Solutions

- Il faut mettre en place un *AI authenticator* qui supervisera les différents outils développés et toujours avoir un *human in the loop* dans la prise de décision.
- On a besoin de plus de sociétés différentes qui développent des outils différents = diversité
- Mise à disposition d'une infrastructure de recherche publique, open source et internationale
- Basée sur le CERN/Agence atomique Mondiale/Oversight Board de Facebook

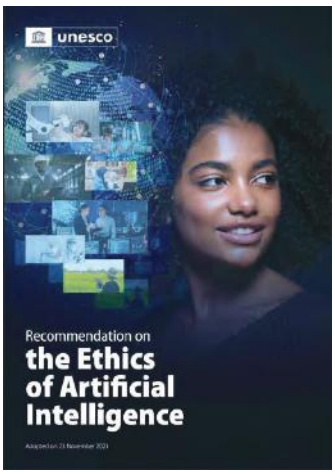
Equipée d'ordinateurs surpuissants, résultats publiés

Dotée de pouvoirs de contrôle

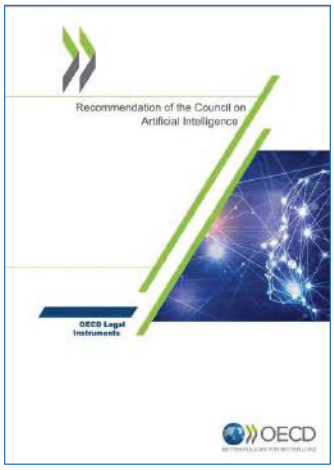
Aux décisions contraignantes

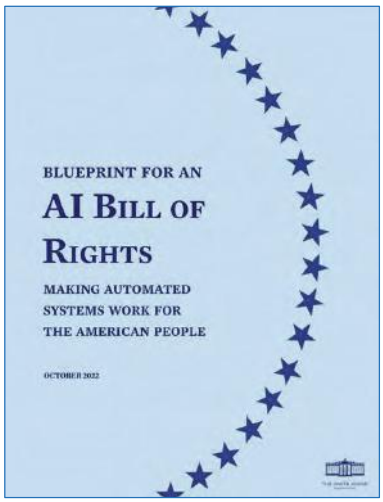
Partie II: Quelques mots à propos des différentes « réglementations » qui arrivent

Autres textes régulatoires **non contraignants** d'importance



Vers une **convention-cadre** aux principes généraux vagues internationale sur l'IA pour novembre 2023 (pour par après, des **conventions sectorielles** des plus précises?).





D'où vient le Règlement européen sur l'intelligence artificielle (IA)?

8 avril 2019

Février 2020

La Commission européenne publie un **livre blanc sur l'IA** et propose de mettre en place un cadre réglementaire européen pour une IA digne de confiance.

Octobre 2020

Le **Parlement européen** adopte trois résolutions législatives sur l'IA couvrant l'éthique, la responsabilité civile et la propriété intellectuelle.

21 Avril 2021

La Commission européenne publie sa **proposition de Règlement IA**.

6 décembre 2022

Le **Conseil** adopte sa position sur la proposition de Règlement IA.

Mai-juin 2023

Il est attendu que le Parlement EU vote sur la proposition de Règlement IA. À la suite de ce vote, les discussions entre les États membres, le Parlement EU et la Commission (=> l'opaque **trilogue**) devraient débuter après ce vote.

Fin décembre 2023

La version finale du texte **pourrait** être adoptée.

Mai 2024

Elections EU.

Toutes les institutions ont donné leur **avis** sur le texte de la proposition:
Résumé de celles en matière de **protection des données personnelles**

- Interdiction de **l'identification biométrique** à distance des personnes dans les espaces publics
- Interdiction des systèmes biométriques utilisés aux fins de **classer les individus**
- Interdiction des systèmes d'IA pour déduire des **émotions**
- Les **autorités de protection des données** devraient être désignées comme autorités de contrôle national de l'intelligence artificielle

Beaucoup appelle à sa mise à jour (pas de référence aux AI génératives par exemple)

Pourquoi un tel Règlement?

Veiller à ce que les **systèmes d'IA mis sur le marché** de l'Union et utilisés soient **sûrs** et respectent les **droits fondamentaux** et les valeurs de l'Union.

