Rapport de TER
Simulation de Réseau Social avec des LLM via Y Social

Lounas Katia, Demeude Edgar, Laurent Maël

23 février 2025

Résumé

Dans ce projet, nous avons exploré la plateforme Y Social [2] en simulant des interactions entre agents à l'aide de différents modèles de langage (LLM). L'objectif principal était de tester l'outil et d'analyser la cohérence des comportements des agents en fonction de leurs publications et interactions. Nous avons collecté et stocké les données sous forme de base relationnelle SQL, appliqué des techniques de fouille de données, et visualisé les interactions à travers divers graphiques. Un Chi-Square Goodness-of-Fit Test a été réalisé pour évaluer si les agents maintenaient une cohérence entre leurs intérêts initiaux et leurs interactions finales. Les résultats ont montré une relative stabilité des préférences des agents, suggérant une certaine cohérence dans leur comportement. Cette étude contribue à une meilleure compréhension du fonctionnement de Y Social et à son évaluation en tant qu'outil de simulation et d'expérimentation.

Mots-clés : Y Social, LLM, fouille de données, interactions simulées, analyse statistique

1 Introduction

La simulation de réseaux sociaux à l'aide de modèles de langage (LLM) est un domaine émergent qui cherche à comprendre les comportements des utilisateurs virtuels dans un environnement contrôlé. Ce projet s'inscrit dans le cadre de la validation et de l'expérimentation de la plateforme Y Social, une plateforme expérimentale qui utilise les LLM pour simuler des interactions au sein de réseaux sociaux. Ces modèles génèrent des échanges entre des agents simulés (utilisateurs fictifs), créant ainsi un environnement contrôlé pour reproduire des événements sociaux réalistes.

Ce travail s'appuie sur des recherches précédentes, notamment celles de Tucker et al. (2018) [3], qui examinent la polarisation sociale et la propagation de la désinformation dans les médias sociaux modernes.

L'objectif principal de ce projet est d'évaluer si les agents agissent de manière cohérente avec leurs intérêts initiaux, en analysant leurs publications et interactions. Pour ce faire, nous avons développé une méthodologie reposant sur l'extraction et l'analyse de données, la visualisation des interactions et l'application de tests statistiques. Notre approche, qui fait appel à des techniques de fouille de données, vise à déterminer si les agents maintiennent une cohérence dans leurs choix et comportements au fil du temps.

2 Méthodologie

2.1 Prise en main de Y Social

La première étape de notre projet a consisté à explorer et à nous familiariser avec la plateforme Y Social. C'est un environnement de simulation permettant la création d'interactions entre agents virtuels, utilisant des modèles de langage pour simuler des comportements sociaux. Y-Social fonctionnant en architecture client-serveur, il est intéressant de pouvoir disposer d'un serveur dédié à la

génération LLM afin d'exécuter les simulations de manière indépendante, bien qu'une utilisation en local hosting soit possible. C'est pourquoi nous avons d'abord tenté d'utiliser un serveur personnel pour réaliser nos simulations.

2.1.1 Configuration du serveur personnel

Nous avons essayé d'utiliser un ancien ordinateur avec un GPU Radeon Pro WX 4100 pour exécuter Y-Social et un modèle LLM. Bien que l'installation ait été réussie, l'usage du GPU a échoué en raison de l'incompatibilité avec ROCm, ce qui a rendu la simulation lente et inefficace. Cela a montré l'importance de vérifier la compatibilité matérielle avant d'entreprendre ce type de configuration.

2.1.2 Analyse des APIs LLM disponibles

Face aux limitations du serveur, nous avons exploré plusieurs APIs LLM : OpenAI, Hugging Face et Anthropic. OpenAI offrait une bonne performance mais était coûteuse et trop lente pour une simulation à grande échelle. Hugging Face, plus économique et flexible, semblait être une meilleure option, bien que non testée en profondeur. Anthropic, axé sur la sécurité, n'était pas adapté pour des générateurs de texte à grande échelle.

2.1.3 Comparaison et conclusion

OpenAI s'est avéré trop cher et lent pour des simulations à grande échelle, tandis que Hugging Face se révèle plus abordable et flexible, bien qu'un test plus poussé soit nécessaire. Anthropic, bien que performant en sécurité, n'était pas pertinent pour nos besoins en génération de texte.(Pour plus de details consulter l'annexe 5)

2.2 Scénarios et simulation

Une fois familiarisés avec la plateforme, nous avons créé plusieurs scénarios pour tester la cohérence des comportements des agents en fonction de leurs intérêts. Les agents ont été initialisés avec un ensemble d'intérêts et de préférences spécifiques. Nous avons conçu des scénarios d'interaction dans lesquels les agents interagissent entre eux, publient du contenu, et réagissent aux publications des autres agents. Les scénarios ont été choisis de manière à refléter une diversité de comportements, allant de simples échanges d'opinions à des discussions plus complexes sur des sujets spécifiques.

