

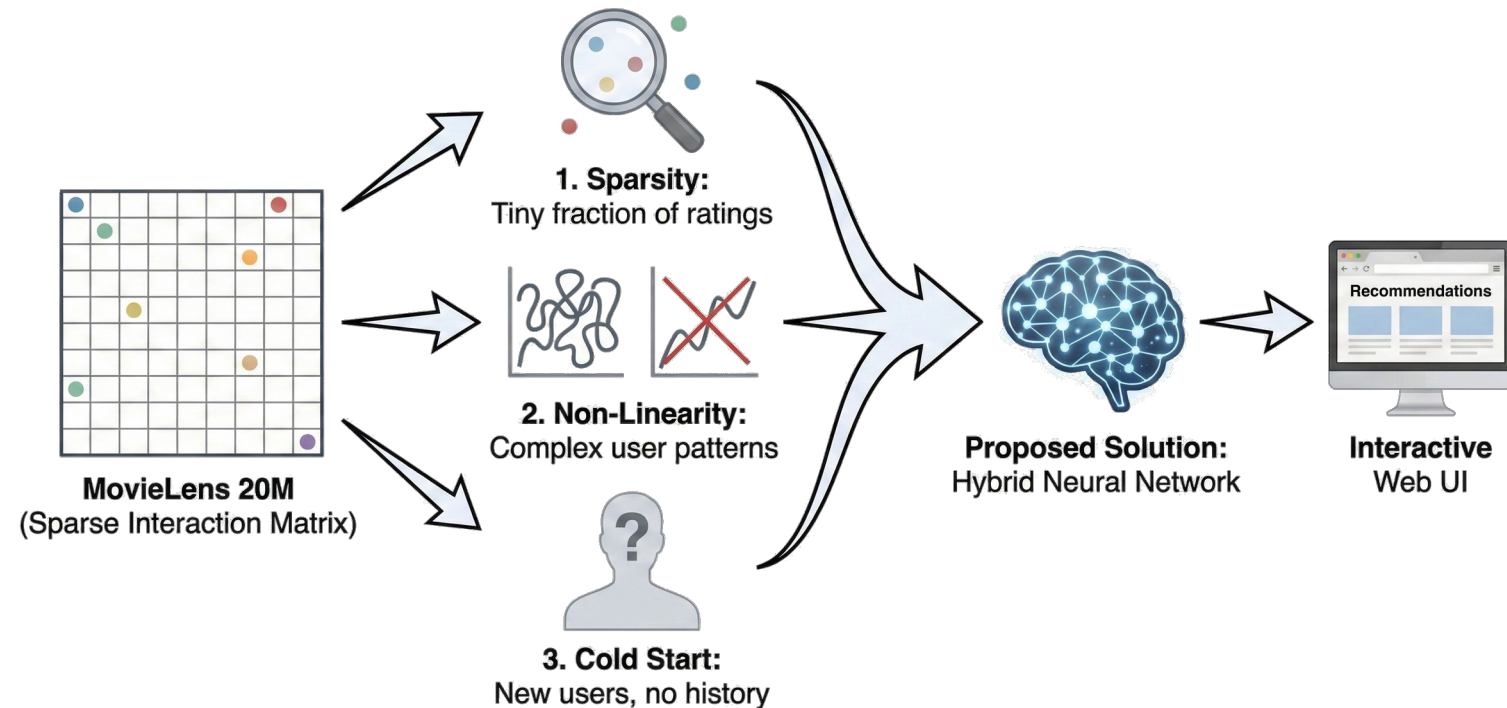


Movie Recommendation

Théodore Billotte, Louis Rollet, Vincent Montero Fontaine
Department of Computer Science and Technology