Faciliter le développement d'un marché unique pour les applications d'IA **légales, sûres et fiables** et prévenir la fragmentation du marché.

Renforcer la **gouvernance** et l'application effective de la législation existante en matière de droits fondamentaux et les exigences de **sécurité** applicables aux systèmes d'IA.

Garantir la **sécurité juridique** afin de faciliter l'**investissement** et l'innovation dans l'IA européenne.

Problème des définitions

Beaucoup de débats

. Territorial : EU

. Matériel: Fournisseurs – Utilisateurs – Distributeurs

Champ d'application

Création d'un *European Artificial Intelligence Board*

Mais aux pouvoirs bien plus faibles que ceux de l'EDPB

Renvoi à :

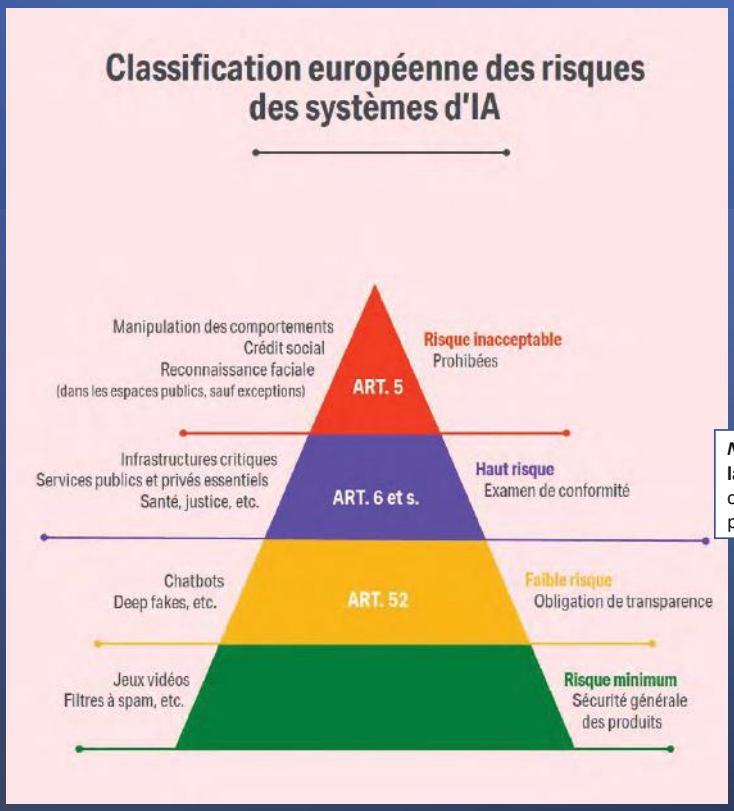
À la modernisation de la Directive relative aux produits défectueux

La future Directive AI Responsabilité

Droit à une réparation

Normes

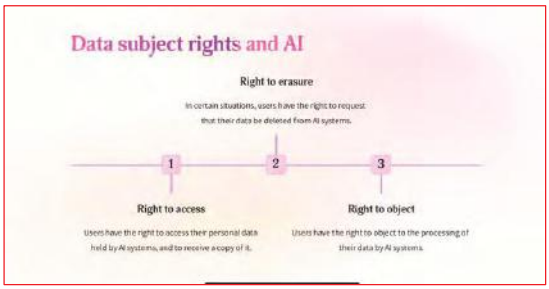
Problématique de la **présomption de conformité** si respect de futurs standards techniques => représentativité de ces organismes privés et qui n'ont pas d'XP en matière de respect des droits fondamentaux

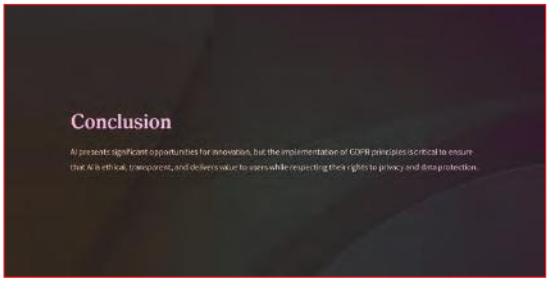


***NEW* - Non repris dans la proposition de texte de la Commission européenne**: l'*AI Act* devrait aussi contenir certaines des règles des AI à haut risques pour les outils comme ChatGPT.

