

PROJET DE MASTER

Nom et prénom : Ouattara Bin'de Ange Emmanuel Désiré

Sujet : *BIG DATA et données personnelles*

I- Information générale

“*Mégadonnées*” ou encore “*données massives*”, le big data désigne en effet des volumes importants de données très diverses, traitées et analysées pour extraire des informations qui seront utilisées dans de nombreux domaines.

Ces mégadonnées sont qualifiées de « *carburant* » ou d’« *or noir* » car leur valeur alimente l’économie numérique. Leur exploitation permet aussi bien d’identifier les causes endogènes et exogènes de certaines maladies (en particulier du cancer), de fluidifier le trafic routier, de préciser le ciblage publicitaire, de réduire la consommation énergétique, etc.

La disponibilité de ces données est facilitée par l’évolution des techniques de transport et de stockage (fibre, cloud, etc.), qui facilitent la croissance exponentielle de la rétention de toutes les données.

Cependant, bien vrai que le BIG DATA traite certains sujets relatifs à la condition humaine (maladies, déplacements, ...), son problème majeur est la présence de données incomplètes. En clair, la qualité des données est loin d’être bonne à cause de certains employés qui commettent des erreurs et qui les accumulent au fur et à mesure, au fil des années. Et même si de nos jours ce problème est identifié, il reste caché et inconnu de l’organisation qui utilise ces Mégadonnées, et il reste sous-estimé.

II- Problème, objectifs, questions

A – Définition du problème

Le problème de ce mémoire est le suivant :

D’abord, nous ferons référence à l’effet GIGO, qui est la tendance à des personnes à accepter la véracité d’un système sans la remettre en question. Lorsqu’on rentre dans le système des données erronées, même si le système est parfait, logiquement les données restent erronées à la sortie. Le risque est alors de tirer de mauvaises conclusions sur de mauvaises informations.

Ensuite, nous ferons référence aux données collectées par les objets connectés. Les objets connectés sont des objets qui captent, traitent et transmettent les données, ils ont donc un lien direct avec le Big Data, et on a également la capacité à se connecter à un réseau d'information afin de recevoir et donner des instructions (smart home, smart phone, smart Watch...).

Ces données collectées vont ensuite être stockées puis grâce au Big Data elles vont pouvoir être analysées. Et plus le temps passe, plus les gens acquièrent ces objets et plus le volume de données augmente.

L'essor du Big Data et des objets connectés vis-à-vis des données personnelles, rend la sécurité de ce dernier de plus en plus compliquée à traiter, car les hackers convoitent les systèmes informatiques.

En somme, les travailleurs du BIG DATA semblent manquer de connaissances qui pourraient prévenir l'effet GIGO, et par conséquent, semble également manquer de protéger les données personnelles des utilisateurs des objets connectés.

B- Objectifs du mémoire

Le but de la recherche sera donc de découvrir quels sont les réels facteurs qui conduisent à l'effet GIGO (Garbage In Garbage Out), pourquoi les erreurs concernant la qualité des données traitées par le BIG DATA restent inconnues des organisations, sous-estimées et par conséquent non traitées, et quel est l'impact que ces deux derniers pourraient produire sur les données personnelles des utilisateurs des objets connectés.

Avec ses connaissances, nous pourrions envisager des solutions qui permettront de réduire, voire d'éradiquer l'effet GIGO afin de traiter en amont les erreurs concernant les données de mauvaises qualités, pour une sécurité plus optimale des données personnelles.

C- Question centrale

La question principale du mémoire est la suivante : ***Dans quelles mesures le BIG DATA impacterait négativement les données personnelles ?***

Pour répondre à la question principale, nous répondrons aux questions suivantes :

- Pourquoi les travailleurs du BIG DATA manquent d'objectivités vis-à-vis de celui-ci ?

- Pourquoi les mauvaises données récoltées et les données erronées identifiées sont quand même traitées ?
- Dans quelles mesures l'effet GIGO et le traitement des données de mauvaise qualité pourraient impacter négativement les données personnelles récoltées par le BIG DATA ?

III- Plan de recherche

A- Méthodes de recherche

Pour déterminer quels facteurs influencent négativement les données du BIG DATA face aux données personnelles, nous ferons dans un premier temps une *recherche documentaire* à travers des données antérieures qui nous permettront de vérifier si cela est effectivement le cas. Cette recherche nous permettra aussi de comparer les données antérieures aux données plus récentes, pour être certains que l'impact négatif que subissent les données personnelles provient exactement des erreurs accumulées par les travailleurs du BIG DATA.

Par la suite, nous effectuerons une *recherche quantitative* qui se basera sur les variables de mesures récoltés dans la recherche documentaire. Ces mesures seront soigneusement analysées, et ses analyses nous permettront d'arriver à une conclusion. Ce qui nous donnerai un aperçu numérique de notre problématique.

B- Plan

Dans ce mémoire, en plus de l'introduction et de la conclusion, les chapitres suivants seront inclus :

Le premier chapitre traitera en profondeur de la véracité du manque d'objectivité des travailleurs du BIG DATA, en se basant sur les données antérieures récoltés lors de la recherche documentaire, et des données analysées lors de la recherche quantitative.

Le *deuxième chapitre* abordera ensuite la raison pour laquelle les données récoltées et les données erronées identifiées sont quand même traitées.

Le *troisième chapitre* fera référence à l'effet GIGO et à l'impact que cet effet pourrait produire sur les données personnelles.

Enfin suivra la conclusion et une réponse sera donnée à la question principale.

IV- SOURCES

Les informations nécessaires à ce mémoire pourront être collectées dans les archives publiques numériques de [digital-mag.ci](#), de [Idh-France.org](#), [H171-dossier 3](#), de [escen.fr](#) ...