



ÉCOLE CENTRALE LYON

MOS  
VEILLE TECHNOLOGIQUE  
RAPPORT

---

## Rapport méthodologique

---

*Élèves :*  
Nathan ETourneau

*Enseignant :*  
Daniel MULLER

21 mars 2022

## Table des matières

<b>1</b>	<b>Démarche générale de veille</b>	<b>2</b>
<b>2</b>	<b>Sources : ciblage et récolte</b>	<b>3</b>
2.1	Google . . . . .	3
2.2	Google Scholar . . . . .	3
2.3	Towardsdatascience et Medium et autres blogs . . . . .	4
2.4	Twitter . . . . .	4
2.5	LinkedIn . . . . .	4
2.6	Github . . . . .	4
<b>3</b>	<b>Curation</b>	<b>5</b>
3.1	Google Alert . . . . .	5
3.2	Nombre de citations et recherche booléenne . . . . .	5
3.3	Zotero . . . . .	6
<b>4</b>	<b>Diffusion de la veille</b>	<b>6</b>

## Introduction

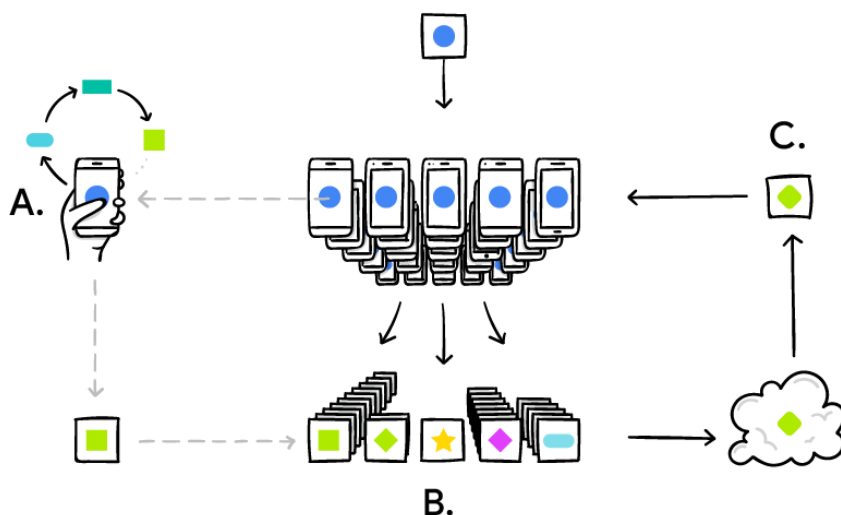


FIGURE 1 – Schéma de l'apprentissage fédéré.

<https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>

Pour la veille technologique de ce sujet de MOS, j'ai choisi de réaliser ma veille sur l'apprentissage fédéré, qui est une technique de Deep Learning qui permet l'apprentissage d'un modèle de réseaux de neurones centralisé, à partir de données décentralisées (par exemple réparties sur des smartphones ou sur des véhicules autonomes), sans jamais avoir à transmettre les données).

J'ai pour projet professionnel de devenir ingénieur-chercheur, et ce sujet me passionne. Pour être parfaitement honnête, le travail que j'ai réalisé n'est pas réellement celui d'une veille technologique, mais plutôt celui d'un premier état de l'art en vue d'une éventuelle thèse dans le domaine, car cela m'est plus utile compte tenu de mon projet professionnel.

## 1 Démarche générale de veille

Mon sujet de veille porte sur un domaine technique assez peu connu du grand public, et les principaux acteurs du sujet sont les GAFAM ainsi que les quelques entreprises qui se sont positionnées sur ce sujet. La première étape de mon travail a été de suivre ces entreprises et de cerner leurs cas d'usages de l'apprentissage fédéré.

Le domaine de ma veille technologique est un champ très actif de recherche, et des articles (ou prépublications) sortent tous les jours. Initialement, je lisais l'intégralité de ces articles, mais ce rythme est insoutenable et il est en plus inutile d'avoir autant d'informations pour une simple veille destinée à informer le grand public.

