Journée ICO – 05/12/2025

Antony Dalmiere

Master cybersécurité

Master psychologie

Thèse CNRS

Encadrants

Vincent Nicomette

Guillaume Auriol

Pascal Marchand

Contexte

23%

taux de réussite de phishing peu ciblé

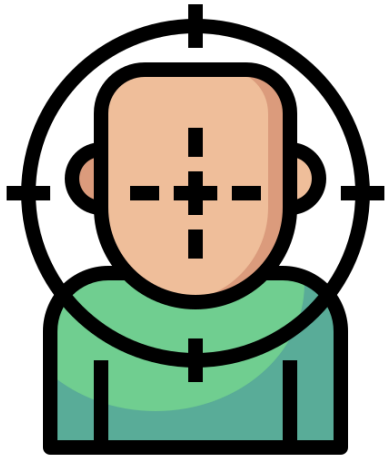
Bakhshi, T., Papadaki, M., & Furnell, S. (2009). Social engineering : Assessing vulnerabilities in practice. Information Management & Computer Security, 17(1), 53-63. https://doi.org/10.1108/09685220910944768

La théorie

PSYCHOLOGIE SOCIALE

PSYCHOLOGIE COGNITIVE

MARKETING

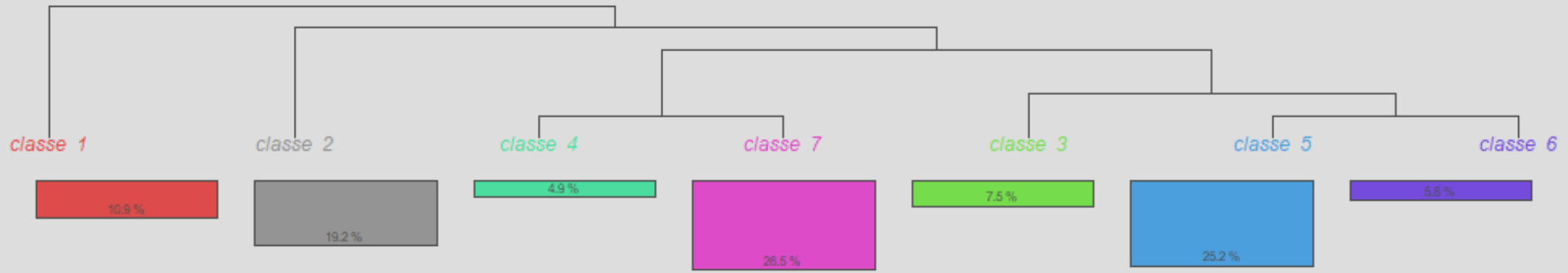



Social
Engineer
Similarity
Include-Half-Confidential-Details
Attractiveness
Authority
Contact-Data-Present
Semantic-Attack
Secure-Communication-Channel
Use-Statistics
Obscurantism
Divergence
Reciprocity
Fake-consistency
Flattery
Scarcity
Foot-in-the-door
Door-in-the-face
Request-Minor-Favours
Curiosity-Appeal
Reverse-Psychology
Threatening
Imply
Need
Emotion
Decline
Request
A classification of manipulation technique used in social engineering attacks and underlying cognitive biases, needs, norms, and emotions
Antony Dalmiere, Vincent Nicomette, Guillaume Auriol, Pascal Marchand
To cite this version:
Antony Dalmiere, Vincent Nicomette, Guillaume Auriol, Pascal Marchand. A classification of manipulation technique used in social engineering attacks and underlying cognitive biases, needs, norms, and emotions. Thecon25, Apr 2025, Toulouse, France. hal-05027416
HAL Id: hal-05027416
https://hal.science/hal-05027416v1
Submitted on 9 Apr 2025
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.
L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DÉTECTION PAR IA DES TECHNIQUES DE MANIPULATION PSYCHOLOGIQUE

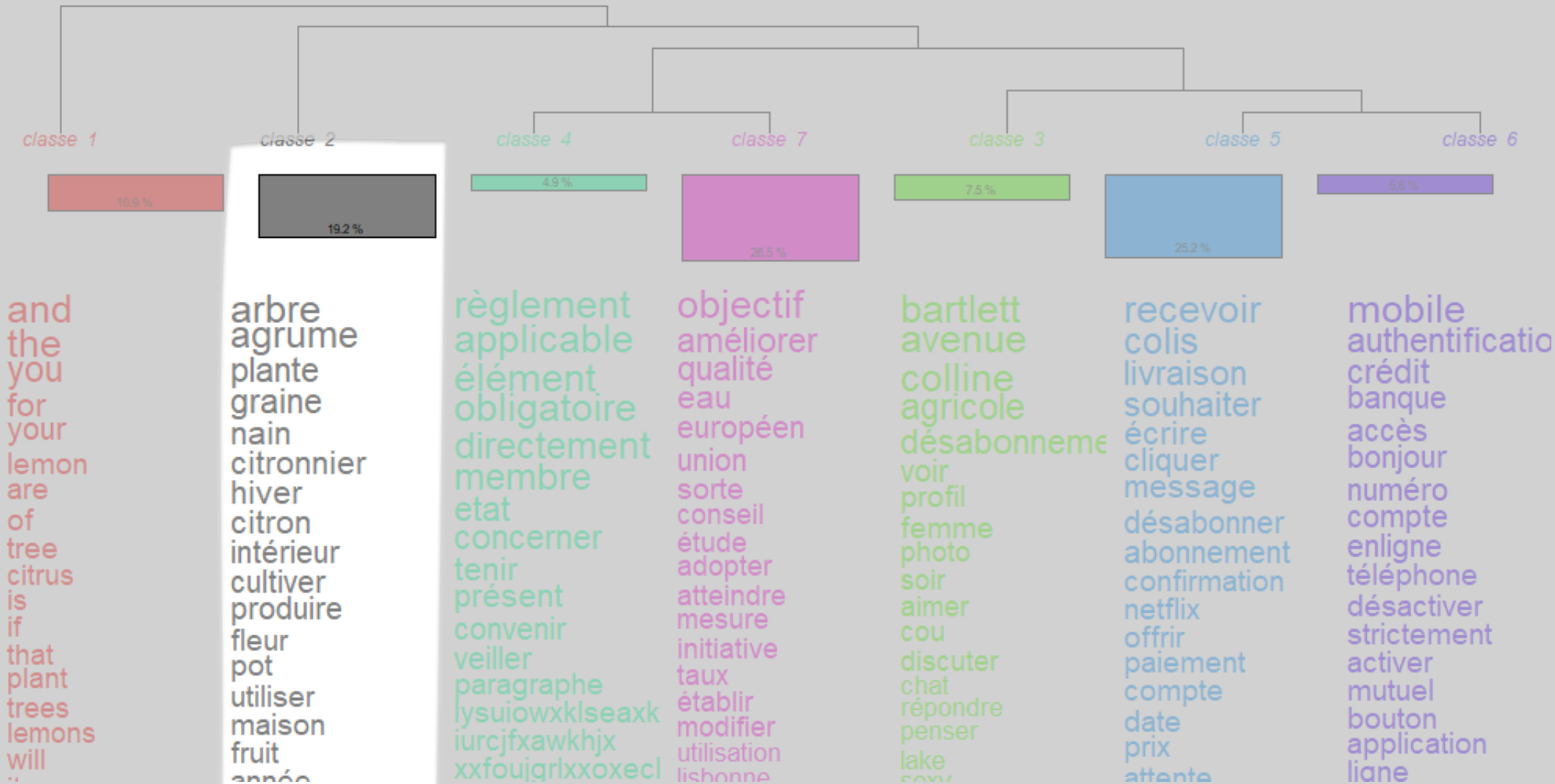
LE GANG DES AGRUMES

Le gang des agrumes

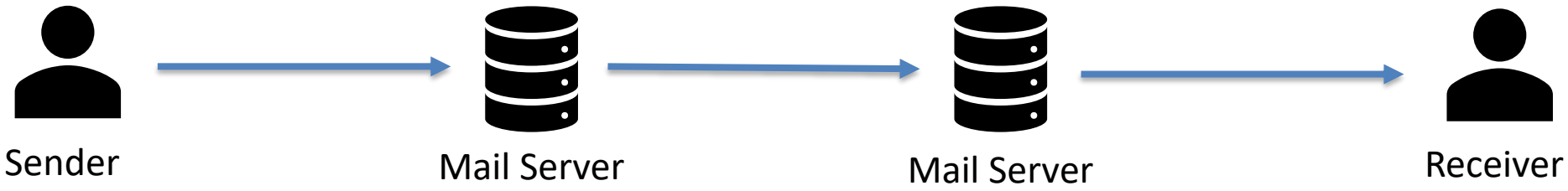


classe 1	classe 2	classe 4	classe 7	classe 3	classe 5	classe 6
and	arbre	règlement	objectif	bartlett	recevoir	mobile
the	agrume	applicable	améliorer	avenue	colis	authentificatio
you	plante	élément	qualité	colline	livraison	crédit
for	graine	obligatoire	eau	agricole	souhaiter	banque
your	nain	directement	européen	désabonneme	écrire	accès
lemon	citronnier	membre	union	voir	cliquer	bonjour
are	hiver	etat	sorte	profil	message	numéro
of	citron	concerner	conseil	femme	désabonner	compte
tree	intérieur	tenir	étude	photo	abonnement	enligne
citrus	cultiver	présent	adopter	soir	confirmation	téléphone
is	produire	convenir	atteindre	aimer	netflix	désactiver
if	fleur	veiller	mesure	cou	offrir	strictement
that	pot	paragraphe	initiative	discuter	paiement	activer
plant	utiliser	lysuiowxklseaxk	taux	chat	compte	mutuel
trees	maison	iurcjfxawkhjx	établir	répondre	date	bouton
lemons	fruit	xxfoujgrlxxoxecl	modifier	penser	prix	application
will	année		utilisation	lake	attente	ligne
			lisbonne	sexy		

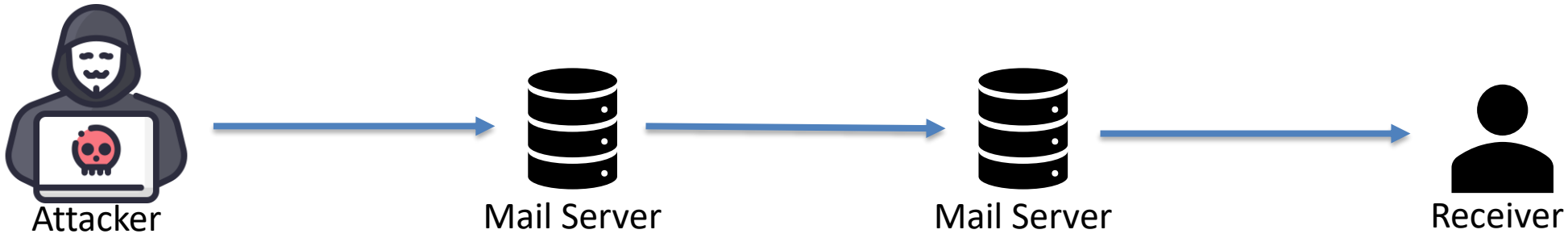
Le gang des agrumes



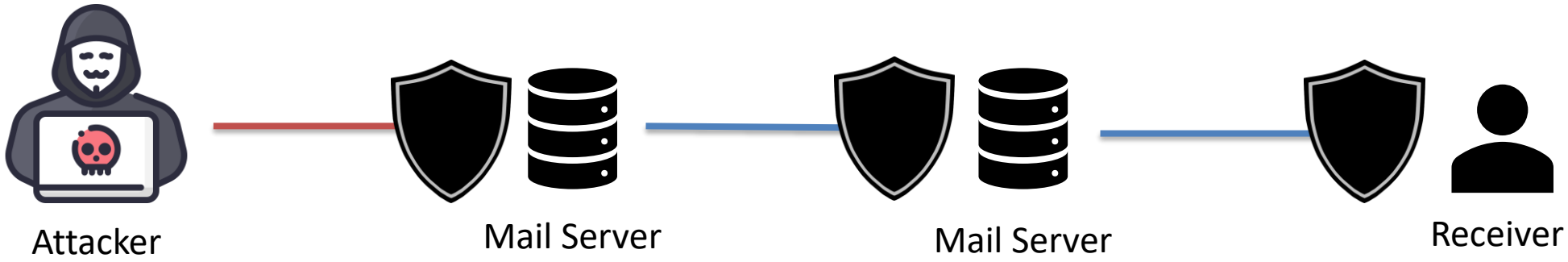
Comment ça arrive dans la boite mail ?



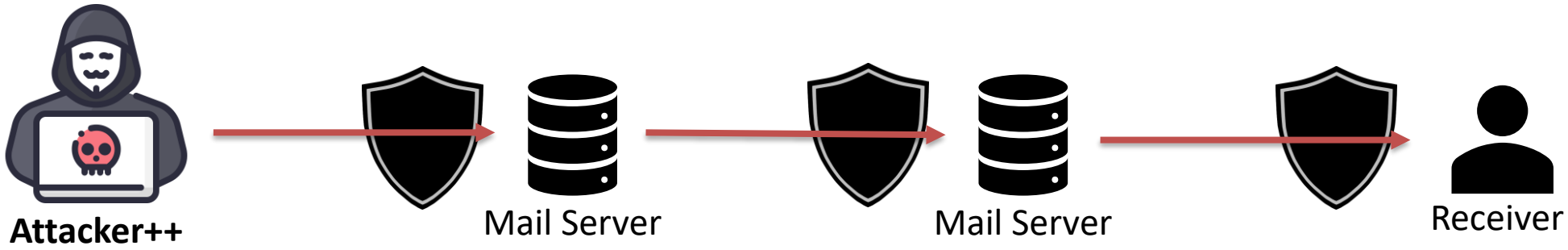
Comment ça arrive dans la boite mail ?



