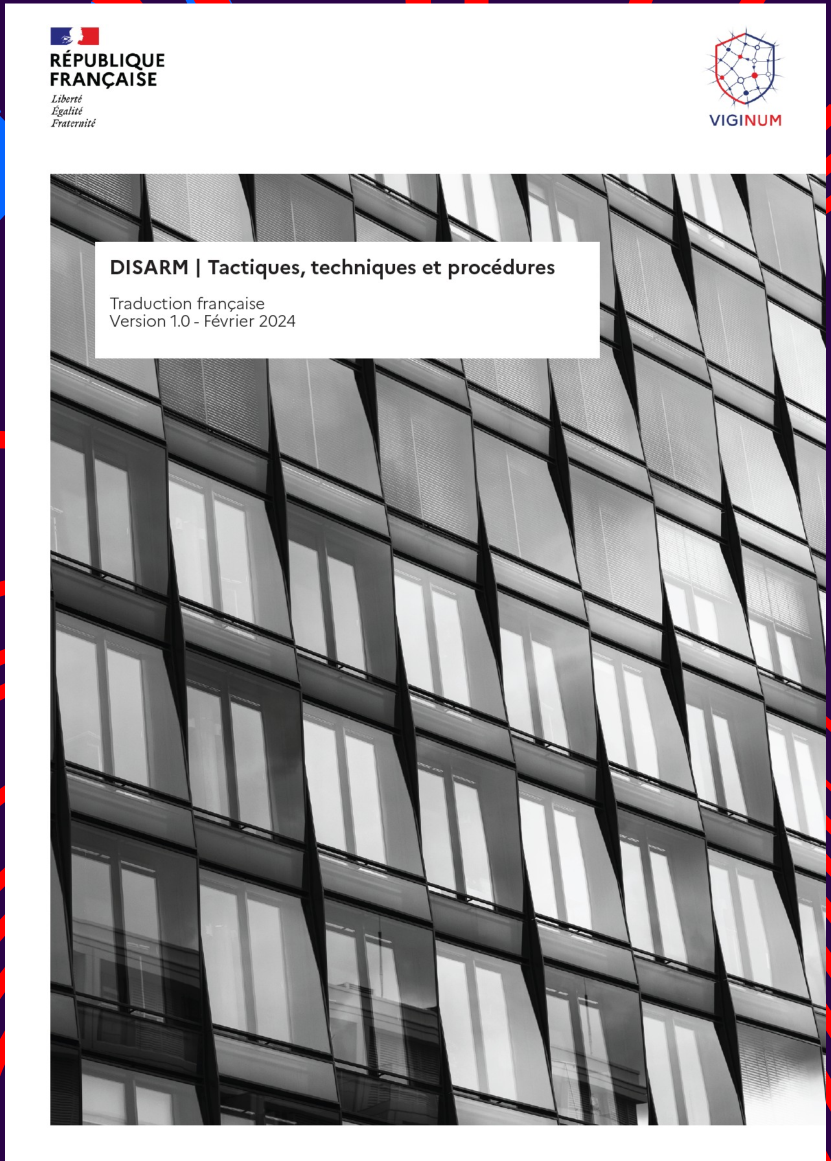
Tactic stage: TA16

Summary: Stories planted or promoted in computational propaganda operations often make use of experts fabricated from whole cloth, sometimes specifically for the story itself.

Relationships

Has counters:

C00008 Create shared fact-checking database
C00011 Media literacy. Games to identify fake news
C00012 Platform regulation

