

Large Language Models for Personalized, Multimodal, and Explainable Recommendation and Information Retrieval

Grands Modèles de Langue pour la Recommandation et la Recherche d'Information Personnalisées, Multimodales et Explicables

Directeur de thèse : Mickey Mouse mickey.mouse@agroparistech.fr

Avril 2025

1 Introduction

Ce sujet de thèse vise à construire et évaluer de nouveaux systèmes de recommandation basés sur le filtrage collaboratif mais bâtis dans des espaces de représentation textuelle.

En combinant des techniques de d'analyse de sentiments et d'apprentissage de profils, nous proposons d'apprendre des représentations d'utilisateur versatiles. Ces représentations auront une double vocation : d'une part, prédire précisément les items susceptibles d'intéresser l'utilisateur ; d'autre part, générer les revues ou avis que l'utilisateur pourrait écrire sur ces items. L'objectif est ainsi d'améliorer la pertinence des recommandations grâce à une compréhension approfondie de l'utilisateur, tout en facilitant l'explicabilité des choix effectués par la génération de textes explicatifs.

Dans un second temps, l'idée de cette thèse est d'exploiter les systèmes de recherche d'information basés sur le dialogue. Un dialogue qui doit permettre d'affiner la requête utilisateur sur le plan de la pertinence mais aussi un dialogue tourné vers le profil lui-même, afin de le rendre plus transparent, contrôlable et modifiable directement par l'utilisateur.

Enfin, ces travaux incluent une importante dimension évaluative, liée aux défis spécifiques à l'évaluation de textes générés automatiquement. Les modèles de langage à grande échelle (LLMs) nécessitent en effet la création et la validation de nouvelles métriques d'évaluation, adaptées aux approches personnalisées proposées dans cette thèse.

2 Contexte

Les systèmes de recommandation et la recherche d'information représentent deux axes complémentaires d'accès à l'information. L'enjeu de ces systèmes est de filtrer et ordonner des contenus pertinents (*produits, articles, pages web, contenu multimédia, etc.*) à proposer aux utilisateurs en fonction de différents paramètres et contextes [Ricci et al., 2010, Manning, 2009]. La recherche d'information consiste à faire correspondre les requêtes de l'utilisateur avec les contenus pertinents, tandis que la recommandation consiste à faire cette correspondance avec l'historique et le profil de l'utilisateur. En particulier, les systèmes de recommandation exploitent les informations disponibles sur les utilisateurs, les items (contenus) et sur les interactions entre utilisateurs et items, pouvant être explicites (*notes, likes, ...*) ou implicites (*clics, visionnage, ...*), pour proposer des suggestions personnalisées, notamment à travers des techniques de filtrage collaboratif [He et al., 2017, Koren et al., 2009] et d'analyse de contenus [Aggarwal and Aggarwal, 2016, Lops et al., 2011]. Parallèlement, la recherche d'information s'est développée en s'appuyant sur des méthodes d'analyse textuelle et de traitement du langage, permettant de répondre aux requêtes des utilisateurs en explorant de vastes bases

de données d'articles, de documents et d'autres sources d'information [Mitra et al., 2018]. Les données exploitées par ces systèmes sont parfois structurées sous forme de connaissances [Hogan et al., 2021] et se présentent désormais sous différentes modalités, incluant le texte et l'image. Le texte issu des documents, des pages web, des revues d'utilisateurs ou encore des descriptions d'items, occupe une place majeure. Les systèmes modernes [Liu et al., 2024, Wei et al., 2023, Yuan et al., 2023] arrivent à tirer parti de plus en plus de l'information multimodale pour affiner les recommandations et la pertinence des résultats de recherche.

