CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

PROTÉGER les données personnelles
ACCOMPAGNER l'innovation
PRÉSERVER les libertés individuelles

IA génératives

Quels enjeux pour la régulation des données personnelles ?

Les lundis de l'IA et de la finance – 5 juin 2023

Félicien Vallet

ChatGPT (OpenAI)
Retour sur la décision italienne

* **Qui ?**
  * Garante per la protezione dei dati personali
  * Equivalent italien de la CNIL
  * Existe depuis 1997

* **Quand et quoi ?**
  * **31 mars 2023** [lien]
    * Mesure d'urgence interdisant à OpenAI de traiter les données des utilisateurs italiens
    * Doutes relatifs à la conformité RGPD concernant :
      – **l'information des utilisateurs** dont les données ont servi à l'apprentissage du modèle
      – **la base juridique** pour la constitution du modèle
      – le caractère **inexact** de certaines données
      – l'absence de tout mécanisme de **vérification de l'âge**

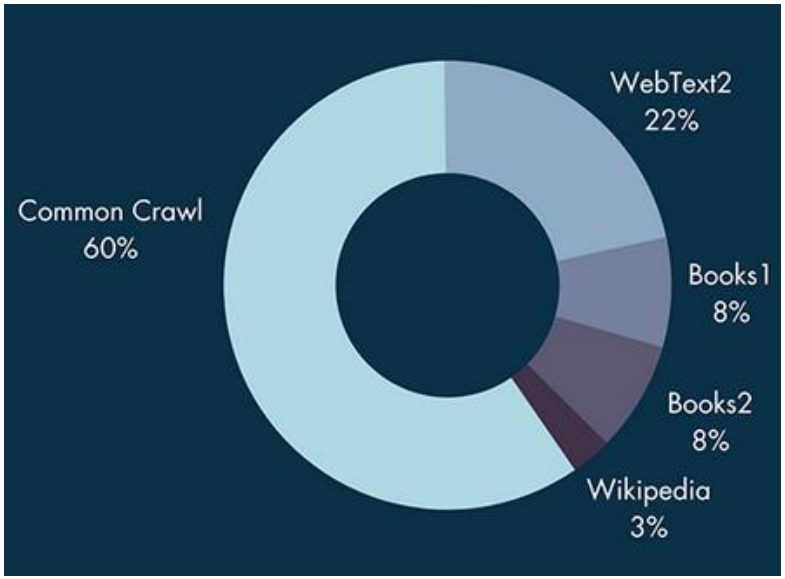
ChatGPT (OpenAI)
Retour sur la décision italienne

* **Quand et quoi ?**
  * **12 avril 2023** [lien]
    * Après discussions avec OpenAI, identification d'un **ensemble d'exigences** auxquelles la société devait se conformer d'ici le 30 avril

  * **28 avril 2023** [lien]
    * Neuf mesures prises par OpenAI :
      - **information** sur les traitements mis en œuvre
      - possibilité de **s'opposer au traitement de ses données** pour les utilisateurs et non-utilisateurs,
      - introduction de **mesures d'effacement des données inexactes** (sachant que la correction des données apparaît impossible aujourd'hui),
      - clarification de la façon dont les **données des utilisateurs peuvent être réutilisées**
      - mise en œuvre des **mécanismes de vérification d'âge** (déclaration)
    * → Rétablissement de l'accès à ChatGPT en Italie

Des questions sur plusieurs niveaux
1) Les données d'apprentissage

* Des volumes de données **considérables**
  * **~570GB**
  * Des données issues de sites :
    * **grand public** (e.g. Reddit, BlogSpot)
    * **institutionnels** (europarl.eu, nasa.gov)
    * **académiques** (mit.edu, cornell.edu, berkeley.edu, cnrs.fr, etc.)
    * **de presse** (euronews.fr, lefigaro.fr, ouest-france.fr, etc.)
    * et bien d'autres…



* Quel **encadrement** des traitements de **données personnelles** impliqués ?

