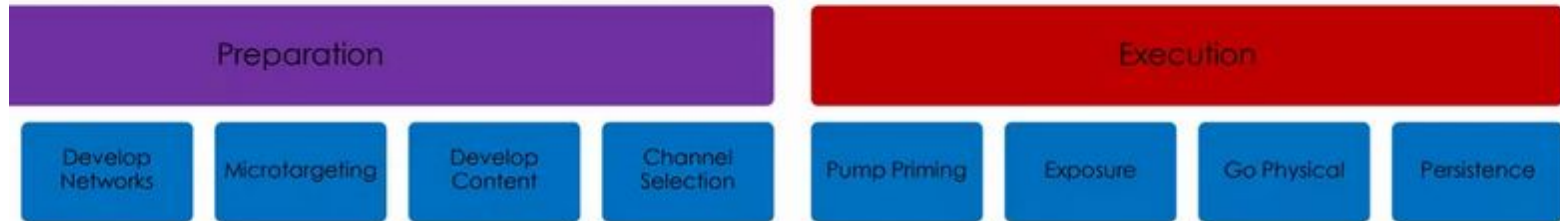


# Présentation projet fil rouge

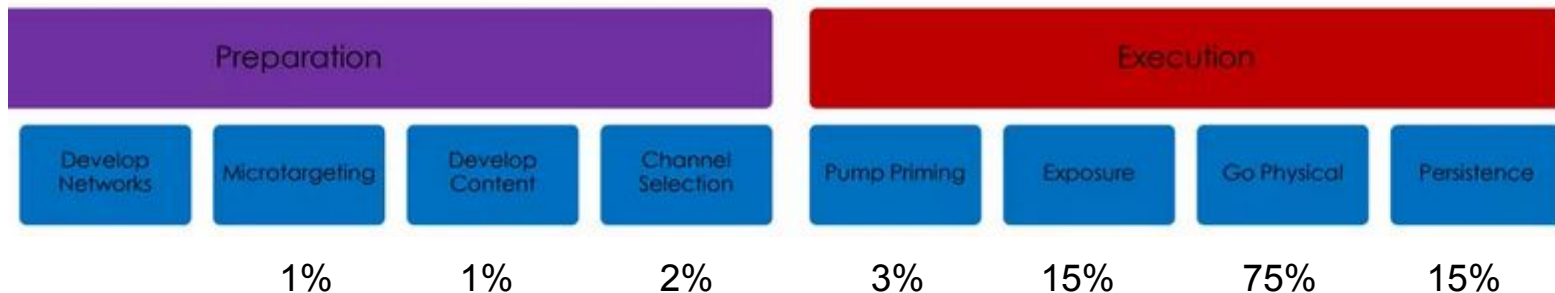
Janvier 22

# Vision du projet

1. Extraire des informations de métadonnée : sentiment, mots-clés, intimité, persuasion
2. Créer un algo utilisant ces données pour classifier ces données avec les TTPs de DISARM
3. Afficher sur OpenCTI les alertes



# DISARM framework



Métadonnées

analyse de sentiments

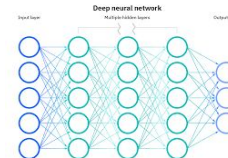
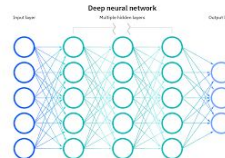
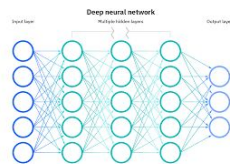
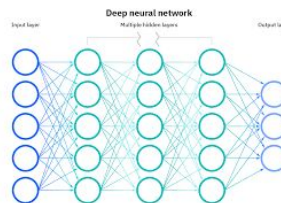
score d'intimité