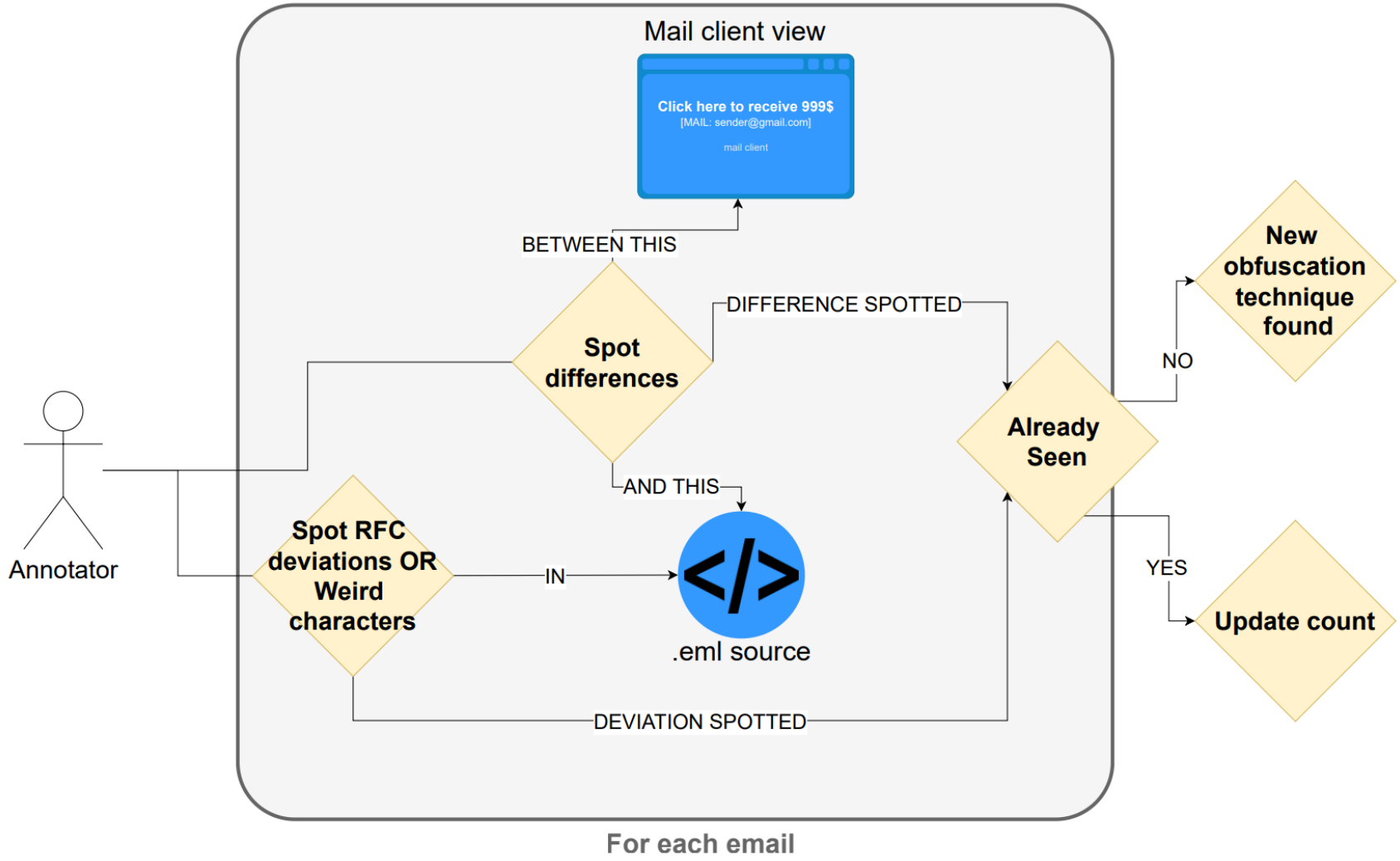
Comment ça arrive dans la boite mail ?



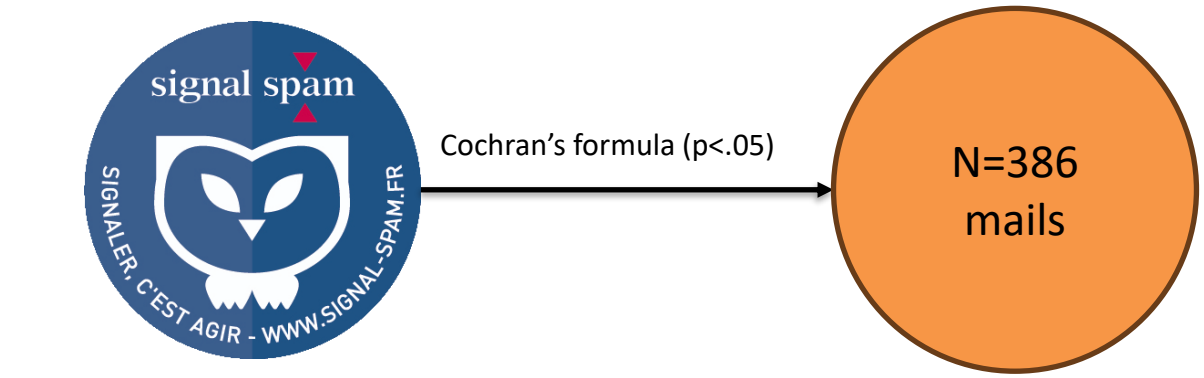
Comment ça arrive dans la boite mail ?