Introduction & Problem Statement

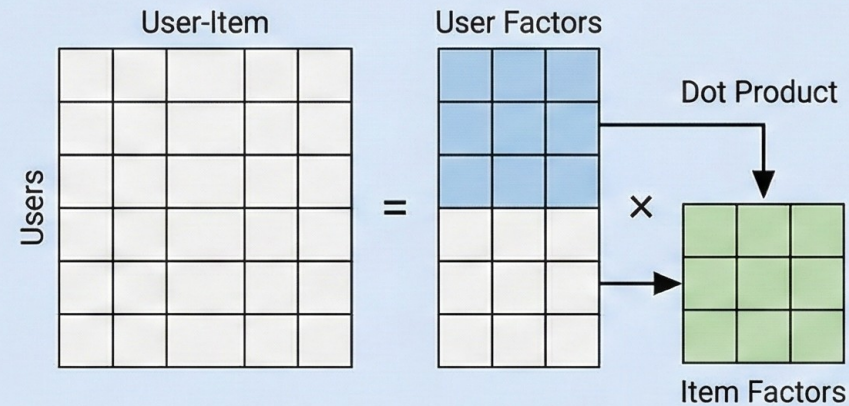
Objective: Build a robust recommendation engine capable of predicting user ratings and solving the "Cold Start" problem.



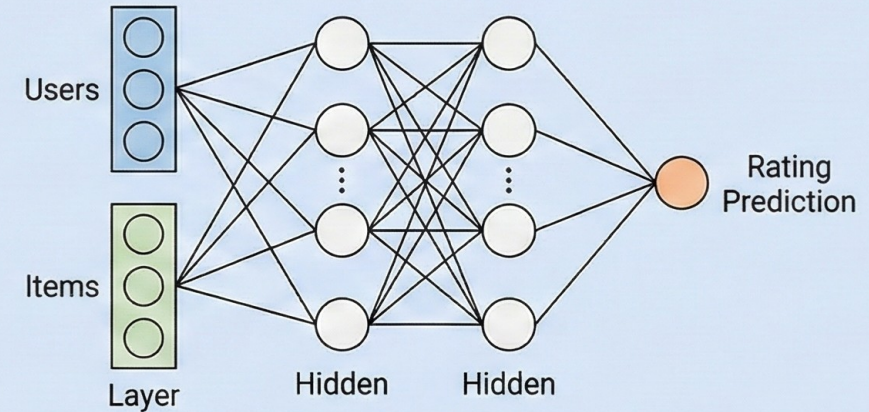
Proposed Solution: A multi-model benchmarking approach culminating in a **Hybrid Neural Network** deployed via an interactive Web UI.

Methodology Overview

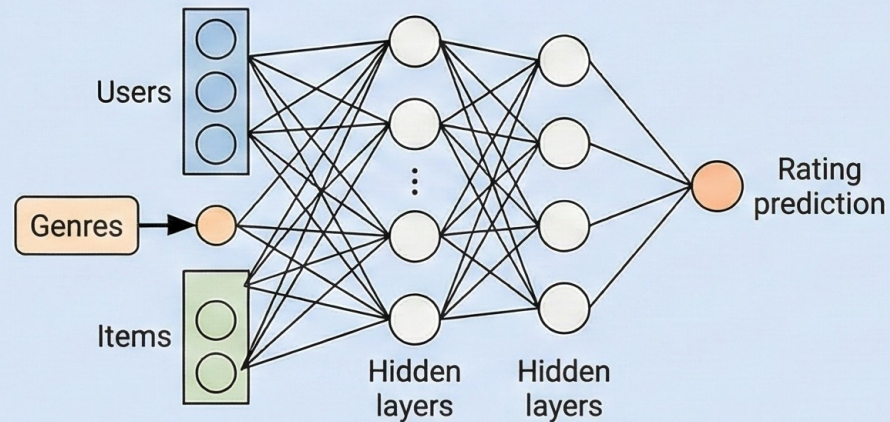
Model 1: Matrix Factorization (Baseline)



Model 2: Neural Collaborative Filtering (NCF)



Model 3: Hybrid NCF (Extension with Genres)



Model 4: LLM Agent



Model 1 - The Baseline (Matrix Factorization)

