DIMA
Un *framework* pour décrire les attaques cognitives

De quoi parle-t-on ?

La « **désinformation** » est généralement définie comme la diffusion d'informations délibérément fausses ou trompeuses. On parle alors de « fausses informations » ou de « fake news ».

La « **manipulation de l'information** » est une action délibérée (intention de nuire), clandestine (les victimes sont inconscientes) de diffusion d'information falsifiées, déformées ou forgées.

Une **opération d'influence** (OI) combine diffusion d'information et actions physiques dans le but de **modifier des comportements**.

Attaque cognitive ?

Les campagnes de manipulation de l'information visent à **modifier les perceptions des opinions publiques** sur des situations particulières. (renforcement cognitif)

La guerre cognitive vise à altérer directement les mécanismes de compréhension du monde réel et **de prise de décision** pour déstabiliser ou paralyser un adversaire.

« Une attaque cognitive correspond à **l'utilisation intentionnelle d'un ou plusieurs biais ou heuristiques** dont l'objectif est de provoquer une réaction de la cible. »

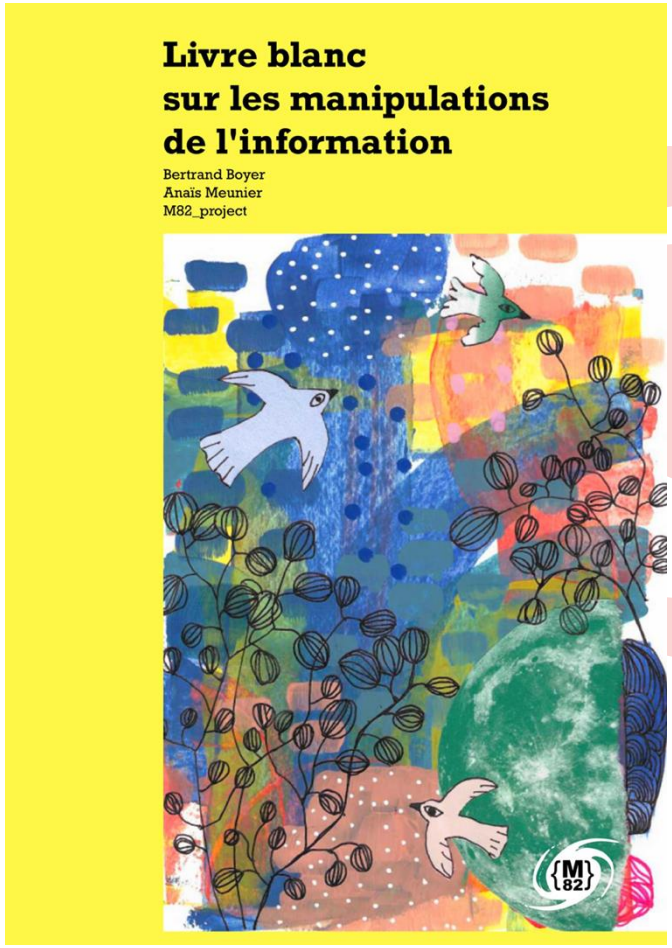
Les cerveaux humains, vulnérables aux **biais cognitifs**, sont directement ciblés pour « modifier des comportements ».

Objectif de la matrice : **caractérisation des attaques cognitives**. Comprendre les biais exploités par un attaquant potentiel pour qu'une information provoque un changement de comportement.

Il ne s'agit pas ici de « détecter des *fake news*, » mais d'évaluer à quel point une information reçue est construite (délibérément) pour exploiter un ou plusieurs biais cognitifs.

Renforcer la défense par la connaissance des vulnérabilités exploitées

Pourquoi DIMA ?



Que faut-il modéliser et pourquoi ?

