

***Computer science***  
— Théorie vs expérimentation —  
**Partie 4 : Ethique des algorithmes**

Christine Solnon

INSA de Lyon - 5IF

2019 / 2020

le temps des  
ILS TRANSFORMENT LES SCIENCES, L'INDUSTRIE, LA SOCIÉTÉ... ILS BOULE-

algorithmes  
VERSENT LES NOTIONS DE TRAVAIL, DE PROPRIÉTÉ, DE GOUVERNEMENT,

serge abiteboul  
DE VIE PRIVÉE... ET D'HUMANITÉ. QUI N'A PAS ENTENDU PARLER DES ALGORITHMES ? ILS

et gilles dowek  
NOUS FACILITENT LA VIE, MAIS NOUS REDOUTONS ÉGALEMENT QU'ILS NOUS ASSERVISSENT...



**DANS QUELLE SOCIÉTÉ  
SOUHAITONS-NOUS VIVRE ?**

# Principes éthiques de la commission Européenne

~ European Commission's High-Level Expert Group on Artificial Intelligence

## 5 principes fondamentaux :

- Beneficence : Do good
- Non-maleficence : Do no harm
- Autonomy : Respect for self-determination and choice of individuals
- Justice : Fair and equitable treatment for all
- Explicability : Operate transparently

# ACM statement on algorithmic accountability (1/2)

[https://www.acm.org/binaries/content/assets/public-policy/2017\\_usacm\\_statement\\_algorithms.pdf](https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf)

**Awareness** : Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

**Access and redress** : Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

**Accountability** : Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.

**Explanation** : Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.

# ACM statement on algorithmic accountability (2/2)

[https://www.acm.org/binaries/content/assets/public-policy/2017\\_usacm\\_statement\\_algorithms.pdf](https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf)

**Data Provenance** : A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

**Auditability** : Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.

**Validation and Testing** : Institutions should use rigorous methods to validate their models and document those methods and results. They should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

# Ethique des algorithmes

## Rapport de la CNIL sur les enjeux éthiques de l'IA (2017)

*(...) besoin d'une réflexion collective sur un pacte social dont certains aspects essentiels (libertés fondamentales, égalité entre les citoyens, dignité humaine) peuvent être remis en question dès lors que l'évolution technologique déplace la limite entre le possible et l'impossible et nécessite de redéfinir la limite entre le souhaitable et le non souhaitable.*

## Peut-on formaliser une éthique et la programmer dans un algorithme ?

Oui pour des situations bien définies, mais quid des situations inédites ?

### Que dit la loi ?

- Article 1 de la loi informatique et libertés de 1978 : *L'informatique doit être au service de chaque citoyen.(...) Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.*
- 3 principes relayés dans le RGPD 2018 :
  - Conditions de collecte et de conservation des données personnelles
  - Interdiction à une machine de prendre seule des décisions emportant des conséquences cruciales pour les personnes
  - Droit pour les personnes d'obtenir des informations sur la logique de fonctionnement de l'algorithme

# Plan du cours

## 1 Validation théorique d'algorithmes

## 2 Evaluation expérimentale d'algorithmes

## 3 Ingénierie algorithmique

## 4 Ethique des algorithmes

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

## 5 Conclusion du cours

# Quelles décisions un algorithme peut-il prendre ?

## Des décisions importantes sont déjà prises par des algorithmes :

- Donner une contravention pour excès de vitesse
- Acheter ou vendre des actions
- Piloter un avion
- Affecter des étudiants à des universités en fonction de leurs vœux
- Mettre en relation des personnes sur des sites de rencontre

Mais ce sont des humains qui définissent ces algorithmes

## Quelques questions :

- Peut-on déléguer à un algorithme tout type de décision ?  
↪ Quid de l'aide à la décision ?
- Comment expliquer une décision prise/conseillée par un algorithme ?
- Comment contester une décision prise par un algorithme ?
- Qui est responsable des conséquences d'une décision prise/conseillée par un algorithme ?



