

BC19 - Opening Internet Monopolies to Competition with Data Sharing Mandates Presentation

Intro

DIAPO 1

Le document que l'on va vous présenter est un « policy brief ». Alors qu'est-ce qu'un policy brief ?

Vous êtes scientifique ou expert dans un organisme de recherche ou de développement. Vous avez publié les résultats d'une recherche ou d'une expertise dans le cadre d'un projet. Ces résultats peuvent être utiles à des décideurs pour faire évoluer les idées ou aider à la décision : vous pouvez en tirer parti pour rédiger un policy brief. Le *policy brief* argumente un message (idée, politique, méthode, solution...) ayant une portée sur une question ou un problème auxquels sont confrontés des décideurs : il propose une analyse et des éléments de réflexion ou d'action. Le *policy brief* est généralement en libre accès sur internet.

Donc celui-ci a pour titre : « Opening Internet Monopolies to Competition with Data Sharing Mandates », il a été écrit par Claudia Blancotti et Paolo Ciocca en 2019, donc c'est un document plutôt très récent, qui traite donc de problématiques très actuelles. Claudia Blancotti est une économiste à la banque d'Italie. Paolo Ciocca est commissaire chez Consob, la commission du stock exchange italien. On a donc deux spécialistes de l'économie comme auteurs.

DIAPO 2

Ce « policy brief » traite du partage des données des utilisateurs des sites Internet aux entreprises des sites qu'ils visitent. Plus précisément, nous allons parler du fait que les GAFAM (Google, Apple, Facebook, Amazon et Microsoft) concentrent en grande majorité ces données. On parle même de « surveillance capitaliste ». Evidemment, cette concentration des données, qui représentent à la fois une source de revenus pour ces entreprises, et un moyen de grandir et de s'améliorer, posent de nombreux problèmes sur différents aspects :

DIAPO 3

- un problème de sécurité sur le stockage de ces données

-le droit à la vie privée, préoccupation croissante des consommateurs en Europe et en Amérique

-le problème de la concurrence, qui est la crainte que la domination des entreprises qui possèdent les données d'une majorité d'utilisateurs d'Internet empêchent les concurrents potentiels de rentrer sur le marché.

Par la suite, nous allons vous montrer comment les auteurs analysent ces problèmes et cherchent des solutions pour les résoudre, bien que celles-ci ne soient pas si simples car il faut à la fois protéger la vie privée et la sécurité des utilisateurs.

DIAPO 4

Collective security

DIAPO 5

Data in unprecedented volume

Depuis 2015, les ordinateurs ont commencé à surpasser les humains dans les tâches de reconnaissance d'objets.

ImageNet a ouvert la porte à l'immense croissance du machine learning (ML), un type d'intelligence artificielle (IA) basé sur des algorithmes qui dérivent des règles de décision à partir d'exemples observés.

Au fil du temps, les applications d'Intelligence artificielles sont passées de tâches simples telles que la classification d'objets statiques à des efforts considérablement plus complexes, qui ont encore accru - et continueront d'augmenter - la pertinence des données pour le progrès technique, social et économique. L'OCDE (Organisation de coopération et de développement économiques) a défini l'innovation fondée sur les données comme « un pilier clé de la croissance économique du XXI^e siècle » (OCDE 2015).

DIAPO 6

Collective security

Evidemment, Les sociétés GAFAM bénéficient d'un avantage significatif en matière de données par rapport à leurs concurrents.

Bien que les sociétés GAFAM diffèrent en termes du type de variables qu'elles collectent, chacune d'elles connaît ses utilisateurs dans la plupart des dimensions suivantes: les informations d'identification personnelle, y compris les caractéristiques physiques; contacts sociaux; localisation géographique; emploi; croyances, opinions et préférences; et les actions effectuées en ligne, qui peuvent inclure les pages Web visitées, les produits achetés,

le montant d'argent dépensé, les liens cliqués, les vidéos regardées et les recherches effectuées.

En effet, cela représente énormément d'informations dont on ne se rend pas toujours compte qu'on les divulgue. Par exemple, on s'inquiète moins lorsque l'on poste la photo d'un paysage que lorsque l'on poste une photo de famille. Cependant, cette photo donne tout de même l'information du lieu où vous vous trouvez et quand vous vous y trouviez.