Notre méthode pour le savoir

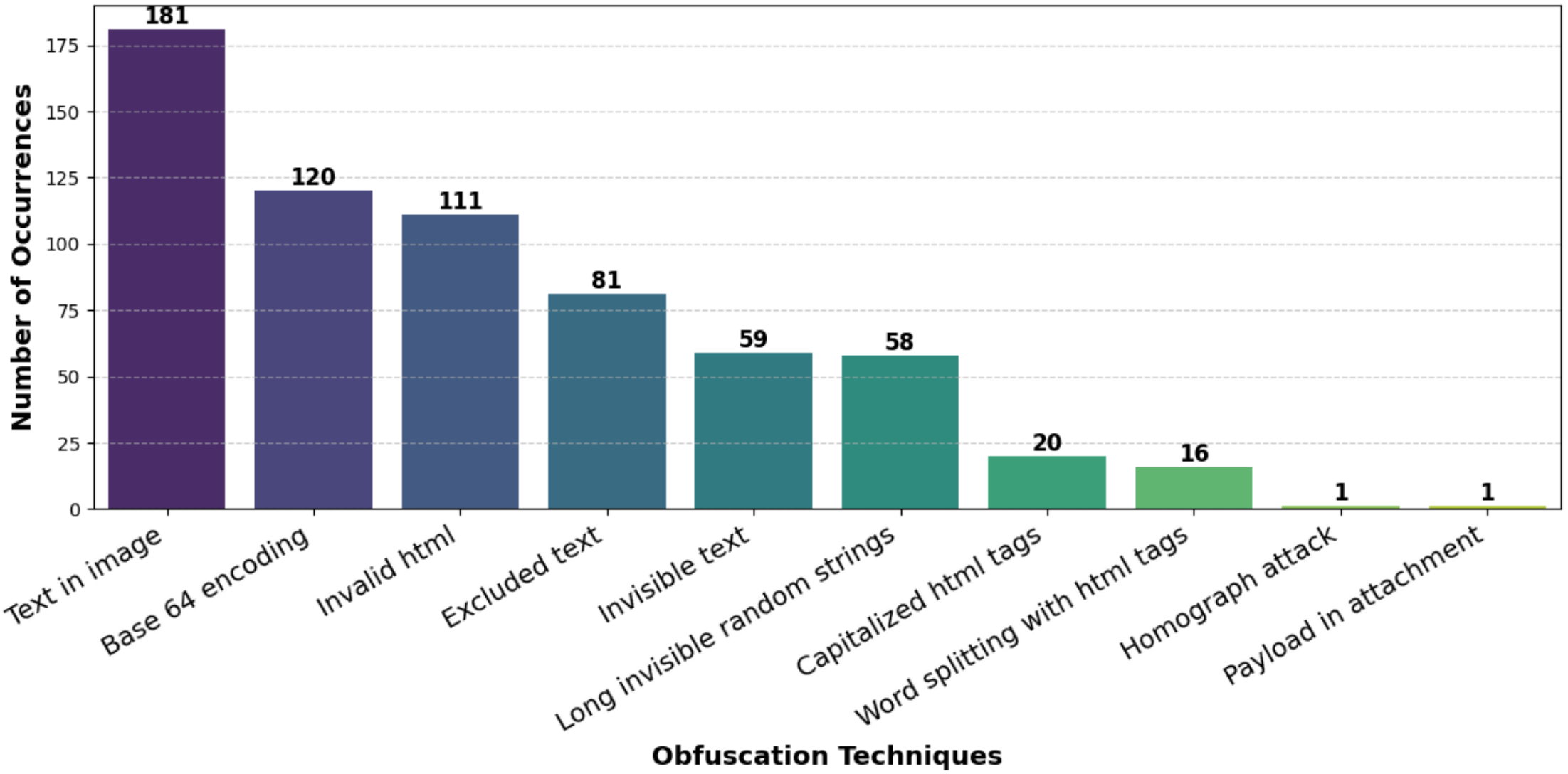


Dataset



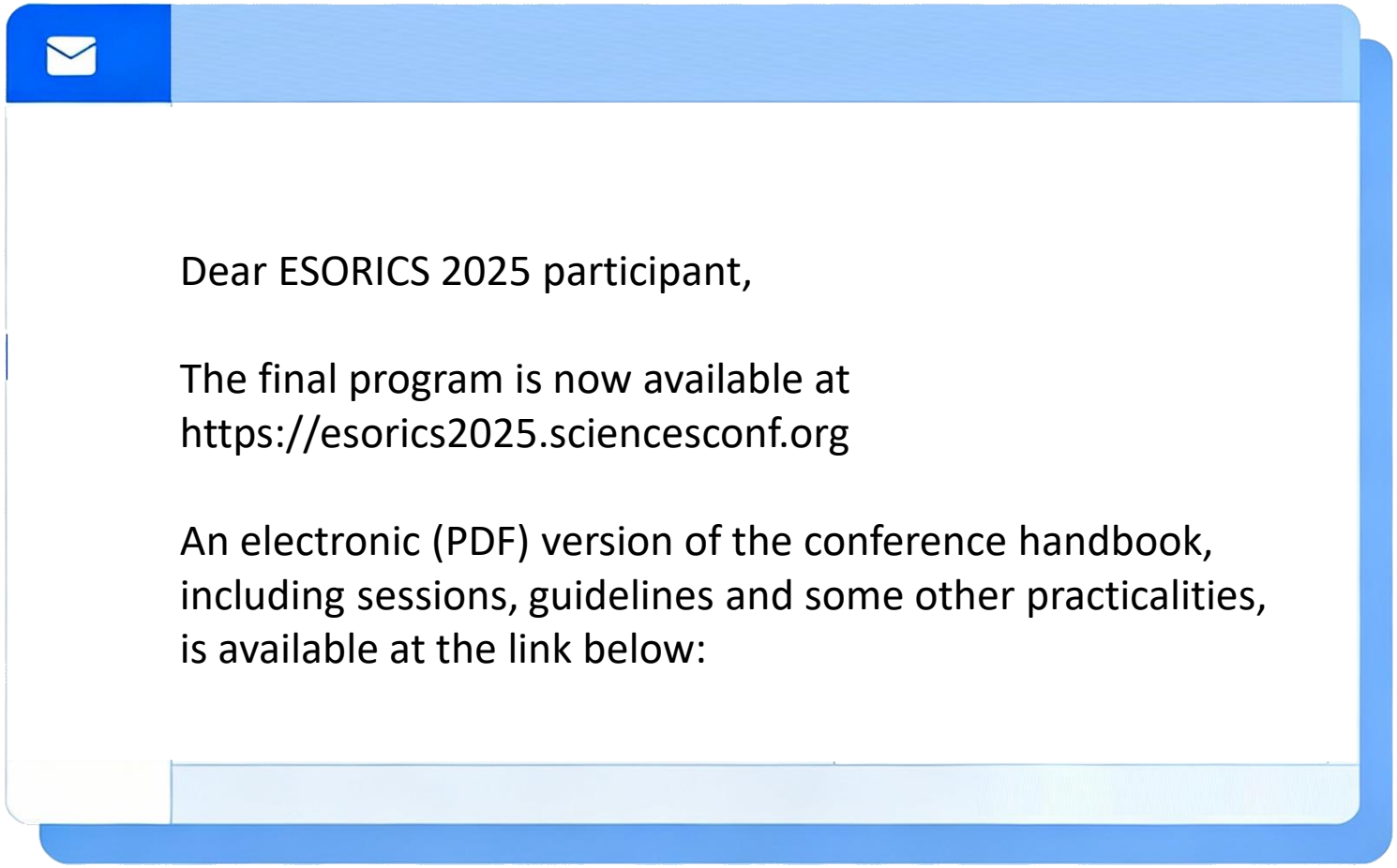
Mail de phishing français

Les techniques d'obfuscation



Technique #1 Text in image

Vu par les filtres



Vu par des humains

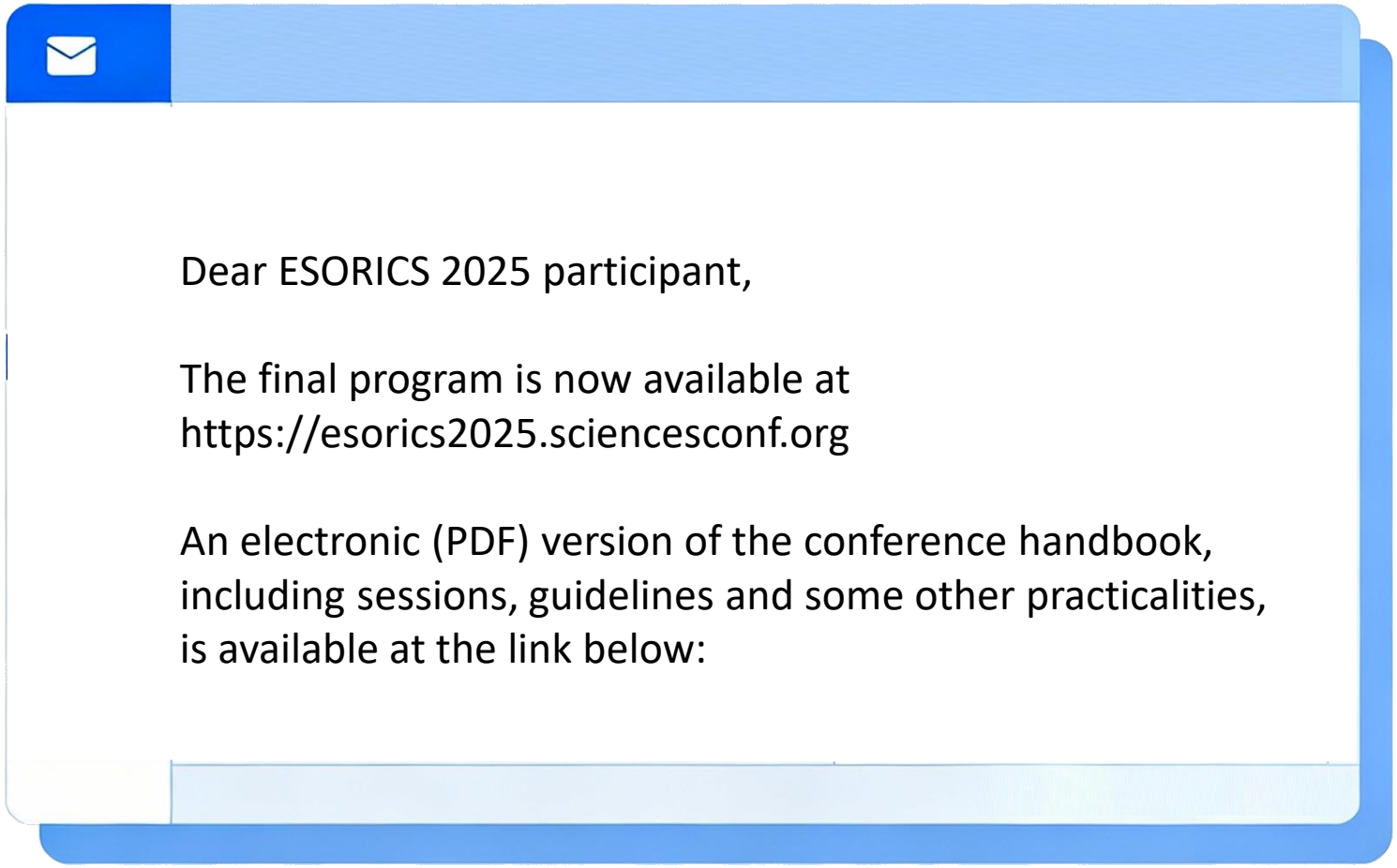
Technique #5 Invisible Text

```
<html>
  <body>
   Dear ESORICS 2025 participant,

The final program is now available at
https://esorics2025.sciencesconf.org
[....]

    <p stype="font-size:0px; font-
color:white">
      Hemoglobin is an iron-containing,
tetrameric protein in red blood cells
that binds oxygen in the lungs and
delivers it to body tissues.
    </p>
  </body>
</html>
```

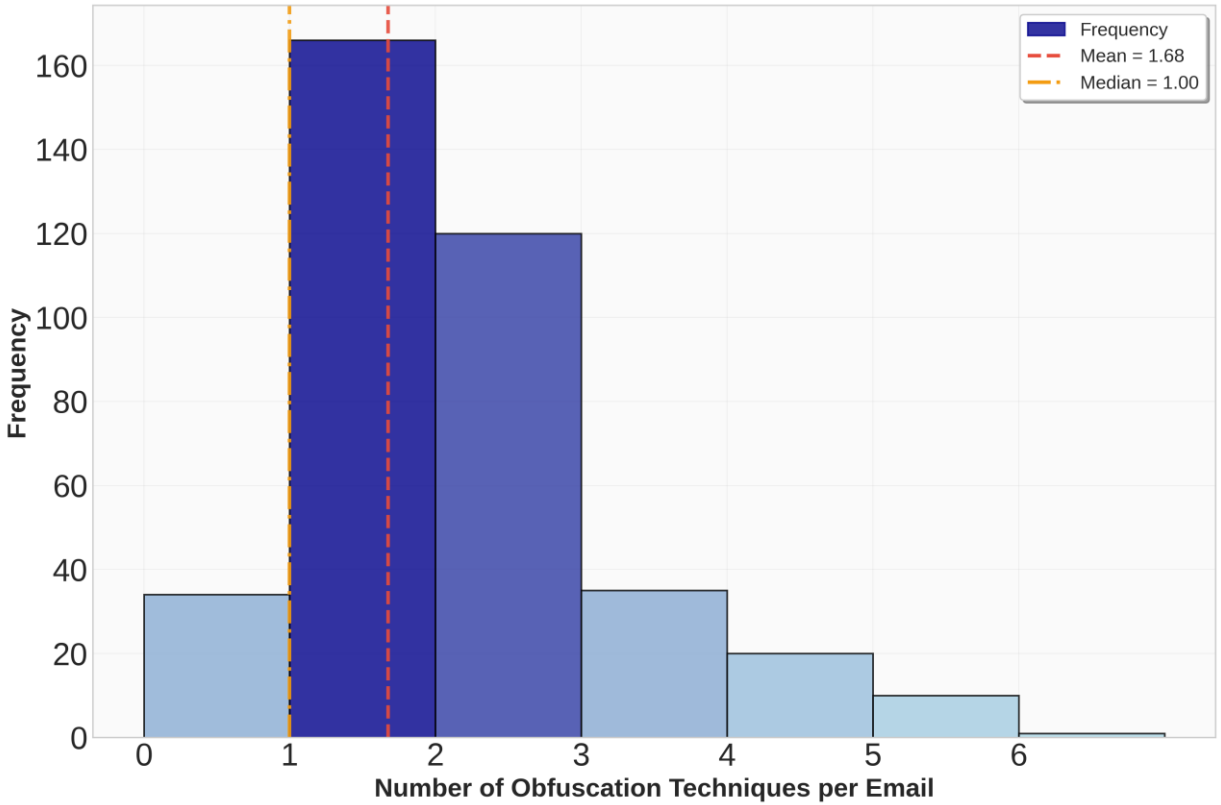
Vu par les filtres



Vu par des humains

COMMENT CES TECHNIQUES D'OBFUSCATION SE COMBINENT-ELLES

Combinaison des techniques d'obfuscation



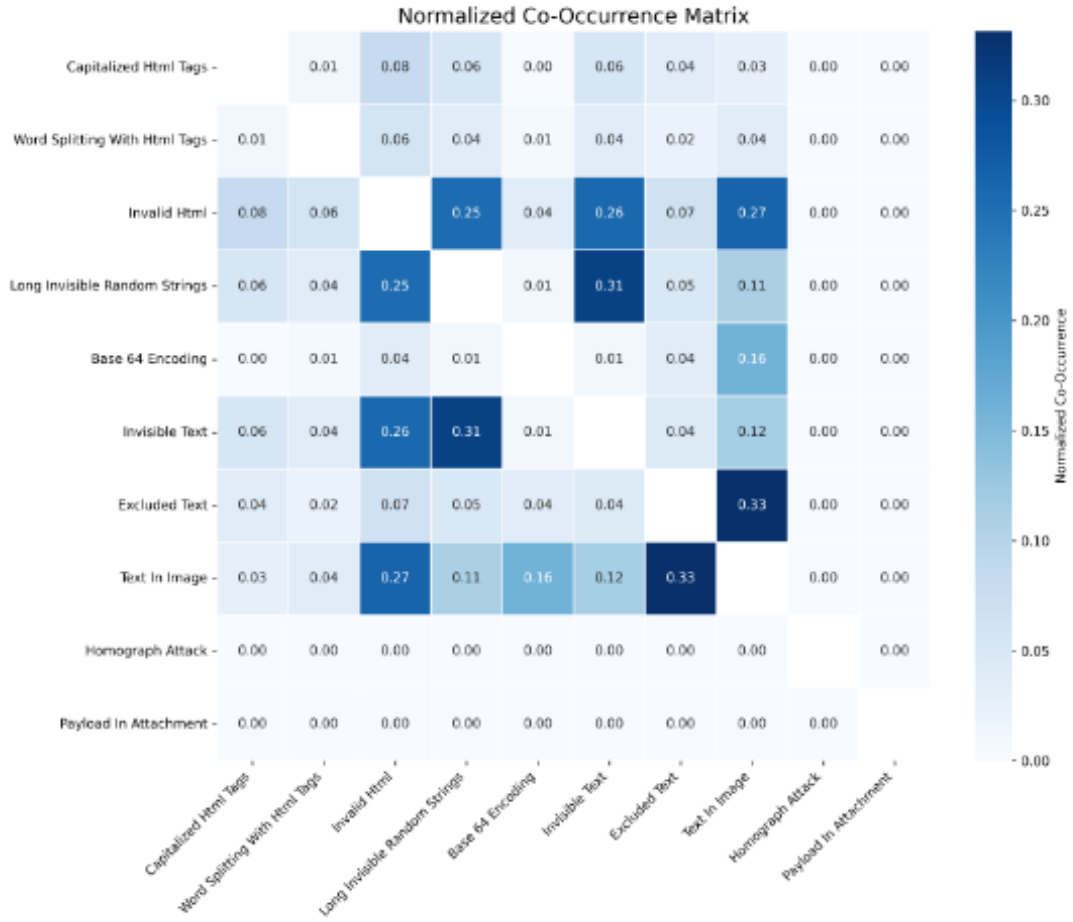
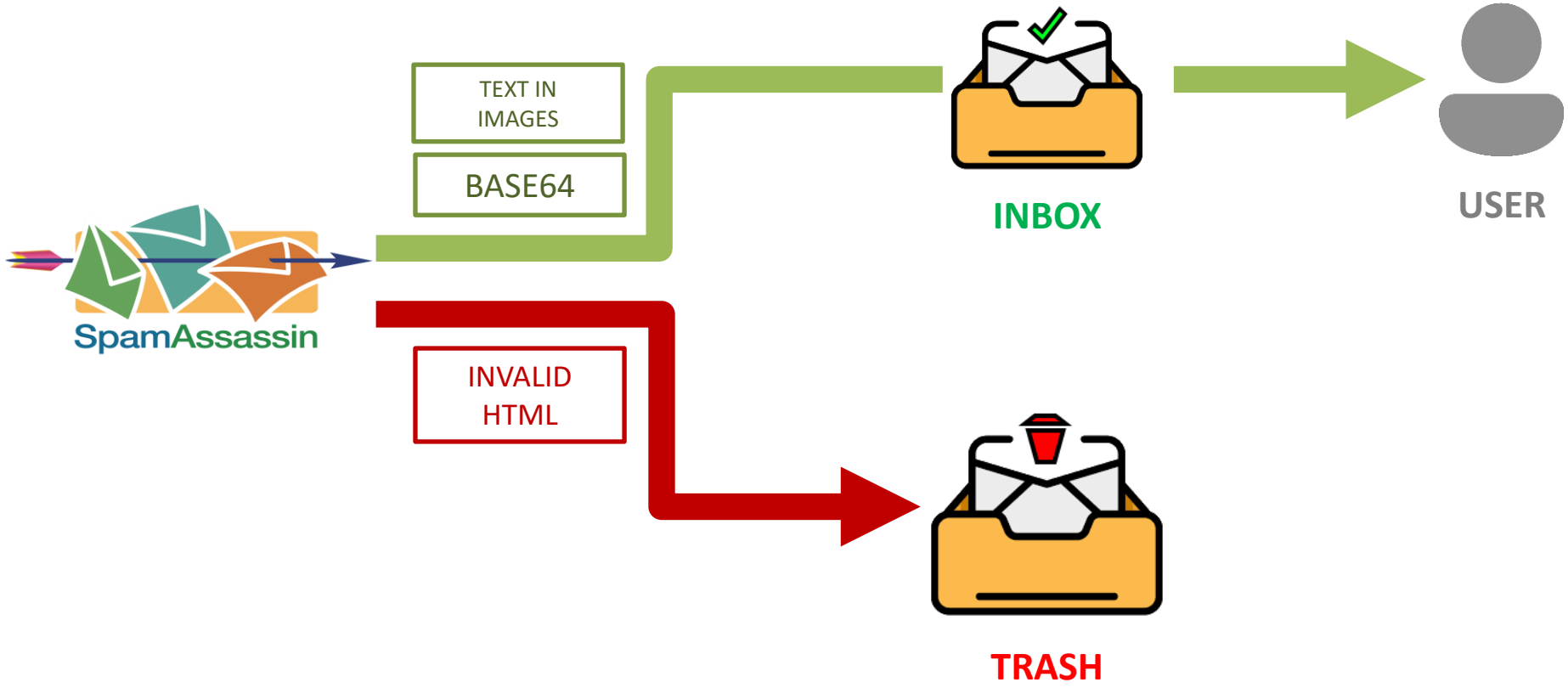
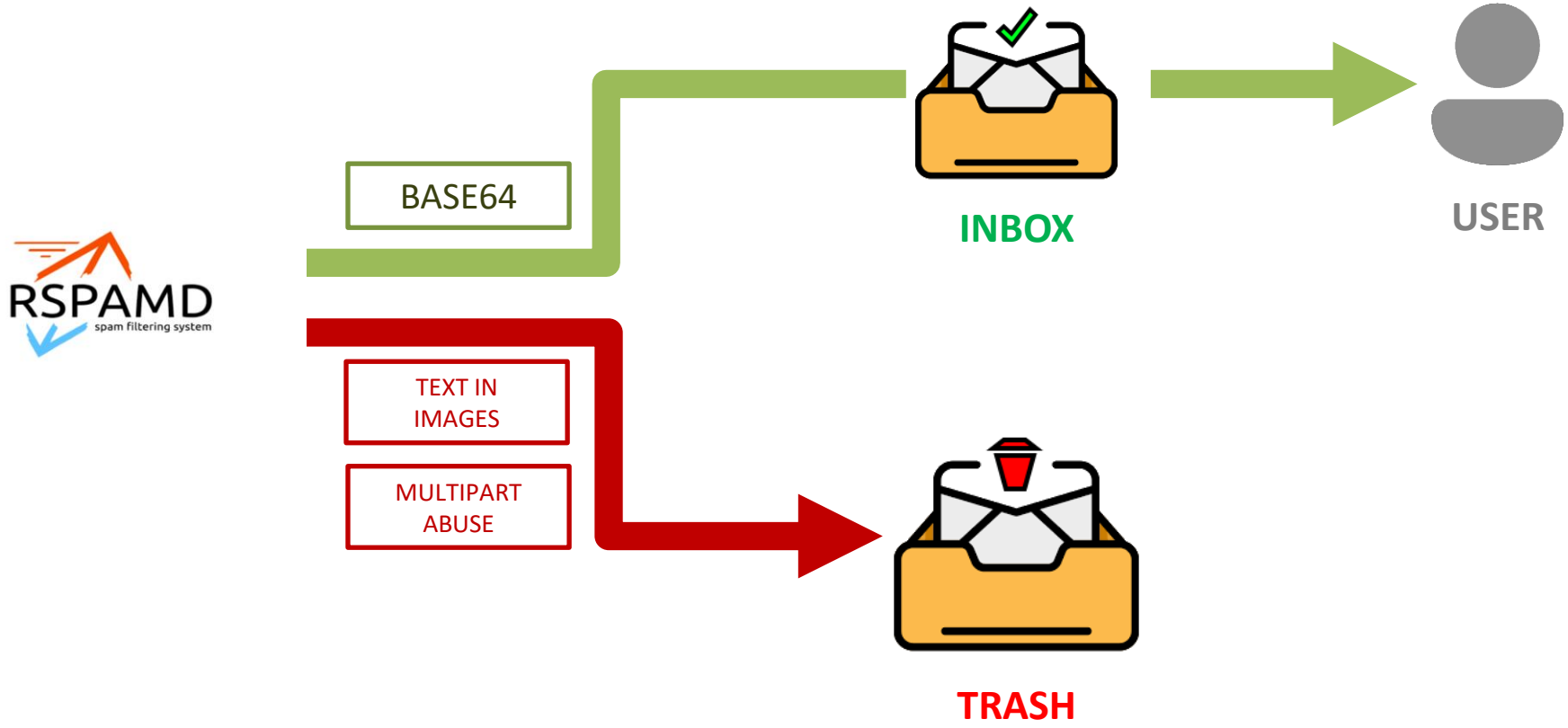