Le traitement du langage naturel (NLP) a connu de grandes avancées grâce à l'émergence des grands modèles de langage (LLMs) [Brown et al., 2020, OpenAI, 2023, Raffel et al., 2020, Touvron et al., 2023]. Ces modèles offrent des capacités impressionnantes dans les tâches de compréhension et de génération de texte, telles que la traduction automatique [Rivera-Trigueros, 2022], la réponse aux questions [Zhuang et al., 2023], la synthèse de texte [Goyal et al., 2022], l'analyse des sentiments [Wankhade et al., 2022] et la génération de code [Li et al., 2022]. Plusieurs travaux se sont penchés sur l'utilisation des LLMs dans les systèmes de recommandation [Zhao et al., 2024]. Certains les utilisent directement comme modèles de recommandation dans divers scénarios, tels que la prédiction de note [Geng et al., 2022, Kang et al., 2023], le ranking [Carraro and Bridge, 2024, Hou et al., 2024], la recommandation séquentielle [Harte et al., 2023, Liao et al., 2023] ; d'autres [Geng et al., 2022, Lubos et al., 2024] les emploient pour générer des explications aux recommandations. En recherche d'information, les LLMs interviennent notamment dans les approches de dense retrieval, en transformant requêtes et documents en représentations vectorielles riches qui facilitent une recherche sémantique dépassant la simple correspondance lexicale [Karpukhin et al., 2020]. Ils sont également employés pour le re-ranking [Nogueira et al., 2019], la synthèse des documents pertinents [Goyal et al., 2022] et la génération augmentée par récupération (RAG) [Lewis et al., 2020]. Enfin, pour la recommandation et la recherche d'information, les LLMs jouent un rôle clé dans l'interaction utilisateur-système, notamment par le biais de systèmes conversationnels dédiés à la recherche d'information [Jannach et al., 2021]. Ainsi, les grands modèles de langage s'imposent comme des outils de choix pour repenser et améliorer à la fois les systèmes de recommandation et les méthodes de recherche d'information.

La multimodalité, permettant de combiner des informations provenant de différentes sources telles que le texte et l'image, joue un rôle essentiel dans la compréhension fine des préférences des utilisateurs pour l'amélioration à la fois des systèmes de recommandation et de recherche d'information personnalisés [Wei et al., 2023, Yuan et al., 2023]. Pour exploiter au mieux ces différentes modalités, il est nécessaire d'extraire les informations pertinentes de chacune d'elles. En ce qui concerne la modalité image, les réseaux de neurones convolutifs (CNN) [He et al., 2016, Simonyan and Zisserman, 2014] et les Transformers visuels [Dosovitskiy, 2020, Touvron et al., 2021], sont largement utilisés pour analyser et extraire des caractéristiques visuelles, telles que les couleurs dominantes, les motifs, ou la composition d'une image. Des méthodes d'extraction d'information, telles que les opinions avec de l'analyse de sentiment basée sur les aspects (ABSA) [Liu, 2022], la modélisation de thématiques [Blei et al., 2003], les méthodes d'extraction d'entités et de relations [Lample et al., 2016], ou encore les LLMs par leur capacité de compréhension de texte, permettent de prendre en main la modalité textuelle. L'intégration de la multimodalité dans les systèmes de recommandation et de recherche d'information permet de combiner la richesse des informations issues de chaque modalité, d'améliorer la compréhension des préférences des utilisateurs et d'augmenter la pertinence des suggestions, offrant également des opportunités d'explicabilité accrue, en justifiant ces suggestions sur la base d'éléments visuels et textuels, permettant ainsi une meilleure transparence et une expérience utilisateur optimisée.

D'autre part, l'explicabilité reste un défi clé pour ces systèmes [Zhang et al., 2020]. L'explicabilité dans les systèmes de recommandation et de recherche d'information consiste à fournir des informations complémentaires expliquant pourquoi un item est recommandé à un utilisateur ou pourquoi un contenu est pertinent sachant la requête de l'utilisateur. En recommandation, les approches traditionnelles d'explication reposent souvent sur la similarité avec des utilisateurs [Resnick et al., 1994, Sharma and Cosley, 2013] ou des items proches [Sarwar et al., 2001]. Avec l'essor des grands modèles de langage (LLMs), de nouvelles approches ont émergé, telles que la génération de textes explicatifs ou la génération de revue personnalisés pour un item recommandé [Geng et al., 2022, Lubos et al., 2024]. En complément, deux autres approches particulièrement prometteuses sont l'explication basée sur les aspects ou les opinions, et l'explication

causale et contrefactuelle. L'explication basée sur les aspects [Li and Xu, 2020, Liu, 2022] nécessite d'extraire les opinions présentes dans les revues des utilisateurs, afin d'expliquer les recommandations futures de manière plus détaillée. Quant à l'explication causale et contrefactuelle [Feder et al., 2022, Gao et al., 2024], elle permet de modéliser les éléments qui influencent les préférences des utilisateurs et d'expliquer les recommandations en explorant des scénarios alternatifs. Ces différentes méthodes d'explication contribuent non seulement à renforcer la transparence des recommandations, mais aussi à améliorer la confiance des utilisateurs dans les systèmes de recommandation.