Et cette information intéresse tout autant les hackers que votre photo de famille car ils peuvent en déduire des informations sur des questions sensibles telles que les opérations militaires ou les initiatives politiques par exemple.

Le problème de la concentration des données entre peu d'acteurs, est que si ses services ne sont pas disponibles, même pendant un court instant, les conséquences sont immenses car elles touchent énormément de personnes.

Nous savons tous que les vulnérabilités techniques touchent tous les systèmes, même ceux des GAFAM. Par exemple Facebook a subi plusieurs violations de données, la dernière a affecté 30 millions d'utilisateurs. Et bien sûr, les cyberattaquants se concentrent sur ces entreprises car elles possèdent énormément de données appartenant à énormément de personnes.

Un des plus grands risques serait que des acteurs hostiles prennent avantage de la portée presque universelle des GAFAM pour manipuler l'opinion publique, notamment grâce au machine Learning. Ceci crée des conflits entre les acteurs publics et privés car les entreprises veulent plus d'espaces publicitaires et le plus ciblés possible alors que les régulateurs veulent protéger l'intégrité du processus démocratique.

Jusqu'à maintenant, les entreprises ont toujours régulé elles-mêmes, ce qui n'est évidemment pas facile car elles reçoivent des accusations de manipulation d'un côté et de l'autre elle fait face à des questions sur les problèmes de censure. Bien sûr, la désinformation émerge mais l'inaction serait dangereuse pour ces plateformes qui risquent de perdre la confiance des utilisateurs.

Il faut donc trouver une solution pour éviter ces attaques.

DIAPO 7

Consumer rights

Le second risque associé à la concentration des données concerne les droits des consommateurs.

Si la loi de protection des consommateurs existe et concerne toutes les entreprises quelles que soient leur taille, il y a encore énormément de zones grises dans ce qui concerne le traitement des données et ce sont donc les entreprises des GAFAM, comme elles touchent des milliards de personnes, qui vont décider implicitement des normes et standards.

En ce qui concerne la vie privée, des consensus ont été trouvés. Par exemple, un utilisateur doit pouvoir séparer la sphère publique de la sphère privée en fixant ses propres limites de qui peut accéder à quelles informations (profil public ou privé sur Instagram ou facebook par exemple). De plus, l'utilisation, le partage et la vente de données personnelles collectées par les entreprises doit nécessiter un consentement de la part de l'utilisateur.

Pour lutter contre la concentration des données, une meilleure connaissance de la part des utilisateurs de où vont leurs données et qui a accès à leurs informations serait nécessaire. Lanier et Weyls proposent en 2018, la coordination des individus pour négocier avec les entreprises. En effet, tant que les personnes concernées essaient de vendre leurs informations individuellement, leur pouvoir de négociation est minime.

Enfin, d'autres affaires concernant les GAFAM et leurs collections de données est qu'elles sont parfois utilisées pour pratiquer la discrimination. En 2018, le ministère américain du Logement et du Développement urbain a constaté que Facebook offrait aux propriétaires et aux promoteurs immobiliers la possibilité de faire de la publicité de manière sélective en fonction de l'ethnie. Il est difficile pour les régulateurs de détecter ce genre de débordements car les preuves dans les algorithmes sont perdues au milieu d'autres facteurs dans les modèles de boîte noire et souvent, cela est fait de manière plus subtile dans les algorithmes où des variables telles que l'ethnicité, le sexe, l'âge et la religion peuvent être mal prises en compte dans les décisions.

DIAPO 8

En conclusion, la concentration des données par les GAFAM met en danger la vie privée des utilisateurs car la décision sur le traitement de leurs données est concentrée entre très peu de mains, ce qui est inquiétant, malgré le fait que de plus en plus de régulations pour protéger la vie privée des utilisateurs soient mises en place.

Notes :

Compétition :

GAFAM = budget énorme → meilleurs centres de recherche → fuite des cerveaux

Mais on ne peut pas vraiment jouer sur cet élément

Si ils proposent toujours les meilleurs services, ce n'est pas très grave

Problème = Si un GAFAM entre sur un nouveau marché et qu'il écrase la concurrence, alors qu'il ne propose pas le meilleur service... (ou qu'il se repose sur ses lauriers).