Fig. 2. Normalized co-occurrence matrix of obfuscation techniques

Efficacité contre les antispams



Efficacité contre les antispams



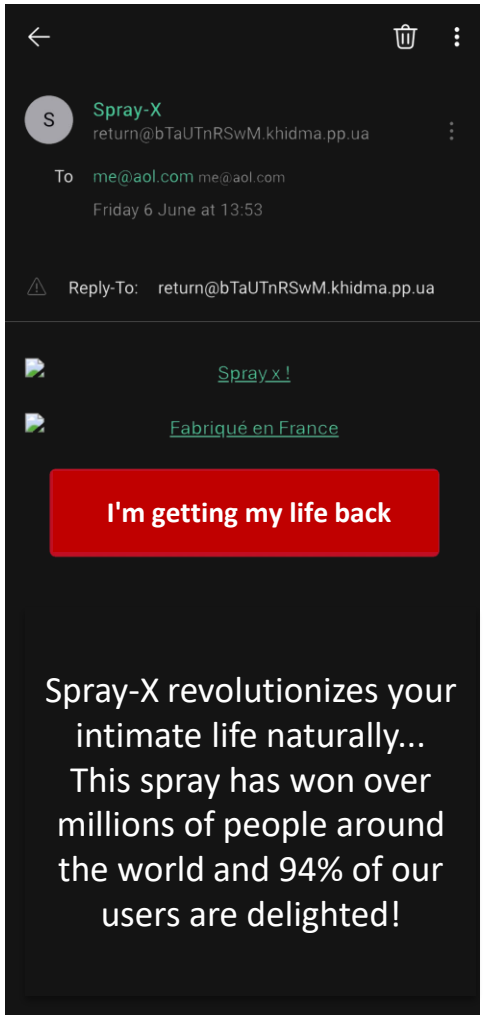
Qu’avons-nous appris ?

Base64 est efficace contre les antispams testés.

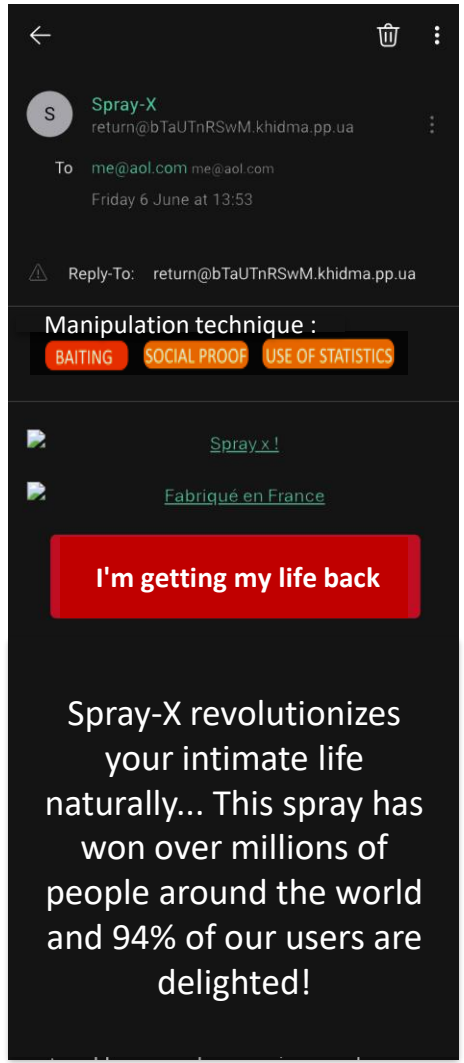
Les techniques d'obfuscation peuvent être contre-productives et augmenter le score de spam.

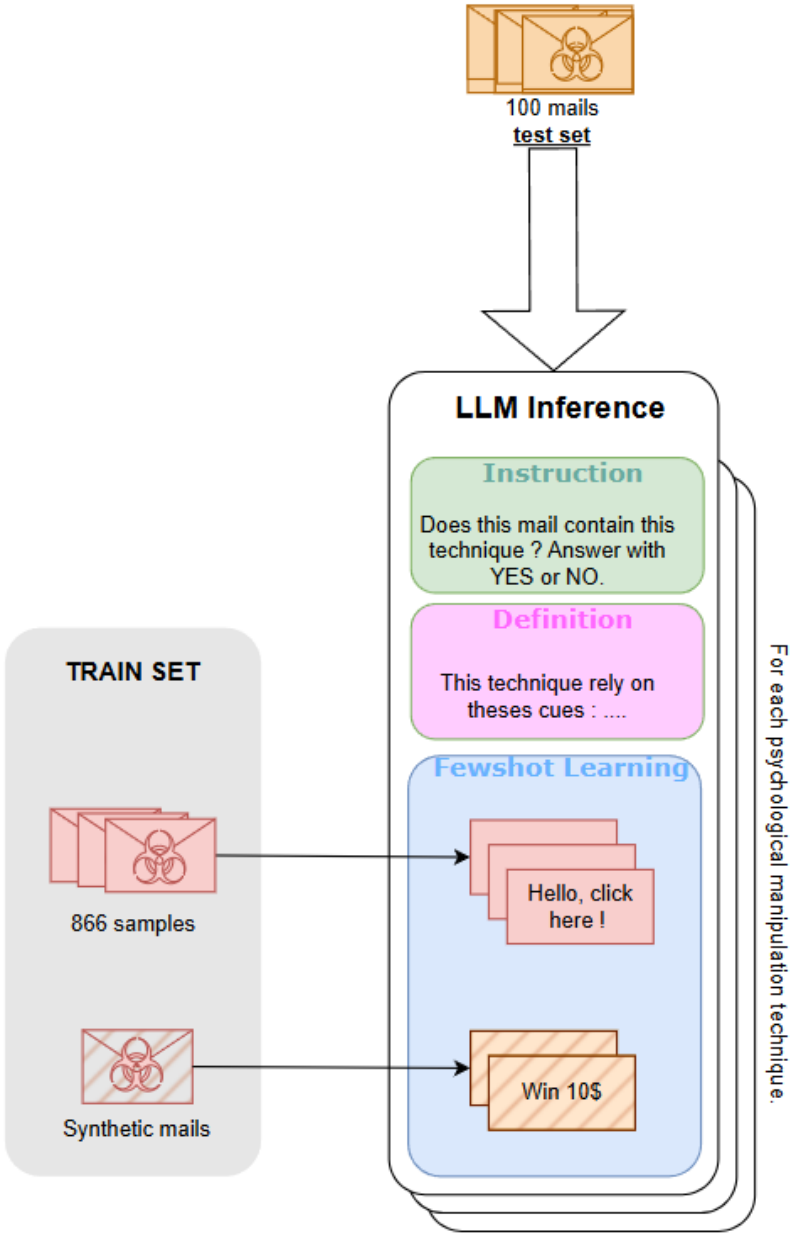
DÉTECTION PAR IA DES TECHNIQUES DE MANIPULATION PSYCHOLOGIQUE

Cas d'application



Comment ?