AI Now Institute
Apr 13, 2023

Partie III: Interactions de l'IA avec le RGPD

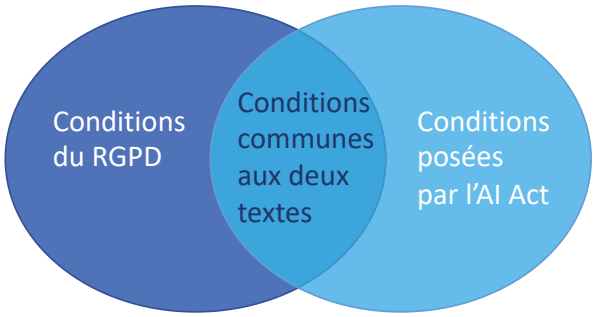




Presentation done in 2 minutes

You like it?

AI Act et le RGPD: des **approches similaires**



Protéger le **consommateur européen** ses droits et ses intérêts (amendes: 30M ou 6%)

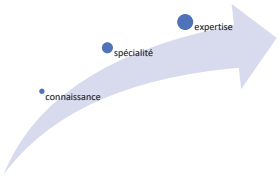
Approche basée sur les **risques** et obligation de réaliser des **analyses d'impact**

Transparence et **explicabilité**

« Accountability » et **documentation** de la moindre décision

Obligation de **coopération** entre plusieurs intervenants

AI Act et le RGPD: Un **rôle** plus étendu pour le **DPO**



Le DPO doit être le IA Champion de sa société

Ses RGPD data stewards => ses RGPD-AI data stewards

Etendre le *privacy by design* aux exigences de l'IA Act

Continuer à suivre l'actualité en y incluant l'IA

Les exigences des IA à hauts risques + respect des droits fondamentaux/éthique seront à inclure incluses dans les DPIAs tout le long de l'IA *use chain*

Il devra adapter les guidances/policies/ procédures internes à l'IA

Inclure les risques IA dans son monitoring annuel

Les **tensions** entre le RGPD et les outils d'IA

Droit à l'information + droit d'accès (input data, target & csq de la décision) + application de l'Art. 22 RGPD

Problématique de la boîte noire – secret des affaires/IP – ML v. decision trees – Explication générale ou individualisée (cons. 71)?

Décision prise (uniquement/partiellement) grâce à la machine

Droit à connaître non seulement la décision mais aussi au raisonnement (« inference »)

Biais de confirmation (suivre trop aveuglément la machine)

Web scraping: le fait qu'une donnée personnelle soit sur le web ne veut pas dire qu'elle n'est plus une donnée personnelle protégée par le RGPD

Principes de *data minimisation* et du *fair processing* non respectés?

Problématique des données d'entraînement

Quid du droit à la transparence quand OpenAI (plus très open d'ailleurs) ne dévoile plus rien?

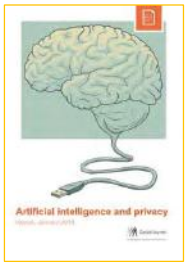
Représentativité du jeu de données.
Quid du droit à l'oubli?
De l'utilisation de données sensibles?

*Respect à tous les instants d'une bonne **sécurité informatique** autour des données (art. 32 RGPD)*

Quid si les résultats de l'outil dévoilent des données personnelles contenues dans ses données d'entraînement?

Une privacy IA pourrait : coûter plus cher, être plus énergivore et (surtout) être moins performante

Data breach de ChatGPT le 24 mars 2023



« Le RGPD et ses principes généraux sont **compatibles** avec le développement de l'IA »

« Les clauses du RGPD sont **vagues** et sujettes à beaucoup d'interprétations »

« Des **guidances** de l'EDPB et des différentes autorités nationales sont urgemment nécessaires »

Analyse de quelques **tensions**: il faut distinguer **deux situations**

Les données personnelles d'une personne sont utilisées dans les **données d'entraînement** pour créer un **modèle**

Principe de la **limitation des finalités** (Art. 5.1.b)

Principe de **minimisation des données** (Art. 5.1.c)

Principe de **limitation de la conservation** (Art. 5.1.e)