Des questions sur plusieurs niveaux
2) Le modèle de langage (LLM)

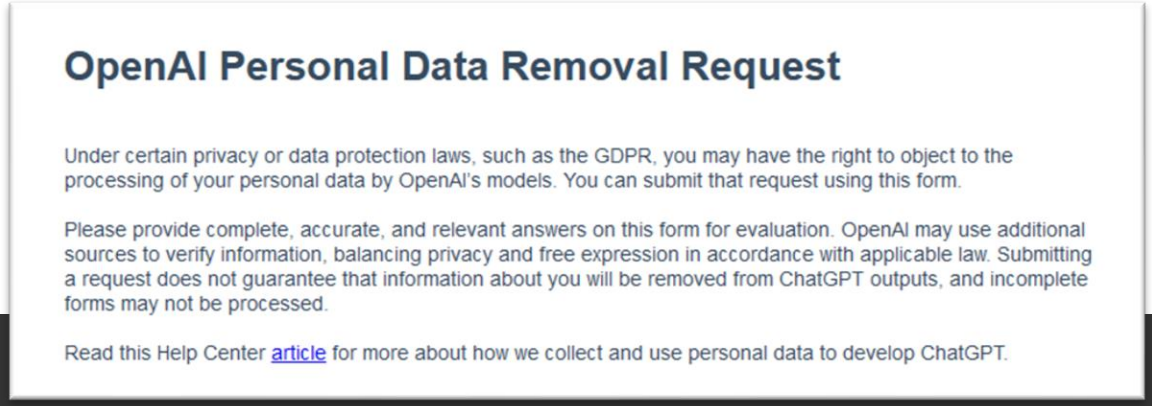
* Quid des **données personnelles** dans les modèles ?
  * Droit d'opposition proposé par OpenAI

* Le **statut** des modèles d'IA pose question (et plus largement que pour les seuls LLMs)
  * Des possibilités d'attaques par :
    * Inversion de modèle
    * Inférence d'appartenance / inférence d'attributs
    * Mémorisation

  → [Article LINC] *Petite taxonomie des attaques des systèmes d'IA*

* Enjeu pour les modèles **déjà entraînés** pour l'opposition/effacement et rectification
  * OpenAI a indiqué à la Garante **l'impossibilité de mettre en œuvre la rectification des données** au niveau du modèle

  → [Article LINC] *Machine Unlearning*



OpenAI Personal Data Removal Request

Under certain privacy or data protection laws, such as the GDPR, you may have the right to object to the processing of your personal data by OpenAI's models. You can submit that request using this form.

Please provide complete, accurate, and relevant answers on this form for evaluation. OpenAI may use additional sources to verify information, balancing privacy and free expression in accordance with applicable law. Submitting a request does not guarantee that information about you will be removed from ChatGPT outputs, and incomplete forms may not be processed.

Read this Help Center article for more about how we collect and use personal data to develop ChatGPT.

Des questions sur plusieurs niveaux
3) L'interface utilisateur

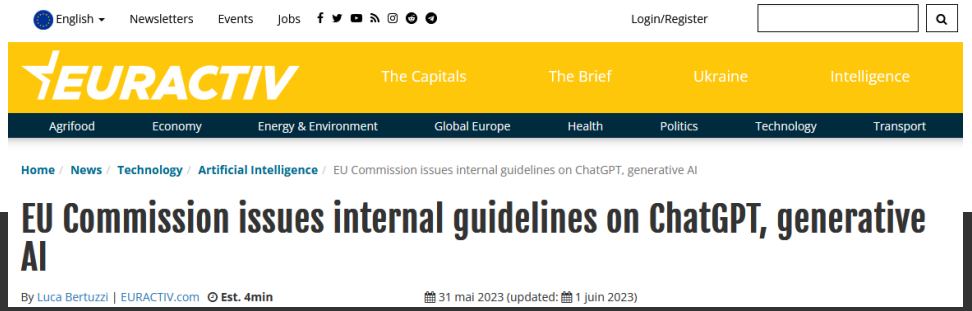
* Permet **d'interagir avec l'utilisateur** (via l'invite ou *prompt*)

* Pose des questions concernant :
  * Les informations **fournies en entrée** par l'utilisateur
  * Les **sorties produites** par le système
    * Si ChatGPT propose un contenu inexact à mon encontre, est-ce un manquement au titre du RGPD ?