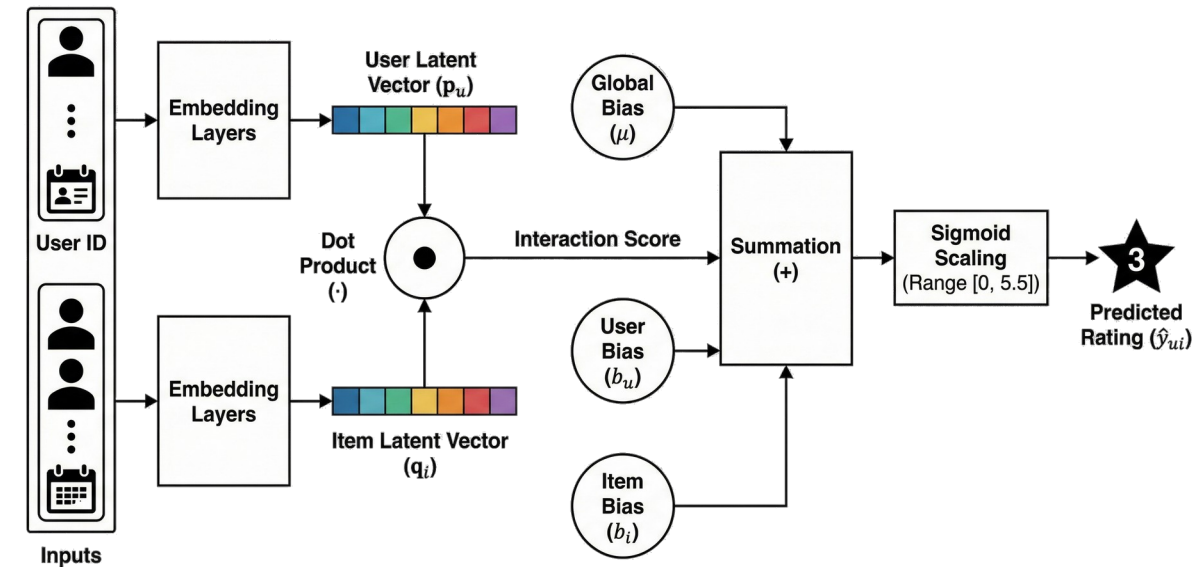
Concept: Decomposes the interaction matrix into latent feature vectors for Users and Items.

Methodology: Standard SVD-style factorization implemented in PyTorch.

$$\hat{y}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^T \mathbf{q}_i$$

Where:

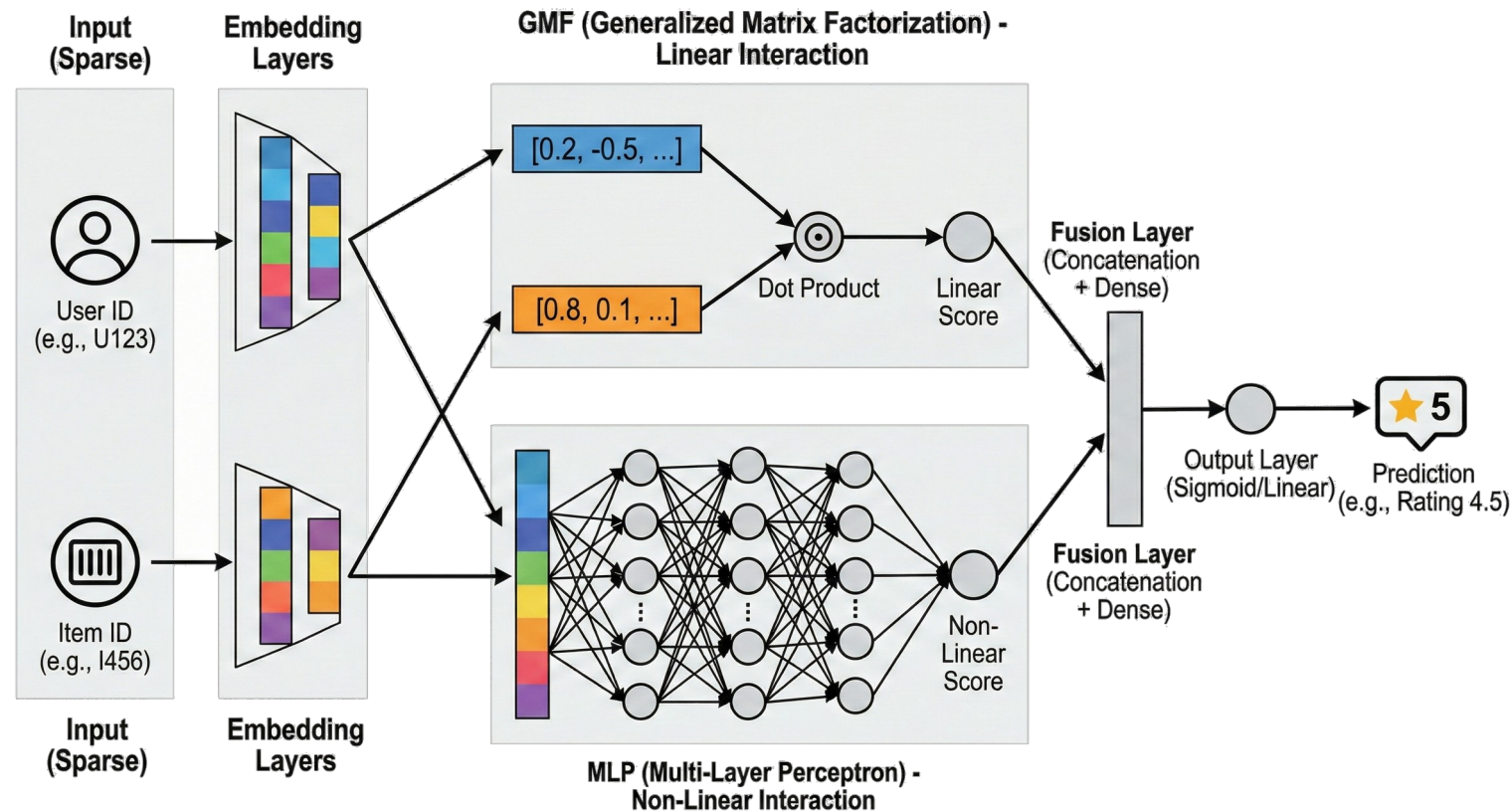
- μ : Global bias
- b_u, b_i : User and Item biases
- $\mathbf{p}_u, \mathbf{q}_i$: Latent vectors for user u and item i