# Plan du cours

## 1 Validation théorique d'algorithmes

## 2 Evaluation expérimentale d'algorithmes

## 3 Ingénierie algorithmique

## 4 Ethique des algorithmes

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

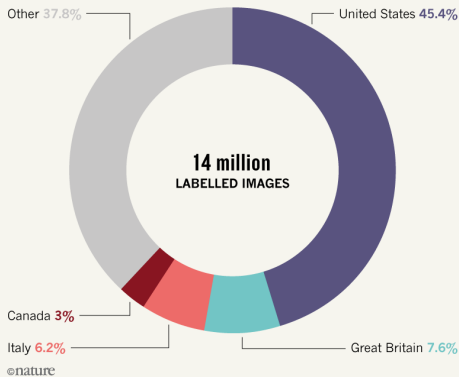
## 5 Conclusion du cours

# Un algorithme est-il toujours objectif et impartial ?

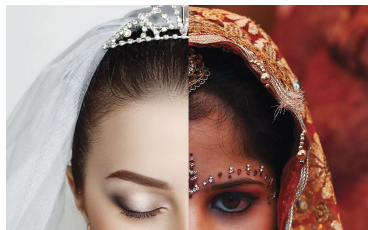
## ~ Exemple 1 : ImageNet

### IMAGE POWER

Deep neural networks for image classification are often trained on ImageNet. The data set comprises more than 14 million labelled images, but most come from just a few nations.



Laquelle de ces deux photos représente une mariée ?



Un modèle appris reproduit les biais de l'ensemble d'entraînement !

Zou and Schiebinger : *AI can be sexist and racist — it's time to make it fair*,  
in Nature 559, 324-326 (2018)

# Un algorithme est-il toujours objectif et impartial ?

~ Exemple 2 : Word Embedding (e.g., Word2vec)

## Principe du "word embedding" :

- Plongement des mots dans un espace vectoriel : les mots apparaissant dans des contextes similaires sont proches dans l'espace
- Permet d'inférer des analogies  
~  $king - man + woman = ?$

## Entraînement sur les articles de *news.google.com* :

- *doctor - man + woman =*
- *computer programmer - man + woman =*

## Question :

Les données d'entraînement devraient-elles être représentatives du monde tel qu'il est, ou du monde que nous aimerions avoir ?

# Un algorithme est-il toujours objectif et impartial ?

~ Exemple 2 : Word Embedding (e.g., Word2vec)

## Principe du "word embedding" :

- Plongement des mots dans un espace vectoriel : les mots apparaissant dans des contextes similaires sont proches dans l'espace
- Permet d'inférer des analogies  
~  $king - man + woman = queen$

## Entraînement sur les articles de *news.google.com* :

- $doctor - man + woman = ?$
- $computer\ programmer - man + woman = ?$

## Question :

Les données d'entraînement devraient-elles être représentatives du monde tel qu'il est, ou du monde que nous aimerions avoir ?

# Un algorithme est-il toujours objectif et impartial ?

~ Exemple 2 : Word Embedding (e.g., Word2vec)

## Principe du "word embedding" :

- Plongement des mots dans un espace vectoriel : les mots apparaissant dans des contextes similaires sont proches dans l'espace
- Permet d'inférer des analogies  
~  $king - man + woman = queen$

## Entraînement sur les articles de [news.google.com](https://news.google.com) :

- $doctor - man + woman = nurse$
- $computer\ programmer - man + woman = homemaker$

~ Les analogies reproduisent les biais présents dans les données

Bolukbasi et al : *Man is to Computer Programmer as Woman is to Homemaker ? Debiasing Word Embeddings*, in NIPS 2016

## Question :

Les données d'entraînement devraient-elles être représentatives du monde tel qu'il est, ou du monde que nous aimerions avoir ?

# Plan du cours

## 1 Validation théorique d'algorithmes

## 2 Evaluation expérimentale d'algorithmes

## 3 Ingénierie algorithmique

## 4 Ethique des algorithmes

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

## 5 Conclusion du cours

# L'informatique : solution pour la transition énergétique ?

## L'informatique peut aider à réduire nos consommations :

- Apprendre des modèles pour prévoir les pics de consommation, et réduire la consommation pendant ces pics en déplaçant certaines demandes
- Optimiser des itinéraires
- Mettre en relation conducteurs et passagers pour covoiturer
- Faire des réunions en visio-conférence ou du télé-travail pour limiter les déplacements
- ...

