

IA Défis Ethique

Alison PATOU
Patou.alison@gmail.com



Programme

- Introduction
- Données personnelles
- Enjeux éthiques
- TP



A savoir

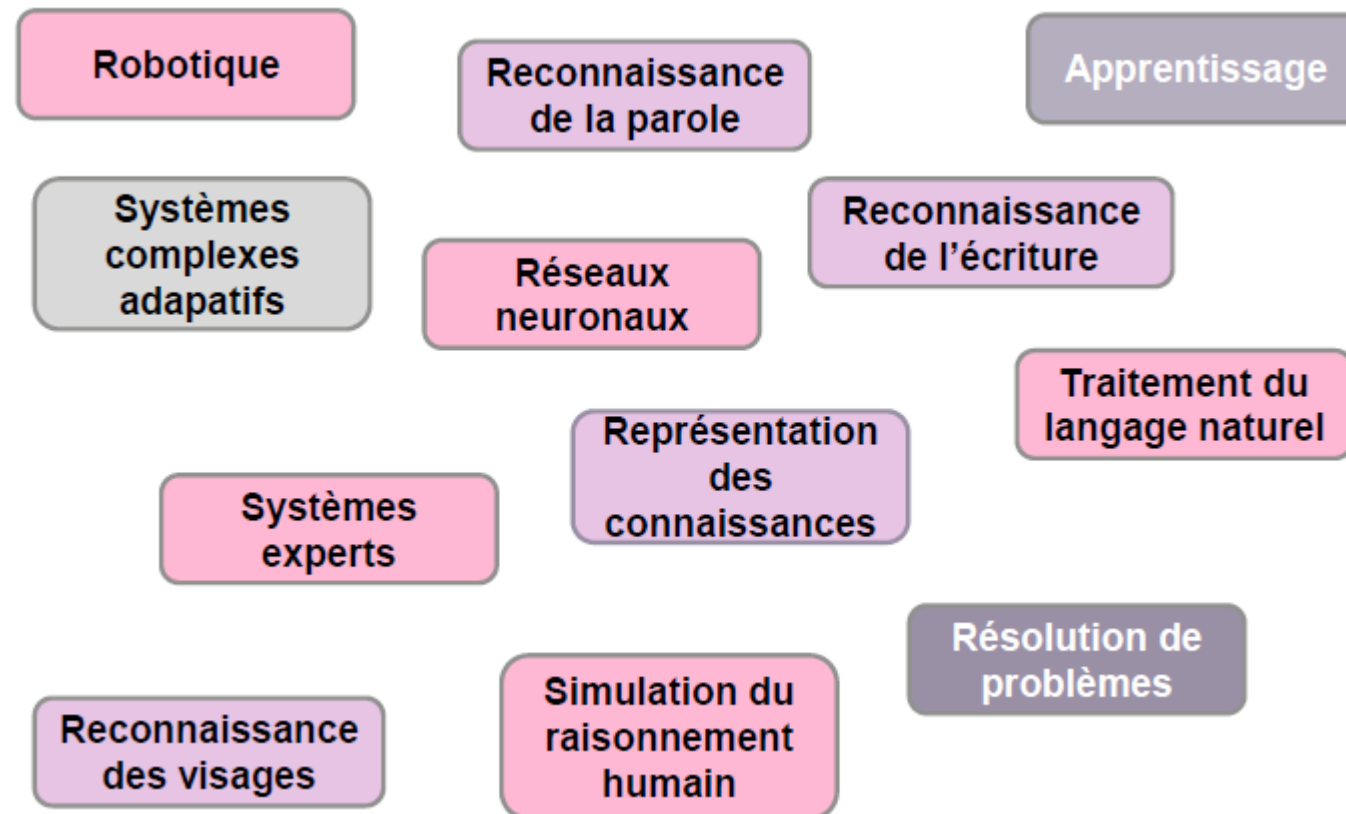
- Tout est sur mon GitHub :
<https://github.com/apatou>
- (L'essentiel à retenir du cours, les dataset, exercices, ...)
- Merci de m'envoyer à chaque fin de séance vos TPs : patou.alison@gmail.com
- 1 projet en groupe

