

# IVRE, IL PRATIQUAIT LE ML SANS REcul

---

HUGO MOUGARD

4 février 2019

NANTES MACHINE LEARNING MEETUP

Je suis

ingé R&D ML à [source{d}](#)

co-organisateur de ce meetup

*Je ne suis pas*

climatologue

sociologue

magistrat

...

Je reviens de la FOSDEM

Je reviens de la FOSDEM, en avion



# PLAN DE LA PRÉSENTATION

Introduction

Le ML est-il important ?

Coût carbone

Technocratie

Asymétrie des acteurs

Biais et ML

Pistes d'action

# INTRODUCTION

---

BUT

Examiner le ML sous quelques angles critiques

# QUELQUES POINTS DE DÉPART

- Collapsologie
- Rapports du GIEC sur le Climat
- Révélations d'EDWARD SNOWDEN
- Climat social
- Épisode de DATA GUEULE *Algocratie*

Étude de l'effondrement de la société thermo-industrielle.

[Meetup](#) sur Nantes.

[Rapport](#) spécial sur le réchauffement d'1,5°C (2018).

# RÉVÉLATIONS D'EDWARD SNOWDEN



Omnipotence de la surveillance (2014).

Gilets jaunes, optimisation fiscale, augmentation des inégalités, ...



Épisode 84, *Algocratie*

LE ML EST-IL IMPORTANT ?

---

Pour X, Y et Z,

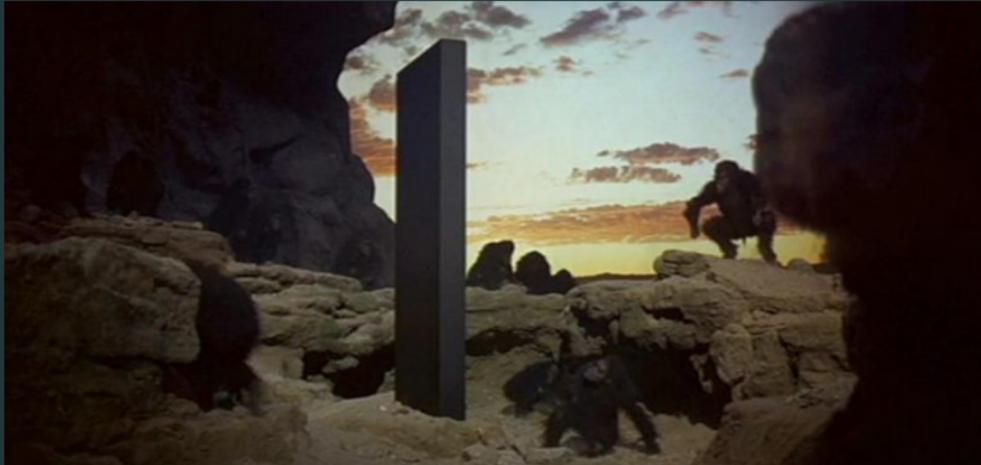
*le ML est-il vraiment important ?*

Replaçons les algorithmes au centre <sup>1</sup>.

---

1. Basé *très vaguement* sur une [présentation](#) de Thierry Caminel.

# ÉVOLUTION GÉNÉTIQUE



Accumulation biochimique des algorithmes (*très lente*).



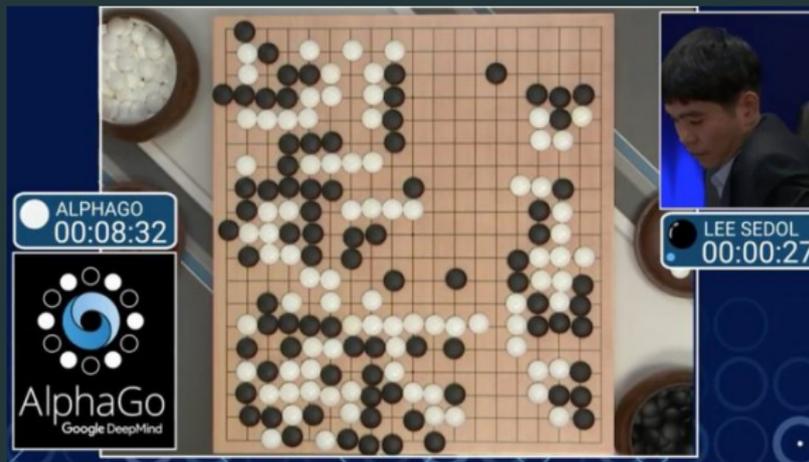
Accumulation culturelle des algorithmes (lente).