De nombreux auteurs (entreprises comme particuliers) ont publié des articles de blog à ce sujet. J'ai également pu trouver les slides de présentations de Google à ce sujet, ainsi

que quelques cours réalisés par certains universitaires sur des sujets connexes.

Enfin, j'ai fréquemment regardé les offres d'emplois et de stage liés à ce sujet, pour mieux identifier les entreprises ayant besoin de l'apprentissage fédéré ainsi que leurs cas d'usage.

La méthode que j'ai employé est plutôt une méthode **PULL** : l'essentiel de ma recherche a été manuel. Toutefois, j'ai essayé la méthode **PUSH** notamment à l'aide d'une alerte Google Scholar sur certains mots-clefs.

## 2 Sources : ciblage et récolte

### 2.1 Google

Un premier outil qui m'a été indispensable est le moteur de recherche Google, et notamment sa fonction de recherche booléenne. Les principaux mots-clefs que j'ai utilisé pour ma recherche sont :

- (federated learning) OR (apprentissage fédéré)
- ((federated learning) OR (apprentissage fédéré)) AND ((differential privacy) OR (homeomorphic encryption))
- ((privacy-preserving) OR (privacy)) AND ((federated learning) OR (apprentissage fédéré))
- ((federated learning) OR (apprentissage fédéré)) AND (applications OR (use cases))
- filetype :pdf intitle :(federated learning)

J'ai également recherché des éléments individuels pour me renseigner, mais cela constituait les principaux mots-clefs de ma recherche.

### 2.2 Google Scholar

L'outil qui m'a le plus servi est Google Scholar : il m'a permis de rechercher des articles de recherche sur l'ensemble des sujets d'intérêt pour l'apprentissage fédéré. Les mots clefs que j'ai utilisé pour Google Scholar sont assez similaires à ceux de Google :

- "federated learning"
- ("federated learning") AND ("differential privacy")
- ("federated learning") AND ("homeomorphic encryption")
- ("federated learning") AND ("secure multi-party computation")
- ("federated learning") AND (personalization)
- ("federated learning") AND (secure multi-party computation)
- ("federated learning") AND ("attack OR backdoor OR leakage")

J'ai filtré par dates, et j'ai énormément utilisé la fonction "voir d'autres articles similaires", qui permet de trouver des articles très pertinents n'ayant pas les mêmes mots-clefs dans son titre. En effet, une tendance pour les noms d'articles de recherche est d'avoir des noms accrocheurs (par exemple : Show, Attend and Tell, Attention is All You Need).

J'ai également mis en place une alerte Google sur les mots-clefs sus-mentionnés. Cela n'a pas été d'une grande aide : les alertes donnaient les articles récemment sortis, qui étaient trop techniques et pas assez révolutionnaires pour mes besoins.

## 2.3 Towardsdatascience et Medium et autres blogs

Une autre source utile pour ma veille a été les blogs communautaires Medium et TowardsDataScience. Ces blogs permettent aux auteurs de partager des connaissances sur un sujet technique. Le principal avantage de ces blogs est que les billets publiés sont d'un niveau bien plus accessible que les articles de recherche.

Via des recherches Google, j'ai trouvé plusieurs autres blogs intéressants, mais également des pages personnelles de chercheurs sur ce thème. J'ai notamment pu y découvrir des cours en PDF, ainsi que les slides de présentations sur ces thèmes.

## 2.4 Twitter

Une autre partie de ma veille a été réalisée sur Twitter. Les chercheurs sont en général assez actifs sur le réseau social, et j'ai pu y trouver des petits tweets de chercheurs ayant réalisé des travaux et souhaitant les partager. La concision du réseau permet de très rapidement filtrer l'information pertinente où non.

Les fils (threads) sont une très bonne façon d'exposer un sujet passionnant. Par exemple, ce thread était très intéressant.

## 2.5 LinkedIn

Une autre partie de ma veille a été effectuée sur LinkedIn. L'idée n'était pas tellement de trouver des publications pertinentes sur le réseau social, mais plutôt de rechercher les entreprises intéressées par le secteur, en regardant notamment les offres d'emploi. En effet, une entreprise qui recherche des chercheurs, des doctorants ou des stagiaires en Federated Learning est ainsi une entreprise très intéressée.

