LITERATURE REVIEW: IMPLEMENTING DEEP LEARNING TECHNIQUES IN MEDIA & ENTERTAINMENT RECOMMENDATION SYSTEMS

Maria Ingold
12693772
Unit 7
Research Methods and Professional Practice
University of Essex Online
15 June 2024

CONTENTS

INTRODUCTION

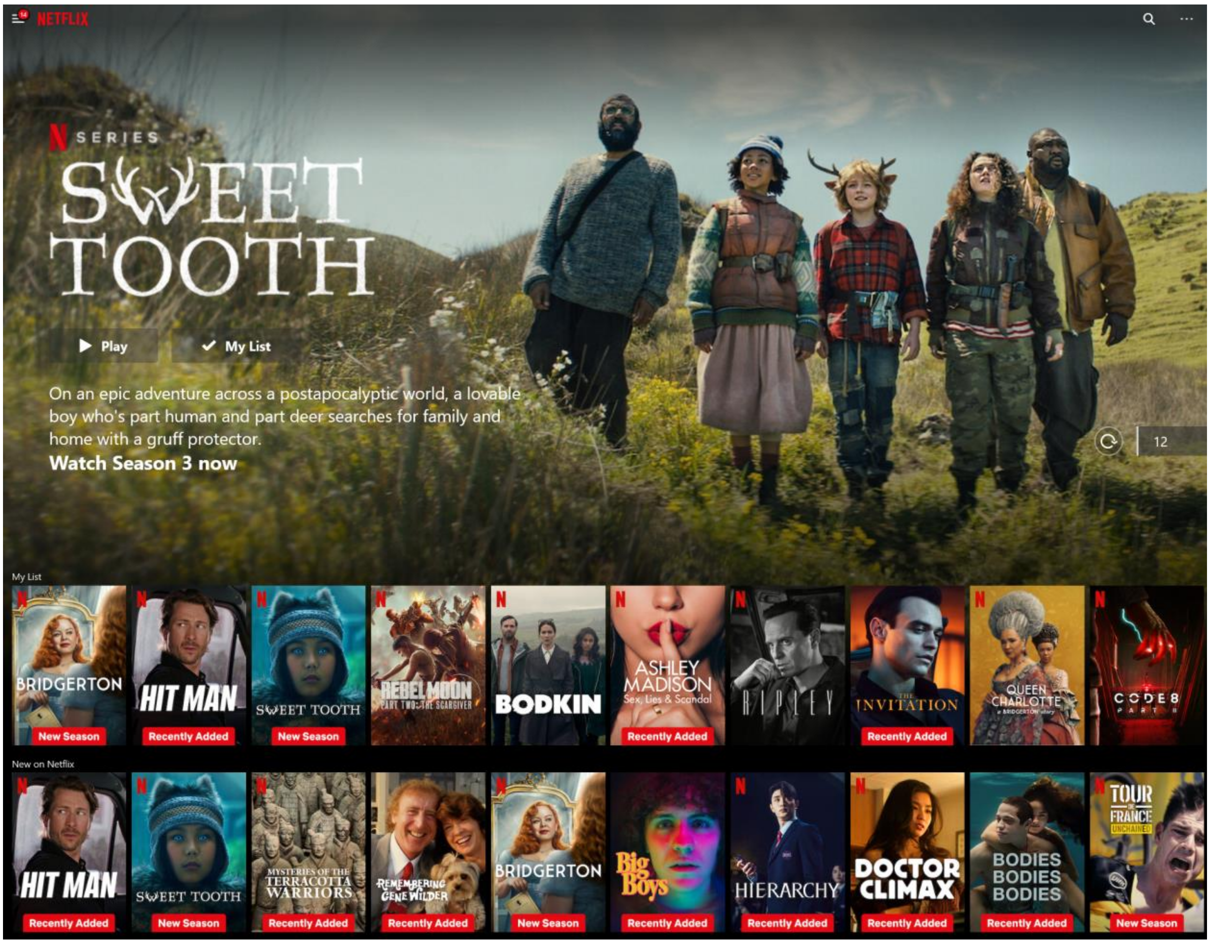

FIGURE 1 | Netflix recommendations personalised matrix

Steck et al. (2021) puts Netflix's perspective succinctly, 'the value of a recommender system can be measured by the increase in member retention'. Member satisfaction drives customer retention which generates revenue (Steck et al., 2021). Personalisation using recommender (or recommendation) systems aims to improve satisfaction by minimising time spent searching and maximising time spent watching appealing content (Frey, 2021).