Pros: Fast and interpretable.

Cons: limited linearity (underfits complex patterns).

Model 2 - Neural Collaborative Filtering (NCF)

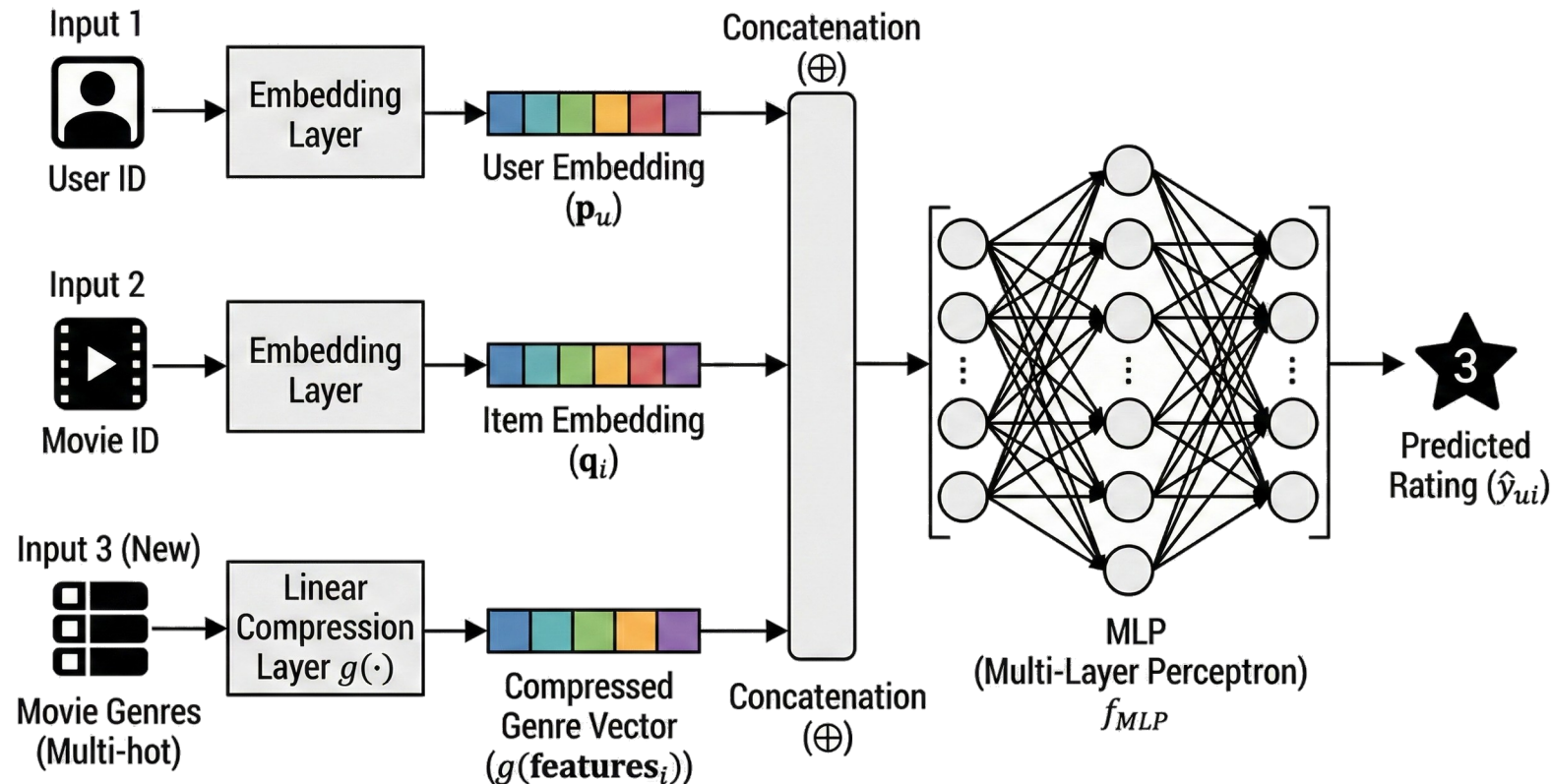


Motivation: Addressing the limitations of the dot product (linear combination) in Matrix Factorization.

