

Rapport

INVESTIGATION DES DONNEES PERSONNELLES SUR INSTAGRAM

Le 26 Mars 2025

Projet-2A

Ayoub GOUBRAIM
Saad TALEMSI
Sylvain GENETIAUX
Anas MSANDA

Tuteur : Tanguy GERNOT
Emmanuel GIGUET



TABLE DES MATIERES

1. Présentation générale du projet	3
2. Les Objectifs	3
2.1. Objectifs à mi-parcours	4
2.2. Objectifs finaux	4
3. Méthodologie de développement	4
3.1. Approche Agile	4
3.2. Étape 1 : Définition du besoin	4
3.2.1. Besoins identifiés	5
3.2.2. User Stories principales	5
3.3. Étape 2 : Planification agile et sprints	5
3.3.1. Veille technologique (travail préalable	5
3.4. Développement par MVP (Minimum Viable Product)	5
4. Travail réalisé et fonctionnalités	6
4.1. Téléversement des données et guide utilisateur	6
4.2. Catégories d'analyse	6
4.2.1. Connections	7
4.2.2. Activité	8
4.2.3. Préférences	9
4.2.4. Temps & Focus	10
5. Répartition des tâches	11
6. Bilan des objectifs non finalisés	11
6.1. Objectifs non atteints	12
6.2. Motifs	12
7. Perspectives d'évolution du projet	13
8. Conclusion	13

1. Présentation générale du projet et du contexte

À l'ère du numérique, la quantité de données personnelles générées et stockées sur les réseaux sociaux n'a jamais été aussi importante. Instagram, Snapchat ou encore TikTok proposent désormais à leurs utilisateurs de télécharger une copie complète de leur historique d'activité, comprenant photos, vidéos, messages, informations de profil, commentaires, et bien plus encore. Ces archives, souvent méconnues ou inexploitées, représentent pourtant une source précieuse pour mieux comprendre l'usage que chaque individu fait de ces plateformes, ainsi que les implications en matière **de vie privée et de traçabilité** numérique.

Ce projet consiste à **étudier et analyser les données personnelles récupérées depuis un compte Instagram**, à travers une approche pratique, technique et critique. Chaque participant télécharge son propre historique d'activité via les outils officiels proposés par Instagram, puis procède à l'**exploration de ces données** dans le but d'en dégager des vues de synthèse pertinentes sur son utilisation du réseau social. Pour ce faire, le projet inclut la **documentation complète** des étapes de récupération des données, l'**analyse des fichiers obtenus**, ainsi que le **développement d'un prototype interactif** permettant de visualiser et d'interroger ces informations. En complément, une **veille technologique** sera réalisée afin d'identifier les solutions existantes dans le domaine, qu'elles soient open source ou commerciales. Le travail est conduit sur un environnement de développement Linux, avec une préférence pour le langage **Python**, reconnu pour sa richesse en bibliothèques d'analyse de données et de traitement de fichiers. Le projet se veut à la fois formateur et révélateur des enjeux actuels liés aux **données personnelles** et à leur **exploitation responsable**.

2. Les Objectifs

L'objectif principal est de développer un **outil d'exploration interactif** capable d'exploiter les archives d'activité récupérées depuis le compte personnel de l'utilisateur. Cela implique plusieurs étapes : la documentation de la procédure de récupération des données, l'analyse structurée des fichiers fournis par Instagram, puis la conception d'un prototype permettant une **visualisation synthétique** des informations.

L'outil développé devra permettre de consulter les différentes catégories de données (photos, messages, informations de profil, interactions, etc.), et d'en extraire des **indicateurs d'activité pertinents** sous une forme lisible et accessible.

2.1. Objectifs à mi-parcours

Dans un premier temps, le projet vise à réaliser une **veille technologique et scientifique** sur l'investigation des données personnelles issues des réseaux sociaux, avec une attention particulière portée aux solutions existantes (open source ou commerciales). En parallèle, une **documentation détaillée** des étapes de récupération des données via Instagram est établie, accompagnée d'une démonstration vidéo.