Video on demand (VOD)—content the user chooses when to watch—can be delivered over-the-top (OTT)—over the internet—or via other methods (Ingold, 2020). Reducing churn and retaining customers is essential for OTT VOD streaming services, such as Netflix, Hulu and Disney+, which rely on subscription (SVOD) and ad-supported (AVOD) payment models (Ingold, 2022; de Zilwa, 2023). Therefore, recommender systems continue to evolve, using a range of artificial intelligence (AI) techniques (Khare & Jhapate, 2022).

'Overview' examines types of traditional recommender systems and deep learning techniques. 'Evaluation' compares published peer-reviewed research on deep learning for recommender systems in OTT VOD streaming services for movies and series between 2018 and 2024, contrasting academic studies with industry applications. 'Future' explores the next era: Large Language Models (LLMs). Google Scholar identified relevant publications using keywords including Netflix, movie, VOD, OTT, recommendation system, recommender system and deep learning (Dawson, 2015).

OVERVIEW

Traditional Recommendation Approaches

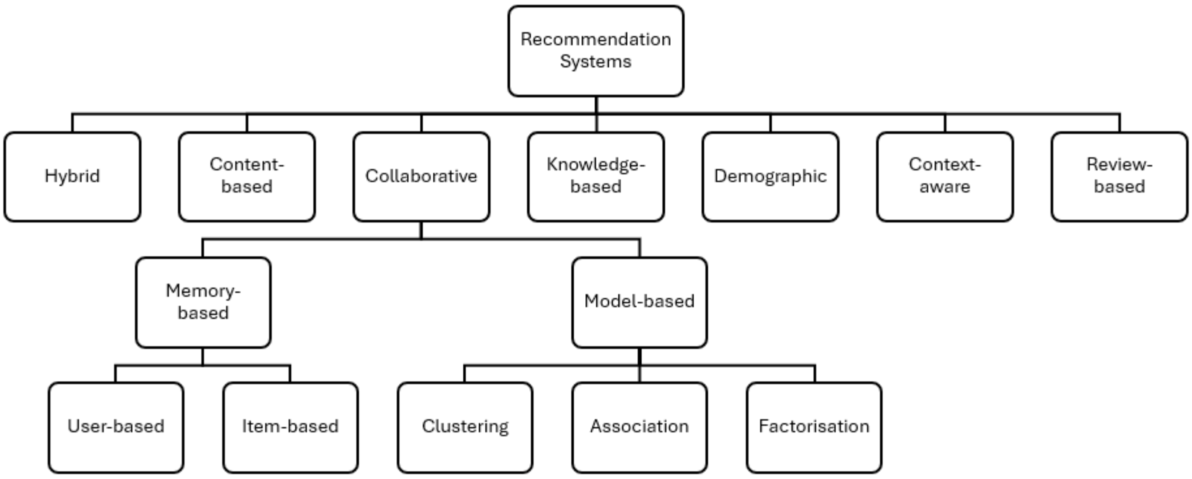

FIGURE 2 | Recommender system approaches

Recommender systems filter information to provide personalised recommendations (Anwar & Uma, 2021). Approaches and subsets are shown in Figure 2. While Zhang et al. (2021) considers knowledge-based and Anwar and Uma (2021) list seven approaches, three are typically mentioned: collaborative filtering, content-based filtering, and hybrid filtering (Goyani & Chaurasiya, 2020; Gupta et al., 2020; Peng et al., 2024).

Content based-filtering (CBF) recommends items similar to those a user has explicitly rated or implicitly chosen in the past by creating item and user profiles (Goyani & Chaurasiya, 2020). It uses traditional machine learning (ML) techniques like Naïve Bayes, nearest neighbour, and decision trees (Zhang et al., 2021).

Collaborative filtering (CF) uses similar tastes—recommending items similar users have liked—via a user-item matrix (Goyani & Chaurasiya, 2020). It can either be memory-based or model-based (Zhang et al., 2021).

- **Memory-based** (heuristic) provides explicit ratings and filters either user-based (user-to-user) or item-based (item-to-item). It commonly uses nearest neighbour to find similarities (Zhang et al., 2021).
- **Model-based** (machine learning or data mining) can use rated or unrated items and may add side information like tags, reviews, and location into the user-item matrix (Zhang et al., 2021). A common approach is matrix factorisation (MF) which uses Singular Value Decomposition (SVD) to reduce dimensionality and improve prediction accuracy (Mu and Wu, 2023).