# HISTOIRE MODERNE

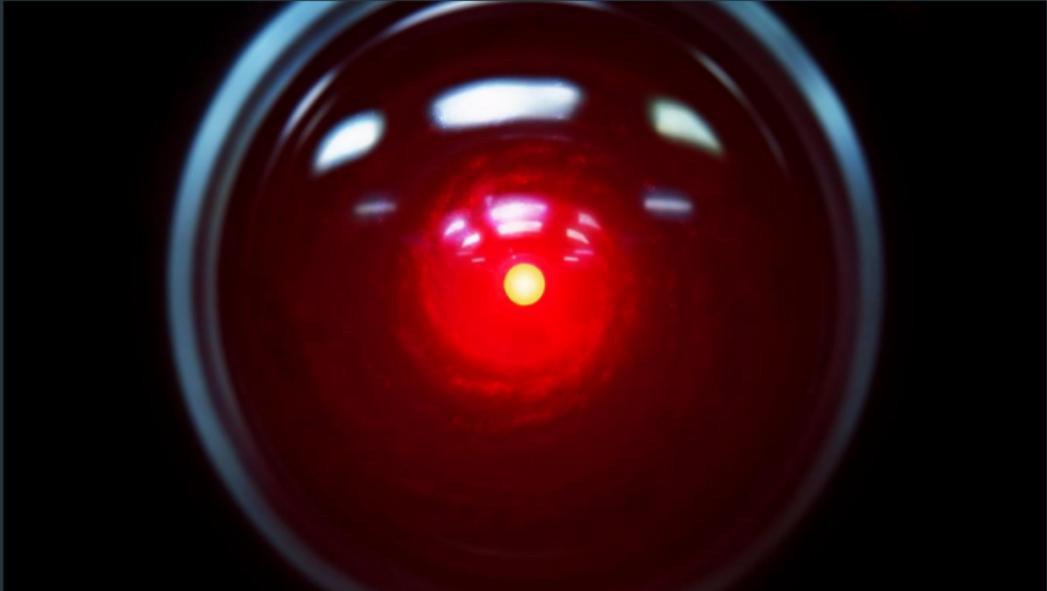


Accumulation explicite des algorithmes — informatique (rapide).

# MAINTENANT



Accumulation semi-automatique des algorithmes — AI/ML (très rapide).



Accumulation automatique des algorithmes — AGI (quasi instantanée).

Capital dominant : algorithmes.

## MEILLEURES VALORISATIONS EN BOURSE

---

2T 2007	2T 2018
PetroChina	<i>Apple</i>
Exxon	<i>Amazon</i>
GE	<i>Google</i>
China Mobile	<i>Microsoft</i>
Bank of China	Berkshire
<i>Microsoft</i>	<i>Facebook</i>
Gazprom	<i>Alibaba</i>
Shell	<i>Tenscent</i>
AT&T	JP Morgan

---

Algorithmes = enjeu majeur

ML = enjeu majeur

# COÛT CARBONE

---

Utilisation de l'impact du numérique comme proxy<sup>2</sup>.