Approche tournée vers l'acteur

Comprendre l'attaque

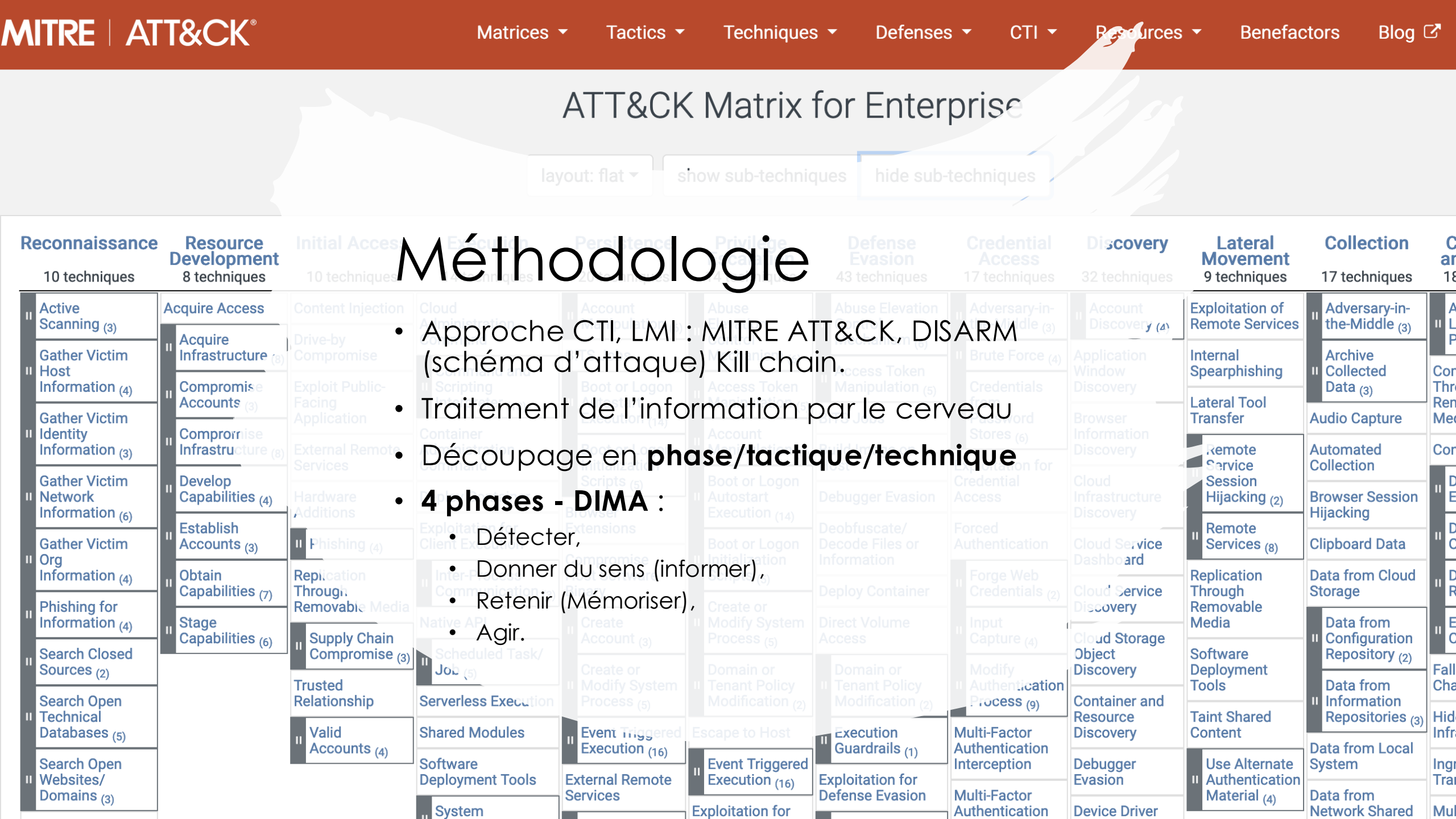
- La « kill chain » pour mieux se défendre
- Modéliser le schéma de l'attaque;

Approche « cible »

Connaitre les vulnérabilités exploitées

- Faire émerger des vulnérabilités à protéger.
- Faire apparaitre des « signatures » pour caractériser un acteur

Pour se défendre, il faut soit « briser l'attaque » soit la rendre inefficace.



Méthodologie

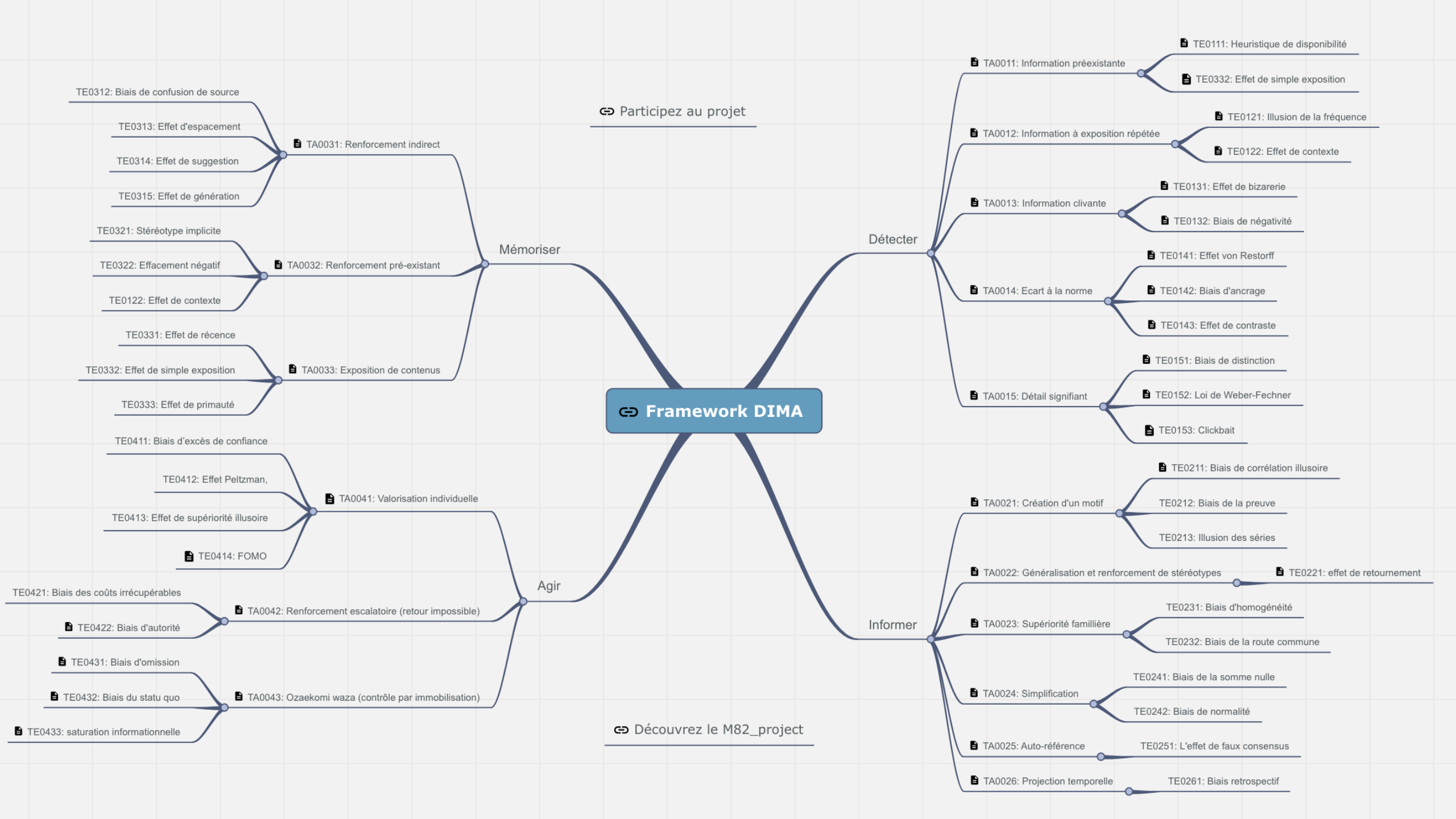
- Approche CTI, LMI : MITRE ATT&CK, DISARM (schéma d'attaque) Kill chain.
- Traitement de l'information par le cerveau
- Découpage en **phase/tactique/technique**
- **4 phases - DIMA** :
  - Détecter,
  - Donner du sens (informer),
  - Retenir (Mémoriser),
  - Agir.