Challenges can include the *cold start* problem for both new users and new data, *data sparsity*, *scalability*, *overspecialisation,* and *transparency* (Goyani & Chaurasiya, 2020).

A **hybrid filtering** approach combines multiple recommendation techniques to solve issues (Anwar & Uma, 2021; Peng et al., 2024). Content-based addresses data sparsity, new-item cold start and transparency, while collaborative model-based matrix factorisation resolves overspecialisation, scalability, and density (Zhang et al., 2021). However, hybrid can increase expense and complexity.

Peng et al. (2024) identify three recommender design classes: traditional, deep-learning, and reinforcement-learning methods. As seen in Figure 3, Pérez Maurera (2023) further categorises traditional as heuristics, statistical and machine learning.

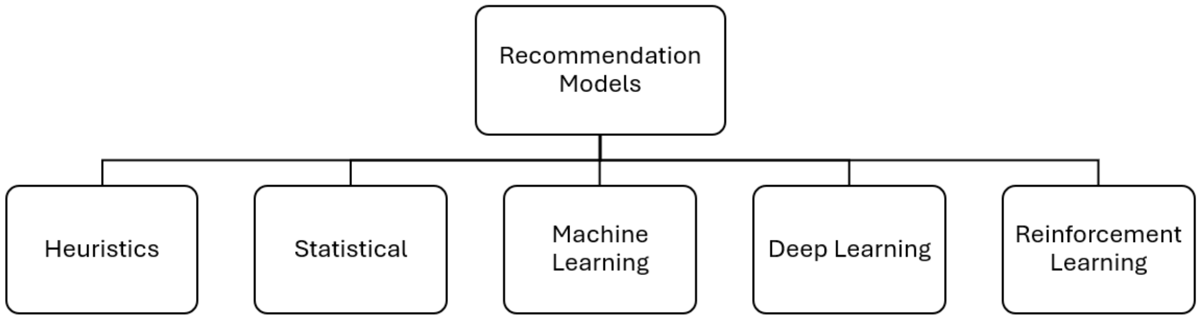

FIGURE 3 | Recommendation models

Deep Learning in Recommendations

Netflix discovered that while no single model best generated all recommendations, deep learning ultimately resulted in significant improvements (Steck et al., 2021). Zhang et al. (2021) identifies some of the more important Artificial Intelligence (AI) deep learning techniques and improvements to recommender systems (Figure 4).

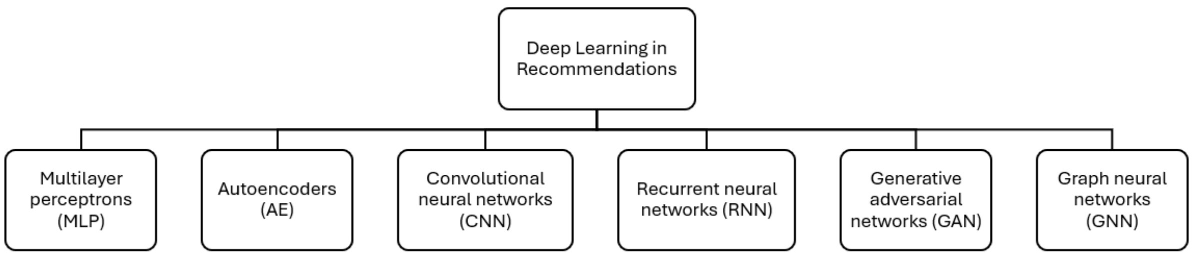

FIGURE 4 | Deep learning

- **Multilayer perceptrons (MLP)** are neural networks used for classification and regression. In recommenders they provide feature engineering in factorisation machines, combine linear and non-linear relationships, and capture high-level abstractions (Zhang et al., 2021; Peng et al., 2024).
- **Autoencoders (AE)** consist of an encoder and decoder and are unsupervised neural networks (Lund & Ng, 2018; Zhang et al. 2021). They can be used for feature learning from item content, learn item representation and user profiling, and combine with matrix factorisation.
- **Convolutional neural networks (CNN)** include convolution layers and pooling, and can recognise and classify images and videos. They address data sparsity and increase rating prediction accuracy if integrated into matrix factorisation (Zhang et al., 2018; Zhang et al., 2021).
- **Recurrent neural networks (RNN)** are directed graphs with memory, so can be used for sequential data, analyse user-item interactions over time, and predict interests (Zhang et al., 2021). RNNs can track user preference changes over time using long short-term memory (LTSM) (Siet et al., 2024).
- **Generative adversarial networks (GAN)** use competing generative and discriminative models to generate new samples (Zhang et al., 2021). They can learn user-item representations from interactions, tags, images, and knowledge graphs.
- **Graph neural networks (GNN)** can learn from neighbourhood graph information to create features for nodes (Zhang et al., 2021).