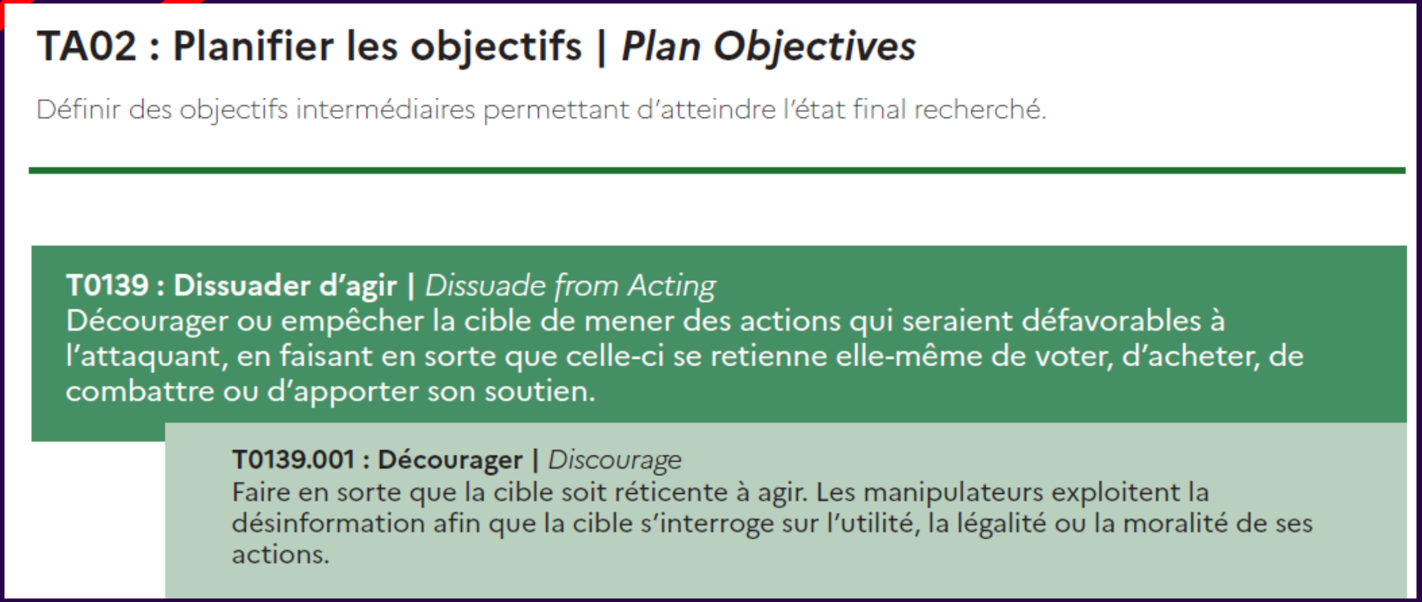


RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

VIGINUM

DISARM | Tactiques, techniques et procédures

Traduction française
Version 1.0 - Février 2024

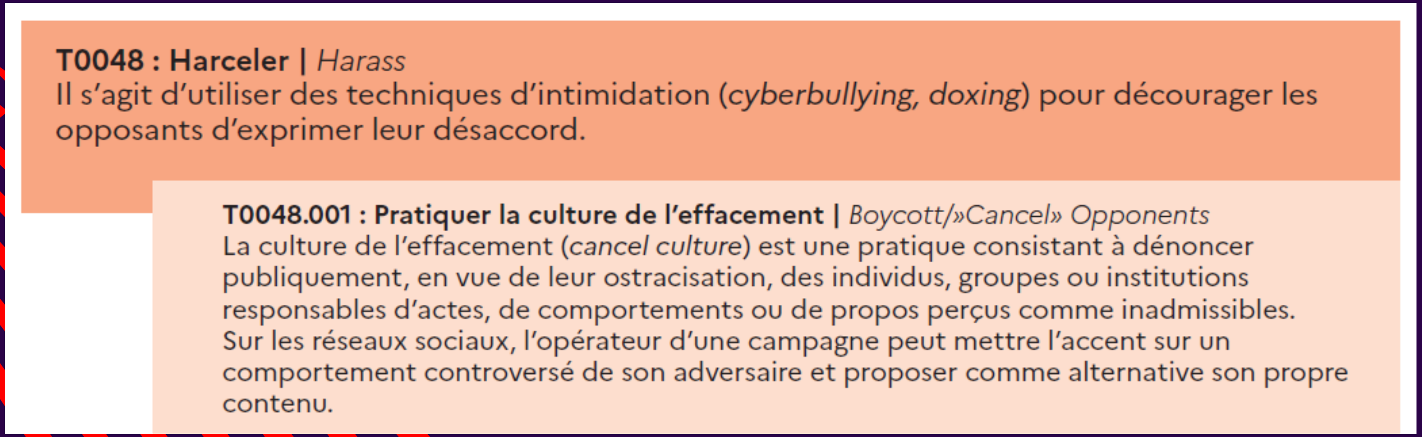


TA02 : Planifier les objectifs | *Plan Objectives*

Définir des objectifs intermédiaires permettant d'atteindre l'état final recherché.