THE CYBER KILL CHAIN

Source :
https://cybotsai.com/introduction-mitre-attck/

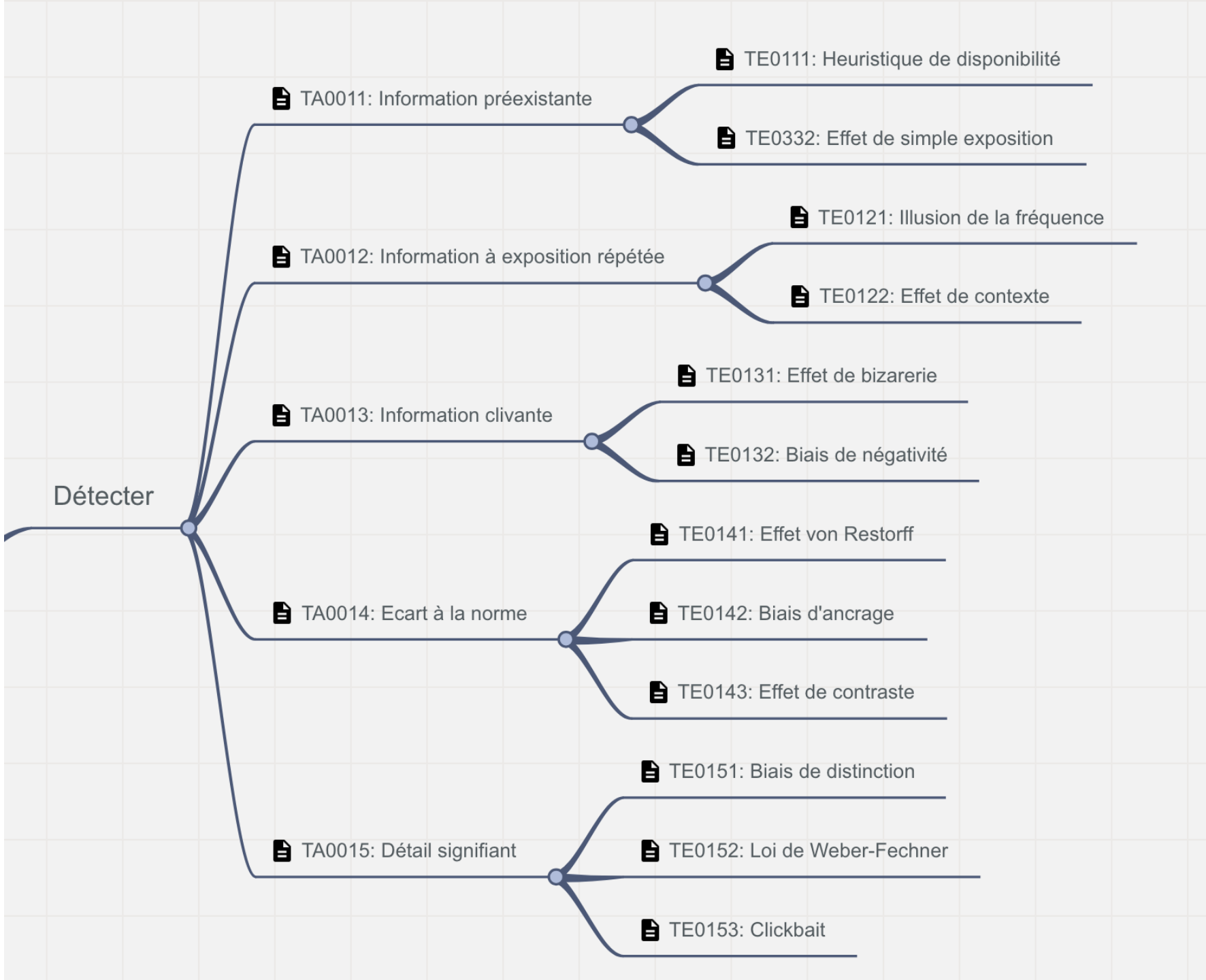
01	Detect / Détecter
02	Inform / Informer (donner du sens)
03	Memorize / Mémoriser (Y/N)
04	Act / Agir

DIMA « Cognitive Kill Chain »


Framework DIMA
Participez au projet
Découvrez le M82_project
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
TE0122: 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
TE0414: FOMO
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
TE0433: saturation informationnelle
Détecter
TA0011: Information préexistante
TE0111: Heuristique de disponibilité
TE0332: 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
TE0153: Clickbait
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 famillière
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

Détecter

POURQUOI CETTE INFORMATION A-T-ELLE ATTIRÉ MON ATTENTION ?



Octobre 2023: les « punaises de lit »

Plusieurs médias évoquent le sujet, rumeurs sur les réseaux sociaux. L'information est détectée par l'utilisation de « l'illusion de fréquence ». Autrement connu sous le nom de phénomène Baader-Meinhof, l'illusion de la fréquence semble être une combinaison de plusieurs biais qui nous amène à pratiquer une forme "d'attention sélective".