Table 1 summarises the strengths and weaknesses of deep learning in recommendation models (Zhang et al., 2019; Steck et al., 2021; Peng et al., 2024).

TABLE 1 | Deep learning strengths and limitations

Strengths	Limitations
• Nonlinear transformation • Representation learning • Sequence modelling • Flexibility • Automates feature extraction	• Interpretability • Fairness and explainability • Data requirement • Extensive hyperparameter tuning

EVALUATION

Comparison of Deep Learning Techniques

This section and Table 2 compare papers on recommendation systems using deep learning for VOD (not linear or live) movie and series (not user generated or short form) streaming services delivered OTT (e.g. not broadcast or cable).

In 2018, deep learning was still emerging in recommender systems (Lund & Ng, 2018). Lund and Ng (2018) noted three prior examples using user content, with their approach instead predicting user ratings. While multilayer perceptron did not deliver results for them, six years later, Siet et al. (2024) found MLP helped mitigate cold-start and scalability issues when used with KMeans. Autoencoders indicated potential, but initial tests took over a week to run. As later papers do not note compute time, it is easy to forget that computational performance for ML hardware roughly doubled every two years (Hobbhahn et al., 2023).

TABLE 2 | Deep learning recommendation systems for OTT VOD

Algorithm	Claimed Result	Dataset	Study	Year
Autoencoders	Lower RMSE, outperformed neighbourhood baseline for collaborative filtering.	MovieLens (45,115 movies by 270,896 users with 26M ratings.) MovieLens BaseLine (943 users, 1,682 movies)	(Lund and Ng, 2018)	2018
Multi-modal heterogeneity Autoencoders, RNN (LSTM), Transformer (BERT)	Significant improvements with deep learning using multimodal heterogeneity and replacing asymmetric matrix factorisation with autoencoders.	Netflix dataset. Medium-sized item set and hundreds of millions of members.	(Steck et al., 2021)	2021
Multi-modal GNN (Graph Attenuation Network)	Statistically significant lowest RMSE for meta information plus description with 5 keyframes trailer video (followed by without video). Suboptimal audio and video embeddings.	MovieLens 100K (extended with text, video, and audio)	(Chakder et al., 2022)	2022
Multi-modal heterogeneity CNN (ResNet-50), GNN (GraphSAGE), Transformer (BERT), MLP, Split Mixture of Experts (SMoE)	Multi-Modal Multi-Interest Multi-Scenario Matching (M5). Significantly surpassed method comparison. Lift for average global hours per visitor.	Test: Dataset-H 150M (30M users, 70K shows across live and VOD) Test: Dataset-D 180M instances (100M users, 3k shows across regions) Launched: Hulu and Disney+ (hundreds of millions)	(Zhao et al., 2023)	2023
Multi-modal CNN	Lower RMSE than User-CF, Item-CF and SVD. Sparse data partially alleviated.	MovieLens 100K and 1M	(Mu and Wu, 2023)	2023
Deep Reinforcement Learning (DRL) + CF	Did not exceed Surprise library performance but surpassed Cornac library models. Potential to address cold-start issue.	MovieLens 1M	(Peng et al., 2024)	2024
MLP + KMeans	RMSE, MAE, precision, recall, and F1 outperform baseline methods. Mitigates cold-start and scalability.	MovieLens 100K and 1M	(Siet et al., 2024)	2024

Autoencoders also proved useful for Netflix; bag-of-videos treats user video plays as a non-temporal set, with autoencoders able to replace asymmetric matrix factorisation and neighbourhood-based approaches (Steck et al., 2021). However, while Peng et al. (2024) tried deep reinforcement learning without much success, shifts beyond deep learning to transformer technology by Steck et al. (2021) and Zhao et al. (2023) delivered results.