* Les utilisateurs **doivent pouvoir :**
  * Être **informés** de la façon dont leurs données sont traitées
  * **S'opposer** au traitement ultérieur (par exemple pour amélioration du système)
  * Pouvoir exercer leur **droit d'accès**
  * Permettre l'effacement des **données inexactes** (via des mécanismes de *safety*)

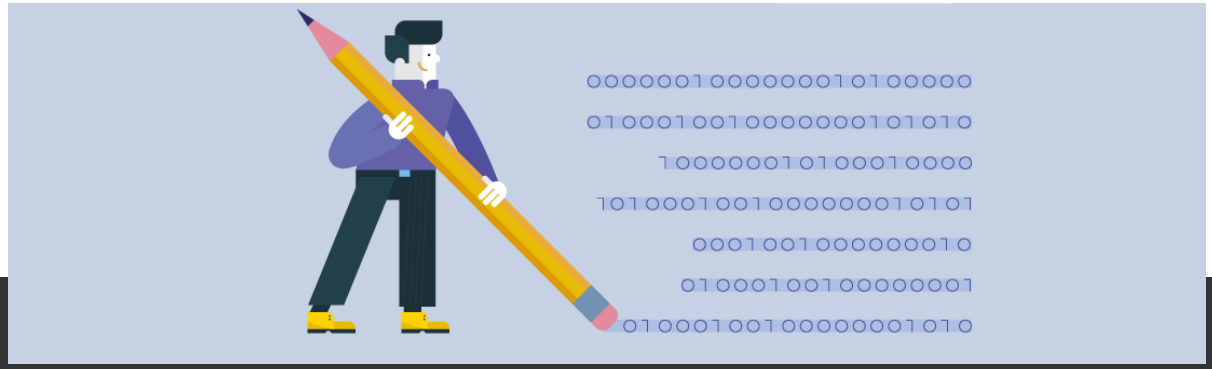
  * → Ce sur quoi se sont focalisés les travaux de la Garante

* Enjeu de l'utilisation de ces outils dans un **contexte professionnel**
  * Exemple de la Commission européenne



Una analogie possible avec les moteurs de recherche ?

* Utilisation de **données moissonnées** (*crawlées*) sur Internet
* Possibilité de **prolonger le droit à l'oubli** aux LLMs ?
  * Empêcherait que des informations concernant les personnes soient générées, sans altérer le modèle sous-jacent (via des mécanismes de filtrage)
* Possibilité d'utiliser des **balises** ou **fichiers** de type « robots.txt » pour s'opposer à ce que son site web alimente des LLMs ?
* → Un parallèle qui ne vaut que **dans une certaine mesure** :
  * multiplicité des usages de ces modèles vs constitution d'un index des sites web
* ***[Réflexions en cours…]***



Quelles suites *en France…*

* **Sur ChatGPT**
  * Des plaintes de déposées (depuis l'action italienne)
  * Une procédure en cours

* **Sur l'IA générative**
  * Publication d'un dossier LINC consacré à l'IA générative
  * Rencontre avec des acteurs de l'écosystème français
  * → **Enjeu : permettre et encadrer** le développement de systèmes d'IA respectueux de la vie privée
  * Publication du plan d'action de la CNIL

Quelles suites
…et en Europe ?

* **Sur ChatGPT**
  * Pas d'application du mécanisme de « guichet unique » du RGPD
  * Taskforce CEPD (ensemble des CNIL européennes) sur ChatGPT et les LLMs

* **Règlement IA** (en cours d'élaboration)
  * Les IA génératives, des systèmes d'IA à « risque limité ». Des exigences concernant :
    * [Com.] La **transparence** : « natural persons are informed that they are interacting with an AI system »
    * [Parl.] La **génération de certains contenus :** *« ensure adequate safeguards against the generation of content in breach of Union law in line with the generally acknowledged state of the art »*
  * Introduction au niveau du Parlement des notions de modèles de fondation et de systèmes d'IA à usage général
  * → **Enjeu :** continuité de la chaîne de responsabilité pour ces systèmes si réutilisés pour des applications à « haut risque »
    * [Parl.] Obligations concernant **l'évaluation des risques, les données d'apprentissage, l'interprétabilité, la prévisibilité, la « corrigibilité »**, etc. ;
    * [Parl.] Documentation des modèles de fondation : « *Description of the **capabilities and limitations** of the foundation model, including the **reasonably foreseeable risks** and the measures that have been taken to mitigate them as well as remaining non-mitigated risks with an explanation on the reason why they cannot be mitigated* ».

* **Une approche UE/US en anticipation**
  * Vers un code de conduite volontaire sur l'IA responsable
  * Une terminologie commune très récemment proposée

CNIL.

Contact : ia@cnil.fr

QUESTIONS