Ce mécanisme est particulièrement à l'œuvre dans la diffusion des thèses complotistes.

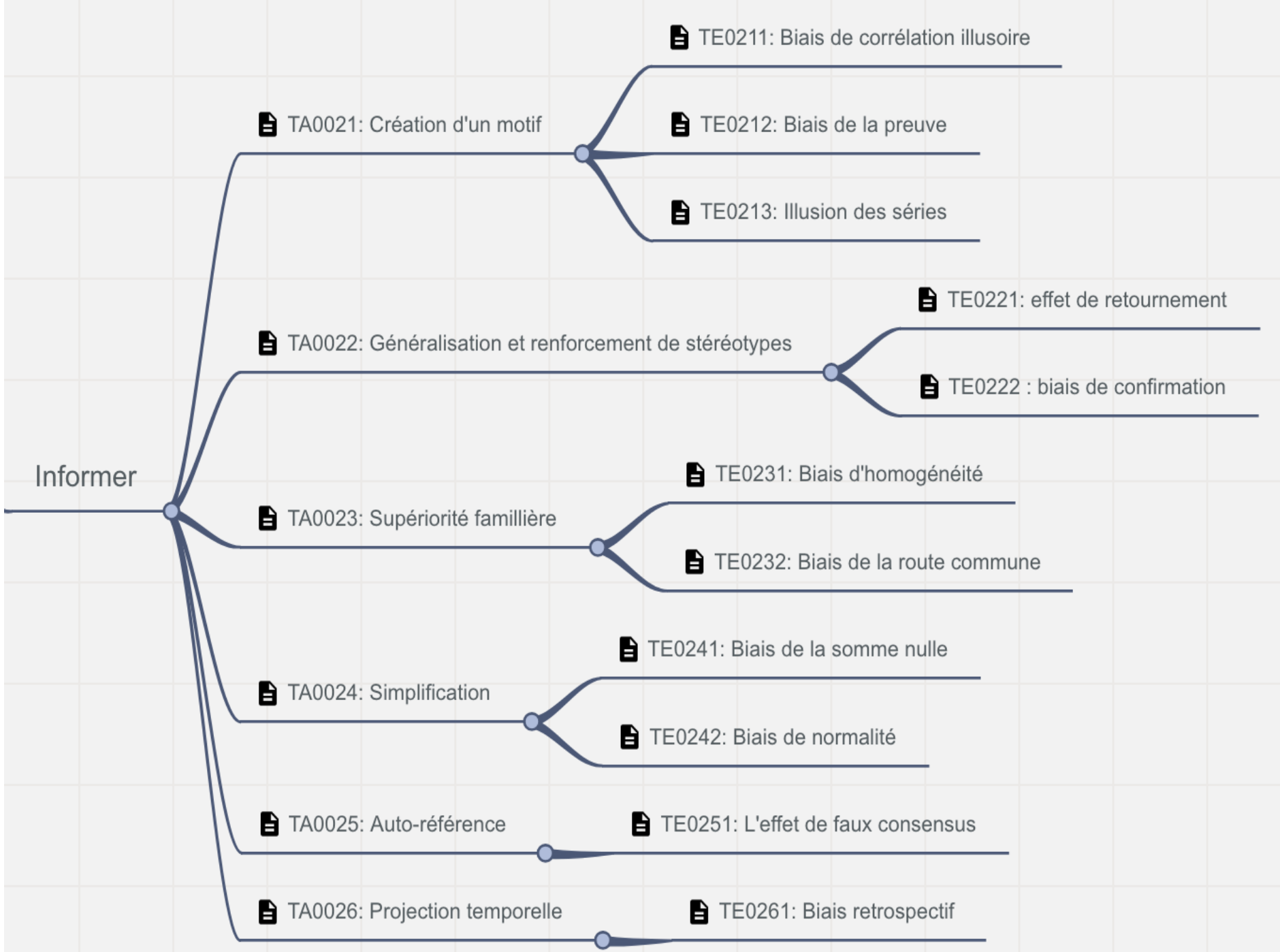
"L'illusion de fréquence consiste, après avoir remarqué une chose une première fois, à avoir tendance à la remarquer plus souvent, ce qui conduit à croire qu'elle se produit plus fréquemment qu'auparavant."

DIMA : TE 0121 « ILLUSION DE FRÉQUENCE »

DISARM : T0042 "Seed Kernel of truth"

Informer

Donner du sens à une information ?

