

# Enjeux numériques du monde contemporain

Une économie de l'attention  
Ce qu'implique la gratuité ?

**Fabien Tarissan**

CNRS – ENS Paris Saclay – ISP

ENUM

# Rappel

Le web en chiffres (2022)

<http://www.internetlivestats.com/>

- 5,5 milliards d'utilisateur
- 1,9 milliards de sites web
- 500 millions de tweets et 5 milliards de vidéos **par jours**
- 5 milliards de requêtes et 200 milliards de mails envoyés **par jours**

⇒ Nécessité d'**algorithmes de classement**

**Popularité** (À côté du web) : **nombre** de références.  
*Clics, views, ..*

**Autorité** (Au dessus du web) : **autorité** des références  
*Links* ⇒ PageRank

**Notoriété** (À l'intérieur du web) : **affinité** (et fraîcheur!) avec l'information issue de son **voisinage**.  
*Tweet, likes, ...* ⇒ EdgeRank

**Prédiction** (Sous le web) : information **inférée** par le comportement d'un utilisateur.  
*Traces* ⇒ Youtube / Spotify / Deezer / Netflix / ...

# Enjeux autour des données

## Ce que nous révèlent les études

Les algorithmes suscitent des inquiétudes :

**The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections.** Robert Epstein and Ronald E. Robertson. *PNAS* 112 (33), 2015 (doi :10.1073/pnas.1419828112).

**Experimental evidence of massive-scale emotional contagion through social networks.** Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock . *PNAS* 111 (24), 2014 (doi :10.1073/pnas.1320040111).

**Exposure to ideologically diverse news and opinion on Facebook.** Eytan Bakshy, Solomon Messing, Lada A. Adamic. *Science*, 348 (6239), 2015 (doi :10.1126/science.aaa1160).

**Measuring the effect of collaborative filtering on the diversity of users' attention.** Augustin Godinot, Fabien Tarissan. In *Applied Network Science*, vol. 8, Springer (2023).

# Enjeux autour des données

## Ce que nous révèlent les études

Les algorithmes suscitent des inquiétudes :

- Impact sur la manière de s'informer

**The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections.** Robert Epstein and Ronald E. Robertson. *PNAS* 112 (33), 2015 (doi :10.1073/pnas.1419828112).

**Experimental evidence of massive-scale emotional contagion through social networks.** Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock . *PNAS* 111 (24), 2014 (doi :10.1073/pnas.1320040111).

**Exposure to ideologically diverse news and opinion on Facebook.** Eytan Bakshy, Solomon Messing, Lada A. Adamic. *Science*, 348 (6239), 2015 (doi :10.1126/science.aaa1160).

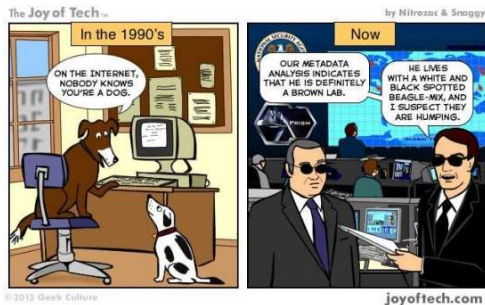
**Measuring the effect of collaborative filtering on the diversity of users' attention.** Augustin Godinot, Fabien Tarissan. In *Applied Network Science*, vol. 8, Springer (2023).

# Enjeux autour des données

## Ce que nous révèlent les études

Les algorithmes suscitent des inquiétudes :

- Impact sur la manière de s'informer
- De plus en plus de collecte de données



# Enjeux autour des données

## Ce que nous révèlent les études

Les algorithmes suscitent des inquiétudes :

- Impact sur la manière de s'informer
- De plus en plus de collecte de données

⇒ Les plateformes détiennent beaucoup d'informations

### Explicites

- profils renseignés
- pages créées
- commentaires
- achats

### Implicites

- sites consultés
- temps passés sur les sites
- recherches effectuées
- publicités efficaces