Cette phase comprend également l'analyse de la structure des fichiers récupérés (formats, types de données, organisation) et la mise en place des **premiers scripts d'analyse en Python**.

Enfin, le **développement du front-end de l'application** est initié, avec la création des premières interfaces pour la visualisation et l'exploration des données.

2.2. Objectifs finaux

L'objectif final est de livrer une **application fonctionnelle et complète**, permettant à tout utilisateur de charger ses archives Instagram et de naviguer dans ses données personnelles de façon fluide et intuitive. L'outil doit offrir des **vues synthétiques de l'activité**, telles que la fréquence de publication, la répartition des types de contenus, l'évolution du profil ou encore les interactions avec d'autres utilisateurs.

Il devra également permettre des **recherches ciblées** dans les messages ou les publications, ainsi que la possibilité de **comparer l'activité entre plusieurs comptes**, dans le cas où l'utilisateur dispose de plusieurs identifiants.

Le projet se conclura par la **livraison de l'outil finalisé**, accompagné de sa documentation utilisateur, et d'une démonstration des fonctionnalités sur des jeux de données personnels.

3. Méthodologie de développement

3.1. Approche agile

Pour la réalisation de ce projet, nous avons adopté une **méthodologie agile**, nous permettant de nous adapter progressivement aux besoins, souvent découverts en cours de développement. Cette approche itérative s'est articulée autour de **sprints courts** visant à livrer des versions fonctionnelles intermédiaires, appelées MVP (Minimum Viable Product).

3.2. Étape 1 : Définition du besoin

N'ayant pas pu rencontrer le client avant le démarrage du projet, nous avons dû définir les **besoins utilisateurs** nous-mêmes, en nous plaçant dans la perspective d'un **utilisateur lambda d'Instagram** souhaitant récupérer et analyser ses données personnelles.

3.2.1. Besoins identifiés :

- Guide pour réclamer ses données à Instagram
- Outils d'analyse des données (interactions, activités...)
- Distribution horaire des activités
- Analyse des préférences (comportements)
- Visualisation du temps passé
- Analyse des messages et émotions
- Interface claire, simple et intuitive
- Export des résultats d'analyse

3.2.2. User Stories principales :

1. En tant qu'utilisateur d'Instagram, je veux pouvoir télécharger mes données personnelles afin de les analyser.
2. En tant qu'utilisateur, je veux visualiser des synthèses de mon activité Instagram (interactions, contenus).
3. En tant qu'utilisateur, je veux pouvoir exploiter mes données sous un format simple et conservable.

3.3. Étape 2 : Planification agile et sprints

3.3.1. Veille technologique (travail préalable)

Avant de coder, nous avons mené une **recherche exploratoire** sur les outils d'analyse open-source, les formats de données JSON générés par Instagram, et les bibliothèques de visualisation adaptées.

Résultat :

Une **base de connaissances** sur les données exportées, et les premiers choix techniques pour le backend et le frontend.

3.4. Développement par MVP (Minimum Viable Product)

Nous avons structuré le développement autour de **trois MVP successifs**, chacun correspondant à une étape fonctionnelle atteinte.

MVP1 – Interface basique + dépôt de fichiers

- Site web statique, non interactif.
- Instructions détaillées pour réclamer les données Instagram.
- Possibilité de téléverser les fichiers ZIP.

- Résultat : Interface fonctionnelle mais sans analyse \Rightarrow **frustration utilisateur**.

MVP2 – Début de l'analyse : distribution horaire

- Implémentation de la **distribution horaire des activités** (graphes interactifs).
- Analyse des préférences et temps à venir.
- Début du traitement des données utilisateurs.

MVP3 – Refonte du frontend et extension

- Abandon du frontend initial (peu flexible).
- Adoption d'un nouvel environnement (React / Next.js).
- Ajout des nouvelles visualisations interactives et filtres FOCUS.