Base légale des **intérêts légitimes** (Art. 6.1.f)

Les données personnelles sont utilisées comme **données d'entrée** (les inputs) pour appliquer l'IA à la personne concernée

Anonymisation – pseudonymisation – mesures de sécurité fortes pour augmenter la conformité

OK pour le **test de compatibilité** (Art. 6.4 + cons. 50):
- La personne n'en serait pas directement affectée
- Distance entre la finalité originale et la nouvelle + les attentes de la personne : ok

OK pour le **test de proportionnalité**
OK si créer un modèle // une **finalité statistique/recherche scientifique** (Art. 89) si les données sont agrégées (forts débats) (+ droit d'opposition)
Intérêts légitimes de la personne non affectés

- **Profiling** où les données d'entrées sont utilisées pour prédire (« to infer ») d'autres données personnelles la concernant.
- Analyse approfondie des **tests de compatibilité** et de **proportionnalité**.
- **Intérêts** de la personne supérieurs (consentement + opt-out)

AI et le RGPD: considérations supplémentaires

L'IA rend **identifiable** des données anonymisées

L'IA **étend** les domaines d'utilisations où sont concernées des données personnelles (reconnaissance faciale)

L'IA a besoin **d'énormément de données** (dont des données personnelles pour fonctionner)

Comment articuler les principes du RGPD avec l'IA lorsque vous commencez votre **mise en conformité** et lorsque vous la continuez année après année?

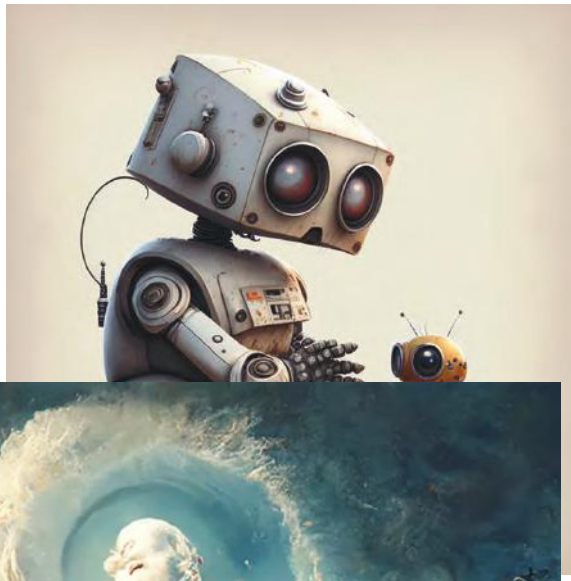
Qui est le **responsable du traitement**? Qui est **sous-traitant**? **Responsables conjoints**?

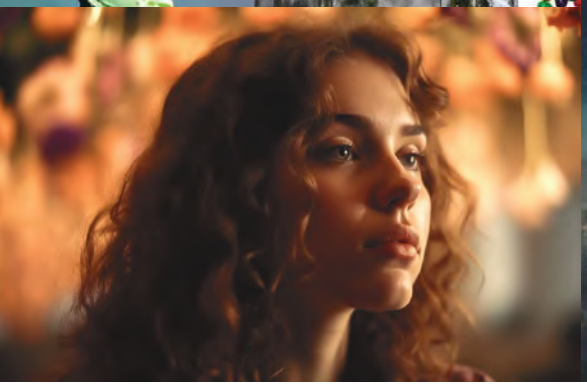
Comment réaliser un **DPIA** complet quand OpenAI ne rend pas public ses données?

Comment rendre responsable des développeurs d'IA situés **hors EU**?

Tant les données personnelles en tant qu'input que les données de résultats (les « inferred data ») sont des **données personnelles**.

Partie IV: Les *IA generatives* (ChatGPT et les autres)





Axel Beelen / axel@axelbeelen.be / 24 avril 2023

ChatGPT is about to revolutionize the economy. We need to decide what that looks like.

New large language models will transform many jobs. Whether they will lead to widespread prosperity or not is up to us.