---

2. Toute cette section est basée sur le [Rapport pour une sobriété numérique de THE SHIFT PROJECT](#)

# BACKGROUND : PLAN

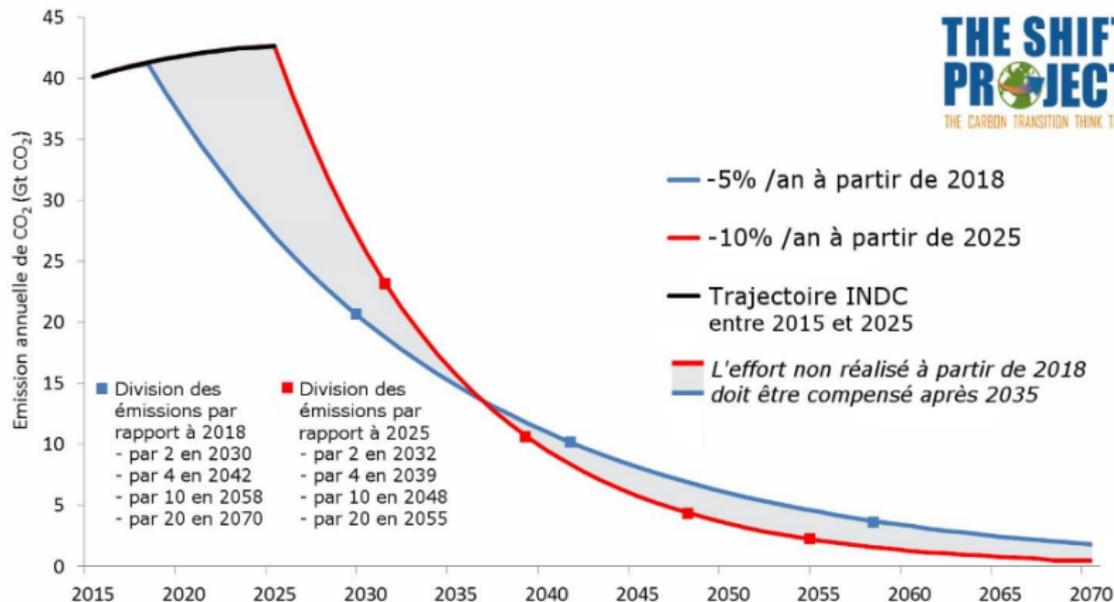
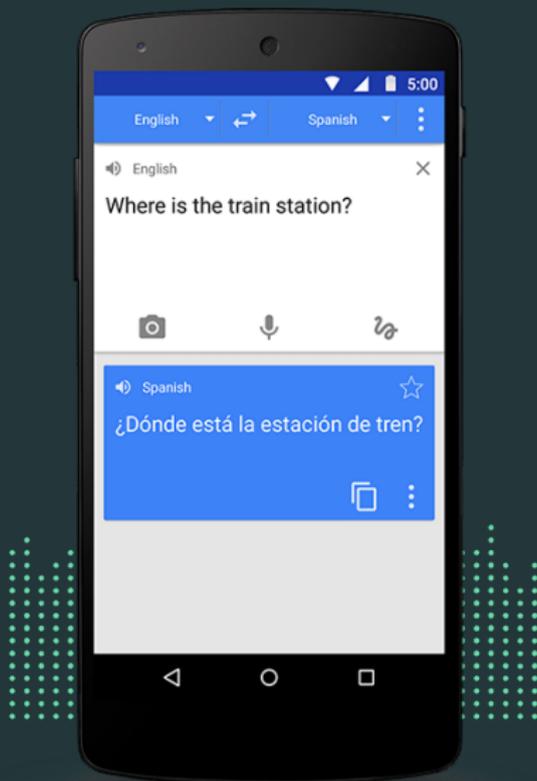


Figure 4 : Trajectoires d'émission compatibles avec une hausse de température limitée à 2°C  
[Source : The Shift Project, 2016]

4% (2% pour l'aéronautique domestique, 8 pour le parc automobile).

Impact énergétique de la vue d'une vidéo : **1500** fois plus grand que la consommation du smartphone

# EFFET YOUTUBE — GOOGLE TRANSLATE VU PAR UN UTILISATEUR



## EFFET YOUTUBE — GOOGLE TRANSLATE VU PAR UN DATA SCIENTIST



Accroître le rendement énergétique d'un objet **augmente** la consommation d'énergie globale dédiée à la fonction technique que remplit cet objet.