Les interactions ont été générées par des modèles de langage tels que llama3.2, dolphin-llama3 et ChatGPT, selon les besoins des simulations. Chaque scénario a été répété plusieurs fois pour garantir la robustesse des résultats et pour analyser la stabilité des interactions au fil du temps. Ces répétitions ont permis de vérifier si les agents maintenaient leurs préférences initiales ou modifiaient leur comportement en fonction des interactions.

2.3 Collecte et analyse des données

Les données générées par les simulations sont stockées dans un fichier de base de données .db, qui est utilisé pour conserver les informations relatives aux agents et à leurs interactions. Chaque enregistrement dans la base de données contient des détails sur les actions de l'agent, ses préférences initiales, ses publications, ses interactions avec d'autres agents, et tout changement observé dans ses comportements au cours des simulations.

Le fichier .db est structuré de manière relationnelle (FIG 1), avec des tables représentant les agents, les publications, les interactions, et les préférences des agents. Par exemple, une table contient les identifiants des agents, leur caractère, niveau d'éducation... Une autre contient les détails d'une publication (contenu, auteur, ordre de publication). Ce format permet de les filtrer selon des critères spécifiques, et de manipuler facilement les données avec par exemple des requêtes SQL pour les analyser sous différents angles.

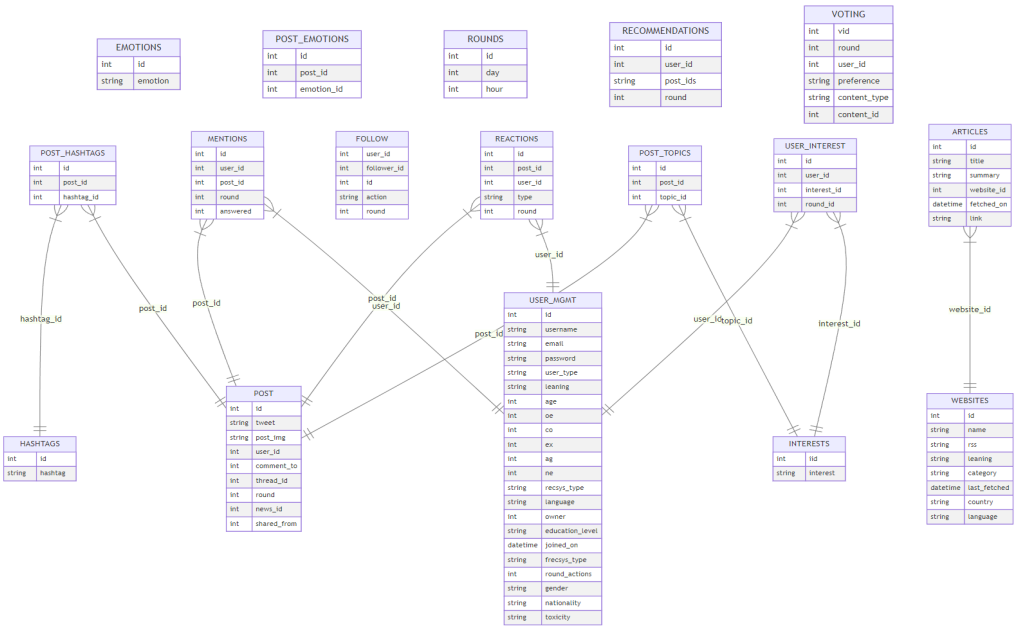

FIGURE 1 – Diagramme de classes

2.4 Interprétation et évaluation des résultats

Dans cette section, nous avons utilisé des techniques de Data Mining pour extraire et analyser les données générées par les simulations. L'objectif était de comprendre les interactions entre les agents, les tendances dans les publications, et la cohérence des agents avec leurs centres d'intérêt initiaux.

Nous avons utilisé des outils tels que Pandas pour la manipulation des données, NetworkX pour l'analyse des réseaux, et Matplotlib/Seaborn pour la visualisation des résultats. Ces outils nous ont permis de créer des graphiques représentant les interactions entre les agents, ainsi que des analyses statistiques pour comparer les différents scénarios simulés.

Pour approfondir notre étude, nous avons cherché à mieux suivre l'évolution des intérêts des utilisateurs. Ainsi, pour chaque utilisateur, nous avons répertorié ses centres d'intérêt à chaque round de la simulation.