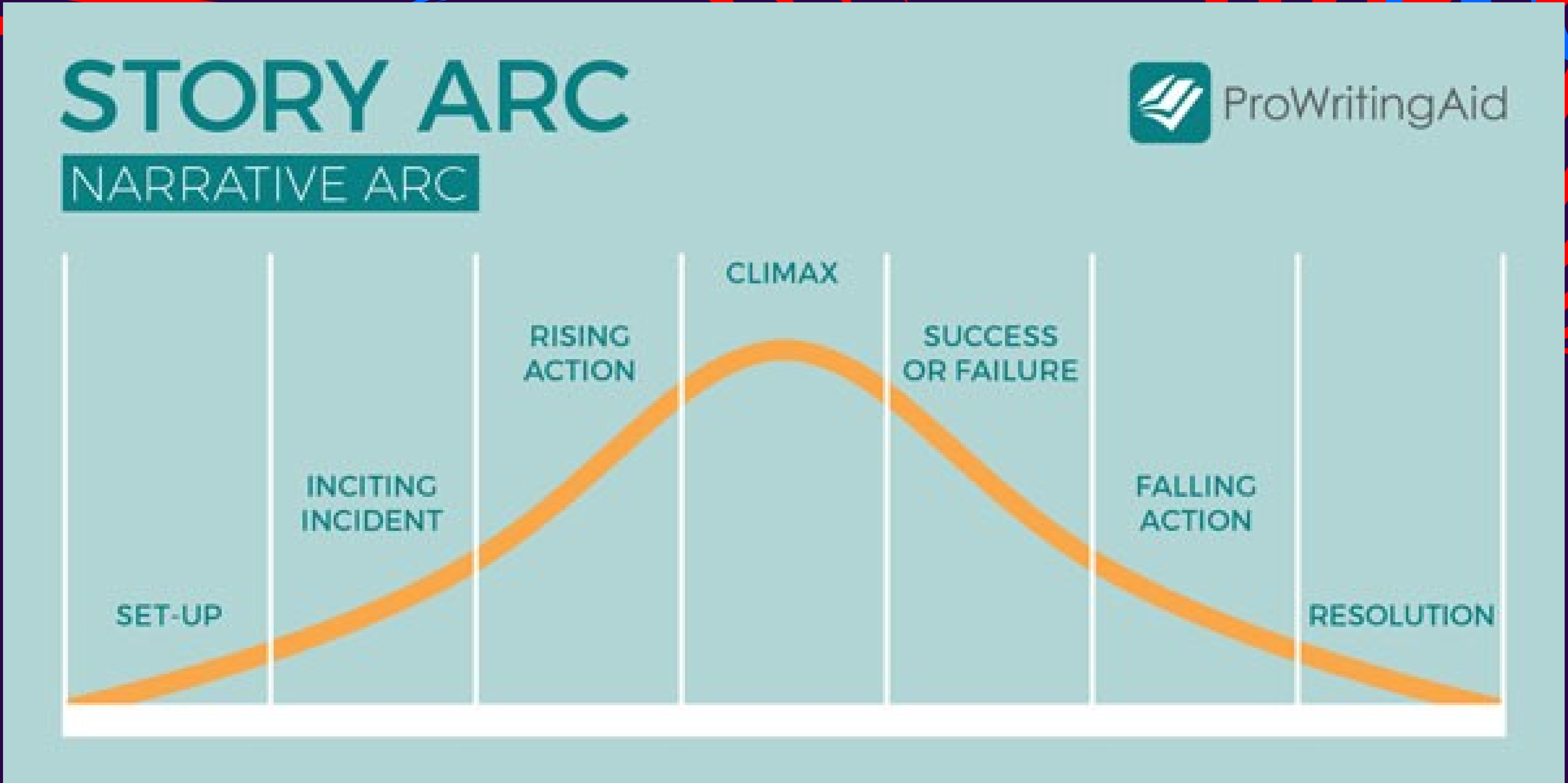
T0139 : Dissuader d'agir | *Dissuade from Acting*
Décourager ou empêcher la cible de mener des actions qui seraient défavorables à l'attaquant, en faisant en sorte que celle-ci se retienne elle-même de voter, d'acheter, de combattre ou d'apporter son soutien.