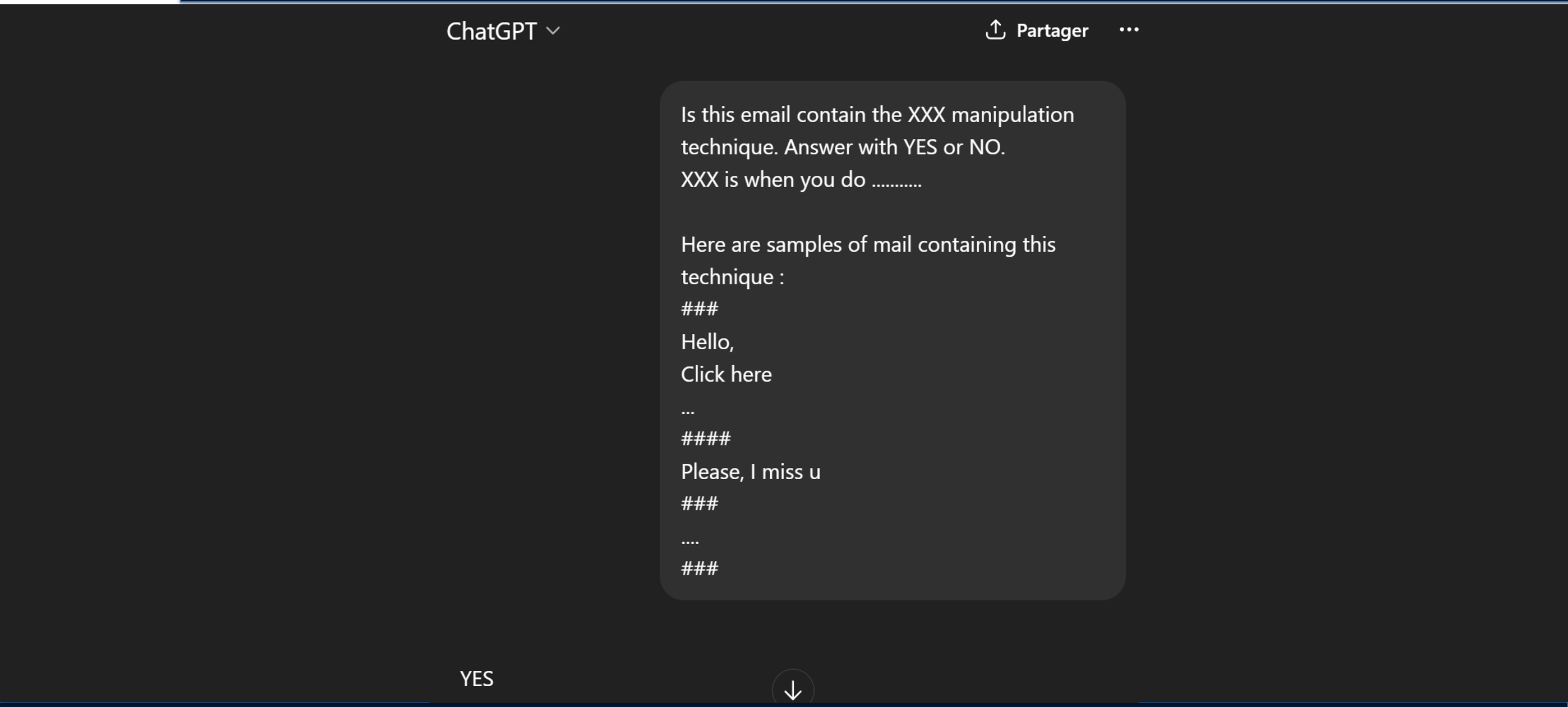
IN-CONTEXT LEARNING

Capacité des LLM à apprendre une nouvelle compétence simplement en la décrivant dans le prompt.

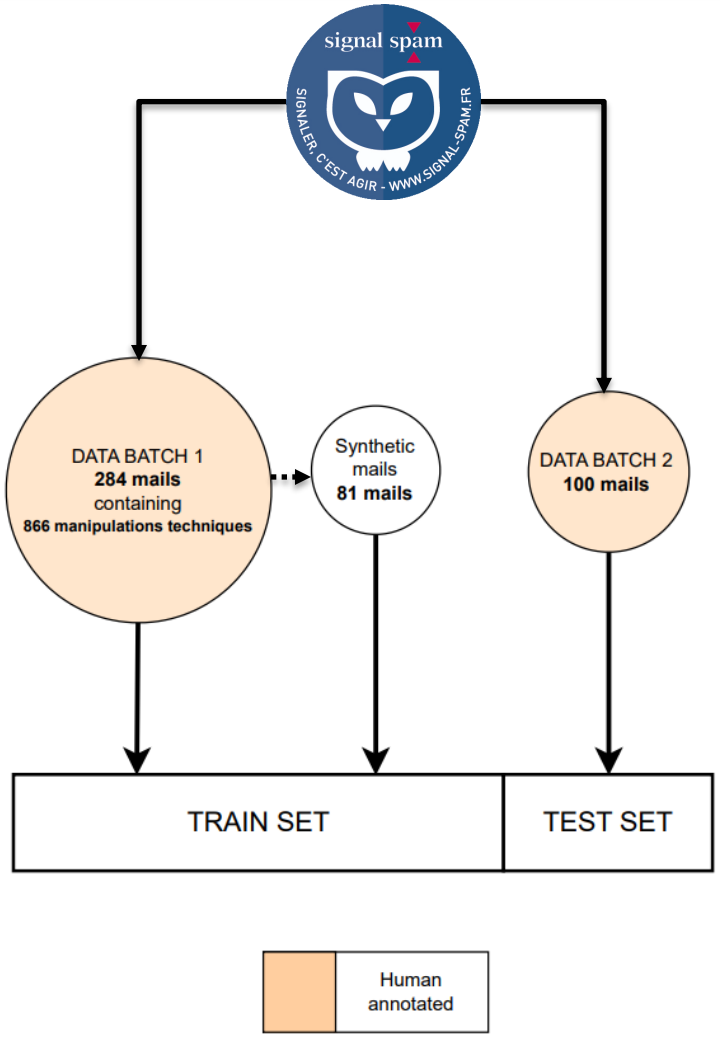
SYNTHETIC DATA

Utilisation d'une IA pour créer les données nécessaires à une autre IA. (Dans notre cas)

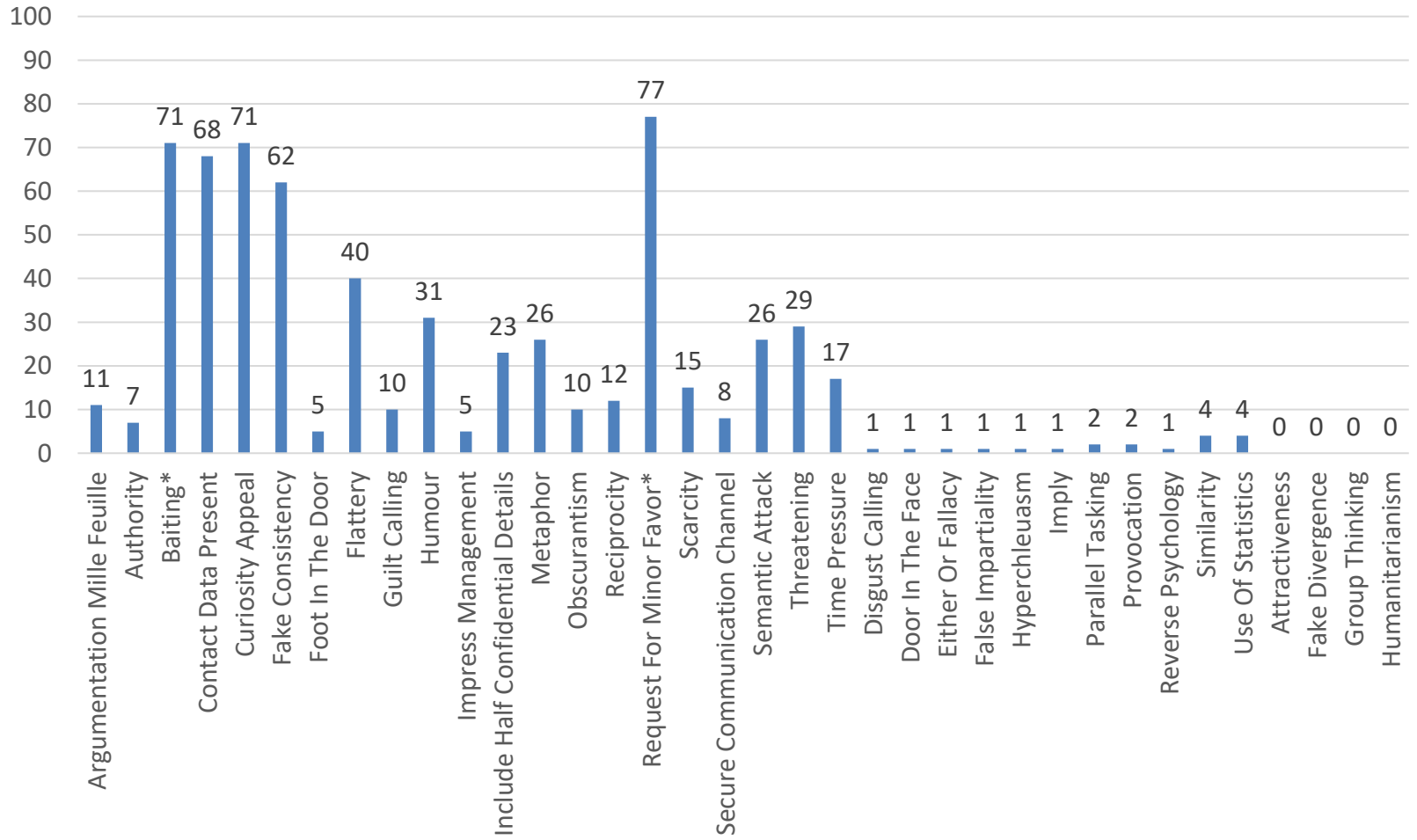
Une approche très simple



Dataset

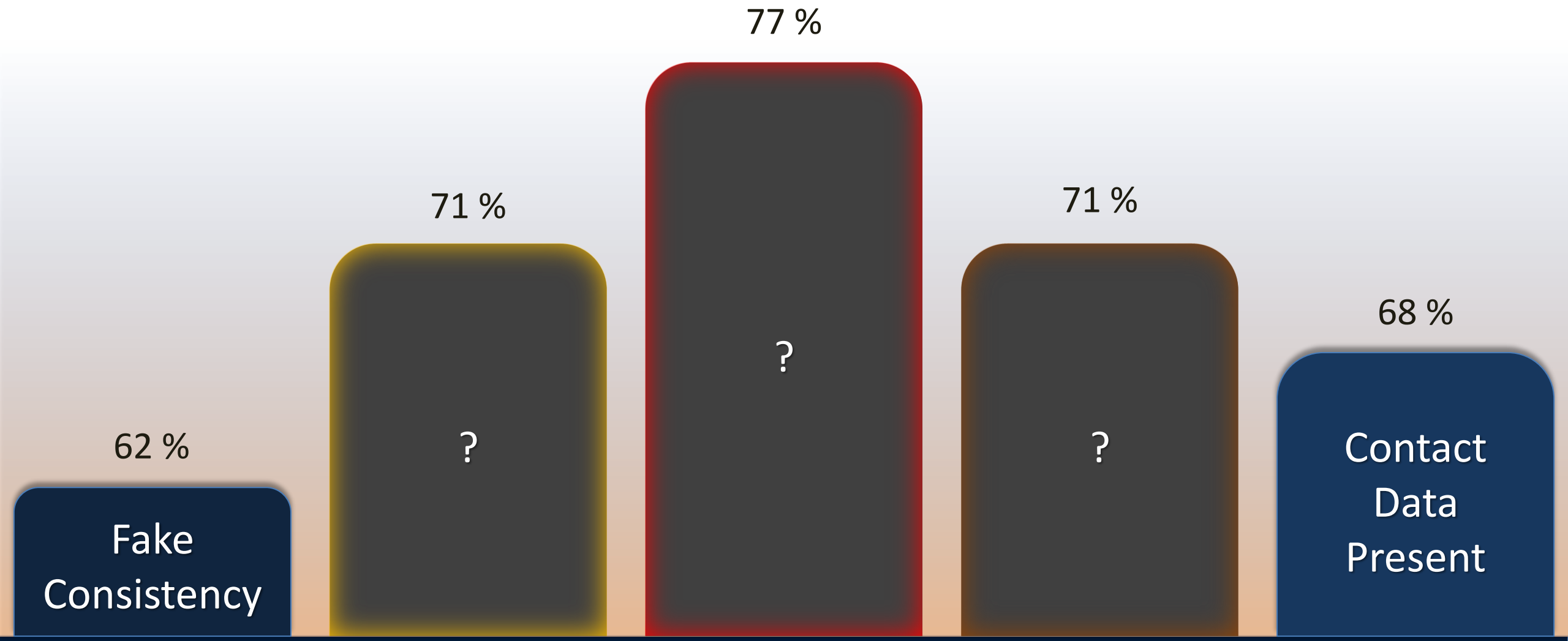

signal spam
SIGNALER, C'EST AGIR - WWW.SIGNAL-SPAM.FR
DATA BATCH 1
284 mails
containing
866 manipulations techniques
Synthetic
mails
81 mails
DATA BATCH 2
100 mails
TRAIN SET
TEST SET
Human
annotated

Utilisation de techniques de manipulation

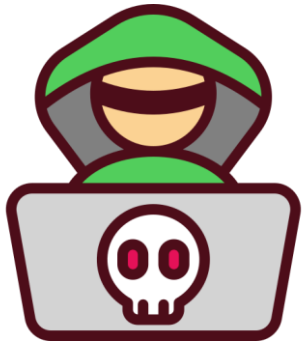


% of use of manipulation techniques in the wild.

Utilisation de techniques de manipulation



TOP 3 : Curiosité



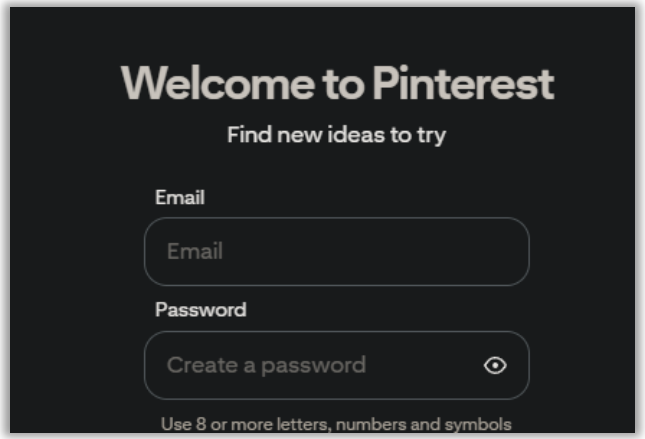
Click here to know more.

secret.pdf

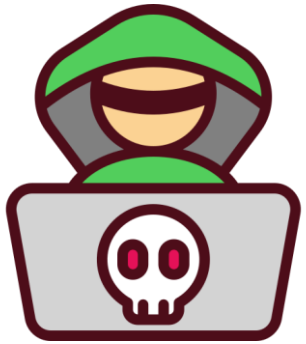
You will never believe it.