3 Résultats et analyse

3.1 Présentation des résultats des simulations

Ce tableau (Tab 1) présente les résultats des simulations réalisées sur Y Social, mettant en évidence l'impact du modèle de langage, de la durée et du sujet traité sur les interactions générées.

LLM	Topic	Nbr de jours	Nbr d'agents	Nbr de posts	Nbr de réactions	Nbr de commentaires
dolphin-llama3	Music Jazz Rock	15	20	189	23	78
llama3.2	Music Jazz Rock	15	20	204	77	83
llama3.2	Music Jazz Rock	25	20	322	20	140
llama3.2	Politique	25	20	322	30	24

TABLE 1 – Synthèse des résultats des simulations sur Y Social

Les simulations réalisées avec Y Social montrent que le modèle de langage, la durée de simulation et le sujet traité influencent les interactions générées. Llama3.2 produit plus de publications et de commentaires que dolphin-llama3. Une durée prolongée (25 jours) augmente significativement le nombre de posts et commentaires, bien que les réactions varient. Le sujet traité impacte l'engagement des agents : la politique suscite moins d'interactions que la musique.

3.2 Analyse des comportements des agents

Dans cette section et ce qui suit, nous analysons uniquement les résultats des simulations portant sur le topic de la musique.

3.2.1 Analyse du contenu généré par les agents

Nous nous intéressons ici au contenu produit par les agents. Les figures 2, 3 présente deux word clouds issus d'une même simulation sur la musique, réalisée sur 15 jours avec deux modèles de langage différents. Ces représentations visuelles permettent d'identifier les thèmes dominants abordés par les agents.

Comme on peut le constater, le modèle dolphin-llama3 génère un contenu moins cohérent que llama3.2. En effet, son word cloud contient plusieurs mots hors contexte et leur grande taille indique une fréquence d'apparition élevée, ce qui suggère un manque de pertinence dans les interactions générées. À l'inverse, le modèle llama3.2 produit un contenu plus structuré et en adéquation avec le sujet de la musique, confirmant ainsi sa meilleure capacité à simuler des échanges cohérents.

FIGURE 2 – Dolphin-llama3 word cloud

FIGURE 3 – llama3.2 word cloud

Les figures suivantes (Fig. 4, 5) illustrent la distribution des différentes longueurs des posts. Il est évident que le modèle dolphin-llama3 présente une proportion importante de posts vides, ce qui compromet la pertinence des résultats.

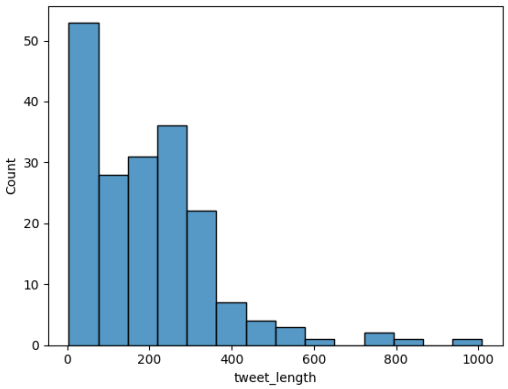

FIGURE 4 – Dolphin-llama3 longueur des posts

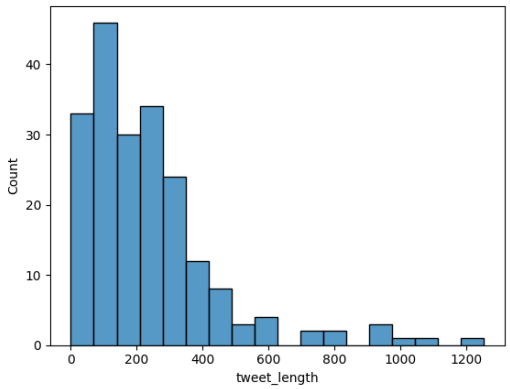

FIGURE 5 – llama3.2 longueur des posts

3.2.2 Analyse des interactions des agents

Pour visualiser le comportement des agents, nous avons généré deux graphiques. Le premier représente les interactions entre agents en fonction de leurs abonnements mutuels. Le second illustre également les interactions, mais cette fois en se basant sur les mentions, c'est-à-dire lorsqu'un agent tague un autre dans son post. De manière générale, les graphiques sont cohérents, que ce soit avec dolphin-llama3 (cf, 6 8) ou llama3.2 (cf, 7 9). On observe une forte interaction entre les agents. À noter que, dans les graphiques des mentions, les chiffres indiquent que l'agent s'est identifié lui-même.