T0139.001 : Décourager | *Discourage*
Faire en sorte que la cible soit réticente à agir. Les manipulateurs exploitent la désinformation afin que la cible s'interroge sur l'utilité, la légalité ou la moralité de ses actions.



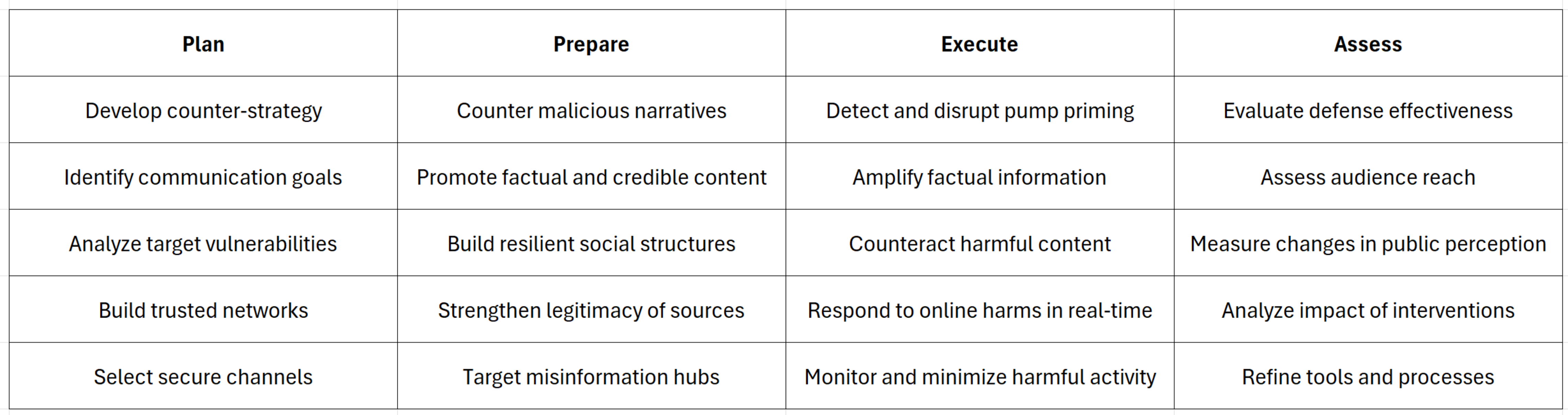
T0048 : Harceler | *Harass*
Il s'agit d'utiliser des techniques d'intimidation (*cyberbullying, doxing*) pour décourager les opposants d'exprimer leur désaccord.

T0048.001 : Pratiquer la culture de l'effacement | *Boycott/»Cancel» Opponents*
La culture de l'effacement (*cancel culture*) est une pratique consistant à dénoncer publiquement, en vue de leur ostracisation, des individus, groupes ou institutions responsables d'actes, de comportements ou de propos perçus comme inadmissibles. Sur les réseaux sociaux, l'opérateur d'une campagne peut mettre l'accent sur un comportement controversé de son adversaire et proposer comme alternative son propre contenu.