1

Introduction

Introduction

Nous sommes définitivement entrés dans l'aire de la Data et de l'Intelligence Artificielle



Bénéfices de l'IA

Augmenter les performances en automatisant ou en perfectionnant certains processus

- Hitachi : Machine Learning sur l'analyse de données et des routines des employés pour tenir le rôle de manager
- Concerne les secteurs d'activités dans lesquels les processus sont pour la plupart automatisés et qui reposent sur des tâches opérationnelles

Augmenter l'expertise humaine d'aide à la décision

- DeepKnowledgeVenture à Honk-Kong possède une IA dans son CA qui fait des recommandations en matière d'investissements ... et qui a un droit de vote

Optimise les produits et les services -Innove

- Amélioration de la connaissance client, de la prise de décision, des processus opérationnels, hyper-personnalisation
- Passage du «push product» avec l'analytique Big Data au «pull client» avec l'IA –entretenir un pouvoir d'attraction

Bénéfices de l'IA

Renforcer la sécurité des systèmes

- Cybersécurité, détection de fraudes, identification des menaces

Aider à découvrir

- En analysant des masses de données considérables, il est possible de chercher de nouvelles propriétés, de détecter des corrélations à partir de signaux faibles

Aider à la créativité

- Design, arts, musique

Enjeux organisationnels

Facilitateurs technologiques :

- Mise à disposition de capacités IA sous forme d'API (vision, prédictif) et du code source dans les plateformes de machine learning

Conséquence organisationnelle :

- Appréhender cette notion dans une optique transversale cross-entreprises ou cross-équipes
- Standardisation des interfaces
- Architecture plus complexe à mettre en place

Enjeux sociétaux

Pas de consensus sur une balance création-destruction

- Apparition de nouveaux métiers
- Emplois menacés (services, ventes, administratif, transports)
- Emplois moins menacés (management, éducation, santé)

Quelles différences entre :

Data Scientist



Data Engineers



Data Analysts



Quelles différences entre :

Data Scientist

also known as Data Managers, statisticians.



A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.

Skills: Mathematics, Programming, Communication



Will use programmes such as:
SQL, Python, R

Data Engineers

also known as database administrators and data architects.



They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.

Skills: Programming, Mathematics, Big data



Will use programmes such as:
Hadoop, NoSQL, and Python

Data Analysts

also known as business Analysts.



They typically help people from across the company understand specific queries with charts.

Skills: Statistics, Communication, Business knowledge



Will use programmes such as:
Excel, Tableau, SQL

Enjeux éthiques

Jusqu'où peut-on laisser aux algorithmes et à ceux qui les développent le contrôle de nos sociétés ?

Comment protéger nos vies privées de l'appétit de machines alimentées par la collecte des données personnelles ?

Comment se préparer aux bouleversements que l'intelligence artificielle va entraîner dans de nombreux secteurs professionnels ?

2

Données personnelles

Si c'est gratuit, c'est que vous êtes le produit !

Les données récoltées

L'utilisation de vos itinéraires, les recherches internet, le wifi gratuit dans un aéroport, ...

On utilise des services numériques au quotidien.

Lorsqu'on utilise ses services, majoritairement gratuits, les données produites peuvent être compilées et analysées, notamment à des fins publicitaires. Votre profil est ensuite analysé au fil des utilisations pour vous proposer les publicités les mieux ciblées et les fonctionnalités correspondant aux habitudes observées.