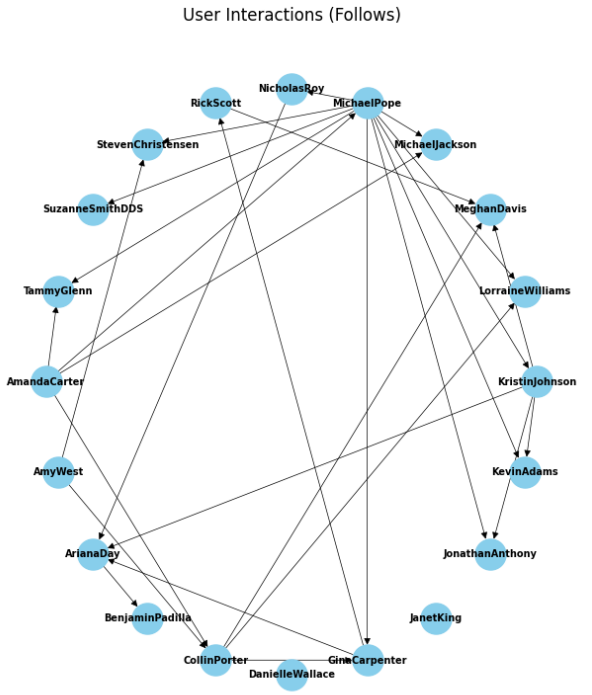

FIGURE 6 – Dolphin-llama3, Graphique d'interaction des agents en fonction de leurs abonnements

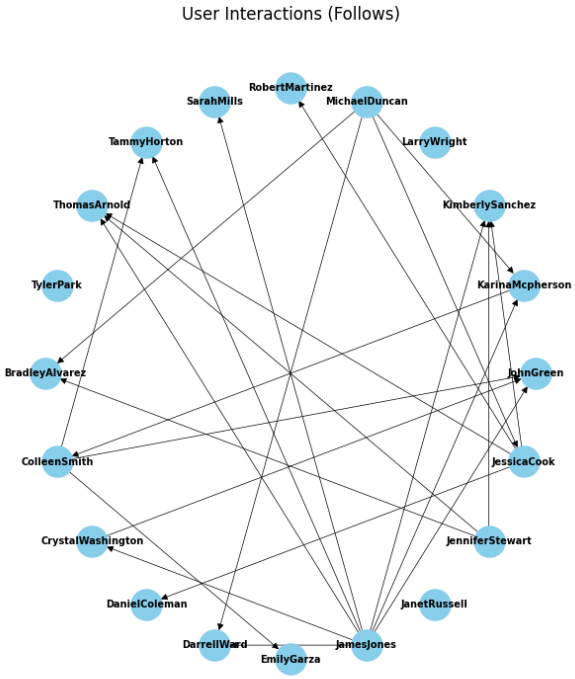

FIGURE 7 – Llama3.2, Graphique d'interaction des agents en fonction de leurs abonnements

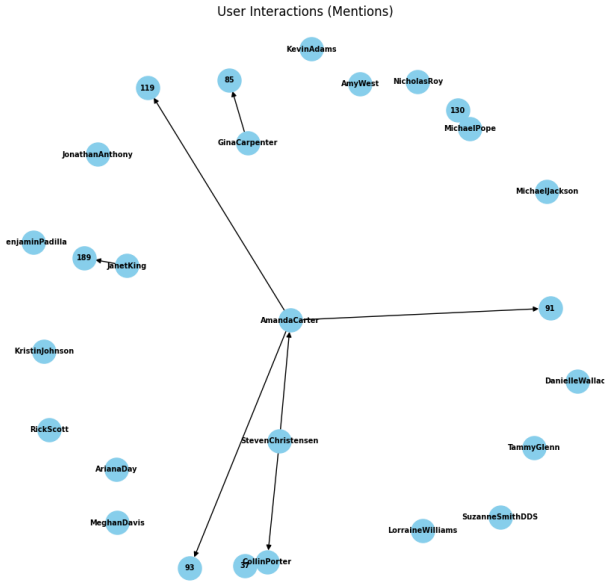

FIGURE 8 – Dolphin-llama3, Graphique d'interaction des agents en fonction de leurs tags

