

Où est Claude ?

À vous de jouer !

Dans ce jeu de cartes, vous allez devoir reconnaître les images générées des vraies images.

Pour chaque paire vous devez motiver vos choix et en discuter avec les autres participants. Soyez **attentifs** !

Prenez deux images proches et tentez de **distinguer le vrai du faux**.

Les **possibilités de modifications d'images**, introduites dans les logiciels et les appareils, sont devenues **si évoluées**, que tenter de repérer les incohérences visuelles que l'on trouvait sur les premières images génératives (comme le doigt de trop, les proportions problématiques, les détails incongrus, les incohérences de lumières) est devenu difficile si ce n'est impossible.

Les **deepfakes** (mot-valise anglais formé de *deep learning*, pour apprentissage profond - désignant une technologie permettant de produire des systèmes capables de générer des contenus - et de *fake*, signifiant « faux » ou « contrefait ») transforment la crédibilité des images et vidéos, renforcent la défiance à leur égard et rendent très **complexe la vérification de l'authenticité** d'une information en facilitant la **propagation de la désinformation**.



Où est Claude ?

Niveau 1

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA

The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in blue. The background is a complex network of black lines connecting various colored geometric shapes (polygons, triangles, and quadrilaterals) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in blue. The background is a complex network of black lines connecting various colored geometric shapes (polygons, triangles, and quadrilaterals) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in blue. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in blue. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 1

C A F É IA



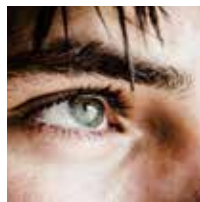
The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colorful, irregular polygons (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, purple, and pink. The overall style is abstract and network-like.

Carte réponse : vérification de réalité

Difficile de distinguer quelles images sont réelles et quelles images sont générées par l'IA. **Les textures, les compositions, les perspectives, les lumières aident peu.** Sur les milliers de joueurs qui se sont vus proposer ces mêmes images sur le site Reality Check, c'est entre 20 et 50% des répondants qui se sont trompés à choisir le faux du vrai. **Croire que nous allons arriver à ne pas nous tromper est certainement le plus important enseignement à réviser.**



VRAI



Où est Claude ? *Niveau 1*

Que retenir ?

Malgré toutes les justifications et explications que vous avez trouvé pour tenter de distinguer le vrai du faux, cet exercice doit surtout être loué pour l'échec où il conduit chacun d'entre nous. En fait, **il est tout à fait normal de se tromper** car nous avons bien du mal à identifier les productions génératives, vu le niveau qu'elles ont atteint.

Les outils de détection d'images génératives eux-mêmes ne parviennent pas toujours à détecter les images qu'ils ont eux-mêmes créées et c'est encore plus vrai des images réelles modifiées qui demeurent plus difficiles à détecter... tant pour les outils que pour les humains.

L'exercice vise à montrer que **la reconnaissance des images génératives est devenue impossible** et qu'en fait, on ne doit pas faire confiance en nos capacités de détection, mais désormais **prendre toute image partagée avec beaucoup de recul**. Lorsqu'elles sont accompagnées d'un contexte, nous devons prendre l'habitude de l'interroger. Mais ce que nous devons d'abord interroger à l'aune de ces images, ce sont nos convictions, nos certitudes, notre croyance à nous sentir plus forts que la machine. Et c'est d'autant plus vrai que lorsque les gens apprennent que le contenu qu'ils consultent est entièrement faux, ils restent influencés émotionnellement par celui-ci.

Si vous re-regardez une paire des images de ce jeu dans quelques jours, vous risquez d'être toujours convaincu que l'image que vous avez identifiée comme vraie alors qu'elle ne l'était pas est réelle !

L'influence survit à la révélation.



Où est Claude ?

Niveau 2

C A F É IA

The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (polygons, triangles, and quadrilaterals) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 2

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 2

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colorful, irregular polygons (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, purple, and pink. The overall style is abstract and network-like.



Où est Claude ?

Niveau 2

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 2

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (polygons, triangles, and quadrilaterals) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



Où est Claude ?

Niveau 2

C A F É IA



The image features a central text 'CAFÉ IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (polygons, triangles, and quadrilaterals) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and digital.



Où est Claude ?