**Pourquoi cette collecte de données ?**

# Économie des plateformes

## Coûts

- 
- 
- 
-

# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes



# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

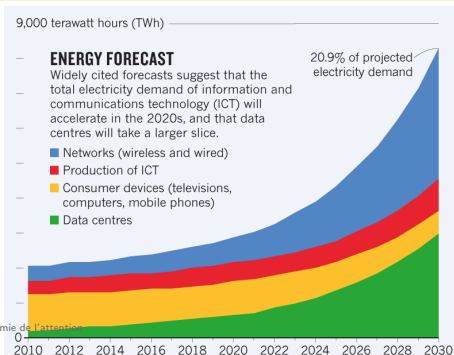


# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

How to stop data centres from gobbling up the world's electricity? Nicolas Jones, *Nature*, 2018

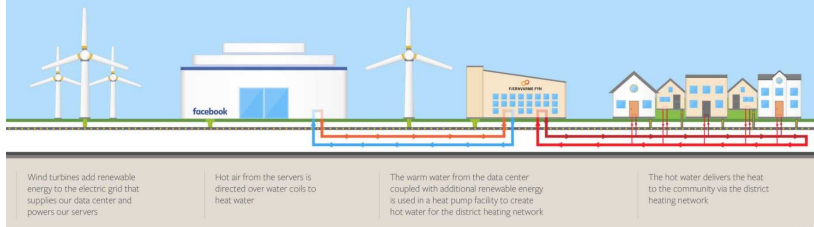


# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

### Odense Data Center: Heat Recovery Process



# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

### Projet de loi portant création d'une taxe sur les services numériques [...] (Taxe Gafa)

<http://www.assemblee-nationale.fr/15/projets/pl1737.asp>

- 1 Quelles sont les motivations politiques pour la mise en place de cette taxe ?
- 2 Quelles sont les motivations économiques pour la mise en place de cette taxe ?
- 3 Quelle forme prendrait cette taxe ?
- 4 Quels secteurs de l'économie du numérique sont visés ?
- 5 Quelles limites sont mises en place ?
- 6 Quelles entreprises en particulier sont concernées par cette taxe d'après vous ?
  - entreprises françaises
  - entreprises étrangères

# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

### Projet de loi portant création d'une taxe sur les services numériques [...] (Taxe Gafa)

<http://www.assemblee-nationale.fr/15/projets/pl1737.asp>

*L'article 1er met en place une taxe sur les produits bruts tirés de certains services numériques fournis par les grands groupes du secteur caractérisés par l'importance de l'internaute dans la création de valeur. Sont concernés les services de ciblage publicitaire, qui permettent d'individualiser les publicités par leur mise en adéquation avec les préférences des internautes, ainsi que les services d'intermédiation numériques, notamment les places de marché du commerce en ligne.*

# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

### Projet de loi portant création d'une taxe sur les services numériques [...] (Taxe Gafa)

<http://www.assemblee-nationale.fr/15/projets/pl1737.asp>

*Les services taxés se caractérisent par une capacité contributive particulière, tirée de la valorisation de l'activité des internautes. Les revenus qui en sont tirés présentent deux caractéristiques principales qui sont déterminantes dans la définition des paramètres de l'impôt. D'une part, les règles de territorialité applicables en fiscalité directe **permettent aux opérateurs de les localiser dans des pays différents de celui où sont situés ces internautes.** D'autre part, le modèle économique sous-jacent repose sur des rendements fortement croissants qui conduisent naturellement à des **positions hégémoniques** de certains acteurs.*

# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

### Projet de loi portant création d'une taxe sur les services numériques [...] (Taxe Gafa)

<http://www.assemblee-nationale.fr/15/projets/pl1737.asp>

*Afin de tenir compte de ces particularités une règle de rattachement spécifique au territoire national est prévue : les revenus mondiaux générés par chaque service seront taxés à **proportion de la part des internautes français** au sein de l'ensemble des utilisateurs de ce service. Seules les entreprises des grands groupes ayant une forte empreinte numérique au niveau mondial (montant annuel **mondial** des produits tirés des services taxés **supérieur à 750 millions d'euros**) et au niveau **national** (montant annuel des produits rattachés à la France tirés des services taxés **supérieur à 25 millions d'euros**) sont concernées car, du fait de la structure concurrentielle des marchés en cause, elles ne sont pas dans une situation comparable à celle des entreprises plus petites.*

# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

### Projet de loi portant création d'une taxe sur les services numériques [...] (Taxe Gafa)

<http://www.assemblee-nationale.fr/15/projets/pl1737.asp>

*Un **taux de 3 %** uniforme sera appliqué à toutes les entreprises redevables.*

↳ Concerne ~30 entreprises en France (revenu 2024 estimé à 800 millions d'euros)

**Vente de biens** Adevinta, Alibaba, Amazon, Apple, Ebay, Google, Groupon, Rakuten, Wish, Zalando.

**Intermédiaire de services** Amadeus, Axel Springer, Booking, Expedia, Match.com, Randstad, Recruit, Sabre, Travelport Worldwide, Tripadvisor, Uber.

**Publicité en ligne** **Criteo**, Facebook, Google, Microsoft, Twitter, Verizon.



# Économie des plateformes

## Quatre galaxies du numérique

PRINCIPALES ACTIVITÉS ET FILIALES DES GAFA

### Valorisation boursière

en milliards de dollars  
au 22 octobre



### Chiffre d'affaires

en milliards de dollars  
en 2018



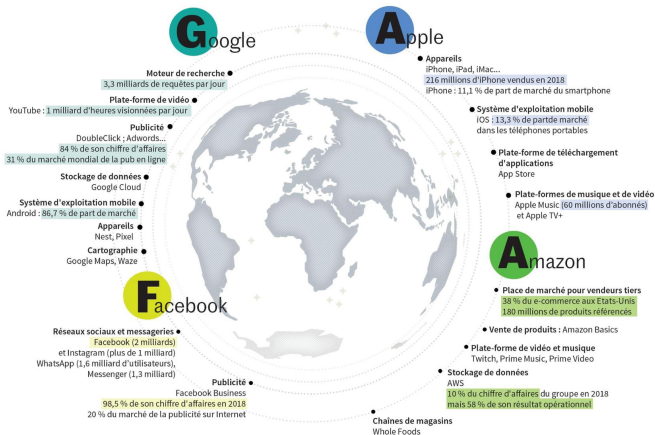
### Résultat net

en milliards de dollars  
en 2018



### Effectifs

en milliers de salariés



INFOGRAPHIE : LE MONDE

SOURCES : IDC ; E-HARKETER; STRATEGY ANALYTICS; ENTREPRISES

# Économie des plateformes

## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

## Recettes

- Ventes matériels et logiciels
- Ventes en lignes (intermédiation)
- Publicités ciblées
- *Freemium*
- Abonnement

⇒ Peut générer énormément de revenus

## Quelques capitalisations

- IMB :  $\leq$  1980
- Apple, MS : années 80
- Amazon : 1997
- Google : 2004
- Facebook : 2012
- Snapchat : 2017
- Spotify : 2018

# Économie des plateformes

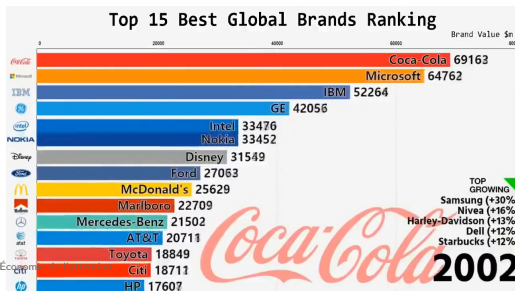
## Coûts

- Matériel (datacenter)
- Employé.e.s
- Énergie
- Taxes

## Recettes

- Ventes matériels et logiciels
- Ventes en lignes (intermédiation)
- Publicités ciblées
- *Freemium*
- Abonnement

⇒ Peut générer énormément de revenus



# Les dérives du web

*I imagined the web as an open platform that would allow everyone, everywhere to share information, access opportunities and collaborate across geographic and cultural boundaries. [...] But over the past 12 months, I've become increasingly worried about three new trends, which I believe we must tackle in order for the web to fulfill its true potential as a tool which serves all of humanity.*

Tim Berners-Lee, 12/03/2017.