Mots clés : Recommandation, Recherche d'information, Grands modèles de langue (LLMs), Traitement du langage naturel (NLP), Personnalisation, Multimodalité, Explicabilité, Interprétabilité, Interaction, Diversité, Analyse des sentiments, Extraction d'informations, Texte, Image, Audio.

3 Problématique générale et Objectifs

La grande idée de ce sujet de thèse est de proposer de nouvelles méthodes de recommandation et de recherche d'information multimodales en tirant parti des grands modèles de langue, afin de fournir des recommandations de contenus à la fois performantes et explicables. La problématique de la recherche se concentre sur les axes suivants : Performance et multimodalité; Explicabilité : traditionnelle, textuelle, basée sur les aspects, contrefactuelle ; Diversité ; Interaction utilisateur-système ; Métriques d'évaluation des IA génératives;

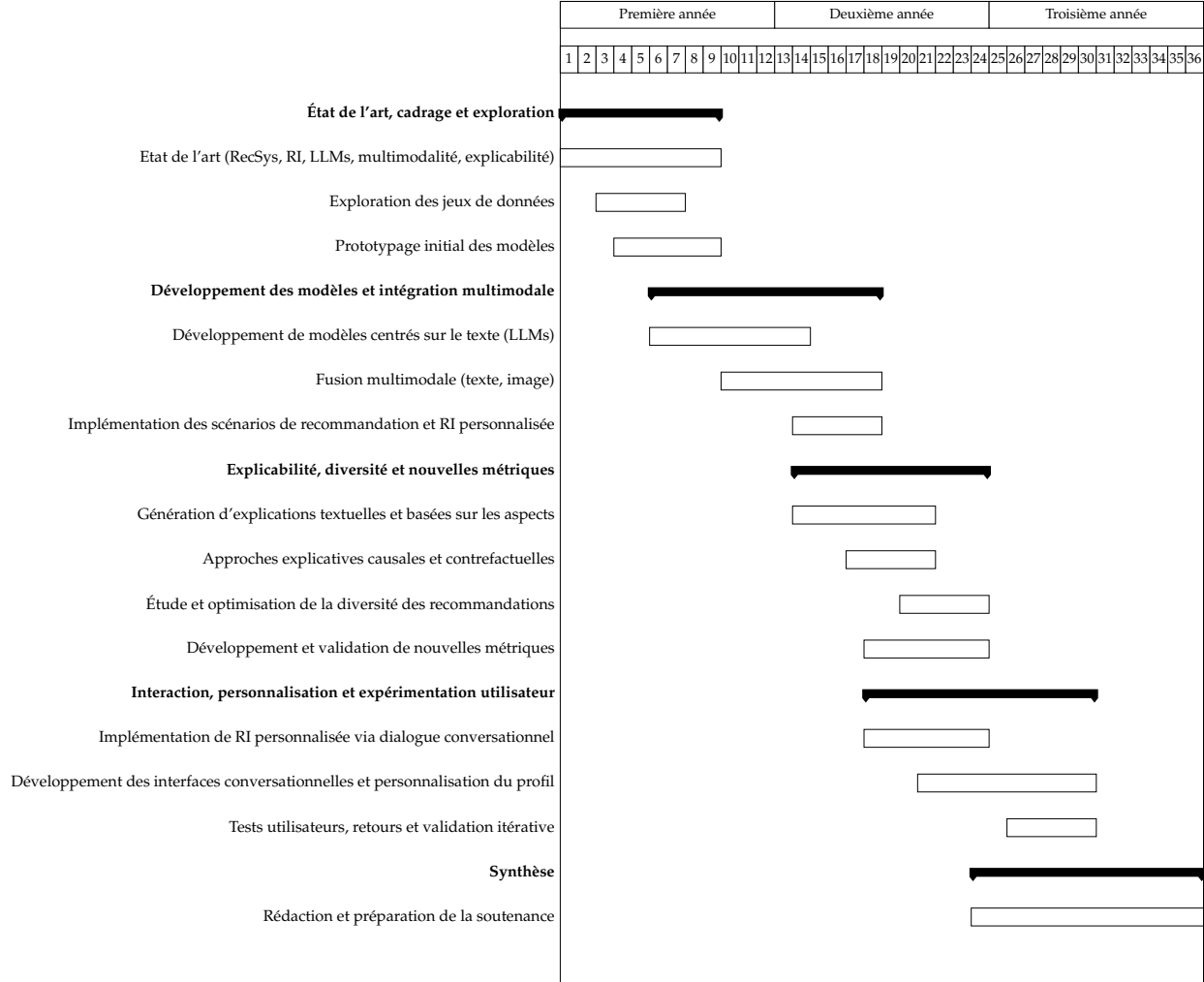
En accord avec les différents axes de recherche énumérés, les objectifs de ce travail sont les suivants :