Crucially, in 2021, Netflix proved the hypothesis that complex models like deep learning benefit from heterogeneous data—supplementary diverse information, including text, metadata, identifiers, genre tags, cast information, images, video and more (Steck et al., 2021; Zhao et al., 2023). Multi-modal deep learning success was demonstrated by Netflix's Steck et al. (2021) and substantiated by academia's Chakder et al. (2022) and Mu and Wu (2023). Subsequently, Hulu and Disney+'s Zhao et al. (2023) deployed the "For You" M5 recommendation to hundreds of millions of subscribers daily, significantly surpassing method comparisons for all scenarios across all regions and all subscription types, resulting in a notable 1.0%/0.8% average global hours per visitor lift.

Strengths and Limitations

Netflix has used recommendations since its 2000 CineMatch algorithm for DVD rentals (de Zilwa, 2023). In 2006, Netflix published a rating dataset to encourage competition for improving recommendation models (Peng et al., 2024). However, according to Narayanan and Shmatikov (2008), it could be de-anonymised.

Academic papers use the public MovieLens datasets of 100,000 or 1 million instances (Lund & Ng, 2018; Chakder et al., 2022; Mu & Wu, 2023; Peng et al., 2024; Siet et al., 2024). MovieLens 100K and 1M are stable benchmark datasets meant for reporting research results. However, 100K was released in 1998, and 1M in 2003 (GroupLens Research, N.D.). Gonzalez et al. (2022) show that predictions in the 1M dataset are biased towards older and male users, possibly due to sociopsychological factors. While GroupLens has larger and more recent stable datasets—up to 20M in 2016—these were not used in the literature. By contrast, industry tests with hundreds of millions of its own recent data, including live A/B tests at scale (Steck et al., 2021; Zhao et al., 2023).

Root Mean Squared Error (RMSE) is frequently used as a success measurement by academia (Lund & Ng, 2018; Chakder et al., 2022; Mu and Wu, 2023; Siet et al., 2024). However, neither Netflix nor Hulu mentioned RMSE. Instead, Netflix described the offline and online mismatch where historical offline datasets (like MovieLens) do not match online A/B tests presented to users (Steck et al., 2021; Zhao et al., 2023). In some cases, Steck et al. (2021) observed worse performance with online tests using deep learning. As a result, Netflix improved metrics as well as models using off-policy evaluation, bandits, and reinforcement learning techniques. However, metrics modification requires online tests not accessible to researchers.

A key limitation is that academia can innovate but only test historically, not online with live users at scale. Conversely, academic research in deep learning, for instance, predates Netflix's user tests. Both mutually advance each other.

Deep learning trends include moving into transformers using BERT (Steck et al., 2021; Zhao et al., 2023). However, Zhao et al. (2024) observes that deep neural networks still have intrinsic limitations. Firstly, the data size and model scale using pre-trained transformer models, like BERT, lack the natural language understanding to sufficiently capture user and item data. Secondly, deep learning architectures use task-specific data and domain knowledge, so fail to generalise. Thirdly, while perhaps more applicable to travel, multi-step reasoning is difficult.