→ Computational Sustainability

**Mais l'informatique consomme aussi des ressources !**

# Exemple 1 : Entraînement des IA

## Quelques chiffres trouvés dans des articles :

- Vainqueur ImageNet 2012 : 6 jours sur 2 GPU Nvidia
- AlphaGo Lee (2016) : plusieurs mois sur 48 Tensor Processing Units
- AlphaGo zero (2017) : 3 jours sur 64 GPU + 19 CPU + 4 TPU pour surpasser AlphaGo Lee, et 40 jours pour surpasser tous les alphaGo
- Skip-gram (word embedding) : 2,5 jours sur 125 cœurs

## Mais ce n'est que la partie visible de l'iceberg...

Consumption	CO <sub>2</sub> e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000

Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Consumer	Renew.	Gas	Coal	Nuc.
China	22%	3%	65%	4%
Germany	40%	7%	38%	13%
United States	17%	35%	27%	19%
Amazon-AWS	17%	24%	30%	26%
Google	56%	14%	15%	10%
Microsoft	32%	23%	31%	10%

Voir Strubell, Ganesh, & McCallum (2019) : Energy and Policy Considerations for Deep Learning in NLP



# Exemple 2 : Bitcoin

Source = <https://digiconomist.net/bitcoin-energy-consumption>

## Annualized Total Footprints

### Carbon Footprint

34.73 Mt CO<sub>2</sub>



Comparable to the carbon footprint of Denmark.

### Electrical Energy

73.12 TWh



Comparable to the power consumption of Austria.

### Electronic Waste

11.32 kt



Comparable to the e-waste generation of Luxembourg.

## Single Transaction Footprints

### Carbon Footprint

291.20 kgCO<sub>2</sub>



Equivalent to the carbon footprint of 727,993 VISA transactions or 48,533 hours of watching Youtube.

### Electrical Energy

613.05 kWh



Equivalent to the power consumption of an average U.S. household over 20.72 days.

### Electronic Waste

94.90 grams



Equivalent to the weight of 1.46 'C'-size batteries or 2.07 golf balls.

# L'Informatique : problème pour la transition énergétique ?

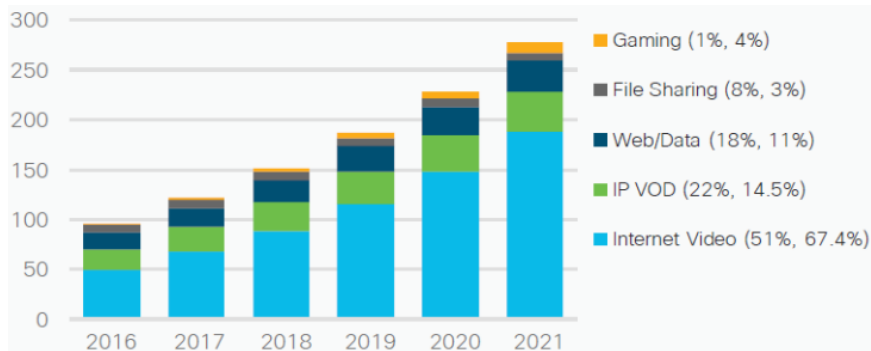
voir <https://theshiftproject.org>

## Consommation électrique des applications informatique :

Environ 10% de la consommation électrique mondiale

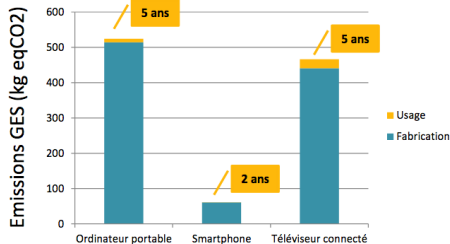
→ + 8% par an (consommation mondiale en hausse de 3%)

## Augmentation du trafic réseau de 25% par an

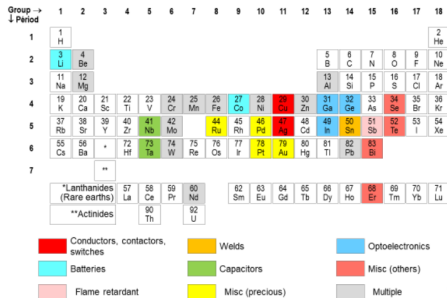


# L'informatique : problème pour la transition énergétique ?

voir <https://theshiftproject.org>



## The main metals of ICTs



Number of connected devices per capita	2016	2021	Annual growth
Asia-Pacific	1.9	2.9	8.3%
Central and Eastern Europe	2.5	3.8	9.1%
Latin America	2.1	2.9	7.0%
Middle East and Africa	1.1	1.4	5.4%
North America	7.7	12.9	11.0%
Western Europe	5.3	8.9	10.9%
Global	2.3	3.5	8.5%