Plusieurs problématiques



Discrimination



Anonymat



Identifiant unique



Propriété de la donnée

Vente/achat d'informations



Fausse prédictions

1. La discrimination

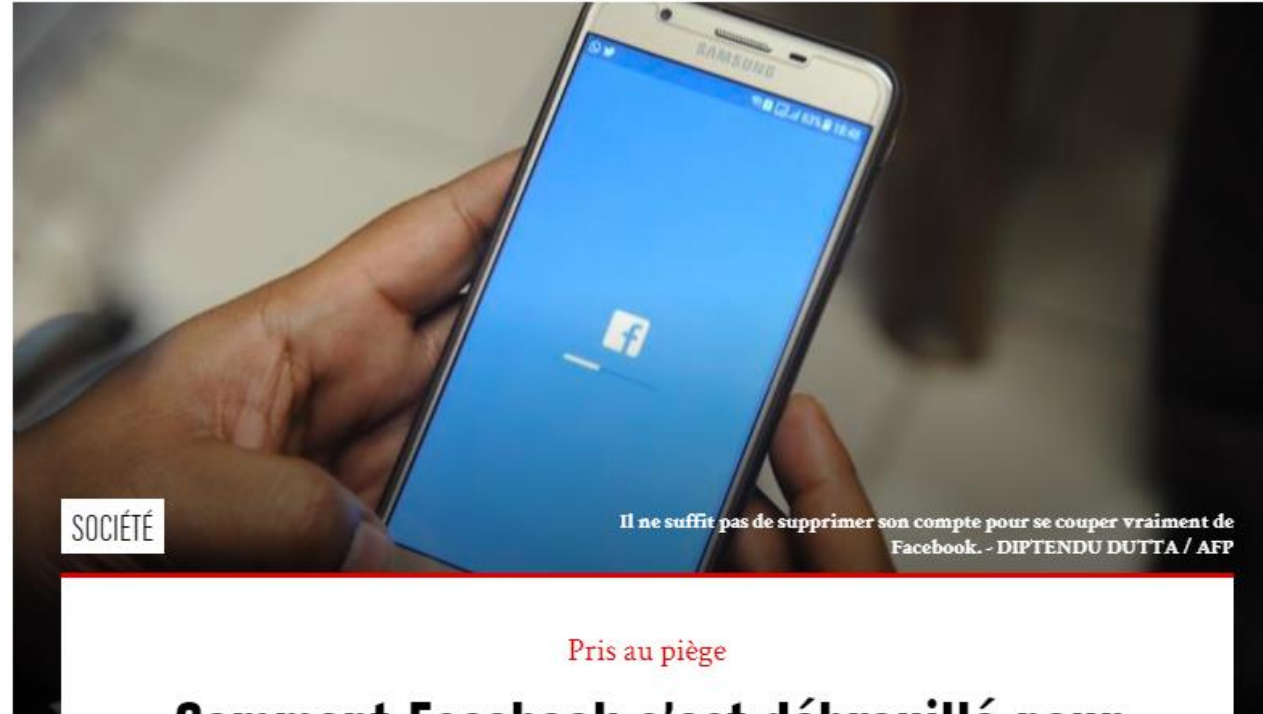


Google Photos labeled black people 'gorillas'

Jessica Guynn | USA TODAY
Updated 2:50 PM EDT Jul 1, 2015



2. Anonymat



**Comment Facebook s'est débrouillé pour
que vous ne le quittiez pas**

3. Identifiant unique

En France, toute personne est identifiée dès sa naissance par son **numéro d'inscription au répertoire national d'identification des personnes physiques** (NIR). Le NIR, communément dénommé « numéro INSEE » ou « numéro de sécurité sociale », se compose de 15 chiffres. Il est généralement considéré comme très signifiant, car il indique le sexe, l'année et le mois de naissance, ainsi que le code chiffré de la commune de naissance et le numéro d'ordre sur le registre des actes de naissance.