Niveau 2

C A F É IA



The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in blue. The background is a complex network of black lines connecting various colored geometric shapes (polygons, triangles, and quadrilaterals) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.



TLG84254

Où est Claude ?

Niveau 2

C A F É IA

The image features a central text 'C A F É IA' where each letter is inside a colored hexagon: 'C' in red, 'A' in green, 'F' in purple, 'É' in yellow, and 'IA' in a blue hexagon. The background is a complex network of black lines connecting various colored geometric shapes (triangles, quadrilaterals, pentagons, hexagons) in shades of red, orange, yellow, green, blue, and purple. The overall style is abstract and network-like.

Carte réponse : Les ambiguïtés de la communication

Sans avoir recours à l'IA générative, la communication produit souvent des images étranges ou improbables dont les codes sont difficiles à prendre pour réels. **Le faux semble probable et la réalité semble fausse.**



Déplacement présidentiel aux Antilles en octobre 2018. En visite sur l'île de Saint-Martin, Emmanuel Macron prend la pose avec deux jeunes hommes, dont l'un, torse nu, adresse un doigt d'honneur. La photo, postée sur Instagram, a créé une vive polémique, Marine Le Pen allant jusqu'à la qualifier « d'impardonnable ».



Opération de communication de Donald Trump en octobre 2024, durant la campagne présidentielle. Il a cuit des frites et a servi des clients dans un fast-food en Pennsylvanie.

VRAI



A Davos en janvier 2026, Emmanuel Macron fait son discours avec des lunettes de soleil miroir (pour dissimuler une hémorragie sous-conjonctivale bénigne) qui vont faire sensation et produire de nombreux mêmes (c'est-à-dire des répliques et détournements parodiques) sur les réseaux sociaux.



En vacances à Brégançon, Emmanuel Macron est photographié en train de faire du jet-ski. L'épisode donnera lieu à beaucoup de commentaires sur l'image qu'un président doit donner.

Où est Claude ? *Niveau 2*

Que retenir ?

La communication publique a toujours cultivé des ambiguïtés, comme le posait, il y a un siècle, le publicitaire Edward Bernays en théorisant l'influence et la manipulation de masse, rappelant que la propagande politique au XXe siècle est d'abord née au cœur même de la démocratie libérale américaine, bien avant d'être exploitée par les régimes autoritaires. **Aujourd'hui, ces ambiguïtés servent de plus en plus à générer des polémiques, des clics sur les réseaux sociaux, à créer des ralliements, des rejets et des adhésions.**

Les images servent bien plus à être émotionnellement compréhensibles qu'à être authentiques. Leur crédibilité nous invite à nier tout besoin de vérification, comme si, finalement, **il était plus amusant d'être dupé que d'être informé.** Ces images fictives servent à ancrer des récits pour orienter nos compréhensions, comme quand Trump nous présente un Gaza en Riviera. Les images génératives semblent simplifier la réalité et permettent finalement de neutraliser le flux des images réelles qui sont peu commodées à comprendre. Elles apaisent l'anxiété de l'actualité en permettant de mieux la maîtriser, de la rendre plus lisible qu'elle n'est, plus conforme à ce qu'elle devrait être ou ce qu'on voudrait qu'elle soit.

Les ambiguïtés de la communication nous rappellent que nous devons douter des images qui circulent, car elles sont toujours orientées - ce qui est particulièrement problématique quand une personne publique ou une institution reconnue mobilise ce type d'images, comme le fait régulièrement Trump en partageant ouvertement des images générées. Cela devrait questionner l'impact que ces fausses images produisent sur chacun de nous et sur nos représentations elles-mêmes. Autrement dit, **même lorsque les gens apprennent que le contenu qu'ils consultent est entièrement faux, ils restent influencés émotionnellement par celui-ci.**

Carte réponse : des deepfakes aux scams

L'IA générative ne permet pas uniquement de générer des images, mais aussi désormais, nombre d'autres documents. **Si les modèles ont longtemps peiné à produire des images intégrant du texte, ce n'est plus du tout le cas aujourd'hui. Il est très facile de produire des documents officiels, des logos, ou de les falsifier.** Ces actions vont démultiplier les tentatives de *scams* (terme anglais pour fraude ou escroquerie) et de *phishing* (hameçonnage, dans le but d'obtenir des renseignements personnels pour usurper une identité) en faisant croire aux victimes qu'elles s'adressent à un tiers de confiance (banque, administration, entreprise).