FUTURE

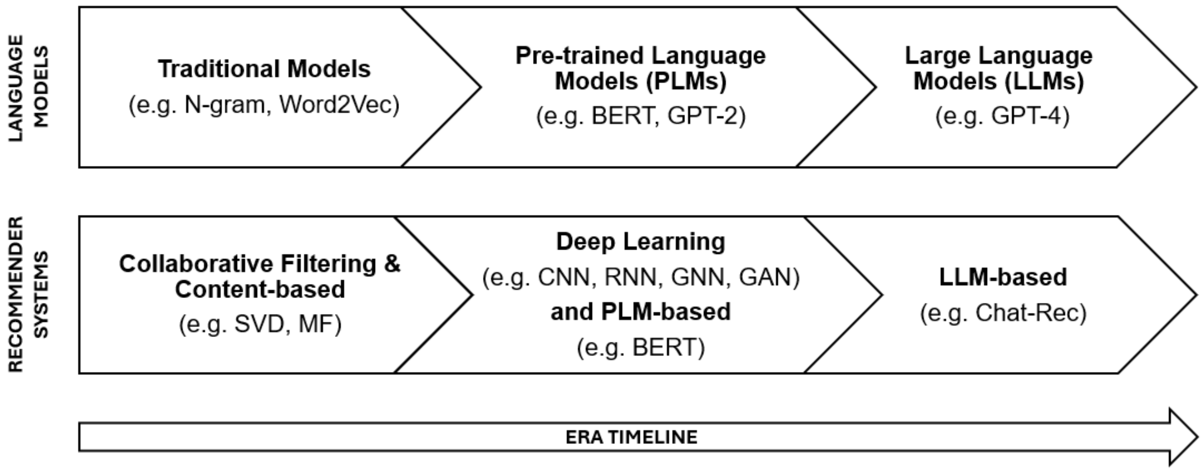

FIGURE 5 | Language Model and Recommender System eras

From predominantly 2023 onwards, recommendation systems have begun to evaluate Large Language Models (LLM) to resolve deep learning issues, as per Figure 5, simplified from Zhao et al. (2024). Neither Netflix nor Hulu have yet released papers on LLM use in recommendations; however, Zhao et al. (2024) show activity, including Chat-Rec, which leverages ChatGPT to refine traditional movie recommendations. This research into LLMs is not without problems. Granada et al. (2023) created VideolandGPT for small Dutch VOD platform Videoland, and while it increased visibility of items not normally recommended, it also introduced recommendations that did not exist on the platform. However, LLMs could address existing issues. Corecco et al. (2024) used LLMs to mimic human behaviour for training, addressing the scarcity of human feedback for reinforcement learning in recommender systems. LLMs are the next era of research, aiming to resolve existing deep learning and pre-trained language model issues.

CONCLUSION

In conclusion, combining heterogeneous data with deep learning in VOD recommendation systems has significantly enhanced the user experience using models like autoencoders, RNNs, CNNs, GNNs, and transformers. However, there is a performance gap between offline historical dataset tests and online user testing. While Netflix addresses internal metrics, there is potential for improved research metrics. The synergy between academic research and industry application continues to drive innovation. Deep learning overcomes some traditional challenges like cold-start, data sparsity and scalability, but lacks the natural language capabilities of LLMs. Although LLMs may address some deep learning constraints, early implementations demonstrate both potential and challenges. Future research is likely to focus on integrating LLMs and multi-modal data to further improve recommendation accuracy, member satisfaction, and retention.

REFERENCES

Anwar, T. & Uma, V. (2021) Comparative study of recommender system approaches and movie recommendation using collaborative filtering, *International Journal of System Assurance Engineering and Management* 12: 426–436. DOI: https://doi.org/10.1007/s13198-021-01087-x.

Chakder, D., Mondal, P., Raj, S., Saha, S., Ghosh, A., & Onoe, N. (2022) Graph Network based Approaches for Multi-modal Movie Recommendation System, *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* 409–414. DOI: https://doi.org/10.1109/SMC53654.2022.9945488.

Corecco, N., Piatti, G., Lanzendörfer, L.A., Fan, F.X. & Wattenhofer, R. (2024) An LLM-based Recommender System Environment. Available from: https://arxiv.org/abs/2406.01631v1 [Accessed 15 June 2024].

Dawson, C.W. (2015) *Projects in Computing and Information Systems*. 3rd ed. Harlow: Pearson.

Frey, M. (2021) *Netflix recommends: algorithms, film choice, and the history of taste*. University of California Press.

Gonzalez, A., Ortega, F., Perez-Lopez, D. & Alonso, S. (2022) Bias and Unfairness of Collaborative Filtering Based Recommender Systems in MovieLens Dataset, *IEEE Access* 10: 68429–68439. DOI: https://doi.org/10.1109/ACCESS.2022.3186719.

Goyani, M. & Chaurasiya, N. (2020) A Review of Movie Recommendation System: Limitations, Survey and Challenges, *Electronic Letters on Computer Vision and Image Analysis* 19(3): 18–37. DOI: https://doi.org/10.5565/rev/elcvia.1232.

Granada, M.G., Zilbershtein, D., Odijk, D., Barile, F. (2023) VideolandGPT: A User Study on a Conversational Recommender System, *CEUR Workshop Proceedings* 3560: 44–49. Available from: https://arxiv.org/abs/2309.03645v1 [Accessed 15 June 2024].

GroupLens Research (N.D.) *MovieLens Datasets*. Available from: https://grouplens.org/datasets/movielens/ [Accessed 15 June 2024].