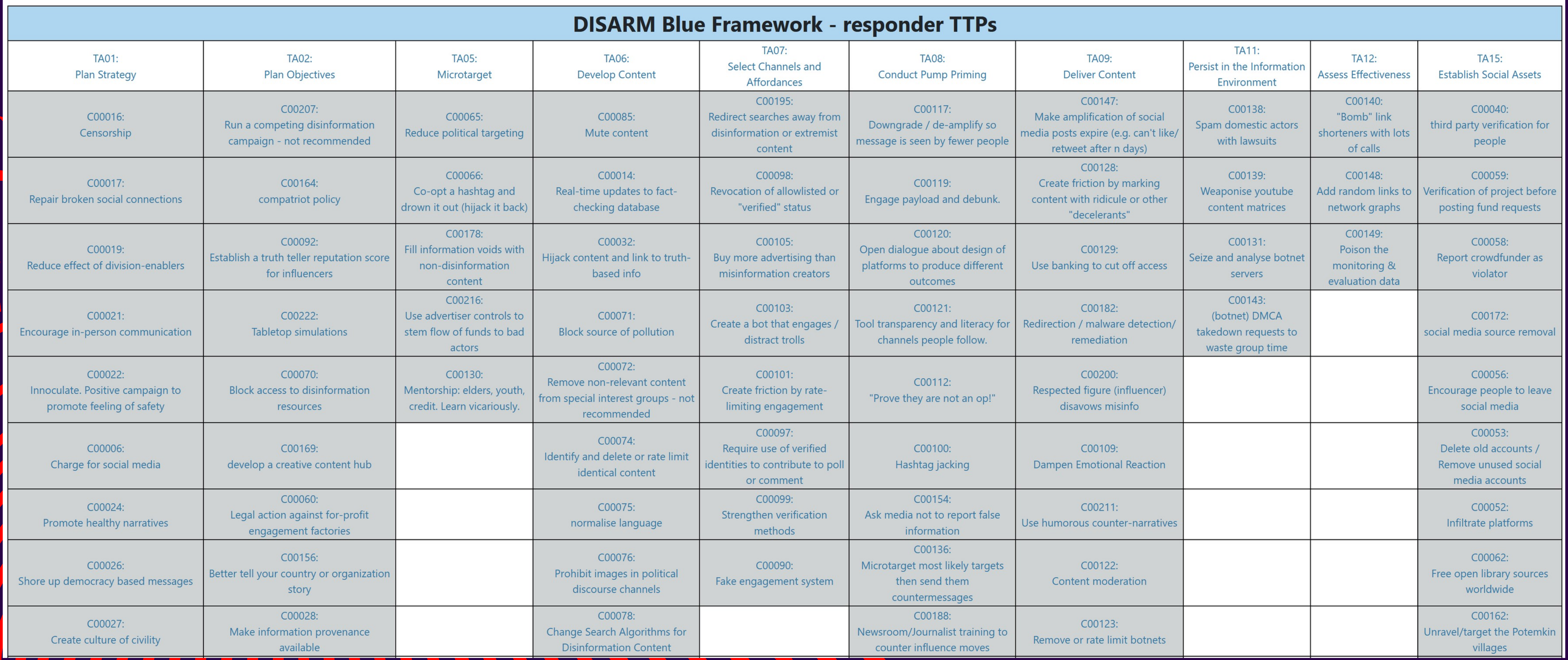
STORY ARC
NARRATIVE ARC
ProWritingAid
SET-UP
INCITING INCIDENT
RISING ACTION
CLIMAX
SUCCESS OR FAILURE
FALLING ACTION
RESOLUTION