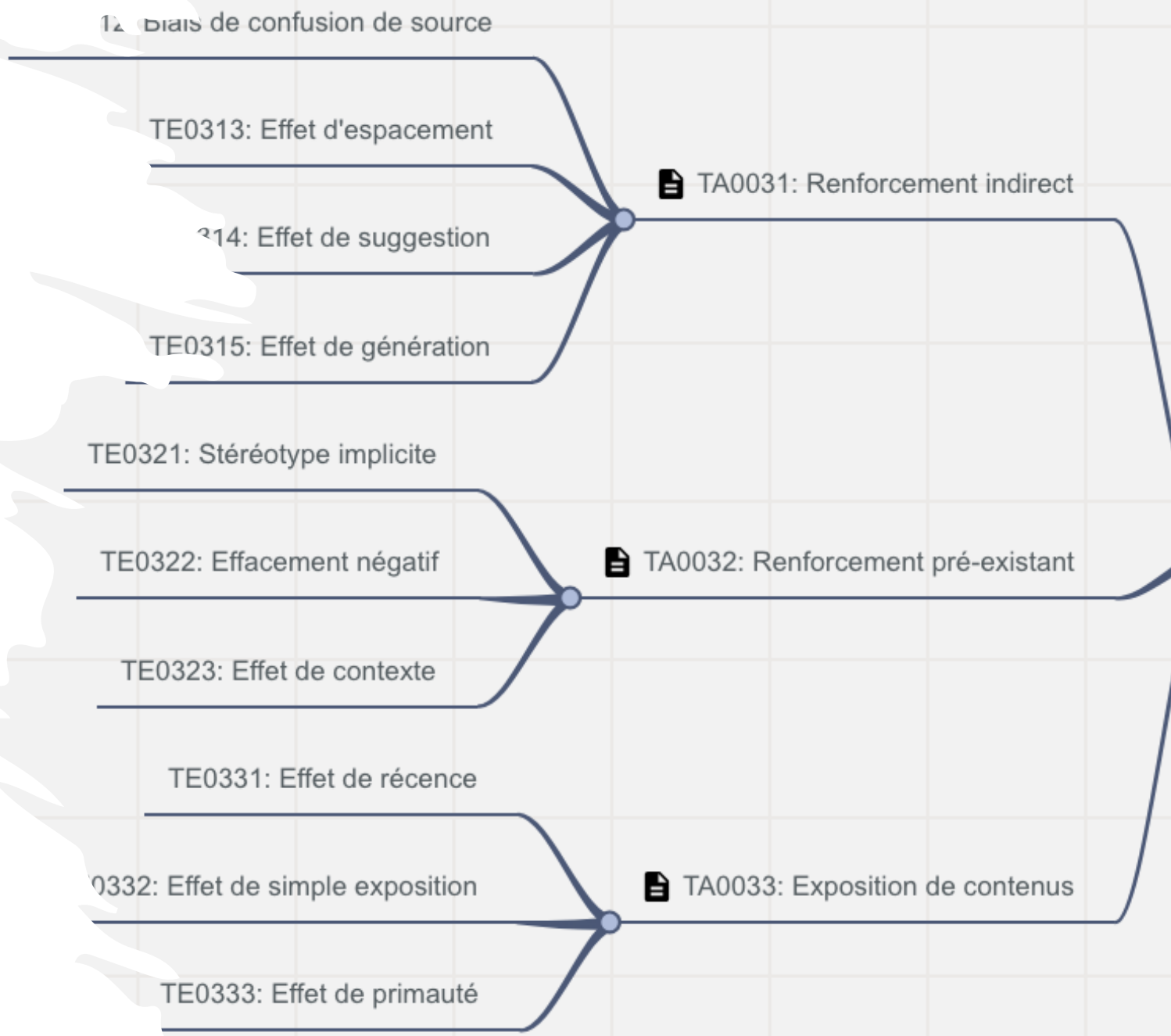


TE 0211: Biais de corrélation illusoire

- Nous donnons du sens à cette information en créant des « corrélations » (i.e – lien entre réfugiés ukrainiens et apparition des parasites)

Mémoriser

- Pourquoi retenir cette information ?