## 3 défis lancés à la communauté des utilisateurs :

- Nous avons **perdu le contrôle** de nos données personnelles
- Les **fausses informations** se répandent trop facilement sur le web (cf cours 1)
- La **publicité politique ciblée** nécessite plus de la transparence

# Données personnelles

## Donnée à caractère personnel (art. 4 du RGPD)

*Toute information se rapportant à une personne physique identifiée ou identifiable ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale ;*

Concernant la perte de contrôle, il faut bien différencier :

# Données personnelles

## Donnée à caractère personnel (art. 4 du RGPD)

*Toute information se rapportant à une personne physique identifiée ou identifiable ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale ;*

Concernant la perte de contrôle, il faut bien différencier :

## Ce qui est illégal ...

### Dangers suite à des attaques

- 2013 : *Yahoo!* (2013) 3 Milliards de comptes
- 2015 : *Ashley Madison* (rencontres adultères) 30 Go de données
- 2017 : *Equifax* (agence de crédit US) 147 Millions de comptes

↳ voir aussi les révélations de Snowden (2013)

## .... de ce qui met en avant la perte de contrôle

### Scandales :

- 2016 : *Geofeedia*
- 2018 : *Cambridge Analytica*

# Publicité politique ciblée

**Computer-based personality judgments are more accurate than those made by humans.** Wu Youyou, Michal Kosinski, and David Stillwell. PNAS 112 (4), 2015 (doi :10.1073/pnas.1418680112).

**These are the political ads Cambridge Analytica designed for you.** Jeremy B. Merrill, Olivia Goldhill, Quartz, 2020.

**Communication politique : quelles sont les règles pour l'utilisation des données issues des réseaux sociaux ?**, CNIL, 2016 (actualisée en 2023)

À l'aide des documents contenus dans les trois articles ci-dessus, vous rédigerez une ébauche de synthèse expliquant **quels enjeux numériques sont soulevés par les techniques de campagnes de publicités politiques ciblées**. Vous analyserez notamment les éléments suivants

**Le phénomène en lui-même :** *En quoi les réseaux sociaux permettent-ils de renouveler les méthodes de campagnes politiques ?, En quoi le ciblage publicitaire peut-être mis en lien avec les campagnes électorales ?, Sur quels éléments scientifiques cette proposition s'appuie-t-elle ?, Que sait-on de l'efficacité de ces techniques dans le contexte politique ?, ...*

**Le contexte éthique et juridique :** *Quelles questions éthiques et juridiques ces campagnes politiques ciblées posent-elles ?, Quels fondements juridiques peuvent invoquer celles et ceux qui les emploient ?, Quelle est la nature des données mises en cause ?, Quelle est la position des instances sur ces questions et quels arguments sont mis en avant ?, ...*

# Publicité politique ciblée

**Computer-based personality judgments are more accurate than those made by humans.** Wu Youyou, Michal Kosinski, and David Stillwell. PNAS 112 (4), 2015 (doi :10.1073/pnas.1418680112).

**These are the political ads Cambridge Analytica designed for you.** Jeremy B. Merrill, Olivia Goldhill, Quartz, 2020.

**Communication politique : quelles sont les règles pour l'utilisation des données issues des réseaux sociaux ?**, CNIL, 2016 (actualisée en 2023)

## Révélation de Cambridge Analytica (CA)

L'exploitation des données récupérées par CA :

- Contrat avec l'équipe de campagne de D. Trump
- Morcelage de l'image du candidat
- Réutilisation des technique de publicités ciblées

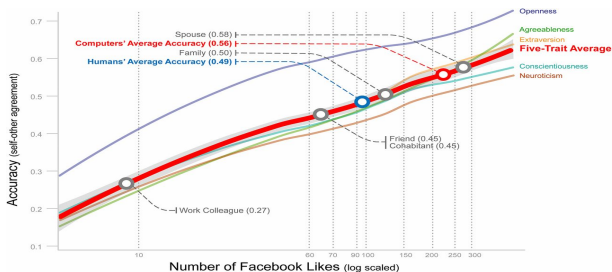


# Publicité politique ciblée

**Computer-based personality judgments are more accurate than those made by humans.** Wu Youyou, Michal Kosinski, and David Stillwell. PNAS 112 (4), 2015 (doi :10.1073/pnas.1418680112).