By David Rotman March 25, 2023

13 DE ABRIL DE 2023

La AEPD inicia de oficio actuaciones de investigación a OpenAI, propietaria de ChatGPT

C‘est un exploit : il n’a fallu que deux mois à ChatGPT pour atteindre les 100 millions d’utilisateurs dans le monde. À côté, les neuf mois d’Instagram, le précédent record,

10 milliards

C’est, en dollars, la somme investie en janvier 2023 par Microsoft dans l’entreprise OpenAI, à l’origine de ChatGPT. L’entreprise de Bill Gates y avait déjà investi 1 milliard de dollars en 2019.

2 millions

C’est le nombre d’images générées quotidiennement par Dall-E, une IA gratuite capable de créer une image à partir d’une description écrite. Son concurrent payant, Midjourney, en génère, lui, 275 000 par jour.

13 millions

C’est, en moyenne, le nombre de “visiteurs uniques” comptabilisés chaque jour sur ChatGPT pendant le mois de janvier 2023.

12 millions

C’est le nombre d’images soupçonnées de plagiat par Getty Images, l’une des plus grandes banques d’images au monde, qui ont été générées par l’IA Stable Diffusion.

AI génératives: **kaséko**?

- **Synthèse** de document, création de texte, d'images (parfois de haute qualité), de graphiques ou de ligne de code, de musique, détection de bug informatique, outil de recherches avancées, **etc.**
- Basés sur modèle de langage (GPT-4 d'OpenAI pour ChatGPT+, LaMDA de Google pour Bard) – font partie du deep learning – on les appelle aussi des « **general purpose AI** » ou des « foundation models »
- Ces larges modèles de langages (LLM) sont entraînés sur des **milliards** de données (tout internet) et de paramètres
- Utilise des quantités massives de données d'apprentissage pour créer des **modèles de prédiction** incroyablement complexes qui leur permettent de produire un nouveau contenu réaliste en réponse à des invites spécifique (les célèbres *prompts*) – chaque nouveau prompt (même identique) produira un nouveau résultat
- Outils **faciles d'utilisation** à la Google (c'est peut-être là la VRAIE (r)évolution) mais qui n'ont AUCUNE (mais alors là vraiment aucune!) compréhension de ce qu'ils font (« écrivent » ou « dessinent »)

Déjà dépassé!

Premières plaintes auprès de la Cnil contre ChatGPT

5 avril 2023

Annonce

Le 4 avril 2023

Commissariat à la protection de la vie privée du Canada | Office of the Privacy Commissioner of Canada

Le Commissariat ouvre une enquête sur ChatGPT

Le Commissariat à la protection de la vie privée du Canada a ouvert une enquête sur l'entreprise qui est à l'origine de ChatGPT, un robot conversationnel doté d'une intelligence artificielle (IA).

« Les répercussions sur la vie privée de la technologie d'IA sont une priorité pour le Commissariat, a déclaré le commissaire à la protection de la vie privée du Canada, Philippe Dufresne. Nous devons suivre la rapide évolution des avancées technologiques, et même conserver une longueur d'avance sur ce plan. Il s'agit d'ailleurs de l'un de mes principaux secteurs d'intérêt en tant que commissaire. »

L'enquête sur OpenAI, l'entreprise qui exploite ChatGPT, a été lancée à la suite d'une plainte selon laquelle des renseignements personnels ont été recueillis, utilisés et communiqués sans consentement.

Comme il s'agit d'une enquête en cours, aucune précision supplémentaire ne peut être communiquée pour l'instant.

It's Way Too Easy to Get Google's Bard Chatbot to Lie

BY VITTORIA ELLIOTT

WIRED

The company's policy bars use of the AI chatbot to "misinform." A study found that it readily spouted untruths on topics from Covid-19 to the war in Ukraine.

04.05.23

Trois d'artistes américains ont porté plainte contre les intelligences artificielles génératrices d'images, Midjourney, Stability AI et DeviantArt, pour pratiques anticoncurrentielles et violation du droit d'auteur. | 30 Jan 2023

Artificial intelligence: stop to ChatGPT by the Italian SA

GPDP | GARANTE PER LA PROTEZIONE DEI DATI PERSONALI

Personal data is collected unlawfully, no age verification system is in place for children

Roma, 31 March 2023

Australian mayor readies world's first defamation lawsuit over ChatGPT content

1964
Joseph Weizenbaum développe le premier chatbot, **ELIZA**, au laboratoire d'intelligence artificielle du MIT. ELIZA peut simuler une conversation avec un humain en utilisant un algorithme simple pour générer des réponses textuelles aux questions.