Social Engineering

Marketing



TOP 2 : Appatage



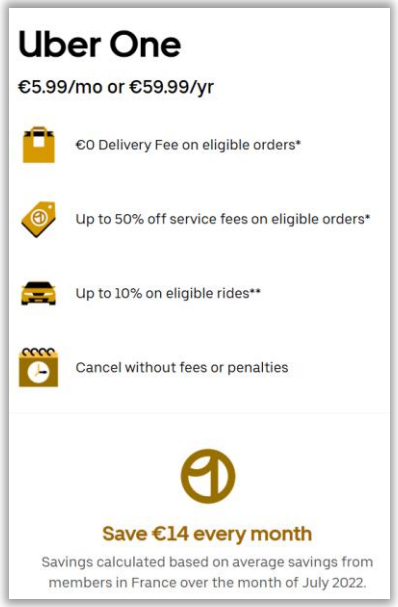
Win 20000$

A hot single near you

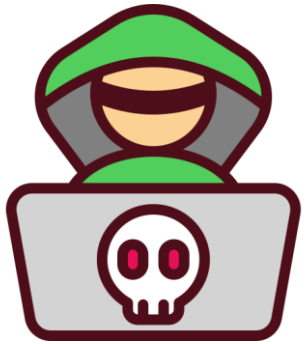
Get a free Iphone

Social Engineering

Marketing



: Demande mineure



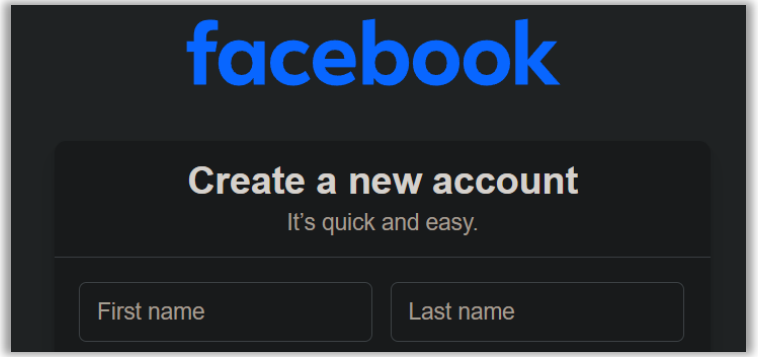
We just need your name

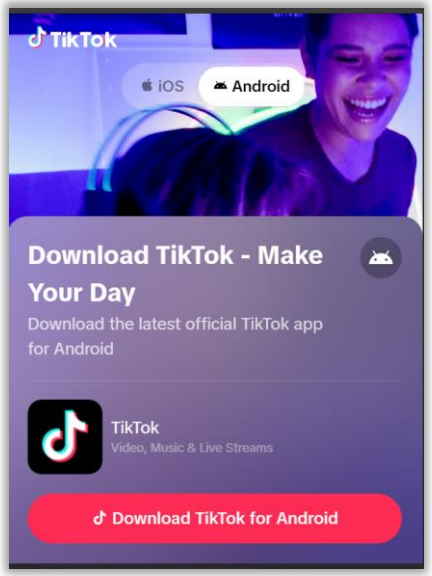
You are one click away from being rich

You just have to reply to this mail

Social Engineering

Marketing





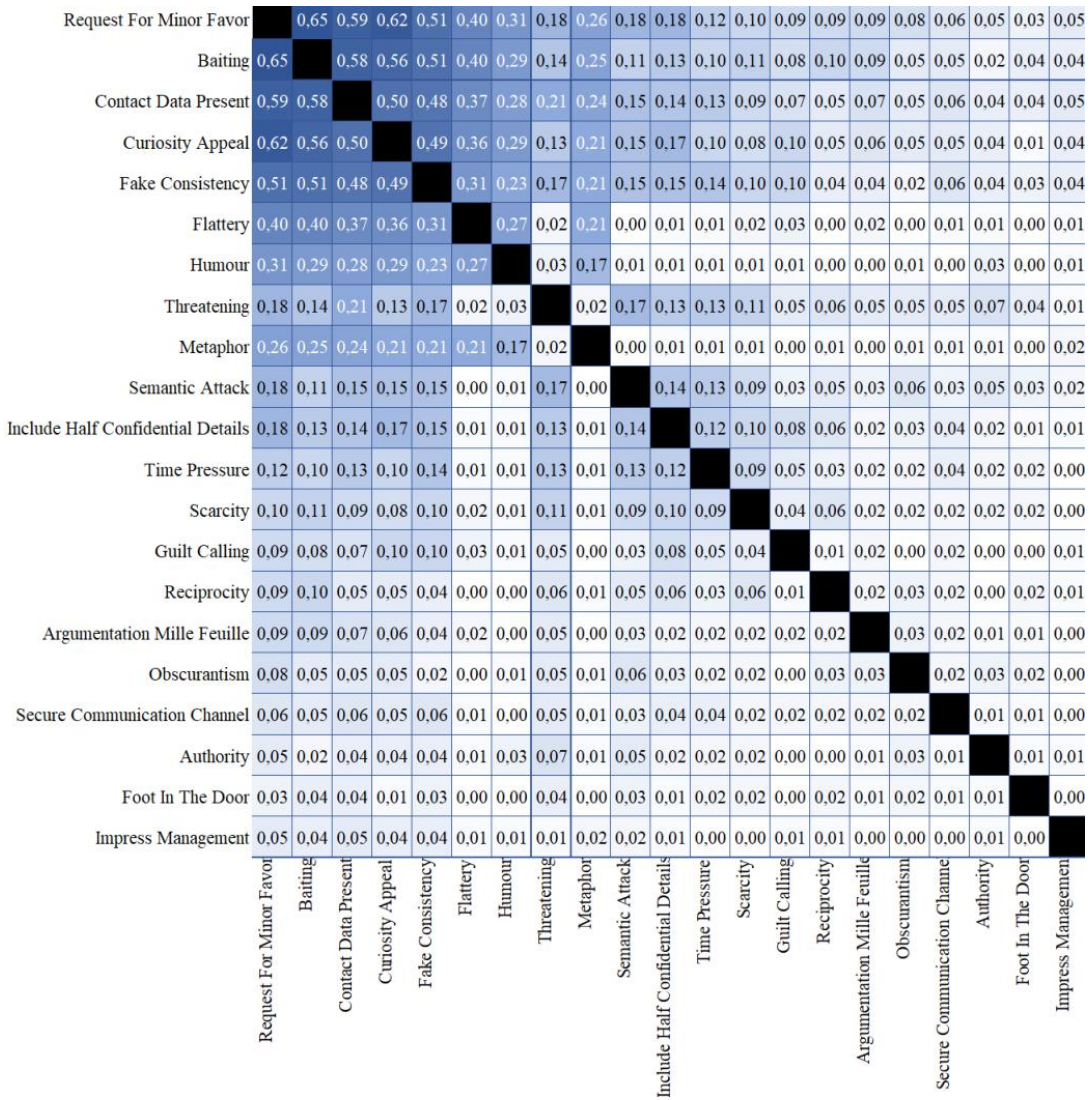
Qu’avons-nous appris ?

Les « vieux scams » sont toujours à la mode.

Les étapes cognitives ciblées en priorité sont la motivation à évaluer et la capacité à évaluer.

COMMENT CES TECHNIQUES DE MANIPULATIONS SE COMBINENT-ELLES

Co-occurrence matrix



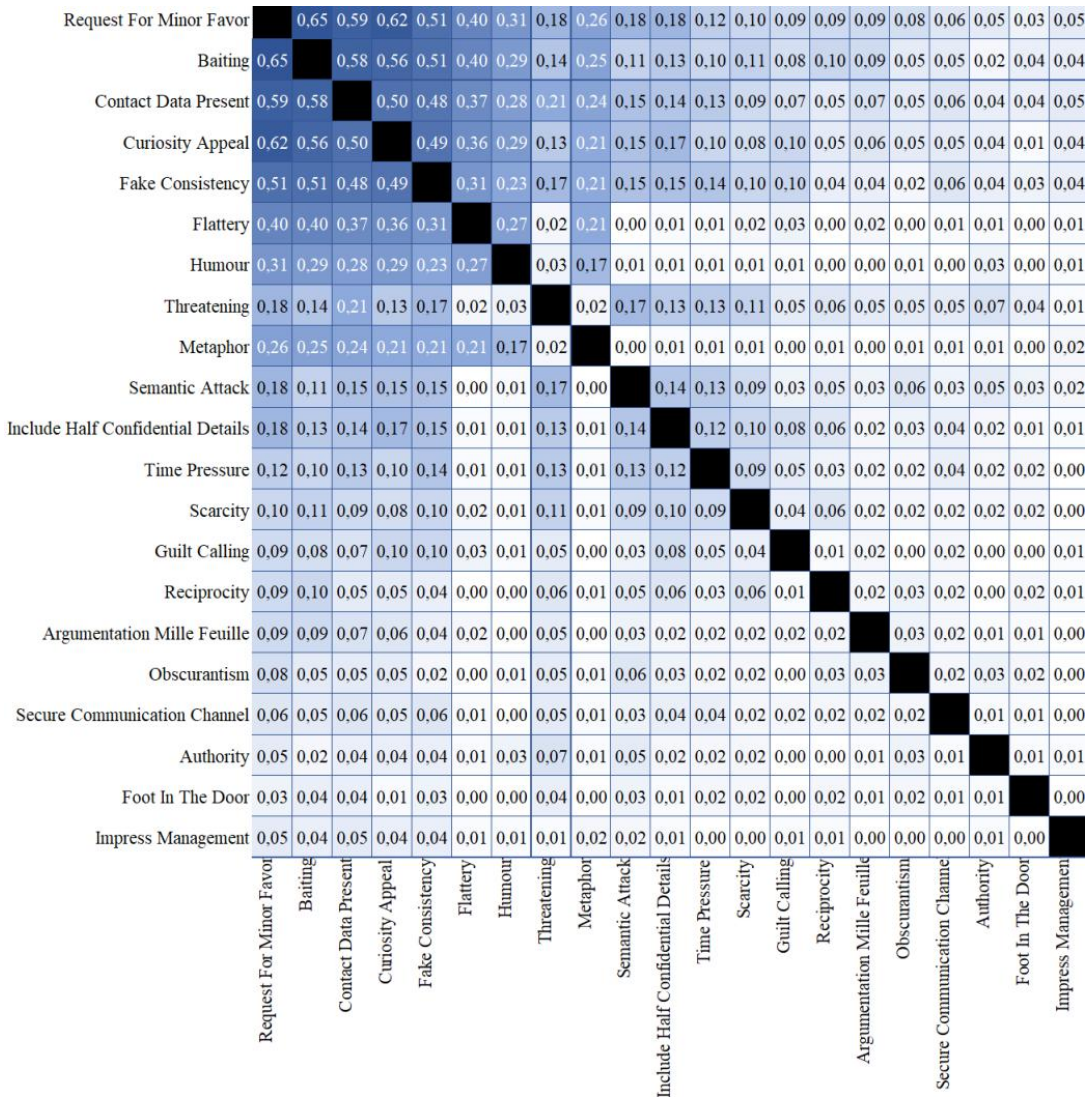
	Request For Minor Favor	Baiting	Contact Data Present	Curiosity Appeal	Fake Consistency	Flattery	Humour	Threatening	Metaphor	Semantic Attack	Include Half Confidential Details	Time Pressure	Scarcity	Guilt Calling	Reciprocity	Argumentation Mille Feuille	Obscurantism	Secure Communication Channe	Authority	Foot In The Door	Impress Management
Request For Minor Favor		0,65	0,59	0,62	0,51	0,40	0,31	0,18	0,26	0,18	0,18	0,12	0,10	0,09	0,09	0,09	0,08	0,06	0,05	0,03	0,05
Baiting	0,65		0,58	0,56	0,51	0,40	0,29	0,14	0,25	0,11	0,13	0,10	0,11	0,08	0,10	0,09	0,05	0,05	0,02	0,04	0,04
Contact Data Present	0,59	0,58		0,50	0,48	0,37	0,28	0,21	0,24	0,15	0,14	0,13	0,09	0,07	0,05	0,07	0,05	0,06	0,04	0,04	0,05
Curiosity Appeal	0,62	0,56	0,50		0,49	0,36	0,29	0,13	0,21	0,15	0,17	0,10	0,08	0,10	0,05	0,06	0,05	0,05	0,04	0,01	0,04
Fake Consistency	0,51	0,51	0,48	0,49		0,31	0,23	0,17	0,21	0,15	0,15	0,14	0,10	0,10	0,04	0,04	0,02	0,06	0,04	0,03	0,04
Flattery	0,40	0,40	0,37	0,36	0,31		0,27	0,02	0,21	0,00	0,01	0,01	0,02	0,03	0,00	0,02	0,00	0,01	0,01	0,00	0,01
Humour	0,31	0,29	0,28	0,29	0,23	0,27		0,03	0,17	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,01	0,00	0,03	0,00	0,01
Threatening	0,18	0,14	0,21	0,13	0,17	0,02	0,03		0,02	0,17	0,13	0,13	0,11	0,05	0,06	0,05	0,05	0,05	0,07	0,04	0,01
Metaphor	0,26	0,25	0,24	0,21	0,21	0,21	0,17	0,02		0,00	0,01	0,01	0,01	0,00	0,01	0,00	0,01	0,01	0,01	0,00	0,02
Semantic Attack	0,18	0,11	0,15	0,15	0,15	0,00	0,01	0,17	0,00		0,14	0,13	0,09	0,03	0,05	0,03	0,06	0,03	0,05	0,03	0,02
Include Half Confidential Details	0,18	0,13	0,14	0,17	0,15	0,01	0,01	0,13	0,01	0,14		0,12	0,10	0,08	0,06	0,02	0,03	0,04	0,02	0,01	0,01
Time Pressure	0,12	0,10	0,13	0,10	0,14	0,01	0,01	0,13	0,01	0,13	0,12		0,09	0,05	0,03	0,02	0,02	0,04	0,02	0,02	0,00
Scarcity	0,10	0,11	0,09	0,08	0,10	0,02	0,01	0,11	0,01	0,09	0,10	0,09		0,04	0,06	0,02	0,02	0,02	0,02	0,02	0,00
Guilt Calling	0,09	0,08	0,07	0,10	0,10	0,03	0,01	0,05	0,00	0,03	0,08	0,05	0,04		0,01	0,02	0,00	0,02	0,00	0,00	0,01
Reciprocity	0,09	0,10	0,05	0,05	0,04	0,00	0,00	0,06	0,01	0,05	0,06	0,03	0,06	0,01		0,02	0,03	0,02	0,00	0,02	0,01
Argumentation Mille Feuille	0,09	0,09	0,07	0,06	0,04	0,02	0,00	0,05	0,00	0,03	0,02	0,02	0,02	0,02	0,02		0,03	0,02	0,01	0,01	0,00
Obscurantism	0,08	0,05	0,05	0,05	0,02	0,00	0,01	0,05	0,01	0,06	0,03	0,02	0,02	0,00	0,03	0,03		0,02	0,03	0,02	0,00
Secure Communication Channel	0,06	0,05	0,06	0,05	0,06	0,01	0,00	0,05	0,01	0,03	0,04	0,04	0,02	0,02	0,02	0,02	0,02		0,01	0,01	0,00
Authority	0,05	0,02	0,04	0,04	0,04	0,01	0,03	0,07	0,01	0,05	0,02	0,02	0,02	0,00	0,00	0,01	0,03	0,01		0,01	0,01
Foot In The Door	0,03	0,04	0,04	0,01	0,03	0,00	0,00	0,04	0,00	0,03	0,01	0,02	0,02	0,00	0,02	0,01	0,02	0,01	0,01		0,00
Impress Management	0,05	0,04	0,05	0,04	0,04	0,01	0,01	0,01	0,02	0,02	0,01	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,00	