1. **Développement de modèles de recommandation performants et multi-modaux :** L'objectif est de proposer des modèles capables de prédire avec précision les préférences des utilisateurs pour divers items, en exploitant non seulement les données numériques (notes) mais aussi les données multimodales, dont le texte.
2. **Amélioration de l'explicabilité des recommandations :** Pour rendre les recommandations plus transparentes et compréhensibles pour les utilisateurs, l'idée est d'exploiter la génération de textes explicatifs pour justifier les recommandations. L'utilisation d'explications basées sur les aspects ainsi que des approches d'explications causales et contrefactuelles doivent permettre de mieux contrôler la dimension générative du système et d'obtenir des garanties sur le respect des préférences et l'absence d'hallucinations.
3. **Etude de la diversité dans les recommandations :** Cet objectif consiste à développer des méthodes qui permettent de générer des recommandations variées pour éviter la répétitivité et favoriser la découverte de nouveaux items. Il s'agit de trouver un équilibre entre la pertinence des recommandations et leur diversité, afin de maintenir l'intérêt des utilisateurs tout en proposant des suggestions adaptées à leurs goûts. De nouveau, l'ancrage textuel de l'espace de représentation permet une meilleure évaluation et optimisation de la diversité.
4. **Interaction utilisateur-système et personnalisation du profil :** Concevoir des interfaces et des outils permettant aux utilisateurs de mieux comprendre leur profil devient possible une fois que le profil est bâti dans un espace textuel. L'idée est d'aller plus loin et de permettre l'affinage et la correction de leur profil à travers un dialogue. Ce contrôle ouvre des perspectives sur la performance du système et sa transparence.
5. **Surmonter les limites des métriques d'évaluation :** Cet objectif consiste à proposer de nouvelles approches pour évaluer la performance des systèmes de recommandation, en tenant compte des limitations des métriques traditionnelles et les scores de similarité pour la génération de texte. Il s'agit de développer des métriques qui capturent la pertinence, la diversité, et la qualité des explications des recommandations.

Ces objectifs s'inscrivent dans une démarche globale visant à développer des systèmes de recommandation et de recherche d'information à la fois personnalisables, performants, explicables et capables de tirer parti de la diversité des données disponibles. En répondant à ces défis, cette recherche entend contribuer

à l'amélioration continue des systèmes de recommandation et de recherche d'information de nouvelle génération, à l'ère des grands modèles de langue et de la multimodalité.

4 Organisation



L'organisation de cette thèse est classique et se structure en plusieurs phases interdépendantes qui s'étalent sur une période de 36 mois. Une première phase de bibliographie permettra de consolider l'état de l'art, d'identifier les axes de recherche pertinents et de définir le cadre méthodologique du projet. Par la suite, l'exploration et le prétraitement des jeux de données viendront alimenter des premiers prototypes, qui serviront de socle pour le développement ultérieur des modèles. Le projet s'articule ainsi autour de plusieurs axes :

- **État de l'art, cadrage et exploration** : Une revue exhaustive de la littérature sur les systèmes de recommandation, la recherche d'information, les grands modèles de langue, la multimodalité et l'explicabilité permettra de dresser un panorama des approches existantes. Parallèlement, une exploration des jeux de données disponibles sera effectuée afin de mettre en place les prototypes initiaux.
- **Développement des modèles et intégration multimodale** : À partir des résultats de la phase exploratoire, des modèles centrés sur le texte (basés sur les LLMs) seront développés et optimisés. Ces modèles seront ensuite enrichis par l'intégration de l'image, via des techniques de fusion multimodale. Des scénarios spécifiques de recommandation et de recherche d'information personnalisée

seront également implémentés à ce stade.

