

# RAPPORT DU MINI-PROJET DE BASE DE DONNÉES

## Partie 2 - DERNIER RENDU

I) Thème choisi pour notre mini projet.....	2
II) Le Prompt.....	2
III) La réponse de l'IA.....	3
Règles de gestion (version métier).....	3
Dictionnaire de données (brut, sans préjuger du modèle).....	4
IV) NOTRE MCD.....	6

## I) Thème choisi pour notre mini projet

-Réseau social pour les étudiants membres d'un établissement, admettant que nous sommes une organisation de développement de réseaux sociaux utiles et éducatifs

Afin de :

- les rassembler par groupes
- leur permettre de se connaître plus rapidement en début d'année et de s'entraider
- leur permettre de partager leurs informations, ou des haut faits, ou même de faire s'organiser pour des travaux de groupe

## II) Le Prompt complété et utilisé pour récupérer la réponse de l'IA.

Tu travailles dans le domaine de la data et de la création de réseaux sociaux. Ton organisation éducative a comme activité de mettre en relation les étudiants d'un même établissement (IUT, École d'ingé, Université) pour favoriser la communication, l'intégration, l'entraide et le travail collaboratif au travers de publications publiques, commentaires, réactions de groupes et messagerie privée ou collectives. C'est une organisation éducative comme une grande école. La base devrait contenir, pour rester complète et minimaliste :

Les établissements (nom, code, date de création), les profils étudiants (nom, prénom, pseudo, email, mot de passe, avatar, bio, promo), les groupes (promo,type (TD, TP etc), projet), avec un lien entre établissements et groupes.

Les profils auront des liens avec les groupes,les publications (texte + média optionnel), avec commentaires et réactions. Les médias seront liés aux publications, aux conversations (privées ou de groupe) et aux messages.

Aussi, il y aura des associations créées avec la participation des profils aux conversations et les messages envoyés (relation n aire), la relation suivi/suiveur entre profils (qui sera une association récursive), et enfin les date de création et/ou date de mise à jour et/ou date d'entrée pour toutes les entités où ça serait nécessaire..

Inspire-toi des plateformes et réseaux sociaux suivants : <https://monlycee.net/> , <https://www.linkedin.com/> et <https://www.instagram.com/>

Ton organisation éducative veut appliquer MERISE pour concevoir un système d'information. Tu es chargé de la partie analyse, c'est-à-dire de collecter les besoins auprès de l'entreprise. Elle a fait appel à un étudiant en ingénierie informatique pour réaliser ce projet,

tu dois lui fournir les informations nécessaires pour qu'il applique ensuite lui-même les étapes suivantes de conception et développement de la base de données.

D'abord, établis les règles de gestion des données de ton organisation éducative, sous la forme d'une liste à puce. Elle doit correspondre aux informations que fournit quelqu'un qui connaît le fonctionnement de l'entreprise, mais pas comment se construit un système d'informations.

Ensuite, à partir de ces règles, fournis un dictionnaire de données brutes avec les colonnes suivantes, regroupées dans un tableau : signification de la donnée, type, taille en nombre de caractères ou de chiffres. Il doit y avoir entre 25 et 35 données. Il sert à fournir des informations supplémentaires sur chaque données (taille et type) mais sans a priori sur comment les données vont être modélisées ensuite.

### III) La réponse de l'IA

## Règles de gestion (version métier)

- L'organisation relie des **étudiants** d'un même **établissement** (IUT, école d'ingénieur, université) pour faciliter communication, intégration, entraide et travail collaboratif.
- Un **établissement** est identifié par un **nom** et un **code** propres à l'établissement, et possède une **date de création**.
- Un **profil étudiant** comprend au minimum **nom**, **prénom**, **pseudo**, **email**, **mot de passe**, **avatar** (facultatif), **bio** (facultatif) et **promo** (année d'entrée ou de sortie selon l'usage interne).
- **L'email est unique** à l'échelle de la plateforme. **Le pseudo doit être unique** au sein d'un même établissement (deux établissements différents peuvent avoir le même pseudo).
- Un étudiant **appartient à un et un seul établissement** à un instant donné.
- Les **groupes** existent pour structurer la vie pédagogique et de promo : ils ont un **intitulé**, un **type** (ex. TD, TP, Projet, Club/Asso), et une **promo** cible. Chaque groupe est **rattaché à un établissement**.
- Les étudiants peuvent **adhérer à plusieurs groupes**; l'**adhésion** enregistre la **date d'entrée** et un **rôle** (membre, responsable, tuteur...).
- Un **profil** peut **suivre** d'autres profils (relation **suivi/suiveur**, orientée et non réciproque).