Gupta, M., Thakkar, A., Aashish, Gupta, V. & Rathore D.P.S. (2020) Movie Recommender System Using Collaborative Filtering, *Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020* 415–420. DOI: https://doi.org/10.1109/ICESC48915.2020.9155879.

Hobbhahn, M., Heim, L. and Adyos, G. (2023) *Trends in Machine Learning Hardware*, *Epoch AI*. Available at: https://epochai.org/blog/trends-in-machine-learning-hardware [Accessed 14 June 2024].

Ingold, M. (2020) *Tech Expert: Inside OTT*, *IBC365*. Available from: https://www.ibc.org/tech-expert-inside-ott/5399.article [Accessed 10 June 2024].

Ingold, M. (2022) *The John Logie Baird Lecture: The State of Video Technology and the Next 5 Years*, *International Broadcasting Convention (IBC)*. Available from:

https://www.ibc.org/video/the-john-logie-baird-lecture-the-state-of-video-technology-and-the-next-5-years/9086.article [Accessed 10 June 2024].

Khare, N. & Jhapate, A. (2022) Review on Collaborative Filtering Machine Learning Approach for Recommendation Systems, *Research Journal of Engineering Technology and Medical Sciences* 66(03): 2582–6212. Available from: www.rjetm.in/ [Accessed 10 June 2024].

Lund, J. & Ng, Y.K. (2018) Movie recommendations using the deep learning approach, *Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018* 47–54. DOI: https://doi.org/10.1109/IRI.2018.00015.

Mu, Y. & Wu, Y. (2023) Multimodal Movie Recommendation System Using Deep Learning, *Mathematics 2023, Vol 11, Page 895* 11(4): 895. DOI: https://doi.org/10.3390/MATH11040895.

Narayanan, A. & Shmatikov, V. (2008) How To Break Anonymity of the Netflix Prize Dataset. DOI: https://doi.org/10.48550/arXiv.cs/0610105.

Peng, S., Siet, S., Ilkhomjon, S., Kim D.Y. & Park, D.S. (2024) Integration of Deep Reinforcement Learning with Collaborative Filtering for Movie Recommendation Systems, *Applied Sciences 2024, Vol 14, Page 1155* 14(3): 1155. DOI: https://doi.org/10.3390/APP14031155.

Pérez Maurera, F.B., Ferrari Dacrema, M., Castells, P. & Cremonesi, P. (2023) Impressions in Recommender Systems: Present and Future Discussion Paper. Available from: http://ceur-ws.org [Accessed 12 June 2024].

Siet, S., Peng, S., Ilkhomjon, S, Kang, M. & Park, D.S. (2024) Enhancing Sequence Movie Recommendation System Using Deep Learning and KMeans, *Applied Sciences 2024, Vol 14, Page 2505* 14(6): 2505. DOI: https://doi.org/10.3390/APP14062505.

Steck, H., Baltrunas, L., Elahi, E., Llang, D., Raimond, Y. & Basilico, J. (2021) Deep Learning for Recommender Systems: A Netflix Case Study, *AI Magazine* 42(3): 7–18. DOI: https://doi.org/10.1609/AIMAG.V42I3.18140.

Zhang, Q., Lu, J. & Jin, Y. (2021) Artificial intelligence in recommender systems, *Complex & Intelligent Systems* 7: 439–457. DOI: https://doi.org/10.1007/s40747-020-00212-w.

Zhang, S., Yao, L. Tay, Y.I. Sun, A. & Tay, Y. (2019) Deep Learning based Recommender System: A Survey and New Perspectives, *ACM Computing Surveys* 52(1). DOI: https://doi.org/https://doi.org/10.1145/3285029.

Zhao, P., Gao, X., Xu, C. & Chen, L. (2023) M5: Multi-Modal Multi-Interest Multi-Scenario Matching for Over-the-Top Recommendation, *KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data* 5650–5659. DOI: https://doi.org/10.1145/3580305.3599863.

Zhao, Z., Fan, W., Li, J., Liu, Y, Mei, X., Wang, Y., Wen, Z., Wang, F. & Tang, J. (2024) Recommender Systems in the Era of Large Language Models (LLMs), *IEEE Transactions on Knowledge and Data Engineering* [Preprint]. DOI: https://doi.org/10.1109/TKDE.2024.3392335.

de Zilwa, D.K. (2023) Netflix: rise, fall and recovery, *Journal of Business Strategy* [Preprint]. DOI: https://doi.org/10.1108/JBS-08-2023-0177/FULL/PDF.