OR

Pour éviter toute interconnexion des fichiers et toute utilisation des données personnelles à des fins autres que celles qui ont justifié leur enregistrement, **la Commission nationale de l'informatique et des libertés (CNIL) refuse l'utilisation d'un numéro unique d'identification des personnes physiques.**

4. Propriété de la données

Le Règlement sur la protection des données personnelles (RGPD) consacre certains droits sur nos données personnelles, tels que le droit d'accès et le droit d'opposition. Ces droits relèvent d'un système d'opt-out qui oblige la personne à demander au responsable de traitement de retirer, rectifier ou déréférencer ses données.



Ce que vous partagez, renseignez, liké, commenté sur Facebook : est-ce que c'est votre propriété ?

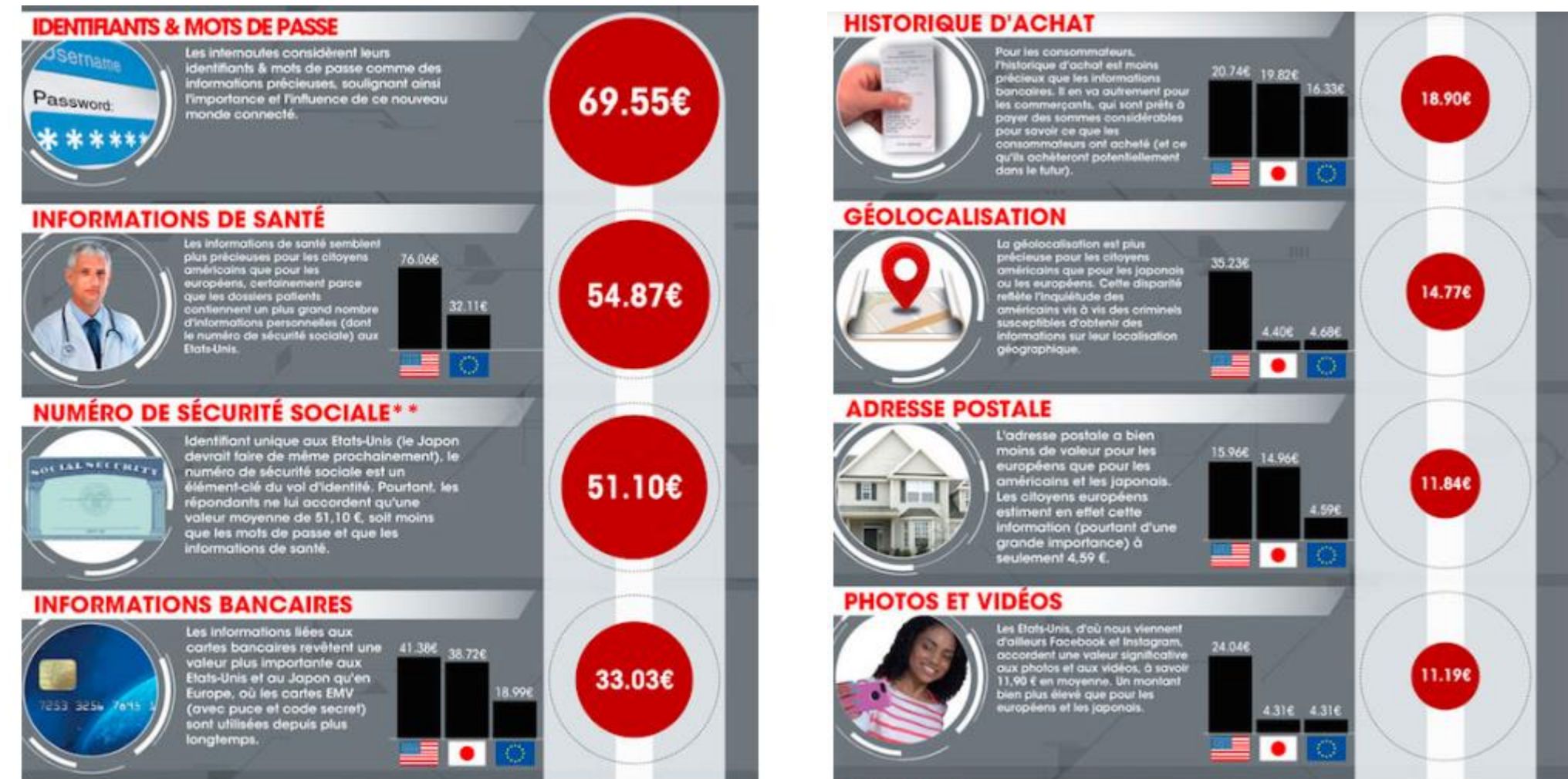
5. Vente/achat d'informations

Exemple du scandale Avast : L'éditeur de l'antivirus gratuit récoltait une masse de données impressionnantes à des fins marketing. Tout est enregistré et envoyé à Jumpshot, une filiale d'Avast.

Celle-ci revend ces informations très précieuses à des géants tels que... Google, Microsoft, L'Oréal ou encore Sephora.



5. Vente/achat d'informations



6. Fausse prédiction

Un homme en colère car sa fille de 16 ans reçoit des publicités lui vantant des habits de bébé

16 décembre 2015 par [Marc Sigrist](#)



Un crédit accordé car le scoring fait (via un algorithme de Machine Learning) donnait un bon résultat mais finalement le client n'honore pas ses mensualités

Big data & Données personnelles

Les BigData permettent de :

- Construire du contenu personnalisé
- Extraire des tendances
- Enrichir des profils,
- Etc.

Certaines applications sont basées sur un traitement intensif de **données personnelles**

Un haut niveau de protection est requis

MAIS il faut laisser suffisamment de flexibilité pour tirer parti de ces masses de données

Définition données personnelles

Une donnée est dite personnelle si :

- Toute information relative à une personne physique identifiée, ou pouvant l'être directement ou indirectement