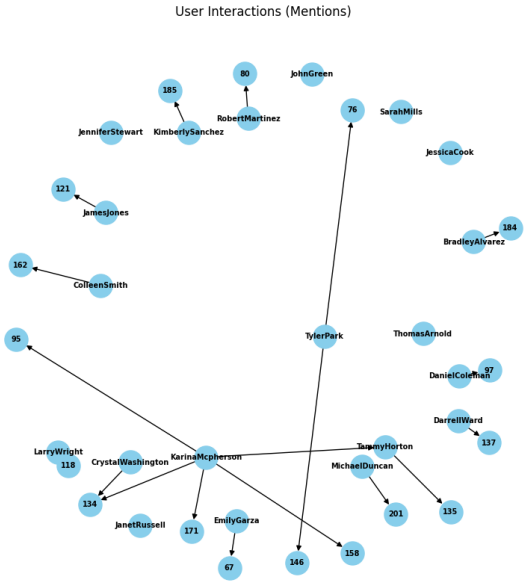

FIGURE 9 – Llama3.2, Graphique d'interaction des agents en fonction de leurs tags

3.3 Analyse de l'évolution des agent

Dans cette section, nous analysons l'évolution des intérêts et l'engagement des agents tout au long de la simulation. Pour chaque agent, nous suivons l'évolution de ses centres d'intérêt à chaque interaction ou engagement. Ces données sont collectées et synthétisées dans le tableau suivant (cf,2), mettant en évidence les tendances et les dynamiques observées au fil du temps.

Agent	Music (init.)	Rock (init.)	Jazz (init.)	Music (% post)	Rock (% post)	Jazz (% post)
BradleyAlvarez	0	1	0	60.00	40.00	0.00
ColleenSmith	1	0	0	33.33	22.22	44.44
CrystalWashington	1	1	0	0.00	75.00	25.00
DanielColeman	0	1	0	50.00	50.00	0.00
DarrellWard	1	1	0	60.00	40.00	0.00
EmilyGarza	1	0	0	33.33	33.33	33.33
JamesJones	0	1	0	33.33	41.67	25.00
JanetRussell	0	1	1	0.00	0.00	100.00
JenniferStewart	0	0	1	0.00	66.67	33.33
JessicaCook	0	1	0	0.00	0.00	100.00
JohnGreen	0	0	1	0.00	50.00	50.00
KarinaMcpherson	1	0	0	45.45	45.45	9.09
KimberlySanchez	0	1	1	0.00	75.00	25.00
LarryWright	0	0	1	0.00	0.00	0.00
MichaelDuncan	1	0	0	100.00	0.00	0.00
RobertMartinez	0	1	0	33.33	66.67	0.00
SarahMills	1	0	0	0.00	100.00	0.00
TammyHorton	1	0	1	0.00	0.00	0.00
ThomasArnold	0	1	1	50.00	50.00	0.00
TylerPark	0	1	1	50.00	50.00	0.00

TABLE 2 – Distribution initiale des intérêts des agents et pourcentage de leurs posts sur chaque sujet

L'analyse montre que l'engagement des agents ne suit pas toujours strictement leurs intérêts initiaux. Certains, comme MichaelDuncan, restent cohérents avec leurs préférences de départ, tandis que d'autres, à l'image de ColleenSmith, diversifient leurs publications. Des cas extrêmes, comme LarryWright et TammyHorton, n'ont produit aucun contenu, suggérant un manque d'engagement. D'autres agents, tels que JanetRussell et JessicaCook, se spécialisent exclusivement sur un sujet. Ces observations indiquent des dynamiques d'évolution des préférences, d'influence et de changements contextuels au cours de la simulation.

La figure qui suit (cf, 10) illustre mieux les percentages initiale des agents dans chaque topic.

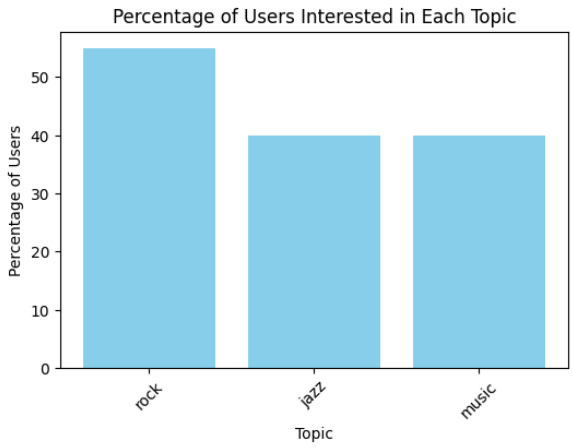

FIGURE 10 – Pourcentage d'utilisateurs initial intéressés par chaque sujet

3.3.1 Chi-square de conformité

Pour explorer l'évolution des centres d'intérêt des utilisateurs avant et après, nous avons appliqué le test du chi-deux de conformité afin d'évaluer la stabilité et la cohérence de ces intérêts au fil du temps, en suivant les étapes suivantes :