Contact Data + Request For Minor Favor

Curiosity Appeal + Contact Data

Baiting+ Request For Minor Favor

Co-occurrence matrix



	Request For Minor Favor	Baiting	Contact Data Present	Curiosity Appeal	Fake Consistency	Flattery	Humour	Threatening	Metaphor	Semantic Attack	Include Half Confidential Details	Time Pressure	Scarcity	Guilt Calling	Reciprocity	Argumentation Mille Feuille	Obscurantism	Secure Communication Channe	Authority	Foot In The Door	Impress Management
Request For Minor Favor		0,65	0,59	0,62	0,51	0,40	0,31	0,18	0,26	0,18	0,18	0,12	0,10	0,09	0,09	0,09	0,08	0,06	0,05	0,03	0,05
Baiting	0,65		0,58	0,56	0,51	0,40	0,29	0,14	0,25	0,11	0,13	0,10	0,11	0,08	0,10	0,09	0,05	0,05	0,02	0,04	0,04
Contact Data Present	0,59	0,58		0,50	0,48	0,37	0,28	0,21	0,24	0,15	0,14	0,13	0,09	0,07	0,05	0,07	0,05	0,06	0,04	0,04	0,05
Curiosity Appeal	0,62	0,56	0,50		0,49	0,36	0,29	0,13	0,21	0,15	0,17	0,10	0,08	0,10	0,05	0,06	0,05	0,05	0,04	0,01	0,04
Fake Consistency	0,51	0,51	0,48	0,49		0,31	0,23	0,17	0,21	0,15	0,15	0,14	0,10	0,10	0,04	0,04	0,02	0,06	0,04	0,03	0,04
Flattery	0,40	0,40	0,37	0,36	0,31		0,27	0,02	0,21	0,00	0,01	0,01	0,02	0,03	0,00	0,02	0,00	0,01	0,01	0,00	0,01
Humour	0,31	0,29	0,28	0,29	0,23	0,27		0,03	0,17	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,01	0,00	0,03	0,00	0,01
Threatening	0,18	0,14	0,21	0,13	0,17	0,02	0,03		0,02	0,17	0,13	0,13	0,11	0,05	0,06	0,05	0,05	0,05	0,07	0,04	0,01
Metaphor	0,26	0,25	0,24	0,21	0,21	0,21	0,17	0,02		0,00	0,01	0,01	0,01	0,00	0,01	0,00	0,01	0,01	0,01	0,00	0,02
Semantic Attack	0,18	0,11	0,15	0,15	0,15	0,00	0,01	0,17	0,00		0,14	0,13	0,09	0,03	0,05	0,03	0,06	0,03	0,05	0,03	0,02
Include Half Confidential Details	0,18	0,13	0,14	0,17	0,15	0,01	0,01	0,13	0,01	0,14		0,12	0,10	0,08	0,06	0,02	0,03	0,04	0,02	0,01	0,01
Time Pressure	0,12	0,10	0,13	0,10	0,14	0,01	0,01	0,13	0,01	0,13	0,12		0,09	0,05	0,03	0,02	0,02	0,04	0,02	0,02	0,00
Scarcity	0,10	0,11	0,09	0,08	0,10	0,02	0,01	0,11	0,01	0,09	0,10	0,09		0,04	0,06	0,02	0,02	0,02	0,02	0,02	0,00
Guilt Calling	0,09	0,08	0,07	0,10	0,10	0,03	0,01	0,05	0,00	0,03	0,08	0,05	0,04		0,01	0,02	0,00	0,02	0,00	0,00	0,01
Reciprocity	0,09	0,10	0,05	0,05	0,04	0,00	0,00	0,06	0,01	0,05	0,06	0,03	0,06	0,01		0,02	0,03	0,02	0,00	0,02	0,01
Argumentation Mille Feuille	0,09	0,09	0,07	0,06	0,04	0,02	0,00	0,05	0,00	0,03	0,02	0,02	0,02	0,02	0,02		0,03	0,02	0,01	0,01	0,00
Obscurantism	0,08	0,05	0,05	0,05	0,02	0,00	0,01	0,05	0,01	0,06	0,03	0,02	0,02	0,00	0,03	0,03		0,02	0,03	0,02	0,00
Secure Communication Channel	0,06	0,05	0,06	0,05	0,06	0,01	0,00	0,05	0,01	0,03	0,04	0,04	0,02	0,02	0,02	0,02	0,02		0,01	0,01	0,00
Authority	0,05	0,02	0,04	0,04	0,04	0,01	0,03	0,07	0,01	0,05	0,02	0,02	0,02	0,00	0,00	0,01	0,03	0,01		0,01	0,01
Foot In The Door	0,03	0,04	0,04	0,01	0,03	0,00	0,00	0,04	0,00	0,03	0,01	0,02	0,02	0,00	0,02	0,01	0,02	0,01	0,01		0,00
Impress Management	0,05	0,04	0,05	0,04	0,04	0,01	0,01	0,01	0,02	0,02	0,01	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,00	

Contact Data Request For Minor Favor Curiosity Appeal Baiting

You will never guess what happened to him.

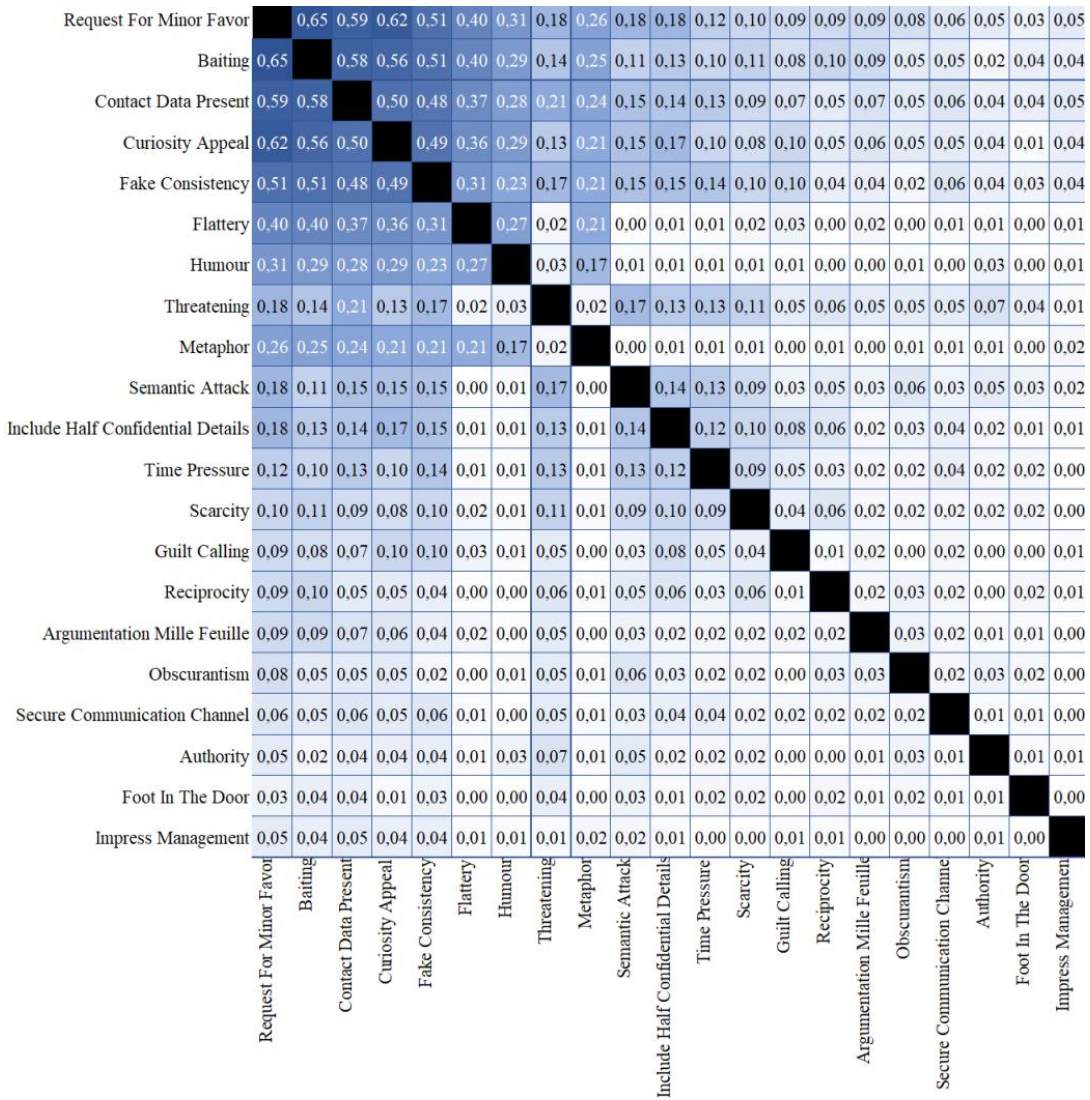
John,
+33 137118419
7 Rue de Paris, France

Please just answer me.

Judith,
+33 137118419
7 Rue de Paris, France

Save 80% on your insurance just by clicking here.

