From Stranger Things to Your Favorite Things:
Netflix's Recommendation Evolution

Mark (Ko-Jen) Hsiao
Workshop on Large-Scale Video Recommendation Systems @ RecSys 2023
09/18/2023

NETFLIX

Agenda

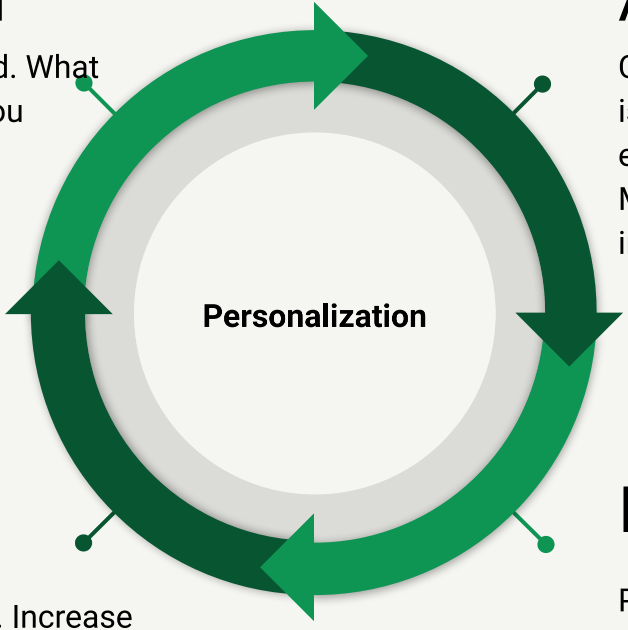
- Overview of Netflix Recommendations
- Sharing of 4 aspects/lessons
- Conclusion

Unification

Objective and Reward. What you design is what you learn and have.

Adaptation

Catalog is changing. Members is changing. Product and external trends are changing. Model adaptation and in-session signal.



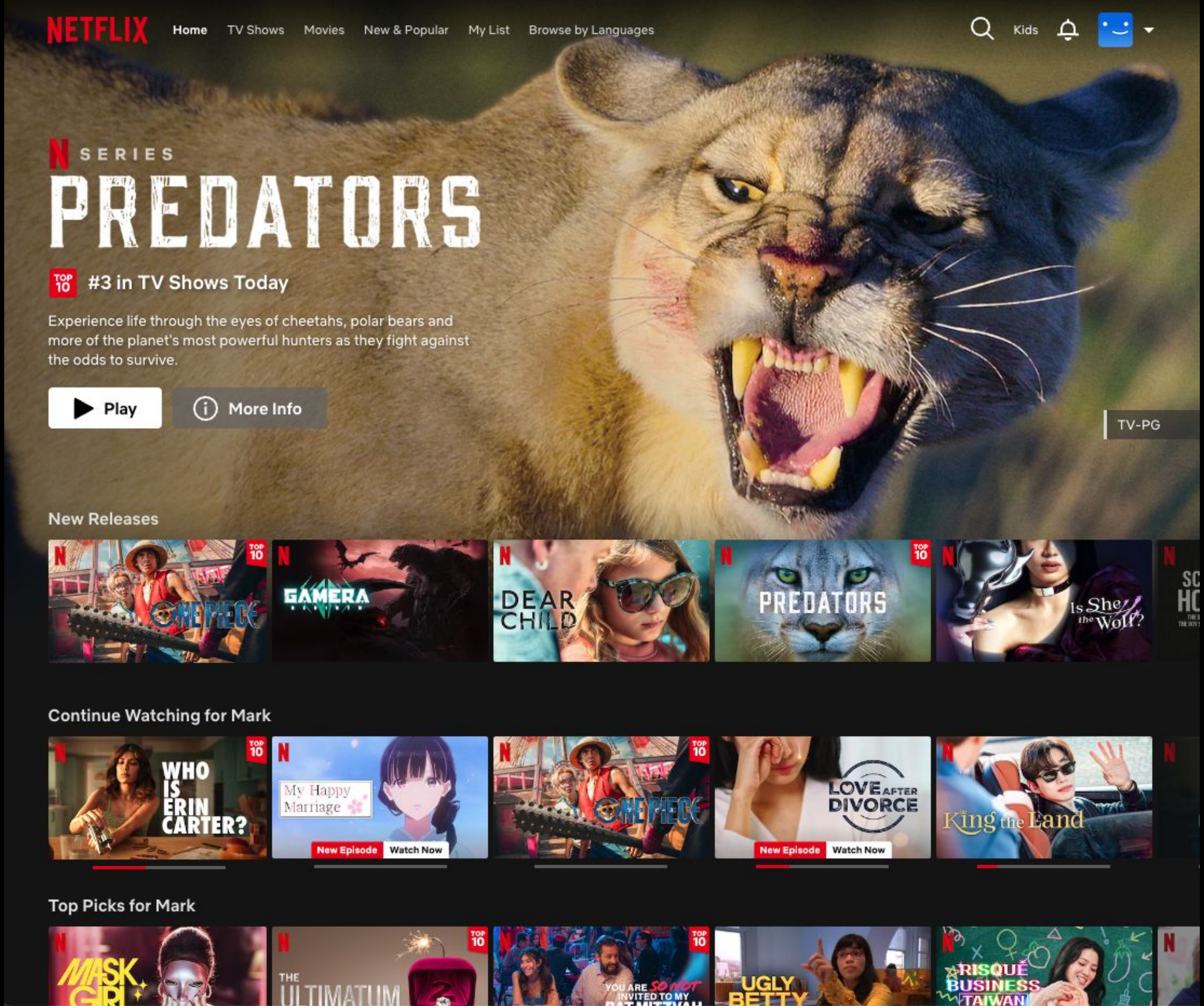
Simplicity

Model Consolidation. Increase flexibility and extensibility of recommendation systems

Innovation

Personalization power-up. Foundation model and beyond.