4. Travail réalisé et fonctionnalités

L'application développée permet une **exploration complète et interactive** des données personnelles extraites d'Instagram. Elle se compose de plusieurs modules fonctionnels organisés autour de la **visualisation des activités**, la **catégorisation des centres d'intérêt**, et l'**analyse temporelle détaillée**.

4.1. Téléversement des données et guide utilisateur

L'utilisateur peut importer directement son **archive Instagram au format ZIP** grâce à une interface intuitive de glisser-déposer. Une **vidéo tutorielle intégrée** l'accompagne dans la procédure de récupération des données depuis le site officiel d'Instagram.

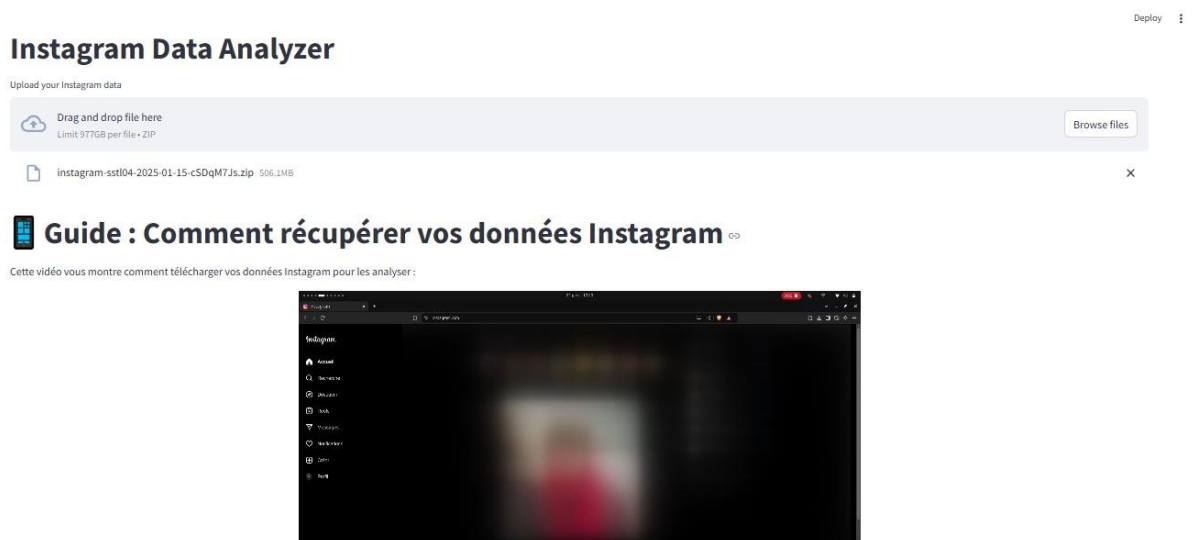


FIGURE 1 : Interface de téléversement des données Instagram et guide vidéo explicatif

4.2. Catégories d'analyse

Une fois les données importées, l'utilisateur accède à l'outil d'exploration via une interface organisée en quatre grandes catégories d'analyse :

Connections, Activité, Préférences et Temps & FOCUS.

Cette organisation permet une navigation claire et rapide entre les différentes fonctionnalités proposées, en regroupant les analyses par thématique.

- **Connections** : Analyse des abonnés et abonnements, avec suivi temporel des connexions.
- **Activité** : Étude des habitudes d'utilisation, interactions par heure et sessions journalières.
- **Préférences** : Répartition des centres d'intérêt détectés dans les données personnelles.
- **Temps & FOCUS** : Indicateurs sur le temps passé, les meilleures heures pour publier et filtres avancés.



FIGURE 2 : Menu de navigation – Accès aux quatre catégories d'analyse de l'application

4.2.1. Connections

Cette section présente une **analyse complète des connexions de l'utilisateur** sur Instagram, incluant les abonnés (followers) et les abonnements (followings), avec une visualisation de leur évolution dans le temps.

- **Vue d'ensemble des connexions** : nombre total de followers et abonnements.
- **Évolution des followers dans le temps** (courbe temporelle).
- **Historique des abonnements** (courbe following).
- **Détails des abonnés et abonnements** : tableau récapitulatif avec nom d'utilisateur, lien de profil, et date d'interaction avec possibilité de télécharger ces données.