2003
Yoshua Bengio et son équipe développent le premier modèle de langage de réseau neuronal *feed-forward*, qui **prédit le mot suivant** lorsqu'on lui donne une séquence de mots.

2014
Ian Goodfellow développe le premier réseau antagoniste génératif (**GAN**) qui peut générer de nouvelles données basées sur un ensemble d'entraînement donné. Par exemple, un GAN entraîné sur les photos peut créer de nouvelles photos qui semblent authentiques pour les humains (si vous ne regardez pas de trop près).

2017
Une équipe de chercheurs de Google propose une nouvelle architecture de réseau simple, le **Transformer**, basée uniquement sur des mécanismes d'attention et supprimant les réseaux de neurones récurrents.

30 nov. 2022
L'article d'Alec Radford sur la pré-formation générative (**GPT**) d'un modèle de langage est republié sur le site Web d'OpenAI, montrant comment un modèle de langage génératif peut acquérir des connaissances et traiter des dépendances non supervisées sur la base d'une pré-formation sur un ensemble de données vaste et diversifié.

30 nov. 2022
OpenAI rend public ChatGPT (basé sur GPT 3.5)

14 mars 2023
OpenAI rend public GPT 4.0 (multimodal)

Mars 2023
Annonces: Google va intégrer Bard dans son moteur de recherches et dans ses autres outils, Microsoft GPT dans sa suite Office & dans Bing, etc.

22 mars 2023
Lettre de « Future of Life Institute » pour avoir une **pause** dans la recherche et le déploiement de l'intelligence artificielle.

30 mars 2023
CAIDP demande à la US FTC d'analyser **OpenAI**

30 mars 2023
La CNIL italienne interdit ChatGPT en Italie pour non respect du **RGPD**

Avril 2023
Dépôt de plaintes auprès de la **CNIL** – L'Autorité canadienne lance une **enquête** sur OpenAI ainsi que l'Autorité espagnole – L'EDPB crée une **taskforce** spécifique

GPDP | GARANTE PER LA PROTEZIONE DEI DATI PERSONALI

La DPA italienne a ordonné la limitation temporaire du traitement à OpenAI (propriétaire de ChatGPT) avec **effet immédiat** en raison de plusieurs violations présumées du **RGPD**, notamment les articles 5, 6, 8, 13 et 25:

1) **Manque de transparence** : ni les utilisateurs ni les autres personnes concernées n'ont été informés du traitement
2) **Absence de base légale** pour la collecte et le stockage massifs de données personnelles, ainsi que la formation des algorithmes pour fournir ChatGPT
3) **Traitement inexact** des données personnelles, car le résultat du traitement ne correspond pas toujours à la réalité.
4) **Absence d'un système de vérification de l'âge** pour empêcher les mineurs (utilisateurs de moins de 13 ans) d'utiliser le service

La DPA a alors ordonné une **interdiction temporaire** en vertu de l'art. 58(2)(f) RGPD avec effet uniquement sur le territoire italien.

Focus

31 mars 2023
La CNIL italienne interdit ChatGPT en Italie pour non respect du RGPD

- Comunicato del 13 aprile 2023
- Comunicato del 12 aprile 2023
- Comunicato dell'8 aprile 2023
- Comunicato del 6 aprile 2023
- Comunicato del 4 aprile 2023
- Comunicato del 31 marzo 2023
- Provvedimento del 30 marzo 2023

Après quelques réunions avec l'entreprise, la DPA a publié un communiqué de presse dans lequel elle a déclaré que la suspension du traitement serait levée fin avril 2023 aura mis en place les **mesures suivantes**:

1) **Transparence**: L'entreprise doit fournir des informations sur les techniques et la logique sous-jacentes au traitement des données personnelles, soit avant l'enregistrement (pour les nouveaux utilisateurs), soit avant le premier accès (pour les utilisateurs enregistrés)
2) **Base juridique du traitement**: au lieu d'un contrat, l'entreprise doit indiquer qu'elle s'est appuyée sur le consentement ou ses intérêts légitimes (art. 6, paragraphe 1, point a) ou f) du RGPD)
3) **Droits des personnes concernées** : Mettre en œuvre des moyens de rectifier ou d'effacer les données des personnes concernées
4) **Mineurs**: Mettre en place un système de vérification de l'âge pour restreindre l'enregistrement des mineurs
5) **Campagne d'information**: OpenAI lancera une campagne de sensibilisation sur différents médias pour informer les individus sur l'utilisation de leurs données personnelles concernant la formation de l'algorithme.