Pour un smartphone, 33 fois supérieur à sa consommation électrique annuelle.

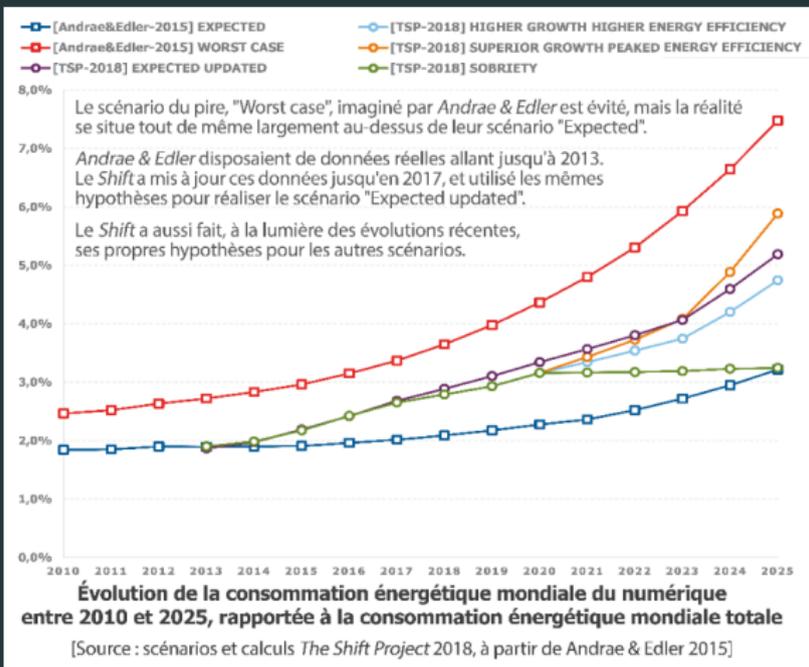
# CONSOMMATION DU NUMÉRIQUE EN CROISSANCE

<b>Consommation d'énergie en Twh</b>	<b>2015</b>	<b>2020</b>	<b>2025</b>	<b>CAGR<sup>8</sup> 2015/2020</b>	<b>CAGR 2020/2025</b>
Expected - 2015	2312	2878	4350	4,5%	8,7%
Worst - 2015	3677	5976	12 352	10%	15,5%
Expected updated	2389	3834	6254	9,9%	10,2%
Higher growth higher EE	2373	3622	5716	8,9%	9,5%
Superior growth peaked EE	2373	3622	7096	8,9%	14,5%
Sobriety	2373	3622	3909	8,9%	1,6%

Tableau 1 : Consommation d'énergie mondiale du Numérique en TWh

[Source : [Lean ICT Materials] Forecast Model. Produit par The Shift Project à partir des données publiées par (Andrae & Edler, 2015)]

# PART DU NUMÉRIQUE EN CROISSANCE



# TECHNOCRATIE

---

- par les experts

- par les experts
- méritocratie

- par les experts
- méritocratie
- productivité

# COMPARATIF RÉDUCTEUR, SUBJECTIF & SUPERFICIEL

	Technocratie	Démocratie
Question optimisée	Comment	Quoi
Décideurs	Experts	Citoyens / représentants
Valeur optimisée	Intérêt sectoriel	Intérêt général

NAONEDIA



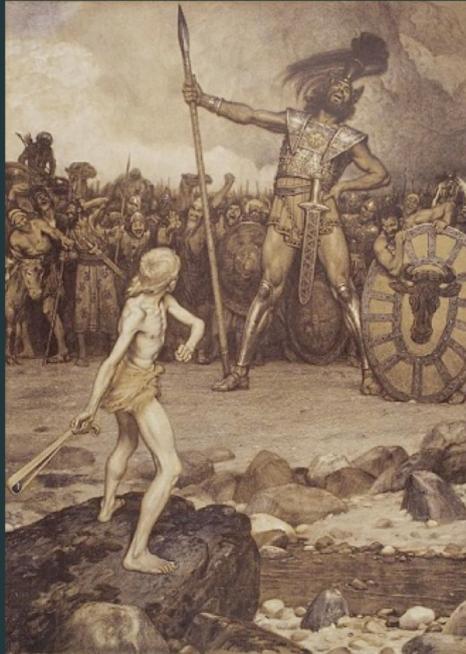
ML = expert parfait → renforcement de la technocratie.