Advantage: Significantly lower RMSE than baseline (~ 0.82 vs ~ 1.16) due to higher model capacity.

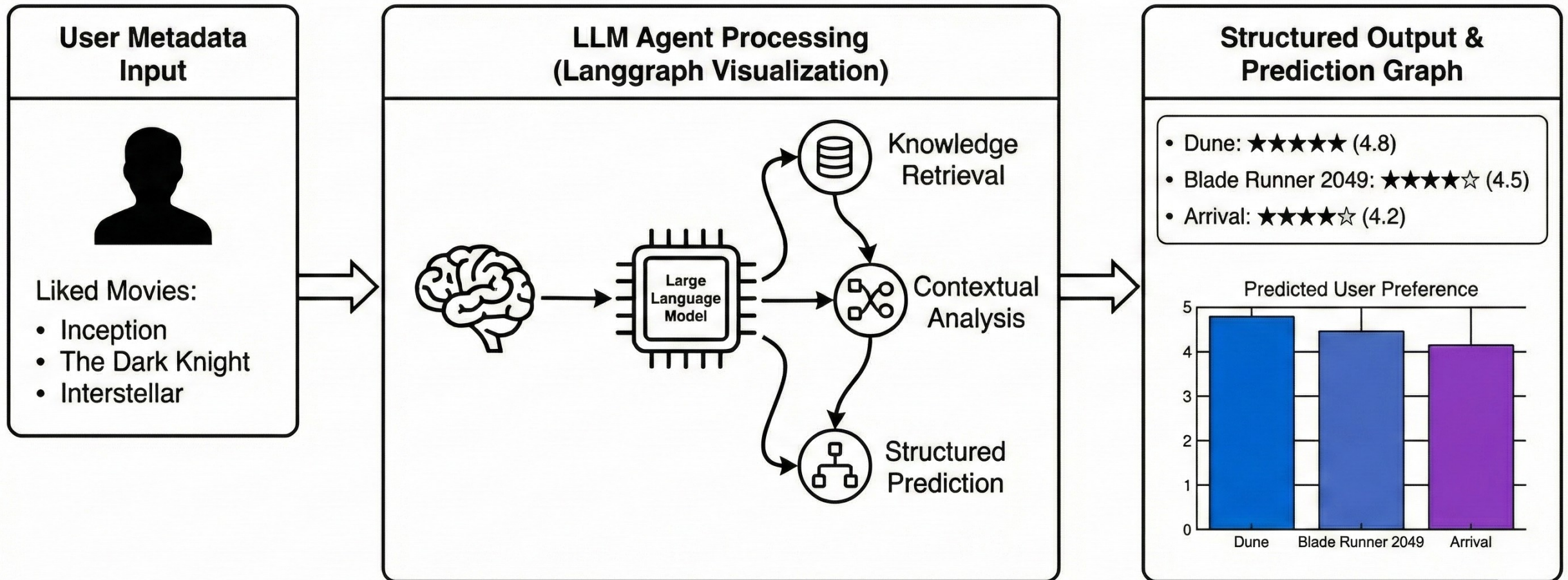
Model 3 - Proposed Hybrid NCF

Innovation: Extending NCF to include Content-Based Features (Genres).