Le suivi a un **statut** (actif, retiré, bloqué) et une **date de début**.

- Les **publications** sont créées par des profils, **portent un texte** (et **zéro ou un média** associé) et ont une **visibilité** (publique à l'établissement, limitée à un groupe, ou publique inter-établissement si l'école l'autorise).  
On conserve la **date de création**.
- Les **médias** (image, vidéo, document) sont **optionnels** et rattachés à une **publication**; on conserve le **type** de média.
- Les **commentaires** sont rattachés à une publication et à leur **auteur**; on conserve la **date de création**.
- Les **réactions** (ex. "J'aime", "Bravo", "Utile") sont rattachées à une publication et à leur **auteur**; **un profil ne peut réagir qu'une fois par type** de réaction à une même publication; on conserve la **date**.
- La messagerie permet des **conversations privées** (entre 2+ profils) et des **conversations de groupe** (rattachées à un groupe).  
Une conversation peut avoir un **titre** (facultatif) et un **type** (privée/groupe).
- La **participation à une conversation** enregistre qui participe et **depuis quand** (date d'entrée).
- Un **message** appartient à une **conversation**, à un **auteur**, un **contenu texte** et une **date d'envoi**. Les messages ne sont visibles que par les **participants**.
- Les **responsables de groupe** et/ou les **administrateurs d'établissement** peuvent **modérer** les publications, commentaires et conversations de leur périmètre.
- Pour toutes les entités pertinentes (profil, groupe, publication, commentaire, réaction, conversation, message, adhésion, suivi), on conserve des **dates de création** et/ou **de mise à jour** (ou **date d'entrée** pour les associations).
- Les contenus doivent respecter la charte de l'établissement; des mécanismes de **signalement** et **blocage** peuvent être activés par la gouvernance (hors du périmètre minimal si non demandé).

## Dictionnaire de données (brut, sans préjuger du modèle)

Donnée (signification)	Type	Taille
Établissement – nom	Texte	150

Établissement – code (code interne/official)	Texte	15
Établissement – date de création	Date	—
Profil – nom	Texte	100
Profil – prénom	Texte	100
Profil – pseudo (unique par établissement)	Texte	30
Profil – email (unique plateforme)	Texte (email)	254
Profil – mot de passe (empreinte/hash)	Texte	255
Profil – avatar (URL)	Texte (URL)	2048
Profil – bio (courte présentation)	Texte	500
Profil – promo (année)	Numérique (entier)	4
Profil – date de création	Date/Heure	—
Profil – date de mise à jour	Date/Heure	—
Groupe – intitulé	Texte	120
Groupe – type (TD/TP/Projet/Club...)	Texte (énumération)	20
Groupe – promo (année cible)	Numérique (entier)	4
Lien groupe–établissement – date de création	Date	—
Adhésion groupe – rôle (membre, responsable...)	Texte (énumération)	20
Adhésion groupe – date d’entrée	Date/Heure	—
Publication – texte	Texte	2000
Publication – visibilité (établissement/groupe/publique)	Texte (énumération)	20
Publication – date de création	Date/Heure	—
Média – URL (fichier associé à une publication)	Texte (URL)	2048
Média – type (image/vidéo/document)	Texte (énumération)	20
Commentaire – texte	Texte	1000
Commentaire – date de création	Date/Heure	—
Réaction – type (Like/Bravo/Utile...)	Texte (énumération)	20
Réaction – date de création	Date/Heure	—