Connections Overview

Total Followers
200

Total Following
358

Connections Timeline

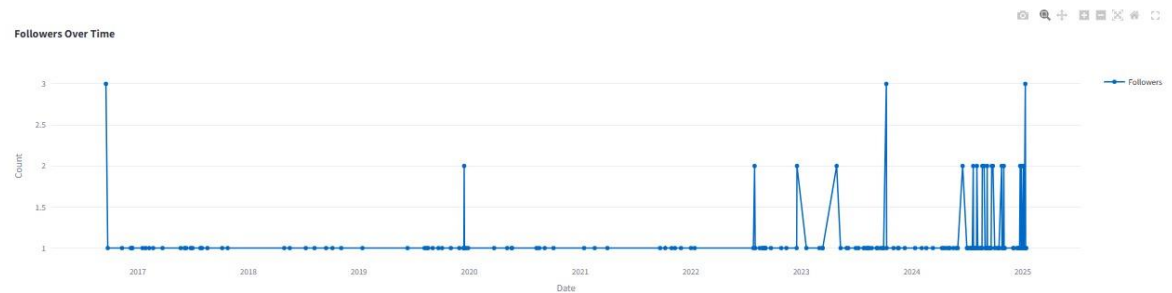


FIGURE 3 : Historique des followers : évolution du nombre d'abonnés

View Followers Details

	timestamp	username	profile_url	date
0	2025-01-13 22:36:37	[redacted]	https://www.instagram.com/[redacted]	2025-01-13
1	2025-01-11 11:06:00	[redacted]	https://www.instagram.com/[redacted]	2025-01-11
2	2025-01-10 21:19:39	liste1.bds	https://www.instagram.com/liste1.bds	2025-01-10
3	2025-01-10 21:19:31	liste_2.bds	https://www.instagram.com/liste_2.bds	2025-01-10
4	2025-01-10 02:35:11	[redacted]	https://www.instagram.com/[redacted]	2025-01-10
5	2025-01-06 00:05:12	[redacted]	https://www.instagram.com/[redacted]	2025-01-06
6	2025-01-05 03:59:01	[redacted]	https://www.instagram.com/[redacted]	2025-01-05
7	2025-01-05 03:45:36	[redacted]	https://www.instagram.com/[redacted]	2025-01-05
8	2025-01-04 23:15:14	[redacted]	https://www.instagram.com/[redacted]	2025-01-04
9	2025-01-04 03:49:12	[redacted]	https://www.instagram.com/[redacted]	2025-01-04

Download followers data

FIGURE 4 : Historique des followers : évolution du nombre d'abonné

4.2.2. Activité

La section **Activité** fournit une **analyse approfondie des comportements d'usage** de l'utilisateur sur Instagram. Elle permet de visualiser les moments de forte et faible activité, ainsi que le détail des sessions journalières.

Une **analyse horaire globale** présente la distribution de toutes les interactions (publications, likes, commentaires, reels) au fil de la journée. L'application met en évidence l'**heure de plus faible activité** par une barre rouge dans l'histogramme, permettant d'identifier clairement les périodes où l'utilisateur est le **moins actif**.

Analyse Horaire des Activités

Cette section montre à quelles heures de la journée vous êtes le plus actif sur Instagram. L'heure avec le minimum d'activité est mise en évidence en rouge.