- **Explicabilité, diversité et nouvelles métriques** : Afin d’assurer la transparence et la robustesse des systèmes, des mécanismes de génération d’explications textuelles, basées notamment sur les aspects, la causalité et des approches contrefactuelles, seront mis en place. Parallèlement, une attention particulière sera portée à l’optimisation de la diversité des recommandations et au développement de nouvelles métriques d’évaluation adaptées aux spécificités des modèles génératifs et multimodaux.
- **Interaction, personnalisation et expérimentation utilisateur** : Le développement d’interfaces conversationnelles permettra d’instaurer un dialogue entre l’utilisateur et le système, facilitant ainsi l’affinage du profil utilisateur et la personnalisation des résultats. Des tests utilisateurs itératifs seront conduits pour recueillir des retours, affiner les modèles et valider l’efficacité des métriques d’évaluation proposées.
- **Synthèse et préparation de la soutenance** : La dernière phase consistera à synthétiser l’ensemble des travaux réalisés, à rédiger la thèse et à préparer les supports de la soutenance.

5 Encadrement

Mickey Mouse Professeur, AgroParisTech, DATAIA, Laboratoire UMR MIA-Paris-Saclay. Vincent Guigue est travaille sur les systèmes de recommandation et l’extraction d’information. Sa recherche porte sur le développement de systèmes de recommandation plus transparents [Afchar et al., 2023], notamment à travers l’exploitation des données textuelles [Dias et al., 2019] et sur l’extraction d’information [Taillé et al., 2020] et sur l’évaluation des textes générés par des modèles de langue [Herseant and Guigue, 2025, Ispas et al., 2025].

References

- [Afchar et al., 2023] Afchar, D., Hennequin, R., and Guigue, V. (2023). Of spiky svds and music recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 926–932.
- [Aggarwal and Aggarwal, 2016] Aggarwal, C. C. and Aggarwal, C. C. (2016). Content-based recommender systems. *Recommender systems: The textbook*, pages 139–166.
- [Blei et al., 2003] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- [Brown et al., 2020] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.
- [Carraro and Bridge, 2024] Carraro, D. and Bridge, D. (2024). Enhancing recommendation diversity by re-ranking with large language models. *arXiv preprint arXiv:2401.11506*.
- [Dias et al., 2019] Dias, C.-E., Guigue, V., and Gallinari, P. (2019). Personalized attention for textual profiling and recommendation. In *Proc. EARS@ SIGIR*.
- [Dosovitskiy, 2020] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [Feder et al., 2022] Feder, A., Keith, K. A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., Eisenstein, J., Grimmer, J., Reichart, R., Roberts, M. E., et al. (2022). Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *Transactions of the Association for Computational Linguistics*, 10:1138–1158.
- [Gao et al., 2024] Gao, C., Zheng, Y., Wang, W., Feng, F., He, X., and Li, Y. (2024). Causal inference in recommender systems: A survey and future directions. *ACM Transactions on Information Systems*, 42(4):1–32.

- [Geng et al., 2022] Geng, S., Liu, S., Fu, Z., Ge, Y., and Zhang, Y. (2022). Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM Conference on Recommender Systems*, pages 299–315.
- [Goyal et al., 2022] Goyal, T., Li, J. J., and Durrett, G. (2022). News summarization and evaluation in the era of gpt-3. *arXiv preprint arXiv:2209.12356*.
- [Harte et al., 2023] Harte, J., Zorgdrager, W., Louridas, P., Katsifodimos, A., Jannach, D., and Fraggoulis, M. (2023). Leveraging large language models for sequential recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1096–1102.
- [He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [He et al., 2017] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182.
- [Herserant and Guigue, 2025] Herserant, T. and Guigue, V. (2025). Seval-ex: A statement-level framework for explainable summarization evaluation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*.
- [Hogan et al., 2021] Hogan, A., Blomqvist, E., Cochez, M., d’Amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., et al. (2021). Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37.
- [Hou et al., 2024] Hou, Y., Zhang, J., Lin, Z., Lu, H., Xie, R., McAuley, J., and Zhao, W. X. (2024). Large language models are zero-shot rankers for recommender systems. In *European Conference on Information Retrieval*, pages 364–381. Springer.
- [Ispas et al., 2025] Ispas, A.-R., Simon, C.-E., Caspani, F., and Guigue, V. (2025). Towards lighter and robust evaluation for retrieval augmented generation. In *ICLR Workshop: Quantify Uncertainty and Hallucination in Foundation Models: The Next Frontier in Reliable AI*.
- [Jannach et al., 2021] Jannach, D., Manzoor, A., Cai, W., and Chen, L. (2021). A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5):1–36.
- [Kang et al., 2023] Kang, W.-C., Ni, J., Mehta, N., Sathiamoorthy, M., Hong, L., Chi, E., and Cheng, D. Z. (2023). Do llms understand user preferences? evaluating llms on user rating prediction. *arXiv preprint arXiv:2305.06474*.
- [Karpukhin et al., 2020] Karpukhin, V., Oguz, B., Min, S., Lewis, P. S., Wu, L., Edunov, S., Chen, D., and Yih, W.-t. (2020). Dense passage retrieval for open-domain question answering. In *EMNLP (1)*, pages 6769–6781.
- [Koren et al., 2009] Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37.
- [Lample et al., 2016] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.
- [Lewis et al., 2020] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474.
- [Li and Xu, 2020] Li, W. and Xu, B. (2020). Aspect-based fashion recommendation with attention mechanism. *IEEE Access*, 8:141814–141823.
- [Li et al., 2022] Li, Y., Choi, D., Chung, J., Kushman, N., Schrittwieser, J., Leblond, R., Eccles, T., Keeling, J., Gimeno, F., Dal Lago, A., et al. (2022). Competition-level code generation with alphacode. *Science*, 378(6624):1092–1097.