Co-occurrence matrix



	Request For Minor Favor	Baiting	Contact Data Present	Curiosity Appeal	Fake Consistency	Flattery	Humour	Threatening	Metaphor	Semantic Attack	Include Half Confidential Details	Time Pressure	Scarcity	Guilt Calling	Reciprocity	Argumentation Mille Feuille	Obscurantism	Secure Communication Channe	Authority	Foot In The Door	Impress Management
Request For Minor Favor		0,65	0,59	0,62	0,51	0,40	0,31	0,18	0,26	0,18	0,18	0,12	0,10	0,09	0,09	0,09	0,08	0,06	0,05	0,03	0,05
Baiting	0,65		0,58	0,56	0,51	0,40	0,29	0,14	0,25	0,11	0,13	0,10	0,11	0,08	0,10	0,09	0,05	0,05	0,02	0,04	0,04
Contact Data Present	0,59	0,58		0,50	0,48	0,37	0,28	0,21	0,24	0,15	0,14	0,13	0,09	0,07	0,05	0,07	0,05	0,06	0,04	0,04	0,05
Curiosity Appeal	0,62	0,56	0,50		0,49	0,36	0,29	0,13	0,21	0,15	0,17	0,10	0,08	0,10	0,05	0,06	0,05	0,05	0,04	0,01	0,04
Fake Consistency	0,51	0,51	0,48	0,49		0,31	0,23	0,17	0,21	0,15	0,15	0,14	0,10	0,10	0,04	0,04	0,02	0,06	0,04	0,03	0,04
Flattery	0,40	0,40	0,37	0,36	0,31		0,27	0,02	0,21	0,00	0,01	0,01	0,02	0,03	0,00	0,02	0,00	0,01	0,01	0,00	0,01
Humour	0,31	0,29	0,28	0,29	0,23	0,27		0,03	0,17	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,01	0,00	0,03	0,00	0,01
Threatening	0,18	0,14	0,21	0,13	0,17	0,02	0,03		0,02	0,17	0,13	0,13	0,11	0,05	0,06	0,05	0,05	0,05	0,07	0,04	0,01
Metaphor	0,26	0,25	0,24	0,21	0,21	0,21	0,17	0,02		0,00	0,01	0,01	0,01	0,00	0,01	0,00	0,01	0,01	0,01	0,00	0,02
Semantic Attack	0,18	0,11	0,15	0,15	0,15	0,00	0,01	0,17	0,00		0,14	0,13	0,09	0,03	0,05	0,03	0,06	0,03	0,05	0,03	0,02
Include Half Confidential Details	0,18	0,13	0,14	0,17	0,15	0,01	0,01	0,13	0,01	0,14		0,12	0,10	0,08	0,06	0,02	0,03	0,04	0,02	0,01	0,01
Time Pressure	0,12	0,10	0,13	0,10	0,14	0,01	0,01	0,13	0,01	0,13	0,12		0,09	0,05	0,03	0,02	0,02	0,04	0,02	0,02	0,00
Scarcity	0,10	0,11	0,09	0,08	0,10	0,02	0,01	0,11	0,01	0,09	0,10	0,09		0,04	0,06	0,02	0,02	0,02	0,02	0,02	0,00
Guilt Calling	0,09	0,08	0,07	0,10	0,10	0,03	0,01	0,05	0,00	0,03	0,08	0,05	0,04		0,01	0,02	0,00	0,02	0,00	0,00	0,01
Reciprocity	0,09	0,10	0,05	0,05	0,04	0,00	0,00	0,06	0,01	0,05	0,06	0,03	0,06	0,01		0,02	0,03	0,02	0,00	0,02	0,01
Argumentation Mille Feuille	0,09	0,09	0,07	0,06	0,04	0,02	0,00	0,05	0,00	0,03	0,02	0,02	0,02	0,02	0,02		0,03	0,02	0,01	0,01	0,00
Obscurantism	0,08	0,05	0,05	0,05	0,02	0,00	0,01	0,05	0,01	0,06	0,03	0,02	0,02	0,00	0,03	0,03		0,02	0,03	0,02	0,00
Secure Communication Channel	0,06	0,05	0,06	0,05	0,06	0,01	0,00	0,05	0,01	0,03	0,04	0,04	0,02	0,02	0,02	0,02	0,02		0,01	0,01	0,00
Authority	0,05	0,02	0,04	0,04	0,04	0,01	0,03	0,07	0,01	0,05	0,02	0,02	0,02	0,00	0,00	0,01	0,03	0,01		0,01	0,01
Foot In The Door	0,03	0,04	0,04	0,01	0,03	0,00	0,00	0,04	0,00	0,03	0,01	0,02	0,02	0,00	0,02	0,01	0,02	0,01	0,01		0,00
Impress Management	0,05	0,04	0,05	0,04	0,04	0,01	0,01	0,01	0,02	0,02	0,01	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,00	

Contact Data Request For Minor Favor Curiosity Appeal Baiting

You will never guess what happened to him.

John,
+33 137118419
7 Rue de Paris, France

Please just answer me.

Judith,
+33 137118419
7 Rue de Paris, France

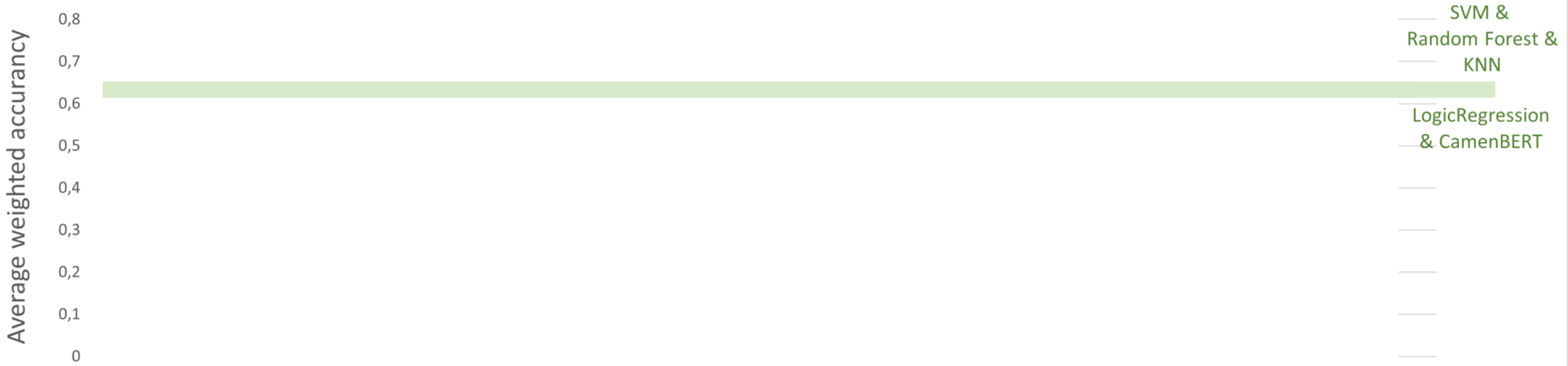
Save 80% on your insurance just by clicking here.

BENCHMARK

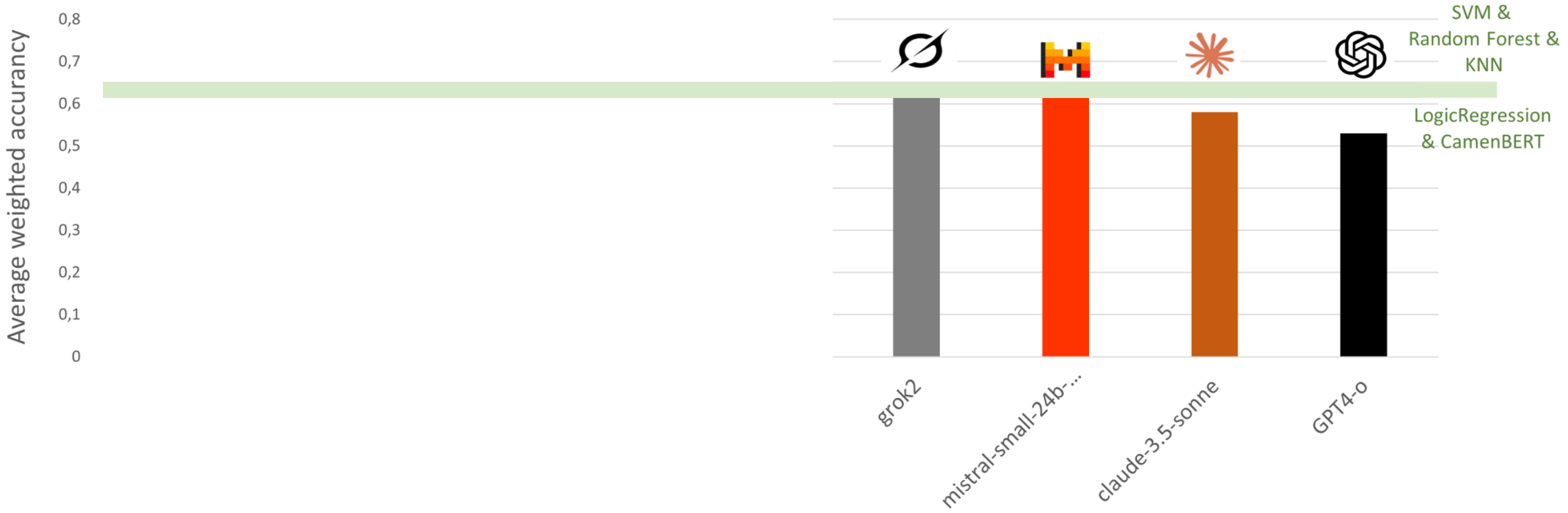
Résultat

Technique	TP	TN	FP	FN	Acc	Rec	Prec	F1
Highly representative techniques								
Argumentation Mille Feuille	4	53	36	7	0.57	0.36	0.10	0.16
Authority	5	71	22	2	0.76	0.71	0.19	0.29
Baiting*	53	26	2	18	0.79	0.74	0.96	0.83
Contact Data Present	48	21	11	20	0.69	0.71	0.81	0.76
Curiosity Appeal	64	11	18	7	0.75	0.90	0.78	0.84
Fake Consistency	46	28	10	16	0.74	0.74	0.82	0.78
Foot In The Door	4	78	17	1	0.82	0.80	0.19	0.31
Flattery	37	54	6	3	0.91	0.93	0.86	0.89
Guilt Calling	1	90	0	9	0.91	0.10	1.00	0.18
Humour	28	52	17	3	0.80	0.90	0.62	0.74
Impress Management	2	54	41	3	0.56	0.40	0.05	0.08
Include Half Confidential Details	14	68	9	9	0.82	0.61	0.61	0.61
Metaphor	26	32	42	0	0.58	1.00	0.38	0.55
Obscurantism	3	70	20	7	0.73	0.30	0.13	0.18
Reciprocity	5	87	1	7	0.92	0.42	0.83	0.56
Request For Minor Favor*	60	13	9	17	0.73	0.77	0.87	0.82
Scarcity	13	56	29	2	0.69	0.87	0.31	0.46
Secure Communication Channel	5	80	12	3	0.85	0.62	0.29	0.40
Semantic Attack	14	59	15	12	0.73	0.54	0.48	0.51
Threatening	12	70	1	17	0.82	0.41	0.92	0.57
Time Pressure	13	67	16	4	0.80	0.76	0.45	0.57

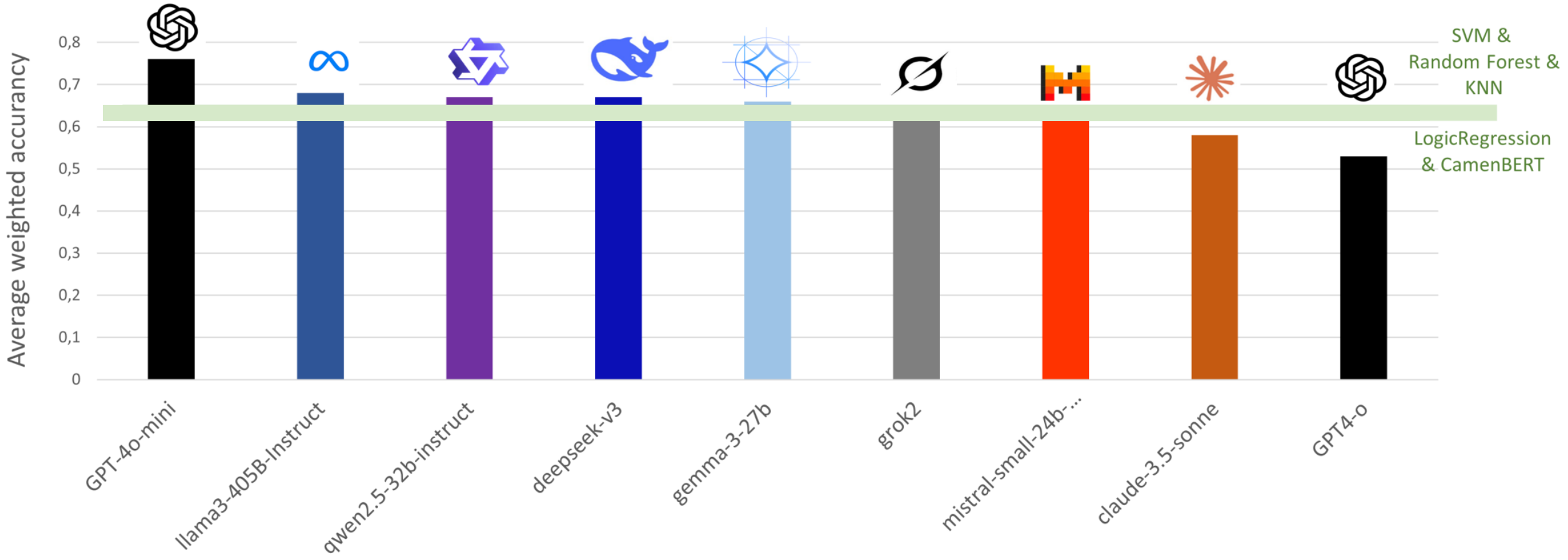
Benchmark


Average weighted accurancy
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
SVM &
Random Forest &
KNN
LogicRegression
& CamenBERT

Benchmark


Average weighted accuracy
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
SVM &
Random Forest &
KNN
LogicRegression
& CamenBERT
grok2
mistral-small-24b-...
claude-3.5-sonne
GPT4-o

Benchmark


Average weighted accuracy
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
SVM &
Random Forest &
KNN
LogicRegression
& CamenBERT
GPT-4o-mini
llama3-405B-Instruct
qwen2.5-32b-instruct
deepseek-v3
gemma-3-27b
grok2
mistral-small-24b-...
claude-3.5-sonne
GPT4-o

Qu’avons-nous appris ?

Les LLM sont de très bons candidats pour détecter des patterns de langages aussi complexes.

Bonne marge de progression.

Data semble être le bottleneck principal.

MERCI

antony.dalmiere@cnrs.fr