DISARM FRAMEWORK BLUE



Plan	Prepare	Execute	Assess
Develop counter-strategy	Counter malicious narratives	Detect and disrupt pump priming	Evaluate defense effectiveness
Identify communication goals	Promote factual and credible content	Amplify factual information	Assess audience reach
Analyze target vulnerabilities	Build resilient social structures	Counteract harmful content	Measure changes in public perception
Build trusted networks	Strengthen legitimacy of sources	Respond to online harms in real-time	Analyze impact of interventions
Select secure channels	Target misinformation hubs	Monitor and minimize harmful activity	Refine tools and processes

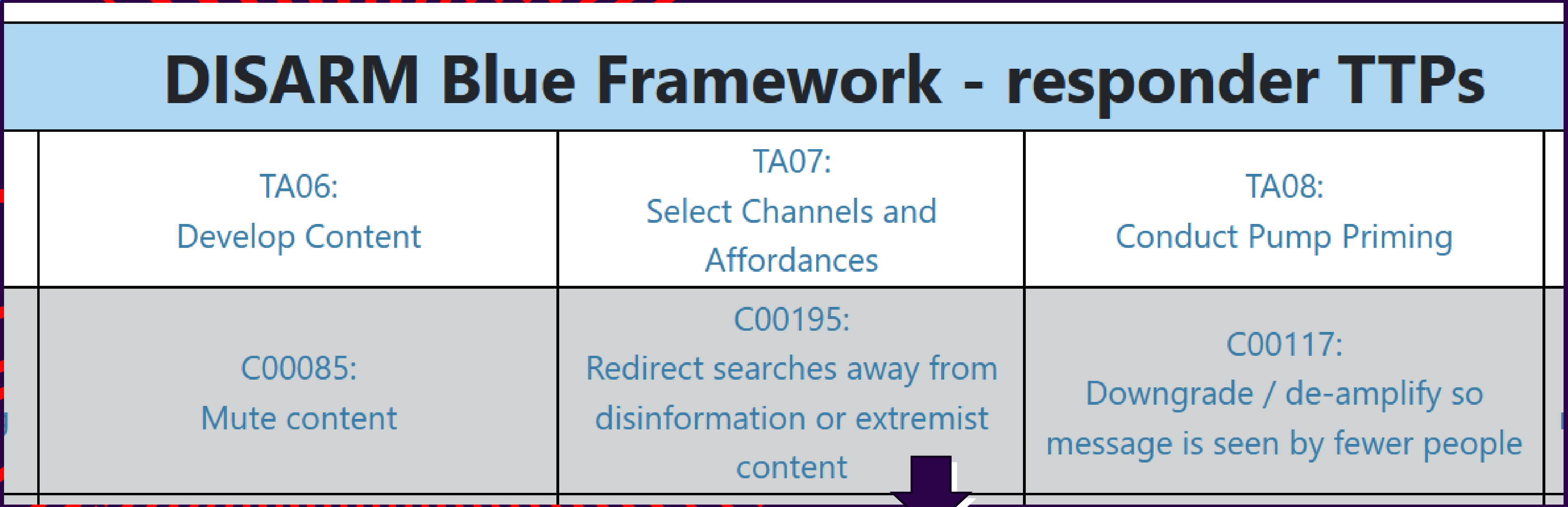
DISARM FRAMEWORK BLUE



DISARM Blue Framework - responder TTPs

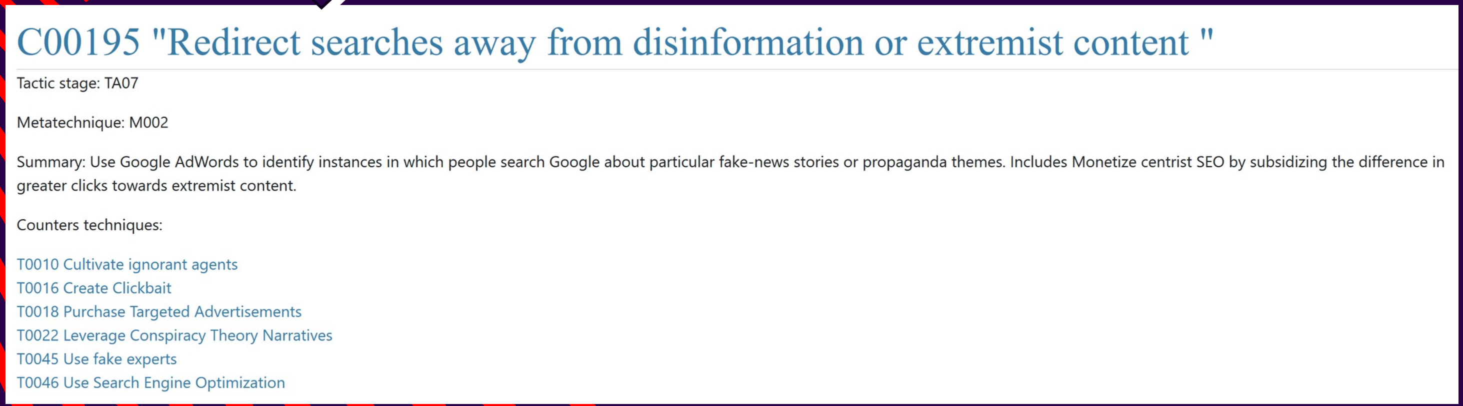
TA01: Plan Strategy	TA02: Plan Objectives	TA05: Microtarget	TA06: Develop Content	TA07: Select Channels and Affordances	TA08: Conduct Pump Priming	TA09: Deliver Content	TA11: Persist in the Information Environment	TA12: Assess Effectiveness	TA15: Establish Social Assets
C00016: Censorship	C00207: Run 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 lots of 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 allowlisted or "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: Innoculate. 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
C00006: 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
C00026: Shore up democracy based messages	C00156: Better tell your country or organization story		C00076: Prohibit images in political discourse channels	C00090: Fake engagement system	C00136: Microtarget most likely targets then send them countermessages	C00122: Content moderation			C00062: Free open library sources worldwide