# Plan du cours

## 1 Validation théorique d'algorithmes

## 2 Evaluation expérimentale d'algorithmes

## 3 Ingénierie algorithmique

## 4 Ethique des algorithmes

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

## 5 Conclusion du cours

# Que changent les plateformes d'intermédiation ? (1/2)

[www.societe-informatique-de-france.fr/wp-content/uploads/2015/12/1024-no7-Grumbach.pdf](http://www.societe-informatique-de-france.fr/wp-content/uploads/2015/12/1024-no7-Grumbach.pdf)

## Qu'est-ce que l'intermédiation ?

- Mise en relation de personnes avec des biens ou des services  
~> Marchés bifaces entre producteurs et consommateurs
- Activité de traitement d'information (appariement l'offre et la demande)  
~> Plus l'intermédiaire a d'informations, et meilleure est l'intermédiation

## Qu'est-ce que la dés-intermédiation ?

Remplacement d'un intermédiaire "traditionnel" par une plateforme

## Matière première des plateformes d'intermédiation : les données

- Données sur les utilisateurs et les services (collectées ou hébergées)  
~> Service initial
- Traces d'utilisation de la plateforme  
~> Nouveaux services (externalités)
- Données de fonctionnement de la plateforme  
~> Amélioration continue

## Que changent les plateformes d'intermédiation ? (2/2)

[www.societe-informatique-de-france.fr/wp-content/uploads/2015/12/1024-no7-Grumbach.pdf](http://www.societe-informatique-de-france.fr/wp-content/uploads/2015/12/1024-no7-Grumbach.pdf)

### Nouveaux modèles économiques :

- Prix d'utilisation bas/nuls pour attirer le plus grand nombre d'utilisateurs
- Bénéfices tirés d'activités secondaires (exploitation des données)

### Nouveaux types de domination :

- Poids économique des "GAFA"
- Le nombre d'utilisateurs d'une plateforme suit une loi exponentielle :  
+ une plateforme a d'utilisateurs, meilleure elle est, et + elle a d'utilisateurs...  
→ Création de situations de monopoles
- Déplacement du pouvoir : Vers le bas ("empowerment" de l'utilisateur) et vers le haut (les plateformes) → Organisation horizontale

### Nouvelles abstractions (couches numériques) du monde physique :

- Service rendu sans être physiquement présent (⇒ Souplesse)
- Nombreuses possibilités de partage : Voitures, logements, nourriture, énergie, connaissance, ... → Solutions à la transition écologique ?
- Partage équitable, continuité de service et adaptabilité

# Plan du cours

## 1 Validation théorique d'algorithmes

## 2 Evaluation expérimentale d'algorithmes

## 3 Ingénierie algorithmique

## 4 Ethique des algorithmes

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

## 5 Conclusion du cours

# L'informatique va-t-elle changer notre façon de travailler ?

## Les outils informatiques changent notre façon de travailler

Télétravail, Visio-conférences, documents partagés, ...

## Certains métiers disparaissent, d'autres apparaissent

- Transition beaucoup plus rapide que lors de la révolution industrielle
- Les nouveaux métiers sont très différents des métiers qui disparaissent  
→ Besoin de formation aux nouveaux métiers

## Les plateformes d'intermédiation changent l'organisation du travail

Facilitent le travail indépendant, mais aussi ponctuel et précaire

## La productivité augmente et le besoin de travail humain diminue

- Faut-il travailler moins pour produire autant, ou travailler autant pour produire plus ?
- Comment répartir les richesses créées sans travail humain ?  
Lire <https://lejournal.cnrs.fr/articles/six-scenarios-dun-monde-sans-travail>



# Vers une économie de la gratuité ?

## Les algorithmes sont des biens non rivaux

Ils sont partagés sans réduire la part de chacun, et à coût (quasi) nul

## Nombreuses ressources numériques gratuites

- Crowdsourcing (Wikipedia, OpenStreetMap)
- Ouverture des données publiques
- Logiciels Open Source

## Comment rétribuer les auteurs de biens non rivaux ?

↪ Crowdfunding, Freemium, Faire payer les services associés, ...