**These are the political ads Cambridge Analytica designed for you.** Jeremy B. Merrill, Olivia Goldhill, Quartz, 2020.

**Communication politique : quelles sont les règles pour l'utilisation des données issues des réseaux sociaux ?**, CNIL, 2016 (actualisée en 2023)



# Publicité politique ciblée

**Computer-based personality judgments are more accurate than those made by humans.** Wu Youyou, Michal Kosinski, and David Stillwell. PNAS 112 (4), 2015 (doi :10.1073/pnas.1418680112).

**These are the political ads Cambridge Analytica designed for you.** Jeremy B. Merrill, Olivia Goldhill, Quartz, 2020.

**Communication politique : quelles sont les règles pour l'utilisation des données issues des réseaux sociaux ?**, CNIL, 2016 (actualisée en 2023)

## Exemples de logiciels de ciblage

- *NationBuilder* : Trump, Mélenchon, Fillon, Juppé, ...
- *Blue State Digital* : Obama, Hidalgo, ...
- *50+1* : Hollande, Macron, ...
- *Corto* : NKM, ...

## Avis de la CNIL (2016) : Communication politique

*La combinaison de données sur chacun des utilisateurs d'un service (comme un réseau social), en l'absence d'outil de contrôle suffisant à leur disposition et de possibilité de s'opposer au profilage, ne peut se fonder sur l'intérêt légitime du responsable de traitement en l'absence de juste équilibre avec les droits et libertés des personnes concernées. Le consentement des internautes est donc nécessaire. Ces combinaisons de données doivent se fonder sur le consentement des personnes et ne sauraient être mises en œuvre uniquement sur le fondement de l'intérêt, même légitime, du candidat ou parti.*

# Quels remparts ?

# Quels remparts ?

Science / outils / bonnes pratiques

- Qwant, DuckDuckGo, Brave, Protonmail, VPN, Signal, ...

**Insuffisant** : Yves-Alexandre de Montjoye, César Hidalgo, Michel Verleysen et Vincent D. Blondel, *Unique in the crowd : The privacy bounds of human mobility*, Scientific Reports, vol. 3, 2013.

# Quels remparts ?

Science / outils / bonnes pratiques

- Qwant, DuckDuckGo, Brave, Protonmail, VPN, Signal, ...

**Insuffisant** : Yves-Alexandre de Montjoye, César Hidalgo, Michel Verleysen et Vincent D. Blondel, *Unique in the crowd : The privacy bounds of human mobility*, Scientific Reports, vol. 3, 2013.

Thomas Lamieux

Fingerprinting  
Une technique de traçage peu connue

# Quels remparts ?

Science / outils / bonnes pratiques

- Qwant, DuckDuckGo, Brave, Protonmail, VPN, Signal, ...

**Insuffisant** : Yves-Alexandre de Montjoye, César Hidalgo, Michel Verleysen et Vincent D. Blondel, *Unique in the crowd : The privacy bounds of human mobility*, Scientific Reports, vol. 3, 2013.

Thomas Lamieux

Fingerprinting

Une technique de traçage peu connue

Légiférer / Réguler / Encadrer

RGPD (2016), Loi pour une République numérique (2016), Loi relative à la lutte contre la manipulation d'information (2018), le DSA/DMA (Europe), la CNIL, ...

# Quels remparts ?

## Science / outils / bonnes pratiques

- Qwant, DuckDuckGo, Brave, Protonmail, VPN, Signal, ...

**Insuffisant** : Yves-Alexandre de Montjoye, César Hidalgo, Michel Verleysen et Vincent D. Blondel, *Unique in the crowd : The privacy bounds of human mobility*, Scientific Reports, vol. 3, 2013.

Thomas Lamieux

Fingerprinting  
Une technique de traçage peu connue

## Légiférer / Réguler / Encadrer

RGPD (2016), Loi pour une République numérique (2016), Loi relative à la lutte contre la manipulation d'information (2018), le DSA/DMA (Europe), la CNIL, ...

## Éducation / formation / médiation

Réforme du bac (2019) :

- Sciences Numériques et Technologie (SNT) : toutes les 2ndes
- Numérique et Sciences Informatiques (ISN) : spécialité de 1ère et Terminale

Enjeux : **Nouveau corps enseignant**

*Questions ?*

<http://tarissan.complexnetworks.fr/>