extraction de mots-clés

Données originales

texte

images



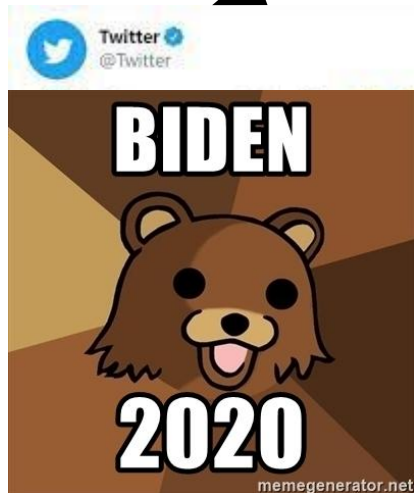
## Execution

Pump Priming

Exposure

Go Physical

Persistence



Traduire le Tweet  
12:13 AM · 23 janv. 2020 · Sprinklr

5 k Retweets 47,7 k J'aime

12:10 AM  
23 octobre 2020



Traduire le Tweet  
12:13 AM · 23 janv. 2020 · Sprinklr

5 k Retweets 47,7 k J'aime

15:30 AM  
23 octobre 2020



Traduire le Tweet  
12:13 AM · 23 janv. 2020 · Sprinklr

5 k Retweets 47,7 k J'aime

17:40 AM  
23 octobre 2020



Traduire le Tweet  
12:13 AM · 23 janv. 2020 · Sprinklr

5 k Retweets 47,7 k J'aime

19:10 AM  
23 octobre 2020



# OPENCTI

docker

2 possibilités :

- VM => fonctionne pas
- Docker => Docker Swarm + reverse proxy (solution retenue en cours)

# Projet européen : MITRE ATT&CK / DISARM



EUROPEAN UNION AGENCY  
FOR CYBERSECURITY



European Union  
EXTERNAL ACTION



## CONTENT

TA06 – Develop Content  
TA02 – Plan Objectives  
TA14 – Develop Narratives

Obtain information that can be used (also altered) in information incidents/operations

Develop content e.g. deep fakes

Carry out actions that support specific narratives e.g. stealing voter registration data



## INFRASTRUCTURE

TA16 – Establish Legitimacy  
TA07 – Select Channels and Affordances  
TA15 – Establish Social Assets

Obtain data (e.g. e-mails addresses) that can be used to disseminate information (e.g. information flooding)

Establishing legitimacy by creating new (fake) accounts or compromising existing accounts

Compromise websites (e.g. news outlets) to display fake information



## DISSEMINATION

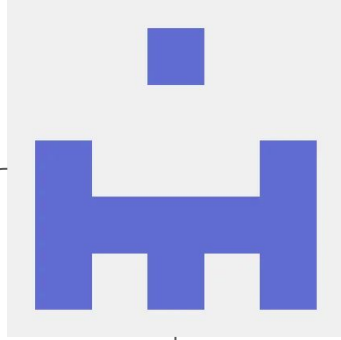
TA17 – Maximize Exposure

**But** : produire un outil de partage pour les différentes communautés regroupant les techniques FIMI/Désinformation corrélées avec quelles TTP

**Comment ?**: en rapprochant le monde du contre FIMI avec celui de la cybersécurité

**Qu'est-ce que l'on récupère ?** : TTP valeurs discrètes : les catégories (primaire / secondaire) - victimes - durée - sévérité - acteur (état, non état, affilié état, ...), motivation ...

# Reconnaissance de texte sur des memes



**The Clod Squad**

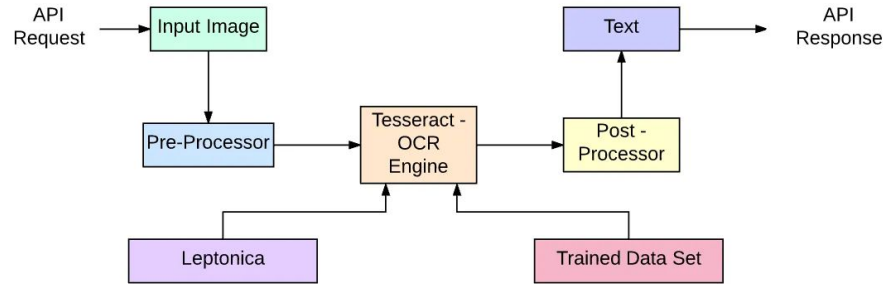


**aka "The Four Donkeys of  
The Islamo-Marxist Apocalypse"**

The Clod Squad aka 'The Four Donkeys of Tho Istamo-Marxist Apocalypse'

Amélioration ?