# ASYMÉTRIE DES ACTEURS

---

# QUESTION



Tous les acteurs peuvent-ils utiliser le ML aussi efficacement ?

Trade-off entre libertés individuelles et contrôle étatique<sup>3</sup>.

**Sûreté** Protection contre l'état

**Sécurité** Protection contre les autres citoyens

---

3. Basé sur l'excellent billet de blog de Maître Eolas [Relisons la notice](#).

Quelques stats sur la NSA

Quelques stats sur la NSA

- $\approx$  11 milliards en 2013 (Snowden)

Quelques stats sur la NSA

- $\approx$  11 milliards en 2013 (Snowden)
- Datacenter en Utah : des Exa-octets en stockage ( $10^{18}$  octets)

Quelques stats sur la NSA

- $\approx$  11 milliards en 2013 (Snowden)
- Datacenter en Utah : des Exa-octets en stockage ( $10^{18}$  octets)
- 30k à 40k employés

Quelques stats sur la NSA

- $\approx$  11 milliards en 2013 (Snowden)
- Datacenter en Utah : des Exa-octets en stockage ( $10^{18}$  octets)
- 30k à 40k employés

→ Fort déséquilibre en faveur de la sécurité.

Et en France

Et en France

- Installation de boîtes noires chez les FAI

Et en France

- Installation de boîtes noires chez les FAI
- Renforcement du pouvoir administratif

Et en France

- Installation de boîtes noires chez les FAI
- Renforcement du pouvoir administratif

→ En sûreté, recours au ML compliqué

Coût d'entrée élevé en ML dû à

Coût d'entrée élevé en ML dû à

- la force de calcul nécessaire (ALPHAGo)

Coût d'entrée élevé en ML dû à

- la force de calcul nécessaire (ALPHAGo)
- la quantité de données nécessaires (CRITEO)

Coût d'entrée élevé en ML dû à

- la force de calcul nécessaire (ALPHAGO)
- la quantité de données nécessaires (CRITEO)
- la quantité d'utilisateurs nécessaire (WAZE)

Coût d'entrée élevé en ML dû à

- la force de calcul nécessaire (ALPHAGO)
- la quantité de données nécessaires (CRITEO)
- la quantité d'utilisateurs nécessaire (WAZE)

→ Effet monopole accentué.

# BIAIS ET ML

---

Dû à l'entraînement depuis les données, risque de biais.

Dû à l'entraînement depuis les données, risque de biais.

- pas biais *statistique*

Dû à l'entraînement depuis les données, risque de biais.

- pas biais *statistique*
- biais *social*

Dû à l'entraînement depuis les données, risque de biais.

- pas biais *statistique*
- biais *social*
- possibilité d'être biaisé statistiquement et pas socialement, ou l'inverse.

Recette classique :

Recette classique :

1. données biaisées → modèle biaisé

Recette classique :

1. données biaisées  $\rightarrow$  modèle biaisé
2. modèle biaisé  $\rightarrow$  nouvelles données biaisées

Recette classique :

1. données biaisées → modèle biaisé
2. modèle biaisé → nouvelles données biaisées
3. **goto** 1., potentiellement avec intérêts

Recette classique :

1. données biaisées → modèle biaisé
2. modèle biaisé → nouvelles données biaisées
3. **goto** 1., potentiellement avec intérêts
4. *Profit*. Biais éternel

Cas YOUTUBE

## Cas YOUTUBE

- optimise le temps passé sur le site

## Cas YOUTUBE

- optimise le temps passé sur le site
- vidéo polémique > vidéo normale pour ce critère

## Cas YOUTUBE

- optimise le temps passé sur le site
- vidéo polémique > vidéo normale pour ce critère
- biais social énorme vers les complot, Trump, etc, ...