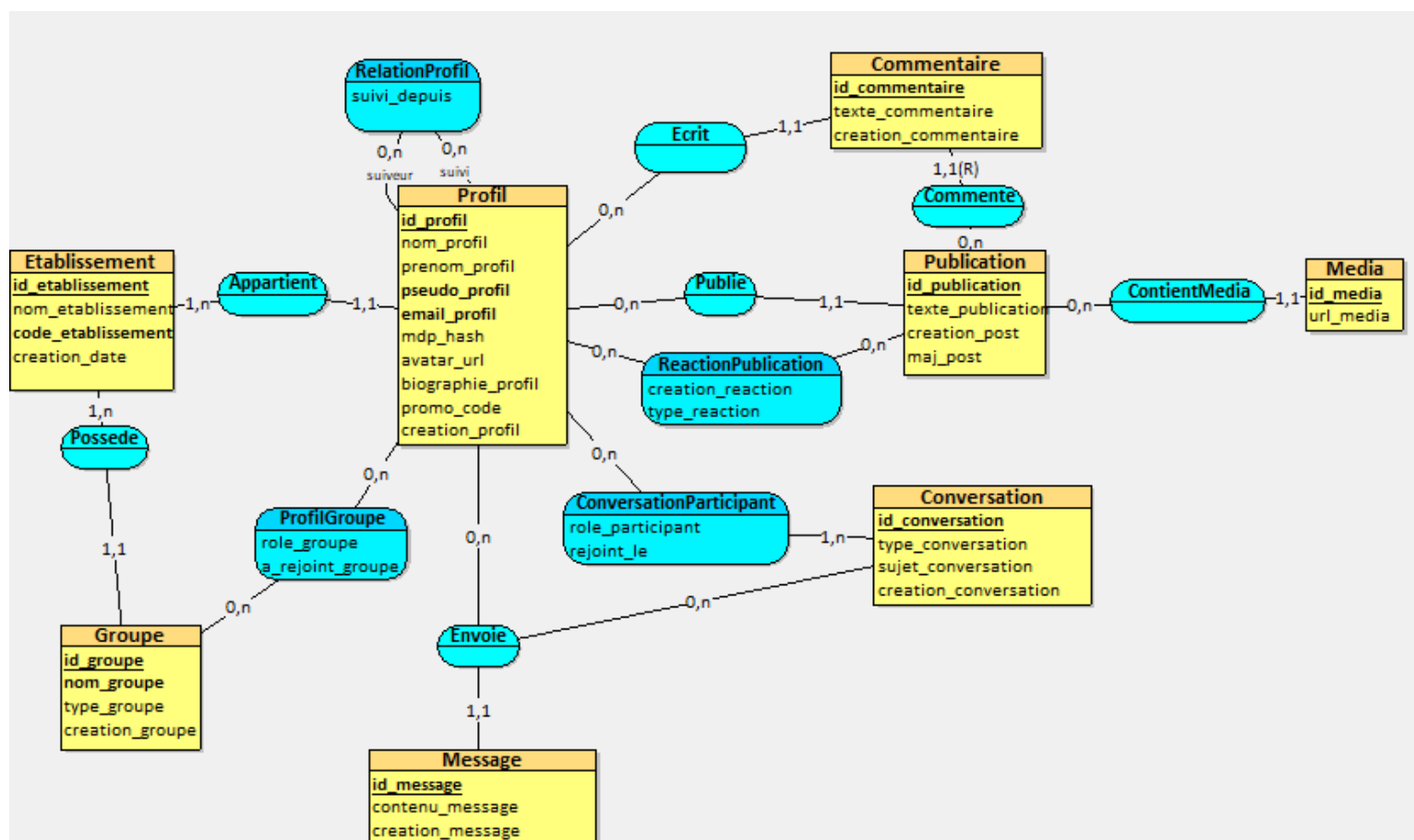
Conversation – titre (facultatif)	Texte	150
Conversation – type (privée/groupe)	Texte (énumération)	20
Participation conversation – date d’entrée	Date/Heure	—
Message – contenu	Texte	2000
Message – date d’envoi	Date/Heure	—
Suivi – statut (actif, retiré, bloqué)	Texte (énumération)	15
Suivi – date de début	Date/Heure	—
Établissement – site web (facultatif)	Texte (URL)	2048

35 données au total, conformes à la fourchette demandée.