- Données illégales : relatives à la santé, aux origines raciales, aux opinions politiques, philosophiques ou religieuses, à l'appartenance syndicale, aux moeurs, aux condamnations pénales

La CNIL - Application

Les champs d'application :

- ✓ Les données personnelles collectées et acquises en France
- ✓ Les données transmises à l'étranger
- ✓ Les traitement des données personnelles en France

La CNIL - Principes

Principe de finalité

- Les données personnelles recueillies ne peuvent être traitées que dans un usage **déterminé et légitime**

Principe de proportionnalité et de pertinence des données

- Seules doivent être traitées les informations **pertinentes et nécessaires** au regard des objectifs poursuivis.

Principe de conservation limitée dans le temps

- La durée doit être proportionnelle à la finalité

Principe de sécurité et de confidentialité

- Garantir la confidentialité des données et éviter leur divulgation à des tiers non autorisés

Principe du respect des droits des personnes

- Toute personne peut s'opposer au stockage et au traitement de ses données personnelles

La CNIL – Pour aller plus loin

Bienvenue sur le MOOC de la CNIL

Vous y trouverez l'ensemble des informations pour vous
initier au RGPD et débiter ainsi
la mise en conformité de votre organisme.

Ce dispositif gratuit est accessible jusqu'au mois de
septembre 2021.
En suivant l'intégralité de ce MOOC, vous pourrez obtenir
une **attestation**.

<https://atelier-rgpd.cnil.fr/>

Accéder au MOOC >

La réalité

Réutilisation des données

- La valeur des données Big Data est liée à leur futures utilisation et à leur réutilisation
- Incompatible avec certains principes clés

Manque de transparence sur la manière dont les données sont traitées

- Par exemple sur internet
- Incompatible avec le fait qu'un utilisateur a le droit d'accéder à ses données personnelles

Le **croisement de données** non sensibles peut générer de l'information sensible

- Risque de **ré-identification**
- Anonymisation robuste ?