→ GAFAM font la loi des prix (prix plus hauts) ; expérience utilisateur dégradée ; moins d'innovation

Raisons :

- Abus de position
- ML
 - = Le machine learning est une technique de programmation informatique qui utilise des probabilités statistiques pour donner aux ordinateurs la capacité d'apprendre par eux-mêmes sans programmation explicite. → plus il y a de données, meilleur est le résultat
 - **Flagship product markets** : internet search for Google and ecommerce for Amazon --> the best user experience because of ML.
 - **Other product market** : for example, text typed in search queries or emails can be used to improve natural language processing capabilities for digital home assistants. --> flagship offering can be used to gain dominant positions in other areas.

Possible Solution :

In 2019, the German competition authority prohibited Facebook from linking user data across different services absent user consent and clearly framed the practice as an abuse of dominant position.

- **Data sharing mandates (DSM) :**
 - = requirements for market leaders to **share anonymous user data** with competitors and with academia
 - **implementing a data sharing mandate** should carefully evaluate the tradeoff between the comprehensiveness of the information shared and re-identification risks, while also **providing new incentives for research on anonymization techniques**.
 - A DSM that obligates companies to sell data to competitors, as opposed to giving it away at no charge, could also be seen as an application of the essential facilities doctrine— which posits that a firm with exclusive control

over a facility that is essential for other firms to effectively compete in a downstream market has an obligation to **grant access to it in exchange for a reasonable price**.

- At the same time, DSM must avoid interventions that feel like arbitrary redistribution of profits that **could discourage both leaders and newcomers from investing**.
- Pros :
 - With respect to markets where the GAFAM companies are expanding but are not dominant yet, a **DSM would reduce risks of future monopolization** by allowing productivity gains from ML to spread throughout the economy.
 - This would especially **benefit small and medium enterprises** (SMEs) that are now unable to employ ML because they do not have enough information. All business owners would be able to obtain information on consumer opinions and decisions relating to their sector of activity, competitors, and products. **Design, production, and marketing choices would improve**. So would overall economic performance.
- Cons :
 - The majority of non-GAFAM companies do not have the means to attract top talent, so they would not be able to compete effectively even if they had much more data.
 - Solution : academic institutions, which still employ some experts, should be included among data recipients
 - Privacy protection :
 - In several jurisdictions, including the European Union and California, it is illegal for companies to share the personal data they collect with third parties without the consent of the data subject. It is legal to share anonymized data. Definitions of anonymization vary across countries, but the common principle is that information should be stripped of all characteristics that make it attributable to a specific individual
 - Many individuals declare that they value their privacy highly, while regularly revealing information online in exchange for small rewards such as “likes” on social platforms.(paradox)
 - Security :
 - Multiplying the number of entities that can access personal information also **increases opportunities for discrimination and exposure to cyberattacks**, as the **protection standards of smaller companies are generally worse** compared to larger ones
 - GAFAM revenues are :
 - Develop products that consumers like

- The availability of a large audience for marketing purposes --> **DSM need to preserve consumer identification**

- Dominance of companies possessing exclusive information on a majority of internet users :
 - Superior ability to recruit analytical talent
 - Prevent market entry of potential competitors who might ultimately deliver better goods and services or foreclose ones before they grow big enough to pose a threat.
 - AI algorithms need to be trained on large datasets to work properly (ML)
 - Ex : ImageNet : a dataset that initially contained 3.2 million digital photographs representing approximately 5.000 real-world objects. --> Machine Learning (ML).
 - Impact on :
 - Flagship product markets : internet search for Google and ecommerce for Amazon --> the best user experience because of ML.
 - Other product market : for example, text typed in search queries or emails can be used to improve natural language processing capabilities for digital home assistants. --> flagship offering can be used to gain dominant positions in other areas.
 - GAFAM have a constant live stream of information on what billions of people choose to do.
 - GAFAM are able to attract scientific talent on par with top research institutions.
 - Entrenched dominance can lead to permanent supracompetitive prices and/or degraded quality for consumers. It may also result in less innovation, since the dominant firm can enjoy rents without having to invest.
 - the potential anticompetitive effects of data concentration in cases relating to firms that do not sell information but rather use it as an input for other products.