- Allows the model to "understand" item nature, not just ID.
- Essential for interpreting why a recommendation was made (e.g., "User likes Sci-Fi").

Model 4 – LLM Agent



Experimental Setup & Training

Framework & Hyperparameters:

- Framework: PyTorch
- Optimizer: Adam (Adaptive Moment Estimation)
- Loss Function: MSE (Mean Squared Error)
- Batch Size: 1024 (Efficient GPU utilization)
- Epochs: 5 (Early convergence observed)

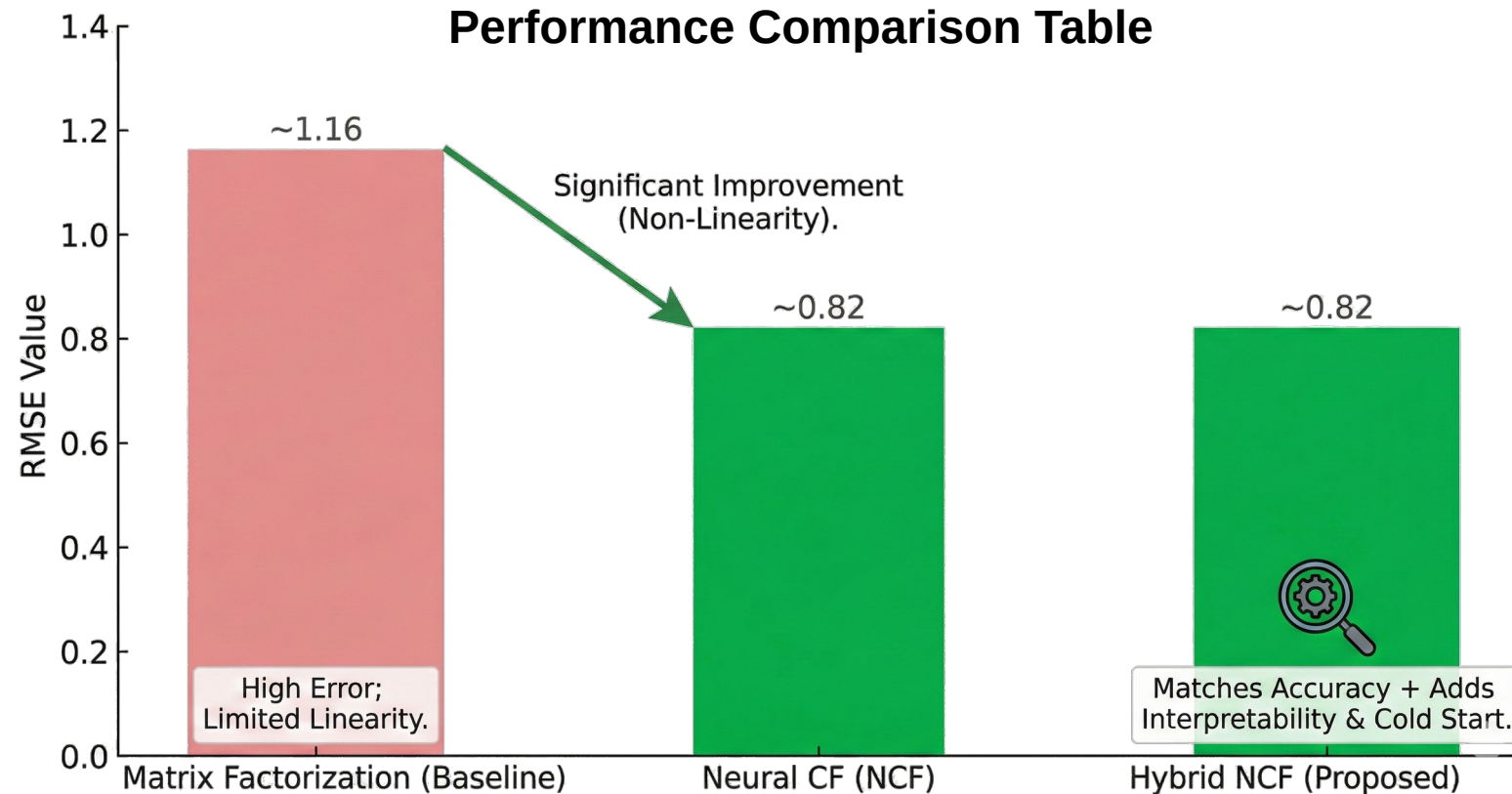


$$\text{MSE} = \overset{\text{Mean}}{\boxed{\frac{1}{n} \sum_{i=1}^n}} \left(\overset{\text{Error}}{Y_i - \hat{Y}_i} \right) \overset{\text{Squared}}{\boxed{^2}}$$

Data & Reproducibility

- Data Splitting: Train/Test split applied to interactions
- Reproducibility: Weights saved as .pth checkpoints in checkpoints/ directory

Results & Evaluation



- Neural models drastically outperform the linear baseline.
- Adding genres (Hybrid) did not lower RMSE further (due to latent factors already capturing genre info implicitly) but enabled Cold Start features.

Application - Interactive Dashboard (Streamlit)



Predicted liked
movies using all
models

User Recommendation Engine

Select Models to Compare

Hybrid x



History for User 1

Detected Tastes: Adventure, Fantasy, Action, Thriller, Horror (Rated 175 movies total)

Lord of the Rings: The Fellowship of the Ring, The (2001)

★ 5.0 | Adventure|Fantasy