Risque de décisions basées sur des **informations incorrectes**

Utilisation intensive du **prédictif** et de **décisions automatisées** :

- Faux positifs (Recherches Google)
- «Dictature» des données

3

Enjeux éthiques

Les 7 principes de l'UE

La Commission Européenne a prodigué sept principes à suivre pour créer des intelligences artificielles » dignes de confiance » :

- Le respect des droits de l'Homme
- **Les algorithmes IA doivent être sécurisés, fiables et suffisamment solides** pour faire face aux éventuelles erreurs et autres inconsistances tout au long du cycle de vie des systèmes IA
- Les citoyens doivent **conserver le contrôle complet sur leurs données**
- **la transparence et la traçabilité** des systèmes IA
- L'accessibilité pour tous
- Le respect de l'environnement
- La mise en place de mécanismes permettant d'asseurer **la responsabilité des créateurs de systèmes IA** en cas de problèmes

Le rapport Villani

CÉDRIC VILLANI

Mathématicien et député de l'Essonne

DONNER UN SENS À L'INTELLIGENCE ARTIFICIELLE

POUR UNE STRATÉGIE
NATIONALE ET EUROPÉENNE

01

Miser sur nos talents

La France compte déjà de nombreux talents. Ce sont eux qui font de la France un pays leader de l'intelligence artificielle, à nous de conforter cet écosystème prometteur.

02

Rassembler nos forces

La France a une particularité : elle possède des bases de données centralisées massives.
Le problème : elles sont sous-exploitées.

03

Poser un cadre éthique

L'intelligence artificielle fait parfois peur.

Permettre le développement de l'IA nécessite de répondre aux enjeux qu'elle pose

Quelques exemples d'actualités

Les tribunaux qui ont recourt à des algorithmes pour établir le profilage des condamnés en termes de « risque » en fonction de la couleur de leur peau à chaque étape de la procédure judiciaire

Les employeurs recrutent à l'aide d'algorithmes qui, par nature, favorisent certains groupes socioéconomiques.

Amazon a continué à commercialiser agressivement son outil de reconnaissance faciale Rekognition en dépit des interrogations sur la vie privée et les biais d'usage.

Comment s'y prendre ?

Les données brutes en tant que telles n'ont peu de valeur tant qu'elles n'ont pas été analysées.

Dans ce cas : est-ce que l'éthique des données ne devrait pas se concentrer moins sur les données et sur les algorithmes qui ont un impact sur la rationalité limitée qui définit les décisions humaines.

Sauf que

Notre recours au scientisme déprécie d'autres formes d'intelligence « humaine », notamment sur :

- L'intelligence émotionnelle (interpersonnelle)
- L'intelligence linguistique (intelligence des mots)
- L'intelligence intrapersonnelle (connaissance de soi)
- L'intelligence spirituelle (existentielle)

Finalement quels sont les risques éthiques ?

1. Risques politiques

Le Big Data peut donc être utilisé afin d'effectuer des campagnes d'influence ciblées pour orienter à son insu l'électeur dans son vote.



Finalelement quels sont les risques éthiques ?

2. Risques sociétaux

L'analyse automatique des données grâce une IA évolue donc constamment dans de multiples domaines, dont il est difficile d'anticiper les risques pour nos sociétés de demain.

D'ailleurs pour l'instant, les nations ne parviennent pas à trouver un accord sur la définition d'une arme mortelle autonome, et le développement des robots tueurs se poursuit sans être réglementé.

Finalelement quels sont les risques éthiques ?

3. Risques environnementaux

le Big Data impacte de manière non négligeable notre environnement. En effet, le traitement et le stockage d'un grand volume de données consomme une énorme quantité d'énergie.

Le Big Data serait à lui seul responsable de 7% de l'énergie électrique consommée et ce chiffre ne fera qu'augmenter.

D'où la nécessité d'encadrer son développement
pour garantir une exploitation sécurisée et éthique
des données.

Place au Travaux Dirigés

Enoncé disponible :

<https://github.com/apatou>