## Cas YOUTUBE

- optimise le temps passé sur le site
- vidéo polémique > vidéo normale pour ce critère
- biais social énorme vers les complot, Trump, etc, ...

→ Facteur important de l'élection de Trump.

## PISTES D'ACTION

---

Directement du [rapport](#) *Pour une sobriété numérique* :

- adopter la sobriété numérique comme principe d'action

Directement du [rapport \*Pour une sobriété numérique\*](#) :

- adopter la sobriété numérique comme principe d'action
- accélérer la prise de conscience des impacts environnementaux du numérique

Directement du [rapport \*Pour une sobriété numérique\*](#) :

- adopter la sobriété numérique comme principe d'action
- accélérer la prise de conscience des impacts environnementaux du numérique
- permettre aux organisations de piloter environnementalement leur transition numérique

Directement du [rapport \*Pour une sobriété numérique\*](#) :

- adopter la sobriété numérique comme principe d'action
- accélérer la prise de conscience des impacts environnementaux du numérique
- permettre aux organisations de piloter environnementalement leur transition numérique
- procéder à un bilan carbone des projets numériques

Directement du [rapport \*Pour une sobriété numérique\*](#) :

- adopter la sobriété numérique comme principe d'action
- accélérer la prise de conscience des impacts environnementaux du numérique
- permettre aux organisations de piloter environnementalement leur transition numérique
- procéder à un bilan carbone des projets numériques
- améliorer la prise en compte des aspects systémiques du numérique dans les secteurs clefs

Directement du [rapport \*Pour une sobriété numérique\*](#) :

- adopter la sobriété numérique comme principe d'action
- accélérer la prise de conscience des impacts environnementaux du numérique
- permettre aux organisations de piloter environnementalement leur transition numérique
- procéder à un bilan carbone des projets numériques
- améliorer la prise en compte des aspects systémiques du numérique dans les secteurs clefs
- mettre en place des mesures à l'échelle européenne

- intégrer le coût équivalent carbone comme métrique décisionnelle

# PRENDRE CONSCIENCE DU TRADE-OFF FEATURES/COÛT ÉCOLOGIQUE

- intégrer le coût équivalent carbone comme métrique décisionnelle
- régulièrement, à plusieurs échelles

# PRENDRE CONSCIENCE DU TRADE-OFF FEATURES/COÛT ÉCOLOGIQUE

- intégrer le coût équivalent carbone comme métrique décisionnelle
- régulièrement, à plusieurs échelles
- en prenant en compte les effets YOUTUBE, rebond, etc

# PRENDRE CONSCIENCE DU TRADE-OFF FEATURES/COÛT ÉCOLOGIQUE



<https://hippocrate.tech/>

<https://hippocrate.tech/>

1. Intégrité scientifique et rigueur

<https://hippocrate.tech/>

1. Intégrité scientifique et rigueur
2. Transparence

<https://hippocrate.tech/>

1. Intégrité scientifique et rigueur
2. Transparence
3. Équité

<https://hippocrate.tech/>

1. Intégrité scientifique et rigueur
2. Transparence
3. Équité
4. Respect

<https://hippocrate.tech/>

1. Intégrité scientifique et rigueur
2. Transparence
3. Équité
4. Respect
5. Responsabilité et indépendance

Lié à l'indépendance du Serment d'Hippocrate du Data Scientist.

Consultation des citoyens, sociologues, magistrats, ...

Très peu d'infos disponibles encore sur le coût carbone et social du ML.

Communiquer à vos utilisateurs l'impact de votre utilisation du ML.

Merci pour votre attention !

Questions / débat time !