Générer n'importe quel document, comme une feuille de salaire ou une fausse facture est désormais assez simple, **peu de détails permettent de différencier la vraie de la fausse.**

VR

Documents n°2, n°3, n°5 et n°7

Les mails et SMS d'arnaques et d'hameçonnages deviennent difficiles à distinguer des sollicitations commerciales, l'un comme l'autre usant de personnalisations convaincantes.

Quant aux deux affiches, l'une est un message du gouvernement rappelant les gestes essentiels à adopter lors de la canicule, l'autre une parodie qui se moque de la communication gouvernementale en reprenant les codes, tout en montrant que les réponses étaient peu adaptées aux enjeux que les gens éprouvent. Elle a été prise au sérieux par d'innombrables personnes, s'offusquant de cette communication trop clairement infantilisante.

Il est désormais très facile de fabriquer des fausses fiches de salaires, de fausses factures, de faux sites, ou même de faux papiers (rappelons que la production de faux est un délit).

Éviter l'escroquerie en ligne est compliqué et s'annonce le devenir de plus en plus.

Ce niveau difficile d'où est Claude a pour objectif de montrer que nous devons apprendre à être très méfiants. Que si nous n'avons pas été trompés, pour l'instant, il est possible que nous le soyons demain. Comme le dit le service de prévention et d'assistance du gouvernement, <https://www.cybermalveillance.gouv.fr> chacun est confronté à une menace omniprésente, massive et diversifiée avec des messages de plus en plus, et de mieux en mieux diversifiés.

Se tromper est humain, nous tromper est lucratif. Soyez attentifs au contexte, aux URL, aux noms de domaines des expéditeurs. Ne donnez jamais vos mots de passe à personne (même à un conseiller bancaire au téléphone), ne cliquez pas sur les liens de messages vous faisant des demandes impératives, même si elles ont l'air officielles, observez les URL, les adresses d'expéditions...

Mais surtout, le propre de l'escroc, c'est d'exploiter nos biais cognitifs (l'urgence, la peur de rater une opportunité, la confiance). À nouveau, **il faut être attentif d'abord aux sensations que ces messages génèrent. Un message qui vous stress est souvent le signe d'un message malveillant.**

Etre attentif

A l'issue du jeu « Où est Claude ? », vous devriez être inquiet sur la dissémination d'images et de vidéos génératives et, plus encore, sur votre capacité à identifier des images génératives. Différentes pistes d'actions sont possibles.

Pistes d'action

► **Doutez, décryptez :** Voilà longtemps que les images sont altérées. Soyez attentifs et toujours vigilants. Les images sont des vecteurs émotionnels forts, c'est pour cela qu'elles sont manipulées. Cherchez à comprendre ce qu'elles veulent vous dire : décryptez-les. Observez la composition, la lumière, ce qu'elle mettent en scène. Zoomez pour voir les défauts. Analysez-les ! Mettre des mots permet de remettre à distance l'émotion qu'elles distillent.

► **Vérifiez :** En cas de doutes, ayez recours à des moteurs d'images inversés disponibles depuis la recherche d'images de la plupart des moteurs ou à des outils d'analyses d'images, comme <https://pimeyes.com> <https://lenso.ai> <https://copyseeker.net> ou des outils d'analyse d'IA dédiés comme <https://imagewhisperer.org>

Informez-vous, formez-vous :

► Le site d'assistance et de prévention contre les intrusions et escroqueries <https://www.cybermalveillance.gouv.fr> donne tous les conseils nécessaires.

Des sites et des comptes de *fact-checking* (de vérification de faits, et notamment de contenus déjà publiés) sont disponibles dans les médias et sur nombre de plateformes, et sont souvent très réactifs, comme la rubrique Factual de l'Agence France Presse : <https://factuel.afp.com> Suivez-les !

Il existe aussi de nombreux guides et ressources en ligne pour apprendre à détecter la désinformation, par exemple le "Guide du journaliste pour détecter les contenus générés par l'IA" (cherchez ces termes sur un moteur).

CAFÉ IA