Tous Types Posts Commentaires Reels

Distribution Horaire - Toutes Activités

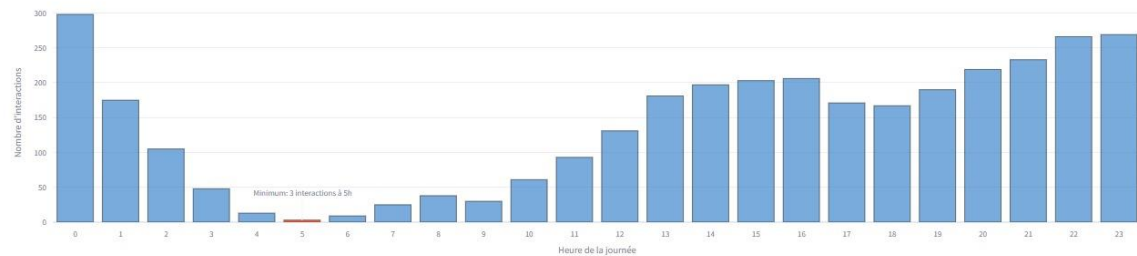


FIGURE 5 : Distribution horaire globale des interactions – Heure minimale d'activité mise en évidence

En complément, l'outil affiche les **meilleures heures pour publier** du contenu, en se basant sur les volumes d'interactions détectés. Les **heures de forte activité** sont mises en avant pour optimiser la portée des publications, tandis que les **heures de faible activité** sont signalées.

Meilleures heures pour poster

Heures de forte activité

- 0h: 298 interactions
- 23h: 269 interactions
- 22h: 266 interactions

Heures de faible activité

- 5h: 3 interactions
- 6h: 9 interactions
- 4h: 13 interactions

FIGURE 6 : Meilleures heures pour publier selon l'activité détectée

4.2.3. Préférences

La section **Préférences** met en avant les **centres d'intérêt détectés** dans les données de l'utilisateur. Elle propose plusieurs représentations graphiques des thématiques abordées et de leur fréquence.

- **Répartition par catégorie** : tableau listant les catégories principales et leur fréquence.
- **Insights & recommandations** : suggestions personnalisées basées sur les préférences détectées.
- **Diagramme circulaire** : visualisation de la part de chaque catégorie.
- **Treemap** : répartition visuelle des sujets.
- **Vue globale des topics** : liste exhaustive.

Categories Breakdown

	category	count
0	Other	128
1	Fashion & Beauty	22
2	Lifestyle	21
3	Food & Drinks	19
4	Entertainment	12
5	Sports	11

Insights & Recommendations

Insights:

- Votre catégorie d'intérêt principale est **Other**
- Instagram utilise ces préférences pour personnaliser votre flux et recommandations

Recommendations:

- Pour diversifier votre expérience, explorez de nouveaux topics
- Pour cibler votre contenu si vous êtes créateur, concentrez-vous sur les topics les plus fréquents

FIGURE 7 : Répartition des données par catégories + recommandations personnalisées

Topics Categories

Répartition des Topics par Catégorie



FIGURE 8 : Répartition des données par catégories + recommandations personnalisées

4.2.4. Temps & FOCUS

La section **Temps & FOCUS** permet une **analyse détaillée du temps passé sur Instagram**, ainsi qu'un **ciblage personnalisé de la période** à analyser. L'utilisateur peut sélectionner librement une **durée spécifique** (ex : les 15 derniers jours) afin d'observer ses comportements sur une fenêtre temporelle précise.

Dans l'exemple ci-dessous, les données analysées couvrent la période **du 1er janvier au 15 janvier 2025**.

Les **métriques clés** sont calculées automatiquement :

- **Nombre de sessions**
- **Durée moyenne par session**
- **Temps total passé**
- **Temps quotidien moyen**

Une **heatmap d'activité** par jour et par heure accompagne ces métriques et permet d'identifier visuellement les moments les plus actifs de la semaine.