## ... et notre façon de penser et d'apprendre ?

### Nouveaux outils pour apprendre :

- Nombreuses données accessibles, à tous les niveaux
- Moteurs de recherche
- Cours en ligne

### Nouvelle forme de connaissance :

La connaissance "algorithmique"

→ On connaît quelque chose si on connaît l'algorithme qui la réalise

### Nouvelles façons de faire de la science :

Computational Sciences (cf cours G. Beslon)

## ... et notre façon de gouverner ?

### Echanges d'informations dans une démocratie représentative :

- Des candidats vers les électeurs : programme électoral
- Des électeurs vers les instances chargées de l'organisation des scrutins : bulletin de vote
- Des élus vers les personnes en charge d'appliquer les décisions

~> Un électeur communique une toute petite quantité d'information !

### Quelques évolutions...

Exemple : Loi pour une république numérique en 2016

- Consultation en ligne des citoyens
- Ouverture des données publiques

### Echanges d'informations horizontales : entre citoyens

~> Peuvent influencer les décisions politiques

# Plan du cours

## 1 Validation théorique d'algorithmes

## 2 Evaluation expérimentale d'algorithmes

## 3 Ingénierie algorithmique

## 4 Ethique des algorithmes

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

## 5 Conclusion du cours

# L'enseignement de l'informatique en France

- 1967 : Plan Calcul
- 1980 : Rapport Simon
  - ↳ Proposition de création d'un Capes d'Informatique
- 1985 : Plan informatique pour tous
  - ↳ Equipement en ordinateurs des établissements scolaires
- 2000 : Brevet Informatique et Internet
  - ↳ Apprendre à utiliser sans comprendre comment ça fonctionne...
- 2012 : Spé ISN en terminale scientifique
- 2016 :
  - Enseignement de l'informatique en primaire et au collège
    - ↳ Intégré aux cours de mathématiques et de technologie
  - Option Informatique et Créativité Numérique en 1ère (L, ES et S)
- 2017 : Option Informatique au Capes de Mathématiques
- 2020 : 1ère session du Capes "Numérique et sciences informatiques"

# Plan du cours

1 **Validation théorique d'algorithmes**

2 **Evaluation expérimentale d'algorithmes**

3 **Ingénierie algorithmique**

4 **Ethique des algorithmes**

- Quelles décisions un algorithme peut-il prendre ?
- Un algorithme est-il toujours objectif et impartial ?
- L'info : solution ou pb pour la transition énergétique ?
- Que changent les plateformes d'intermédiation ?
- L'informatique va-t-elle changer notre façon de vivre ?
- Comment l'informatique est-elle enseignée en France ?

5 **Conclusion du cours**

# Ce qu'il faut retenir ?

## L'informatique est une science...

...où il faut confronter la théorie et la pratique !

- Validation théorique d'algorithmes
  - Etudier la complexité du problème avant de concevoir un algorithme
  - Etudier la complexité théorique des algorithmes
  - Prouver la correction des algorithmes et du code
- Evaluation expérimentale d'algorithmes
  - Bien choisir les jeux d'essai, facteurs, *Design Points*, et indicateurs de performance
  - Analyser soigneusement les résultats
  - Faciliter la reproductibilité
- Ingénierie algorithmique
  - *Algorithm tuning vs code tuning*
    - ↪ Trouver le bon compromis entre efficacité et lisibilité
  - *Parameter tuning*
    - ↪ Utiliser des outils pour automatiser le paramétrage

# Le(s) mot(s) de la fin par Donald Knuth

*If you find that you're spending almost all your time on theory, start turning some attention to practical things ; it will improve your theories. If you find that you're spending almost all your time on practice, start turning some attention to theoretical things ; it will improve your practice.*

*Computer programming is an art, because it applies accumulated knowledge to the world, because it requires skill and ingenuity, and especially because it produces objects of beauty. A programmer who subconsciously views himself as an artist will enjoy what he does and will do it better.*

*We should continually be striving to transform every art into a science : in the process, we advance the art.*

*An algorithm must be seen to be believed.*