C00027: Create culture of civility	C00028: Make information provenance available		C00078: Change Search Algorithms for Disinformation Content		C00188: Newsroom/Journalist training to counter influence moves	C00123: Remove or rate limit botnets			C00162: Unravel/target the Potemkin villages

DISARM FRAMEWORK BLUE



DISARM Blue Framework - responder TTPs

TA06: Develop Content	TA07: Select Channels and Affordances	TA08: Conduct Pump Priming
C00085: Mute content	C00195: Redirect searches away from disinformation or extremist content	C00117: Downgrade / de-amplify so message is seen by fewer people



C00195 "Redirect searches away from disinformation or extremist content "

Tactic stage: TA07

Metatechnique: M002

Summary: Use Google AdWords to identify instances in which people search Google about particular fake-news stories or propaganda themes. Includes Monetize centrist SEO by subsidizing the difference in greater clicks towards extremist content.

Counters techniques:

T0010 Cultivate ignorant agents
T0016 Create Clickbait
T0018 Purchase Targeted Advertisements
T0022 Leverage Conspiracy Theory Narratives
T0045 Use fake experts
T0046 Use Search Engine Optimization

CITIZEN RESPONSIBILITY

- Individuals must verify and share information responsibly.
- Communities play a vital role in promoting media literacy and accountability.
- Personal actions contribute to a healthier information ecosystem.

OPPORTUNITIES FOR GAMIFICATION

- Games like “Bad News” let players experience disinformation tactics firsthand.
- Role-reversal strategies teach how disinformation campaigns are constructed.
- DISARM can be a scenario creation engine and a way for individuals to discover what they are involved in.

FUTURE WORK AND VISION

- Future work focuses on creating scalable, gamified awareness tools.
- Collaboration between public and private sectors enhances impact.
- A unified vision can drive global resilience against disinformation.

THANK YOU FOR YOUR ATTENTION

jp@riant.fr

ugo.peyre42@gmail.com