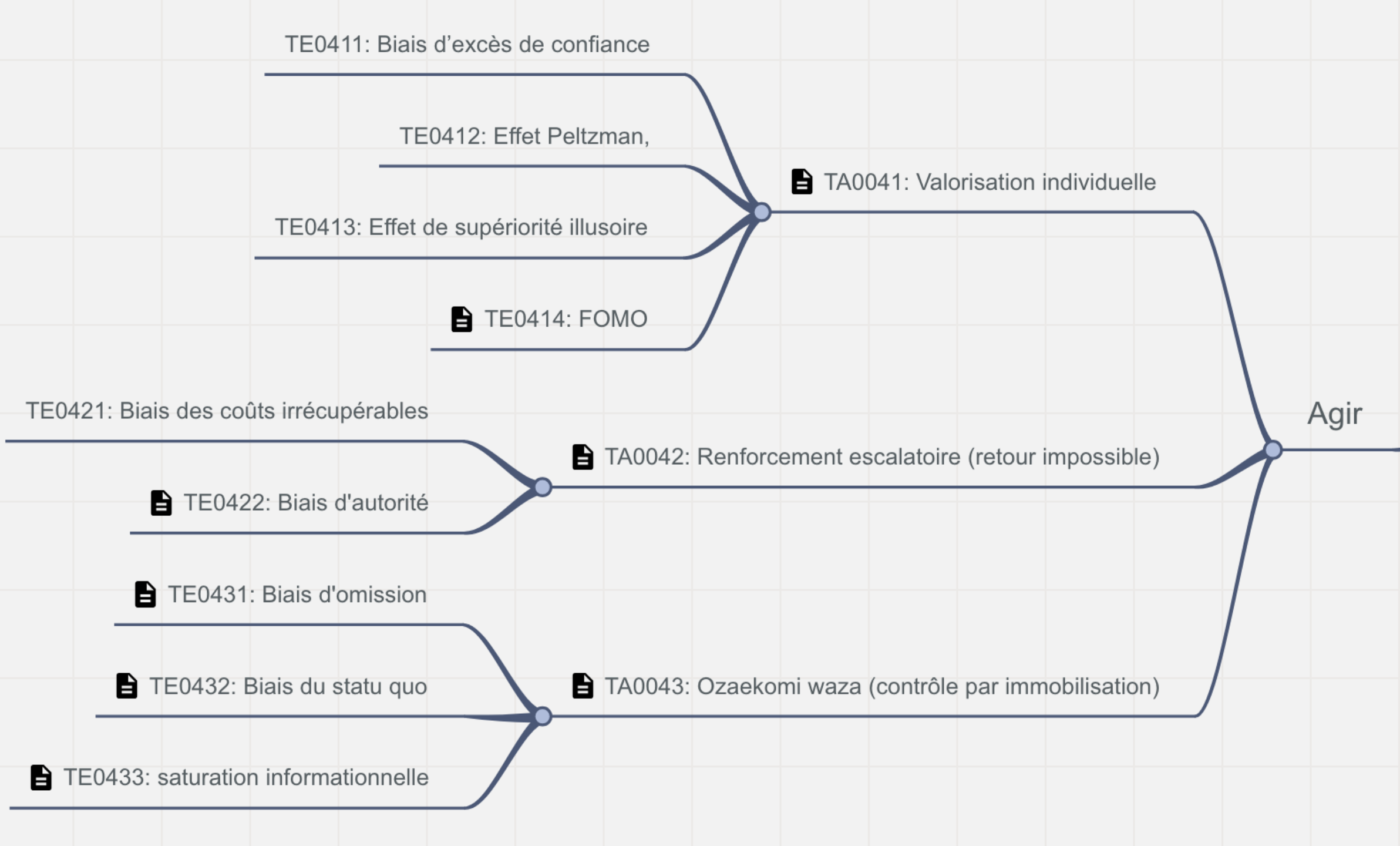
TE 0314: effet de suggestion

TE 0321: Stéréotype implicite

- Usage de l'interrogation
- Amplification de stéréotypes

Agir

POUQUOI CETTE INFORMATION PROVOQUE UNE (RE)ACTION DE MA PART ?


TE0411: Biais d’excès de confiance
TE0412: Effet Peltzman,
TE0413: Effet de supériorité illusoire
TE0414: FOMO
TA0041: Valorisation individuelle
TE0421: Biais des coûts irrécupérables
TE0422: Biais d'autorité
TA0042: Renforcement escalatoire (retour impossible)
TE0431: Biais d'omission
TE0432: Biais du statu quo
TE0433: saturation informationnelle
TA0043: Ozaekomi waza (contrôle par immobilisation)
Agir

TE 0422 : biais d'autorité

Les punaises de lit d'Europe de l'Est se reproduisent à Paris

MIS À JOUR LE 02/10/23 À 08H02

Aude Lorriaux

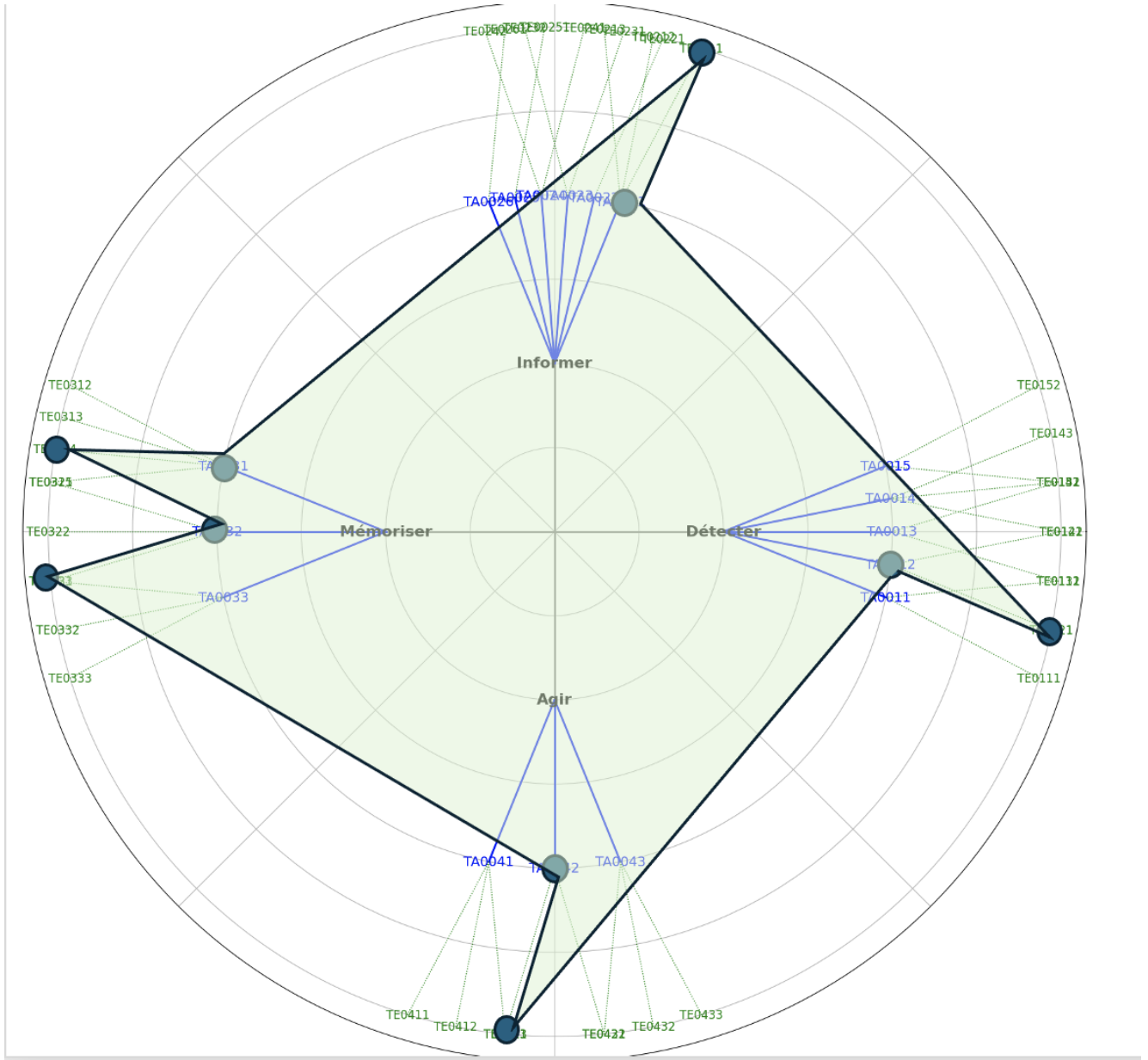
Des réfugiés ukrainiens ont transporté des punaises de lit et d'autres parasites sur leurs vêtements en France. Les experts prédisent une épidémie.