- [Liao et al., 2023] Liao, J., Li, S., Yang, Z., Wu, J., Yuan, Y., Wang, X., and He, X. (2023). Llara: Aligning large language models with sequential recommenders. *arXiv preprint arXiv:2312.02445*.
- [Liu, 2022] Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- [Liu et al., 2024] Liu, Y., Wang, Y., Sun, L., and Yu, P. S. (2024). Rec-gpt4v: Multimodal recommendation with large vision-language models. *arXiv preprint arXiv:2402.08670*.
- [Lops et al., 2011] Lops, P., De Gemmis, M., and Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. *Recommender systems handbook*, pages 73–105.
- [Lubos et al., 2024] Lubos, S., Tran, T. N. T., Felfernig, A., Polat Erdeniz, S., and Le, V.-M. (2024). Llm-generated explanations for recommender systems. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, pages 276–285.
- [Manning, 2009] Manning, C. D. (2009). *An introduction to information retrieval*.
- [Mitra et al., 2018] Mitra, B., Craswell, N., et al. (2018). An introduction to neural information retrieval. *Foundations and Trends® in Information Retrieval*, 13(1):1–126.
- [Nogueira et al., 2019] Nogueira, R., Yang, W., Cho, K., and Lin, J. (2019). Multi-stage document ranking with bert. *arXiv preprint arXiv:1910.14424*.
- [OpenAI, 2023] OpenAI (2023). Gpt-4 technical report. Technical report.
- [Raffel et al., 2020] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.
- [Resnick et al., 1994] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186.
- [Ricci et al., 2010] Ricci, F., Rokach, L., and Shapira, B. (2010). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.
- [Rivera-Trigueros, 2022] Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, 56(2):593–619.
- [Sarwar et al., 2001] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295.
- [Sharma and Cosley, 2013] Sharma, A. and Cosley, D. (2013). Do social explanations work? studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1133–1144.
- [Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [Taillé et al., 2020] Taillé, B., Guigue, V., Scoutheeten, G., and Gallinari, P. (2020). Let’s stop incorrect comparisons in end-to-end relation extraction! In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3689–3701.
- [Touvron et al., 2021] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., and Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR.
- [Touvron et al., 2023] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. (2023). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

- [Wankhade et al., 2022] Wankhade, M., Rao, A. C. S., and Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780.
- [Wei et al., 2023] Wei, W., Huang, C., Xia, L., and Zhang, C. (2023). Multi-modal self-supervised learning for recommendation. In *Proceedings of the ACM Web Conference 2023*, pages 790–800.
- [Yuan et al., 2023] Yuan, Z., Yuan, F., Song, Y., Li, Y., Fu, J., Yang, F., Pan, Y., and Ni, Y. (2023). Where to go next for recommender systems? id-vs. modality-based recommender models revisited. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2639–2649.
- [Zhang et al., 2020] Zhang, Y., Chen, X., et al. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1):1–101.
- [Zhao et al., 2024] Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., Wen, Z., Wang, F., Zhao, X., Tang, J., et al. (2024). Recommender systems in the era of large language models (llms). *IEEE Transactions on Knowledge and Data Engineering*.
- [Zhuang et al., 2023] Zhuang, Y., Yu, Y., Wang, K., Sun, H., and Zhang, C. (2023). Toolqa: A dataset for llm question answering with external tools. *Advances in Neural Information Processing Systems*, 36:50117–50143.