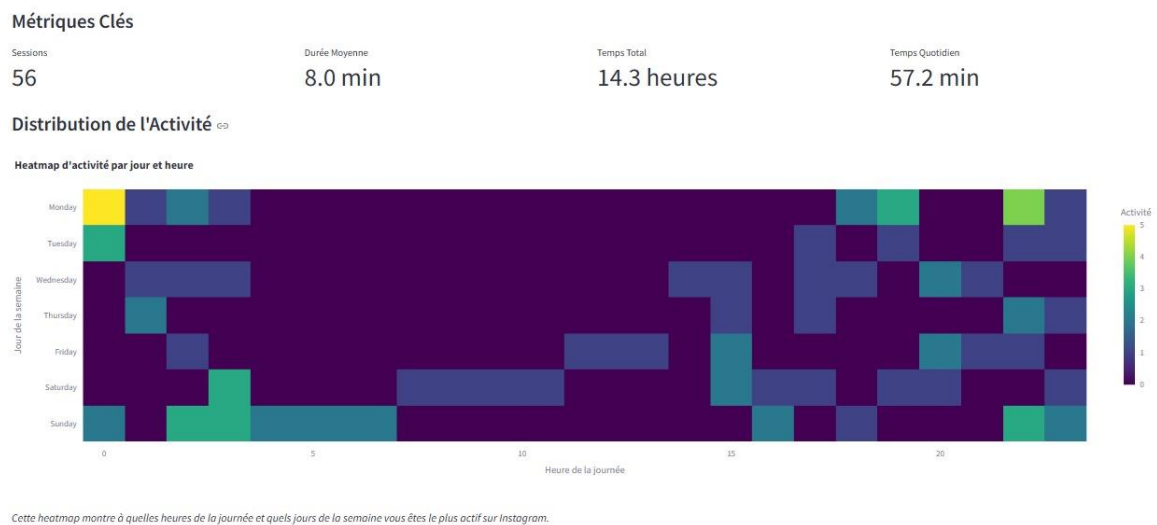


FIGURE 9 : Métriques d'activité et heatmap sur 15 jours (1–15 janvier 2025)

Un **graphique combiné** affiche le **nombre de sessions par jour** et leur **durée moyenne** (courbe), offrant une vue complète de l'intensité d'utilisation au quotidien.

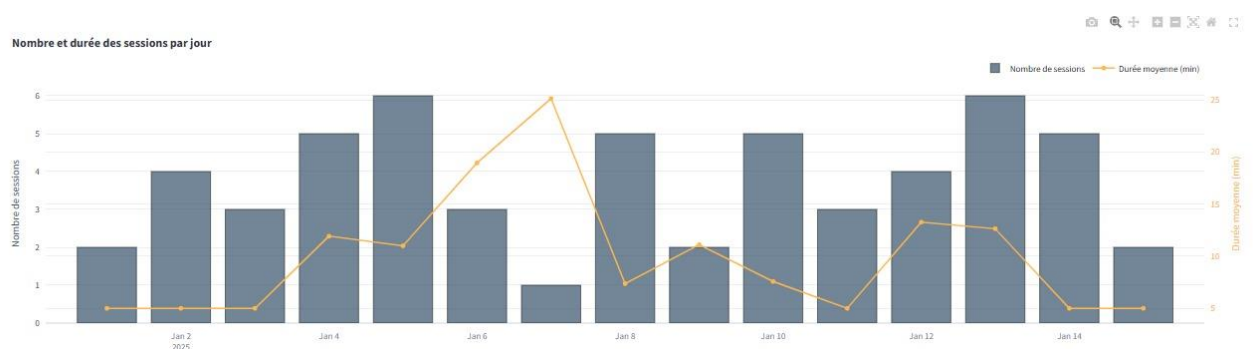


FIGURE 10 : Nombre et durée des sessions par jour

5. Répartition des tâches

Chaque membre de l'équipe s'est vu attribuer des responsabilités spécifiques afin d'assurer une progression cohérente et efficace du projet. Étant donné que la partie **back-end** représentait la charge de travail la plus importante, deux personnes s'y sont consacrées pleinement.

Le développement de l'**interface utilisateur** a été assuré par **Ayoub GOUBRAIM**, qui s'est chargé de la création du **frontend**, de l'intégration des **graphiques interactifs** ainsi que des **tests d'ergonomie** afin de garantir une utilisation fluide et accessible.

La partie **back-end** a été confiée à **Sylvain GENETIAUX** et **Saad Talemsi**. Le premier a pris en charge la gestion des fichiers ZIP fournis par Instagram, la **lecture et structuration des données JSON**, ainsi que la mise en place de la **fonctionnalité d'export des résultats** au format CSV. Le second s'est concentré sur l'implémentation des **analyses de données** (activité horaire, préférences, temps passé), ainsi que sur l'**optimisation des performances** et la gestion des **fichiers volumineux**.