15 COMMENTER PARTAGER SAUVEGARDER UNE FAUTE ?

Les « experts » confirment, ils « prédisent » une épidémie… Il faut donc prendre des mesures.

Bilan

- **Détecter**
  - TA0012: Information à exposition répétée
    - TE0121: Illusion de la fréquence
- **Informer**
  - TA0021: Création d’un motif
    - TE0211: biais de corrélation illusoire
- **Mémoriser**
  - TA0031: Renforcement indirect
    - TE0314: Effet de suggestion
  - TA0032: Renforcement pré-existant
    - TE0321: Stéréotype implicite
- **Agir**
  - TA0042: Renforcement escalatoire
    - TE0422: biais d'autorité

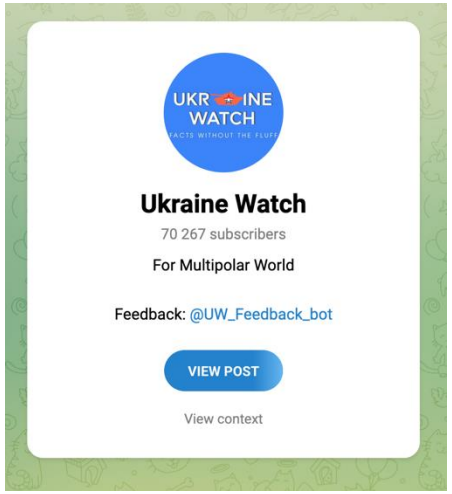


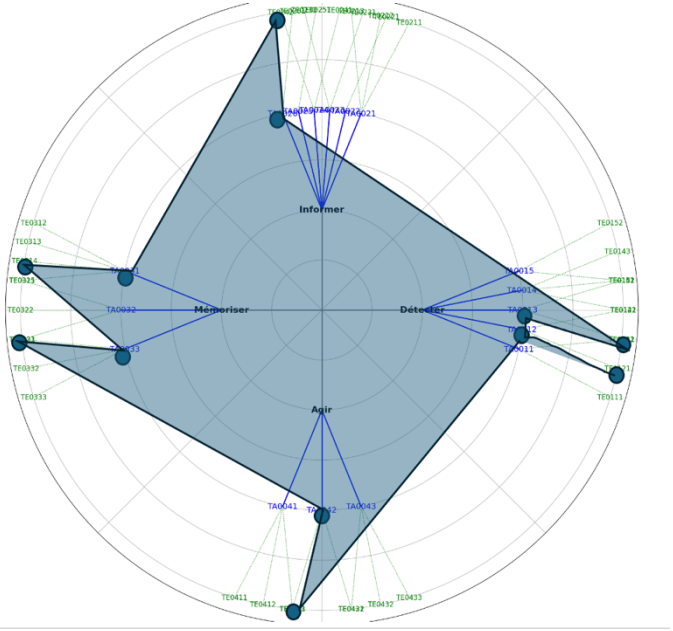
Annulation de la vente de Rafale aux EAU ?

- Arrestation en Fr de Pavel Durov (Telegram). Provoque une reaction officielle des EAU.

- Diffusion et amplification artificielle, vidéo reprenant les codes « Al jazeera ».

Capture d'écran de la fausse vidéo usurpant le logo d'Al-Jazira et diffusée sur X et Telegram. © X





Analyse DIMA

- **DETECTER:**
  - **TE 0121 « illusion de fréquence »**
  - **TE0131 : Effet de bizarrerie**
- **INFORMER**
  - **TE0251: Effet de faux consensus**
- **MEMORISER**
  - **TE0312: biais de confusion de source**
  - **TE0331: Effet de récence**
- **AGIR:**
  - **TE 0422 : biais d'autorité**

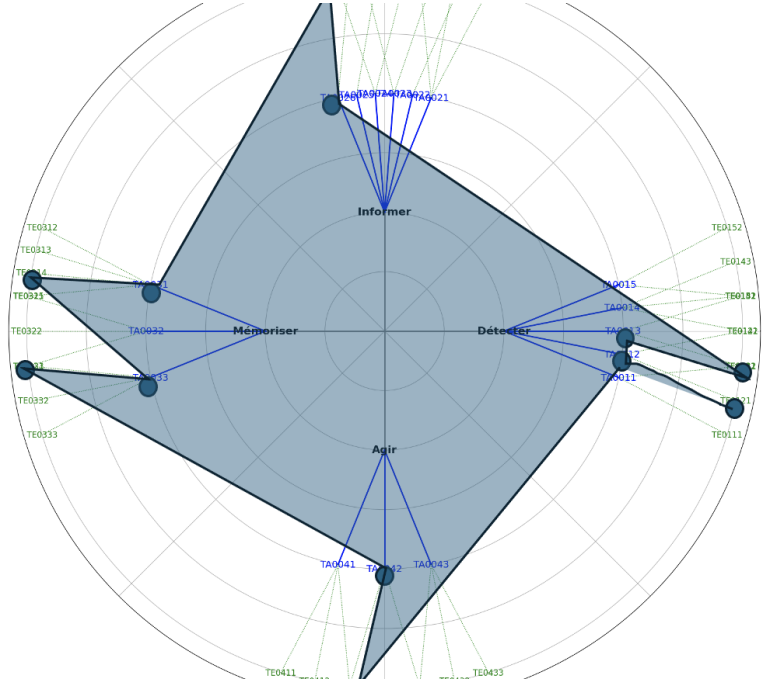
Où allons-nous ?