Freaks (1932)

★ 5.0 | Crime|Drama|Horror

Lord of the Rings: The Return of the King, The (2003)

★ 5.0 | Action|Adventure|Drama|Fantasy

Lord of the Rings: The Two Towers, The (2002)

★ 5.0 | Adventure|Fantasy

Star Wars: Episode V - The Empire Strikes Back (1980)

★ 4.5 | Action|Adventure|Sci-Fi

Load More (5/175 shown)

Predict Recommendations

Hybrid

4.78 - Harry Potter and the Deathly Hallows: Part 2 (2011)

Action|Adventure|Drama|Fantasy|Mystery|IMAX

4.76 - Mummy, The (1999)

Action|Adventure|Comedy|Fantasy|Horror|Thriller

4.76 - Harry Potter and the Deathly Hallows: Part 1 (2010)

Action|Adventure|Fantasy|IMAX

4.71 - Princess Bride, The (1987)

Action|Adventure|Comedy|Fantasy|Romance

4.66 - Rock, The (1996)

Action|Adventure|Thriller

NCF

4.31 - Matrix, The (1999)

Action|Sci-Fi|Thriller

4.25 - Gladiator (2000)

Action|Adventure|Drama

4.25 - Dark Knight Rises, The (2012)

Action|Adventure|Crime|IMAX

4.24 - Harry Potter and the Deathly Hallows: Part 2 (2011)

Action|Adventure|Drama|Fantasy|Mystery|IMAX

4.23 - Life Is Beautiful (La Vita è bella) (1997)

Comedy|Drama|Romance|War

MF

4.44 - Star Wars: Episode VI - Return of the Jedi (1983)

Action|Adventure|Sci-Fi

4.42 - Matrix, The (1999)

Action|Sci-Fi|Thriller

4.40 - Dark Knight, The (2008)

Action|Crime|Drama|IMAX

4.39 - Schindler's List (1993)

Drama|War

4.33 - Inception (2010)