- **Calcul des pourcentages initiaux :** Détermination du pourcentage d'agents initialement intéressés par chaque sujet en comptabilisant le nombre d'agents concernés.
- **Suivi de l'évolution des centres d'intérêt :** Pour chaque utilisateur, nous avons analysé l'évolution de ses intérêts tout au long de la simulation. À la fin, nous avons évalué son engagement envers un sujet en mesurant la proportion de ses publications traitant de ce dernier, afin de déterminer s'il continue à parler de ses centres d'intérêt initiaux.
- **Analyse statistique :** Après avoir calculé et comparé les distributions des intérêts au début et à la fin de la simulation, nous avons appliqué le **test du Chi-deux de conformité** [1] pour évaluer la stabilité et la cohérence des comportements des agents.

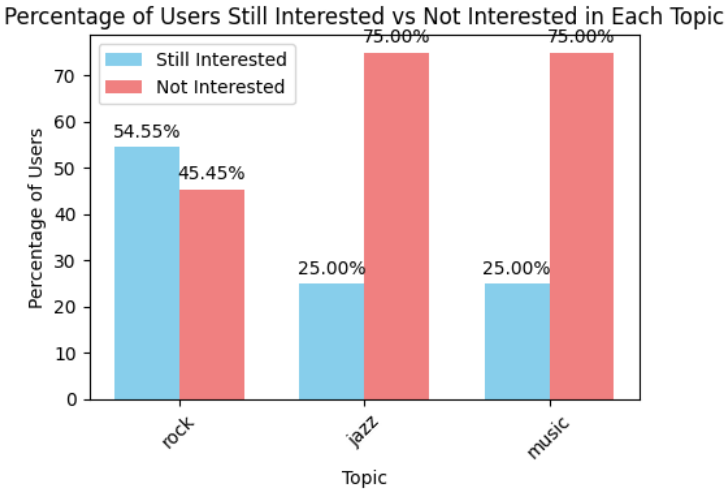

FIGURE 11 – Pourcentage d'utilisateurs toujours intéressés vs non intéressés par chaque sujet

Cette figure 11 illustre la rétention des utilisateurs au sein de chaque catégorie d'intérêt. Nous avons comparé le pourcentage d'utilisateurs ayant conservé leur intérêt à celui des utilisateurs l'ayant abandonné.

Topic	Observed Frequency	Expected Frequency	Chi-Squared Statistic	P-value
Rock	6	11	5.050505	0.024619
Jazz	2	8	7.500000	0.006170
Music	2	8	7.500000	0.006170

TABLE 3 – Distribution initiale des intérêts des agents et pourcentage de leurs posts sur chaque sujet

Les résultats du test du Chi-deux montrent une variation significative de l'intérêt des utilisateurs pour les sujets musicaux. L'intérêt pour le jazz et la musique a fortement diminué par rapport aux attentes ($p = 0.0062$) (cf, 3), tandis que le rock connaît aussi une baisse notable ($p = 0.0246$). Ces écarts suggèrent que les préférences des utilisateurs ne restent pas stables et évoluent au fil du temps, probablement sous l'effet des interactions et du contexte de la simulation.

3.4 Conclusion

Dans ce chapitre, nous avons analysé les résultats des simulations en détail. Les performances des simulations ont montré que les paramètres de configuration jouent un rôle crucial dans l'évolution des résultats, ce qui met en évidence l'importance de bien ajuster les valeurs de ces paramètres pour obtenir des résultats significatifs.

Cependant, malgré une analyse détaillée, les résultats des tests statistiques, en particulier le chi carré, n'ont pas été aussi concluants que prévu. Cette absence d'une valeur significative peut sembler décevante, mais elle est compréhensible lorsqu'on considère la nature changeante des comportements humains. Les individus peuvent ajuster leurs préférences au fil du temps, ce qui perturbe la stabilité des modèles. Par exemple, dans une simulation de réseau social, une personne peut suivre une tendance ou être influencée par un groupe, puis changer d'avis ou d'intérêt, ce qui rend difficile la prévision à long terme des comportements.

Cet aspect d'imprévisibilité est aussi pertinent dans le cadre des Large Language Models (LLM). Bien que ces modèles évoluent en fonction des interactions, ils ne parviennent pas toujours à simuler parfaitement les comportements humains. En effet, ces modèles, souvent conçus pour être polis et respectueux, ont du mal à reproduire la complexité des comportements humains, notamment lorsqu'il s'agit d'interactions plus conflictuelles ou toxiques. Par exemple, un LLM peut être extrêmement poli et courtois dans ses réponses, mais échouer à simuler des comportements humains plus abrasifs ou agressifs, ce qui peut rendre les simulations moins réalistes. Ce décalage entre la politesse du modèle et la diversité des comportements humains dans la réalité souligne les limites actuelles des LLM, notamment dans leur capacité à reproduire des interactions moins positives ou conflictuelles.