Une fois ces entreprises trouvées, une recherche avec des mots-clefs ciblés et ces entreprises permet d'identifier des contenus très intéressants.

En particulier, j'ai découvert par ce biais que Orange était une entreprise très intéressée par l'apprentissage fédéré. Les autres sont Apple/Facebook/Microsoft/Amazon/Nvidia/Intel, l'INRIA, le CEA et Owkin sont intéressés.

## 2.6 Github

Sur des sujets informatiques, GitHub est une mine d'informations : on y trouve des implémentations d'articles de recherches importants (plutôt à destination du chercheur), mais également des pages tenues par des amateurs appelées "Awesome xyz". Ce type de pages est un catalogue de ressources sur le sujet "xyz", et est donc une précieuse source d'information pour la veille ! Par exemple, ce repo Github m'a beaucoup aidé pour ma veille.

## 3 Curation

Je vais présenter les leviers qui m'ont permis d'identifier les informations les plus pertinentes parmi toute l'information disponible.

### 3.1 Google Alert

La première méthode que j'ai utilisé pour sélectionner l'information est Google Alert. Le principal problème est que j'ai reçu une quantité d'alertes trop importantes, j'ai donc souffert d'infobésité.

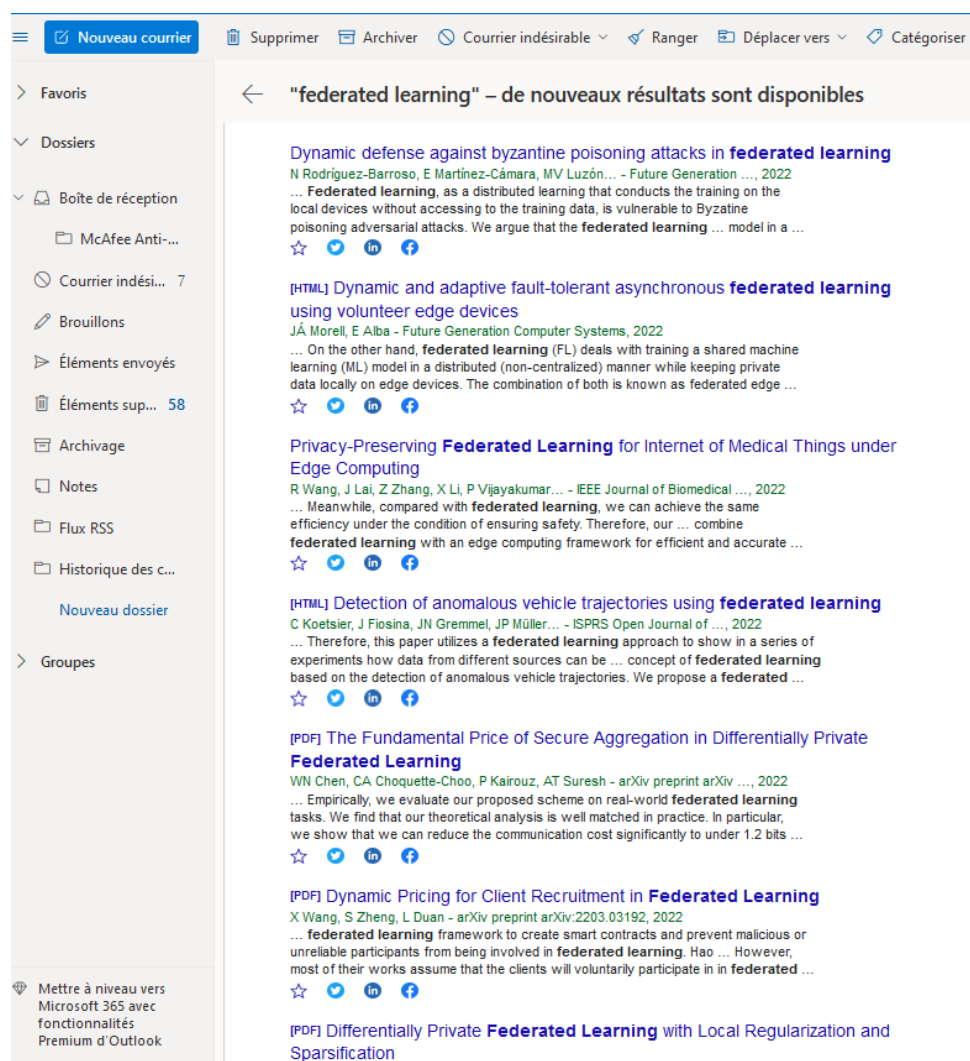


FIGURE 2 – Mail reçu quotidiennement avec Google Alert. Le mail complet est 3× fois plus long

### 3.2 Nombre de citations et recherche booléenne

Une autre méthode que j'ai utilisé pour extraire l'information pertinente est de regarder le nombre de citations des articles de recherche d'intérêt, ainsi que la recherche booléenne sur Google comme sur Google Scholar. Pour les articles de recherche, les mé-

triques sur l'influence des auteurs comme le score  $H_0$  étaient également des indicateurs de la qualité de l'information trouvée.

### 3.3 Zotero

Pour conserver les articles de recherche, j'ai créé un catalogue Zotero contenant les principaux articles du thème. Cette bibliothèque a un double intérêt : centraliser les articles principaux pour les relire si besoin, et constituer une première base bibliographique personnelle en vue d'une thèse dans le domaine.

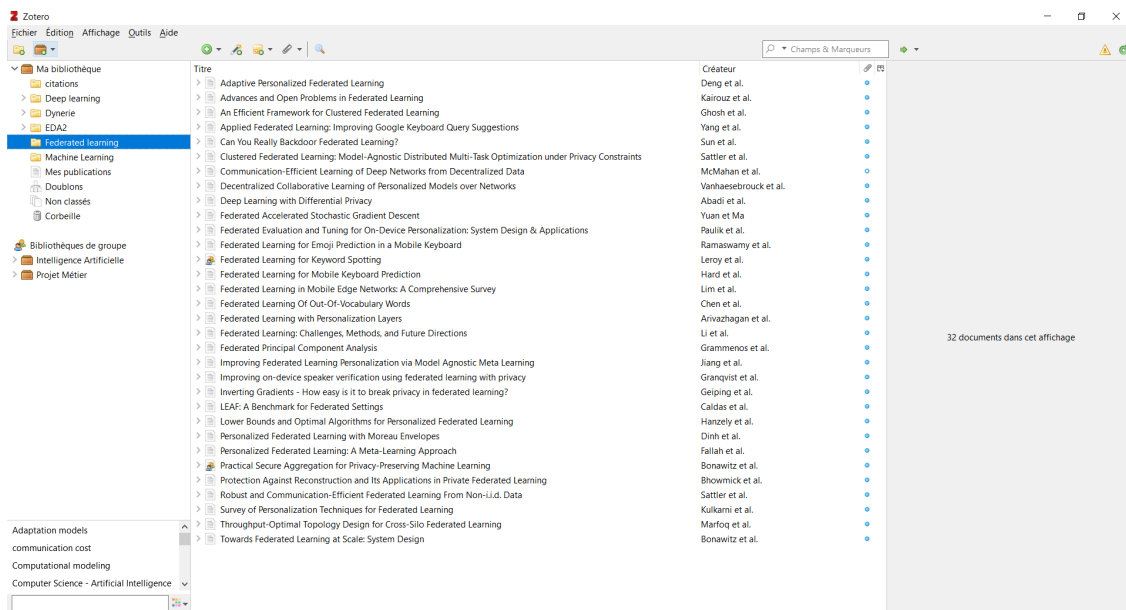


FIGURE 3 – Capture d'écran de ma bibliothèque locale Zotero

## 4 Diffusion de la veille

Pour la diffusion de ma veille, j'ai choisi de la restituer dans un README.md d'un repo Github. Le format Markdown est particulièrement adapté pour un tel travail.