Enfin, **Anas Msanda** s'est occupé de la **veille technologique**, en explorant les formats de données exportés par Instagram et les outils disponibles. Il a également rédigé le **guide de récupération des données** (en version écrite et vidéo), tout en assurant la **coordination générale** du projet, incluant la planification des sprints, la rédaction du rapport et la préparation de la présentation finale.

6. Bilan des objectifs non finalisés

6.1. Objectifs non atteints

- **Analyse détaillée des messages directs (DMs)** : nous avons envisagé d'inclure une fonctionnalité permettant d'analyser les échanges de messages directs, notamment la fréquence des conversations, les contacts les plus sollicités et la répartition temporelle des échanges.
- **Exploration des données multimédias (images, vidéos, audio)** : bien que les fichiers aient été récupérés avec succès dans les archives Instagram, leur intégration et leur visualisation dans l'outil (galerie d'images, visionneuse de vidéos) n'a pas été réalisée.
- **Personnalisation de l'interface utilisateur** : des options d'affichage supplémentaires (choix de thème, export PDF des graphiques) étaient prévues mais n'ont pas été implémentées.
- **Gestion multi-comptes** : Elle nécessitait une adaptation complexe pour traiter plusieurs archives simultanément, ce qui n'a pas pu être réalisé dans le temps imparti. L'outil est donc limité à l'analyse d'un seul compte à la fois.

6.2. Motifs

Les raisons principales expliquant ces objectifs non atteints sont :

- **Volume de données important** : le traitement de certains fichiers complexes (ex : messages ou médias) a nécessité plus de temps que prévu, notamment pour comprendre leur structure et les convertir en formats exploitables.
- **Contraintes de temps** : le temps dédié au projet ne permettait pas de tout intégrer dans les délais, d'autant plus que nous avons priorisé les fonctionnalités clés permettant l'analyse des centres d'intérêt et de l'activité.
- **Complexité technique** : l'analyse des messages et médias nécessite des traitements spécifiques (encodage, extraction) et la gestion de nombreux fichiers, ce qui a engendré des difficultés techniques.

7. Perspectives d'évolution du projet

Plusieurs **axes d'amélioration** sont envisageables pour faire évoluer l'outil vers une solution plus complète et intelligente.

Intégration de l'IA :

L'ajout de l'**intelligence artificielle** permettrait d'enrichir l'analyse, notamment par :

- L'**analyse des messages** avec détection automatique des sujets et sentiments,
- La **reconnaissance d'images** pour classer les photos (lieux, objets),

Extension à d'autres réseaux :

L'outil pourrait être adapté pour analyser aussi les données de **Snapchat** et **TikTok**, offrant une **vue globale** de l'activité numérique.

8. Conclusion

Ce projet nous a permis de **mieux comprendre l'exploitation des données personnelles issues des réseaux sociaux**, en particulier Instagram, tout en développant un outil fonctionnel d'analyse et de visualisation. À travers une approche méthodologique agile, nous avons réussi à **concevoir une application complète**,

capable de traiter les archives utilisateurs et de restituer des **résultats synthétiques, interactifs et pertinents**.

Ce travail nous a également permis de développer des compétences solides en **traitement de données**, en **développement web** et en **gestion de projet**. Malgré certains objectifs non finalisés, les bases posées permettent d'envisager facilement des évolutions futures du projet, notamment par l'intégration de nouvelles fonctionnalités ou l'adaptation à d'autres réseaux sociaux.

En conclusion, ce projet a été **enrichissant tant sur le plan technique que personnel**, et il souligne l'importance croissante pour chaque utilisateur de **comprendre et maîtriser ses données numériques**.



Ecole Publique d'Ingénieurs en 3 ans

6 boulevard Maréchal Juin, CS 45053
14050 CAEN cedex 04