En conclusion, bien que les résultats des simulations ne soient pas aussi robustes qu'espéré, ils mettent en évidence l'évolution constante des comportements humains. Ils illustrent également les défis techniques auxquels les simulations et les systèmes d'IA comme les LLM font face, notamment lorsqu'ils essaient de reproduire des comportements humains complexes et parfois toxiques. Ces résultats nous rappellent l'importance de bien ajuster les paramètres d'entrée et de rester vigilants face aux biais ou aux limitations des modèles utilisés.

4 Conclusion générale

Ce projet nous a permis d'aborder à la fois des aspects techniques et théoriques dans la simulation de réseaux sociaux. Sur le plan technique, nous avons mis en place et exécuté des simulations

complexes, ce qui a impliqué la gestion d'un grand nombre de paramètres et l'optimisation de l'exécution des simulations. Cette partie du projet a révélé l'importance de la bonne gestion des ressources informatiques et de la capacité à adapter le modèle en fonction des contraintes d'exécution.

En ce qui concerne les résultats théoriques, comme discuté dans le chapitre 3, nous avons constaté que malgré l'absence d'un chi carré significatif, les simulations fournissent un éclairage précieux sur l'évolution des comportements dans les réseaux sociaux. Cela démontre la flexibilité des individus et la manière dont leurs préférences peuvent changer au fil du temps. En fin de compte, même si les simulations ne produisent pas toujours des résultats parfaitement conformes aux attentes, elles offrent des insights importants sur la dynamique des interactions sociales et la façon dont les individus réagissent à des événements ou stimuli spécifiques.

En rétrospective, ce projet nous a permis non seulement de mettre en pratique des concepts avancés de simulation et de modélisation, mais aussi de mieux comprendre les défis liés à la modélisation du comportement humain dans des systèmes dynamiques. Ce fut une occasion enrichissante d'appliquer des techniques de simulation dans un contexte de recherche en sciences sociales et d'identifier les limites et opportunités de ces méthodes.

5 Perspectives

Les résultats obtenus avec Y-Social ouvrent plusieurs pistes d'amélioration et d'approfondissement pour affiner les simulations et mieux comprendre les dynamiques des réseaux sociaux.

- **Simulations sur des durées plus longues :** Actuellement, les simulations sont limitées dans le temps, ce qui restreint l'observation de phénomènes à long terme comme l'émergence de bulles de filtres, la polarisation des opinions ou la viralité des contenus. Étendre la durée des simulations permettrait d'obtenir une quantité de données plus conséquente et d'améliorer la représentativité des interactions par rapport aux réseaux sociaux réels.
- **Expérimentation avec Hugging Face :** Notre étude a mis en évidence que les modèles de langage disponibles sur la plateforme Hugging Face sont prometteurs pour ce type de simulation. Tester ces modèles permettrait de comparer leur pertinence avec ceux déjà utilisés et d'évaluer leur capacité à générer des interactions plus cohérentes et variées.
- **Amélioration de la détection des biais et de la toxicité :** Certains LLM ont tendance à éviter les réponses trop tranchées ou conflictuelles, ce qui peut nuire à la simulation réaliste des interactions humaines, y compris des discussions polarisées ou toxiques. Mieux calibrer ces modèles permettrait d'obtenir une représentation plus fidèle des échanges en ligne.

Références

[1] Narayanaswamy Balakrishnan, Vassilly Voinov, and Mikhail Stepanovich Nikulin. *Chi-squared goodness of fit tests with applications*. Academic Press, 2013.

[2] Giulio Rossetti, Massimo Stella, Rémy Cazabet, Katherine Abramski, Erica Cau, Salvatore Citraro, Andrea Failla, Riccardo Improta, Virginia Morini, and Valentina Pansanella. Y social : an llm-powered social media digital twin. *arXiv preprint arXiv :2408.00818*, 2024.

[3] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. Social media, political polarization, and political disinformation : A review of the scientific literature. *Political polarization, and political disinformation : a review of the scientific literature (March 19, 2018)*, 2018.