OCR Process Flow



# Analyse intimacy

- Quantifying Intimacy in Language => on
  - texts (reddit questions)
  - score between  $[-1, 1]$
  - code => Roberta
- SemEval 2023 Task 9: Multilingual Tweet Intimacy Analysis => on tweets
  - data from SemEval 2023 => text with score between  $[1, 5]$  + langue



# analyse de sentiments

Sur des classifieur pré-entraînés

**Vader** (Valence Aware Dictionary and Sentiment Reasoner): Les scores **positif**, **neutre**, et **négatif** sont des ratios pour les proportions de texte qui correspondent à chaque catégorie (ils s'additionner pour être égaux à 1 ou proche)

		TEXT	Negatif	Neutre	Positif	Composé	Sentimen
0	20030219	aba decides against community broadcasting licence	(0.0	1.0	0.0	0.0	'Neutral'
1	20030219	act fire witnesses must be aware of defamation	(0.255	0.745	0.0	-0.34	'Negative'
2	20030219	a g calls for infrastructure protection summit	(0.0	1.0	0.0	0.0	'Neutral'
3	20030219	air nz staff in aust strike for pay rise	(0.293	0.707	0.0	-0.2263	'Negative'

**Happy Transformer** : Labels « **positifs** », « **négatif** », ou « **neutre** » décrivent la polarité du sentiment. Le score = valeur flottante comprise dans [0:1] reflétant l'intensité du sentiment

		TEXT	LABEL	SCORE
0	20030219	aba decides against community broadcasting licence	('neutral'	0.5590684413909912)
1	20030219	act fire witnesses must be aware of defamation	('neutral'	0.9211260080337524)
2	20030219	a g calls for infrastructure protection summit	('neutral'	0.7539046406745911)
3	20030219	air nz staff in aust strike for pay rise	('negative'	0.9383648037910461)

**TextBlob** : Le score de polarité est une valeur flottante comprise dans l'intervalle [-1.0, 1.0], tandis que la subjectivité est une valeur flottante comprise dans l'intervalle [0.0, 1.0], où 0 est très objectif et 1 est très subjectif



		TEXT	Polarité	Subjectivité
0	20030219	aba decides against community broadcasting licence	(0.0	0.0)
1	20030219	act fire witnesses must be aware of defamation	(0.25	0.25)
2	20030219	a g calls for infrastructure protection summit	(0.0	0.0)
3	20030219	air nz staff in aust strike for pay rise	(0.0	0.0)

# SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images

16 articles

3 repo

1 repo fonctionnel

	
<p><b>Predicted:</b> Flag-waving, Name calling/Labeling, Smears, Transfer</p> <p><b>GT:</b> Smears, Appeal to (Strong) Emotions, Name calling/Labeling, Flag-waving, Transfer</p>	<p><b>Predicted:</b> Appeal to fear/prejudice, Exaggeration/Minimisation, Loaded Language, Name calling/Labeling, Smears</p> <p><b>GT:</b> Glittering generalities (Virtue), Loaded Language, Name Calling/Labeling, Smears, Transfer, Whataboutism</p>

100% 12/12 [00:01<00:00, 7.34it/s]

MacroF1: 0.2599678421983516

MicroF1: 0.5791505791505791

DONE!

# Conclusion

**SemEval 23 - Task 3 : Genre,  
Framing, Persuasion Techniques  
in Multi-lingual News**

**SemEval 23 - 4 : Human Value  
Detection**

**SemEval 20 - Task 9: Sentiment  
Analysis for Code-Mixed Social  
Media Text**

## Lifecycle Models

*Figure 1 - The emergency cycle*

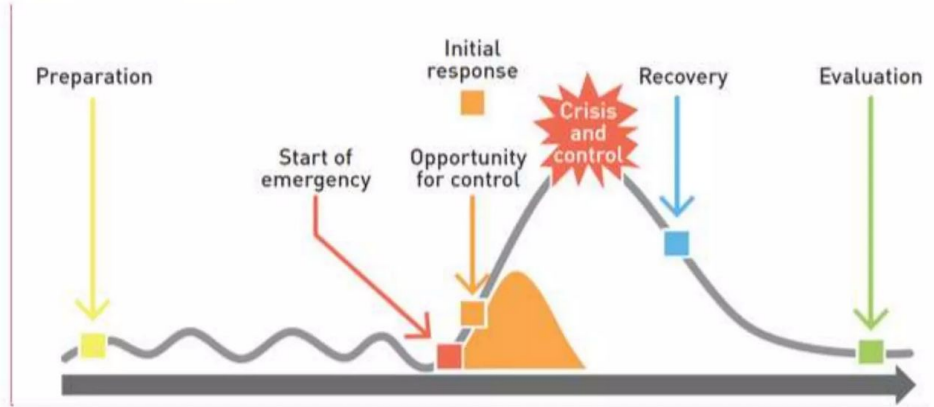


Image: WHO Europe