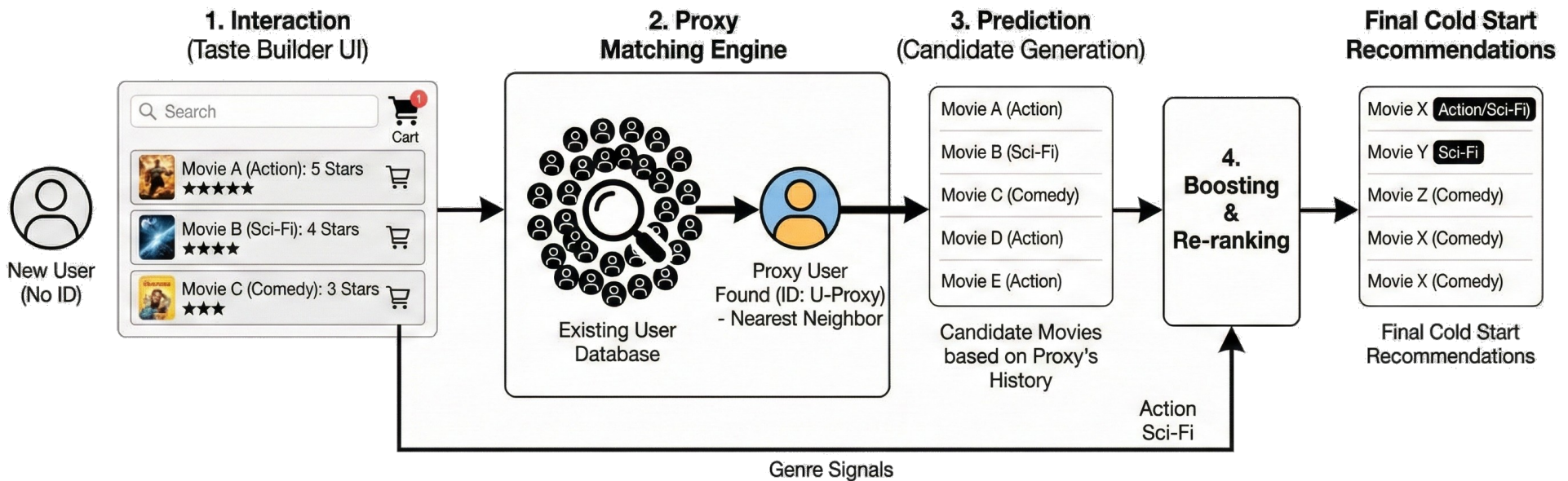
Action|Crime|Drama|Mystery|Sci-Fi|Thriller|IMAX

Choose a User

Solving the "Cold Start" Problem

The Challenge: How to recommend movies to a new user with no ID or history?

Our Taste Builder approach



Conclusion & Future Work

Summary:

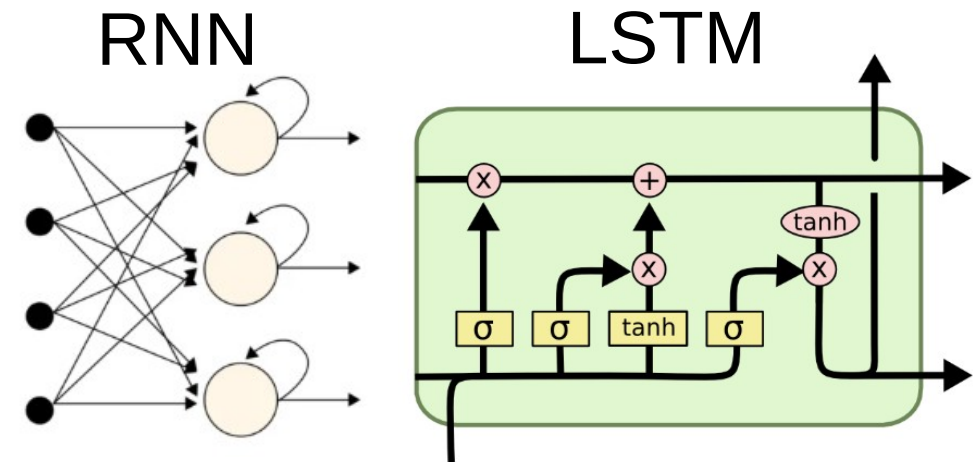
- Successfully benchmarked three architectures.
- Demonstrated the superiority of Neural approaches over Matrix Factorization.
- Built a production-ready UI bridging mathematical embeddings with user needs.

Limitations:

- Dataset size (20M) requires significant RAM.
- Hybrid model complexity increases inference time slightly.

Future Improvements

- Integrate Temporal Dynamics (Time-based RNNs/LSTMs).
- Use Auto-Encoders for better feature compression.
- Deploy model via Docker/Cloud.





Thank you!

Théodore Billotte, Louis Rollet, Vincent Montero-Fontaine
Department of Computer Science and Technology