Annexe : Détails techniques et procédures

Configuration du serveur personnel

Un serveur personnel, basé sur un ordinateur de la faculté équipé d'un processeur Intel i5-8500 et d'une carte graphique Radeon Pro WX 4100, a été utilisé pour héberger Y-Social et exécuter des modèles LLM. L'objectif était d'exploiter le GPU pour générer du texte efficacement.

- **Installation de Y-Social et Ollama :** L'installation du serveur Y-Social et du LLM Ollama s'est déroulée sans problème majeur. Les guides d'installation étaient clairs et bien documentés, ce qui a facilité la mise en place du serveur. Après avoir configuré le pare-feu et mis en place le port forwarding, le serveur était prêt à l'emploi.
- **Problèmes avec le GPU :** Une fois le serveur en place, il est devenu évident que le GPU n'était pas utilisé, malgré des spécifications théoriques adéquates. L'OS (Debian headless) ne disposait pas des pilotes graphiques nécessaires. Nous avons dû installer plusieurs dépendances pour activer l'usage du GPU, ce qui a entraîné des heures de configuration. Le matériel étant ancien et spécifique, les ressources disponibles en ligne étaient limitées, et des rétrogradations de paquets ont été nécessaires pour assurer la compatibilité.
- **Problème de compatibilité ROCm :** Après avoir installé les pilotes nécessaires, il est apparu que la technologie ROCm d'AMD ne supportait pas le GPU utilisé, ce qui a rendu son utilisation pour l'inférence LLM impossible. Ce problème aurait pu être évité si la compatibilité avait été vérifiée en amont.

Analyse des APIs LLM disponibles

Face à la limitation du serveur personnel, plusieurs solutions d'APIs LLM ont été explorées : OpenAI, Hugging Face et Anthropic.

- **OpenAI :** OpenAI propose des modèles performants tels que GPT-3 et GPT-4, et nous avons testé leur API avec un crédit de 100€. Bien que la performance des modèles soit satisfaisante, la simulation a révélé une lenteur significative. Cela pourrait être dû soit à un problème avec l'API d'OpenAI, soit avec la plateforme Y-Social elle-même. De plus, les coûts de l'API étaient élevés, et il est apparu que quelques jours de simulation étaient suffisants pour épuiser les crédits.
- **Hugging Face :** Hugging Face propose une variété de modèles LLM, tels que GPT-2, T5 et BERT, avec une API flexible permettant l'hébergement de modèles personnalisés. Bien que nous n'ayons pas testé cette solution en profondeur, elle semble prometteuse, notamment pour des simulations à grande échelle avec un coût plus maîtrisé.
- **Anthropic :** Anthropic propose le modèle Claude, principalement axé sur la sécurité et l'éthique de l'IA. Bien qu'il soit performant, il n'est pas adapté à des applications nécessitant une génération de texte à grande échelle.

Comparaison des APIs LLM

Coût

OpenAI : Les coûts se sont avérés élevés avec l'utilisation des crédits alloués, ce qui a limité son utilité pour des simulations à grande échelle. Un test de simulation de 20 jours avec 20 agents a rapidement épuisé les crédits.

Hugging Face : Bien que non testé en profondeur, Hugging Face semblait plus économique pour des simulations à grande échelle, offrant un modèle plus flexible.

Anthropic : Bien que performant en matière de sécurité, Anthropic ne semblait pas adapté pour une simulation de grande envergure, du fait de sa focalisation sur des modèles plus éthiques et sûrs.

Performance

OpenAI : Bien que performant sur le plan technique, OpenAI a montré des problèmes de lenteur dans notre simulation, ce qui soulève des questions sur sa capacité à gérer des déploiements à grande échelle.

Hugging Face : Cette API est très flexible et permet une intégration aisée. Sa performance est encore à tester de manière plus approfondie.

Anthropic : La lenteur des simulations n'a pas été un problème majeur ici, mais sa focalisation sur des modèles plus sécurisés pourrait limiter son efficacité pour des besoins de grande échelle.

Facilité d'intégration

Les trois solutions proposent des APIs bien documentées. Cependant, Hugging Face semble offrir plus de flexibilité, notamment pour l'hébergement de modèles personnalisés et l'intégration avec des simulations plus complexes.

Conclusion

L'utilisation d'un serveur personnel a été confrontée à des limitations matérielles importantes, notamment l'incompatibilité du GPU avec ROCm. Les tests d'APIs LLM ont montré que OpenAI, bien que performant, est trop coûteux et lent pour des simulations à grande échelle, tandis que Hugging Face semble offrir une solution plus économique et flexible. Anthropic, bien que performant pour des applications sécurisées, ne correspondait pas à nos besoins pour des générateurs de texte à grande échelle.