Les tailles proposées sont **indicatives métier** (suffisantes pour cadrer le besoin) et laissent à l’étudiant le soin d’ajuster au moment de la conception logique/physique (MERISE : MLD/MPD), tout comme le choix précis des types SQL et des clés/contraintes.

## IV) NOTRE MCD

Fait sur Looping en se basant sur la réponse de l’IA après notre réflexion, puis quelques personnalisations pour que ça soit plus cohérent (en tout cas pour nous, par exemple en ajoutant l’entité faible avec les commentaires)



## V) Informations sur l'étape d'insertion des données

Pour cette étape, nous avons dû remanier notre prompt à plusieurs reprises. Malgré cela, nous constatons encore des erreurs ou des hallucinations dans les réponses de l'IA, probablement dues au grand volume de données demandé.

Même après plusieurs essais, les hallucinations ou non-respect des consignes données dans le script étaient toujours présentes, donc nous avons modifié certaines de ces mauvaises valeurs à la main comme les formats des mails où malgré notre regexp donné dans le script il se trompait en mettant des accents, ainsi que les catégories des conversations (soit "privé" soit "groupe").

D'autres valeurs, comme les messages ou les commentaires, ont été laissées telles quelles, car leur correction manuelle aurait été trop chronophage.

Cette étape nous a permis de constater que l'IA commet facilement des erreurs et qu'il est indispensable de vérifier et valider systématiquement les données générées avant toute utilisation.

## VI) Scénario d'utilisation

### Scénario 1: celui que nous avons retenu pour orienter nos requêtes

Une analyste de contenu et de l'utilisation du réseau veut réaliser un rapport d'utilisation de ce dernier, des contenus populaires et de la bonne utilisation (afin d'estimer s'il sera nécessaire de modifier des contraintes ou d'ajouter de la modération).

Elle regarde :

- L'activité des profils (nombre de publications, commentaires, réactions, messages)
- Les contenus les plus populaires
- La dynamique des groupes (taille des groupes, rôles, ancienneté)
- Les relations sociales
- Les étudiants inactifs (profils sans publications ni groupes)
- Les conversations actives ainsi que la vérification de la qualité des contenus (contenus non nuls, date cohérentes)

### Scénario 2:

Un enseignant-chercheur à l'Université Paris Cité étudie comment les étudiants interagissent dans le cadre de projets collaboratifs via le réseau social.

Il regarde :

- Les publications à contenu académique (sujets, fréquence, engagement)
- Le réseau d'échanges (relations entre profils auteurs de publications similaires)
- L'activité par établissement (comparaison entre universités)
- Les groupes collaboratifs (groupes contenant des profils de plusieurs établissements)

Le but final est d'étudier la collaboration numérique et la propagation de l'innovation entre universités.

### Scénario 3:

Un étudiant-ambassadeur à La Sorbonne est chargé de mettre en avant les clubs étudiants les plus actifs pour la page d'accueil du réseau social.

Il regarde :

- Le classement des clubs (nombre de membres, ancienneté, publications liées)
- L'engagement des membres (taux moyen d'activité des membres par groupe)
- Le profil des administrateurs (informations sur les admins/modérateurs)

Le but final est de produire un top 5 des groupes les plus actifs à mettre en avant sur la plateforme.

## VII) Conclusion

Nous avons dû nous limiter sur le nombre de données à générer, car au-delà d'un certain seuil, les incohérences produites par l'IA devenaient trop importantes.

Elle n'a pas un comportement humain représentatif dans les données générées non plus, et elle s'est également souvent trompée et nous renvoie plus ou moins à chaque fois des données incohérentes, probablement dû au trop grand nombre de données à générer.

Ceci dit, ça nous a aussi fait réfléchir sur la véracité des informations ou des réponses apportées par l'IA à nos questions, ainsi que la confiance qu'on lui donnait. Il faut toujours revérifier les infos et relire les réponses. Il ne faut pas se baser uniquement sur une réponse de l'IA pour répondre à un besoin / une question, mais toujours croiser les résultats avec une vérification humaine.

Malgré ces difficultés, nous avons su les surmonter et répartir le travail efficacement. Nos requêtes répondent bien aux besoins du scénario choisi, et le volume de données obtenu grâce à notre *prompt* d'insertion nous a permis d'obtenir des résultats cohérents et exploitables.