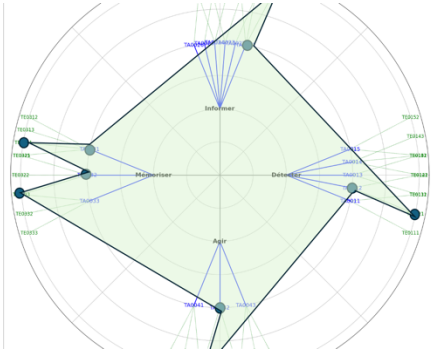
Analyse des cas documentés (projet DIOD);

Evaluer les similitudes par des décompositions de Fourier et les distances de Hausdorff;

Mesurer l'écart dans le temps de comportement d'un acteur par rapport à un modèle de référence.

Identifier les biais les plus systématiquement exploités par un acteur afin de proposer des mesures de remédiation et de résilience.





Digital Influence Operations Database
sponsored by M82

Outiller DIMA

- Une extension navigateur (**Chrome**)

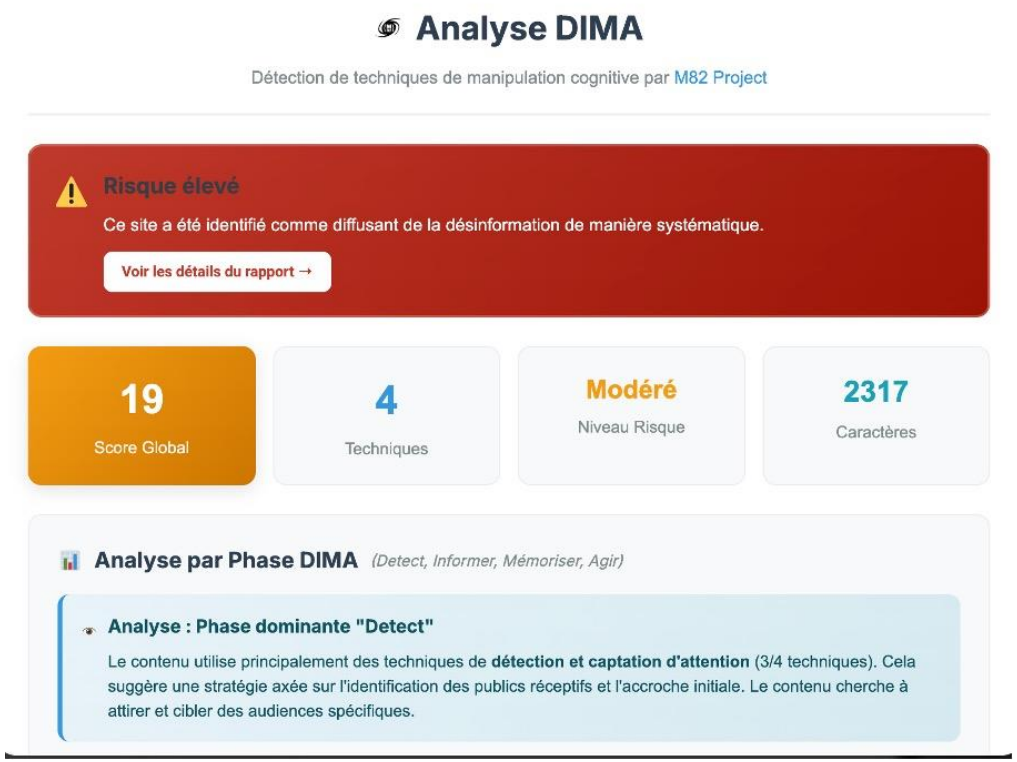
https://github.com/M82-project/DIMA_Plugin_Chrome

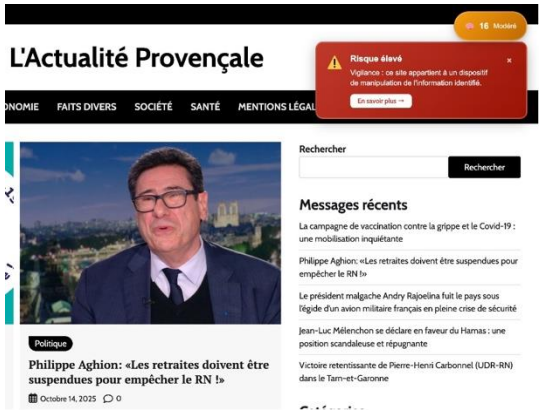
- Une analyse automatique des articles d'un **flux RSS**

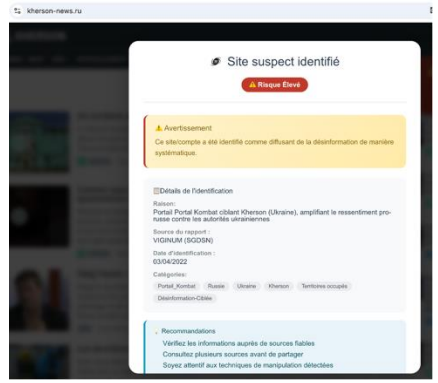
https://github.com/alexbarbaron/DIMAnalyzer

- Des projets en dév (on ne peut pas en parler mais y'a de l'IA dedans)

https://github.com/GenerativSchool-Lab/infoverif-extension.git







DIMA FRAMEWORK

Le Framework est en cours de réalisation,

Soutenu par le **M82 project**

Découvrez la matrice sur le framamind : https://framindmap.org/c/maps/1457115/public

Participez au projet : https://github.com/M82-project