Focus sur les **plaintes** déposées à la CNIL

COMMISSION NATIONALE INFORMATIQUE & LIBERTÉS

1️⃣ La CNIL a déposé deux plaintes contre ChatGPT en France concernant la gestion des données personnelles. Ces plaintes soulèvent des questions sur la conformité de ChatGPT aux régulations européennes, en particulier le Règlement général sur la protection des données (RGPD).

2️⃣ ChatGPT fait face à des blocages et des poursuites dans d'autres pays également. (Italie et l'Allemagne par exemple)

3️⃣ Les IA comme ChatGPT posent des problèmes liés aux données personnelles qui dépassent les régulations actuelles telles que le RGPD.

4️⃣ L'Italie a choisi de bloquer temporairement le service, et certains suggèrent la même chose pour la France.

5️⃣ La première plainte, portée par l'association Janus International, exige l'accès aux données personnelles enregistrées sur le service, ce qu'OpenAI a refusé.

6️⃣ Cette plainte reproche également à OpenAI de ne pas avoir de conditions générales d'utilisation et de ne pas recueillir de consentement explicite concernant la politique de confidentialité.

7️⃣ La deuxième plainte, portée par David Libeau, un développeur engagé dans la protection des données personnelles, s'interroge sur l'exactitude des informations fournies par ChatGPT.

8️⃣ Les deux plaintes questionnent le dataset utilisé pour nourrir l'IA et l'exactitude des informations délivrées par celle-ci.

9️⃣ OpenAI pourrait devoir apporter des modifications pour se conformer à la loi, y compris l'utilisation d'un dataset adapté pour un futur modèle d'IA.

🔟 En cas de blocage, les utilisateurs de ChatGPT pourraient se tourner vers les VPN, comme c'est le cas en Italie.

Avril 2023
Dépôt de plaintes auprès de la **CNIL**

Une conformité au **RGPD** est-elle néanmoins possible?

De OpenAI n'a pas, pas de siège central dans l'un des 27 EM

Peut être soumis à 27 investigations + décisions (amendes ou arrêts)!

Risques particuliers associés à l'IA générative

- L'un des principaux risques associés à cette technologie est son potentiel d'abus. Les images générées par l'IA peuvent être utilisées pour diffuser des informations erronées, tromper les gens ou même pour la création de fausses preuves.
- Certaines préoccupations tournent aussi autour de l'utilisation de cette technologie à des fins malveillantes (création de contenus pornographiques ou simulation de violence).
- Perte totale de vie privée – souvent, ils hallucinent (Sam Altman) à cause de la qualité des données (tout ce qui provient du web n'est pas vrai)
- Risques pour l'**environnement** - non inclusion des langues minoritaires

=> TOUJOURS VERIFIER LES RESULTATS + LES COMPLETER <=

Protections mises en place

- Pour répondre à ces préoccupations, des mesures ont été mises en place pour limiter l'utilisation abusive de cette technologie.
- Tout d'abord, Microsoft a annoncé qu'il limitera l'accès à cette fonctionnalité à un petit nombre de partenaires sélectionnés, notamment des organisations de recherche et des entreprises travaillant sur des projets liés à l'IA.
- L'entreprise a aussi mis en place des restrictions strictes sur les types de contenu qui peuvent être générés à l'aide de cette fonctionnalité. DALL-E utilise une base de données d'images spécifiques pour générer des images, ce qui limite la potentialité de l'IA à créer des images inappropriées.
- De son côté, OpenAI aurait aussi mis en place des protocoles de vérification et d'approbation pour s'assurer que les images générées sont conformes aux normes éthiques et responsables.

FEEL FREE TO CONTACT ME AND LET'S TALK

A→B

+ 00 32 476 223 841
axel@axelbeelen.be
Connect with me on Linkedin