Everything is a recommendation!



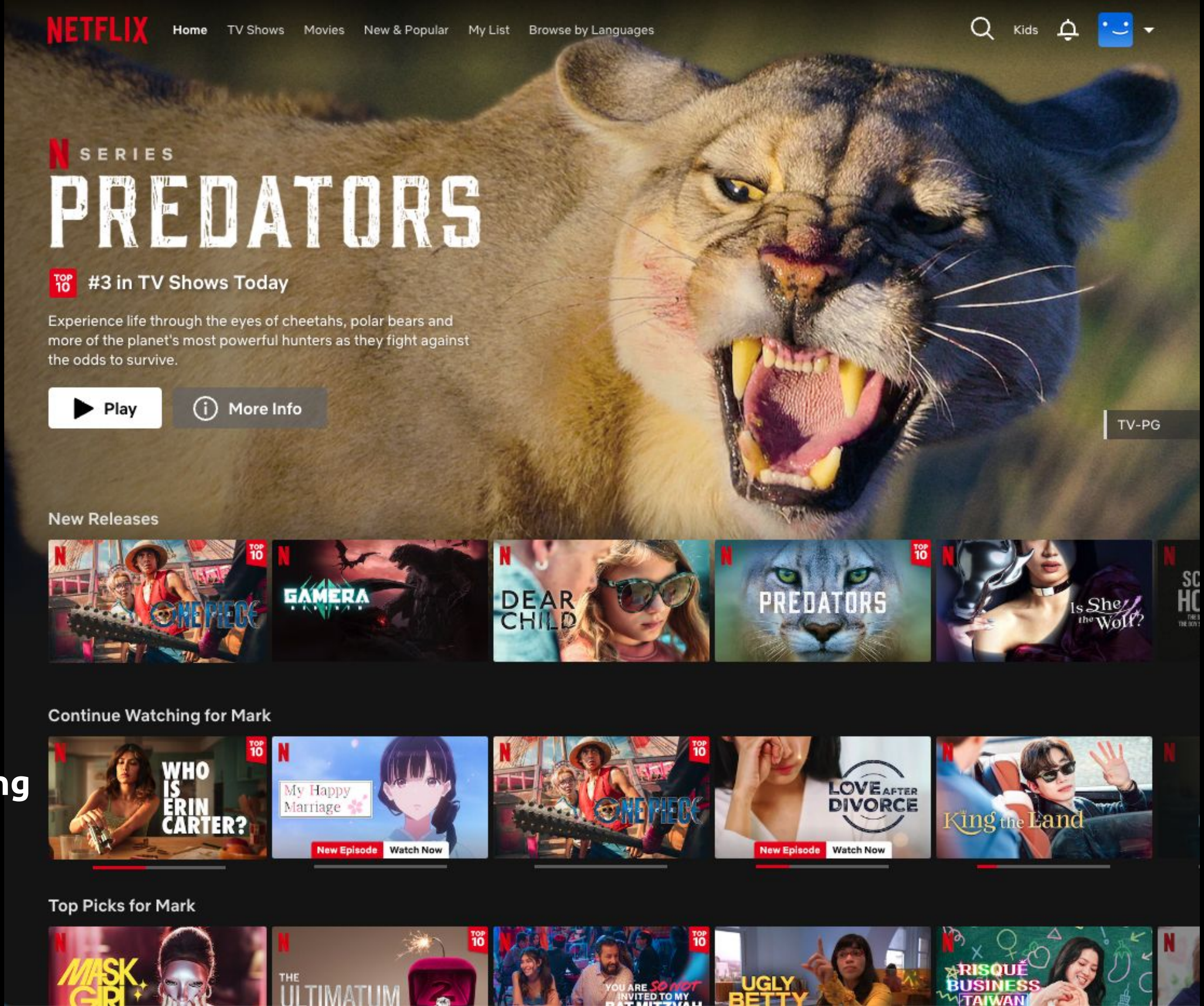
NETFLIX
Home
TV Shows
Movies
New & Popular
My List
Browse by Languages
Kids
SERIES
PREDATORS
TOP 10
#3 in TV Shows Today
Experience life through the eyes of cheetahs, polar bears and more of the planet's most powerful hunters as they fight against the odds to survive.
Play
More Info
TV-PG
New Releases
ONE PIECE
GAMERA
DEAR CHILD
PREDATORS
Is She the Wolf?
Continue Watching for Mark
WHO IS ERIN CARTER?
My Happy Marriage
New Episode
Watch Now
ONE PIECE
LOVE AFTER DIVORCE
New Episode
Watch Now
King the Land
Top Picks for Mark
MASK GIRL
THE ULTIMATUM
YOU ARE SO NOT INVITED TO MY BAT MITZVAH
UGLY BETTY
RISQUÉ BUSINESS TAIWAN

Billboard

New Releases

Continue Watching

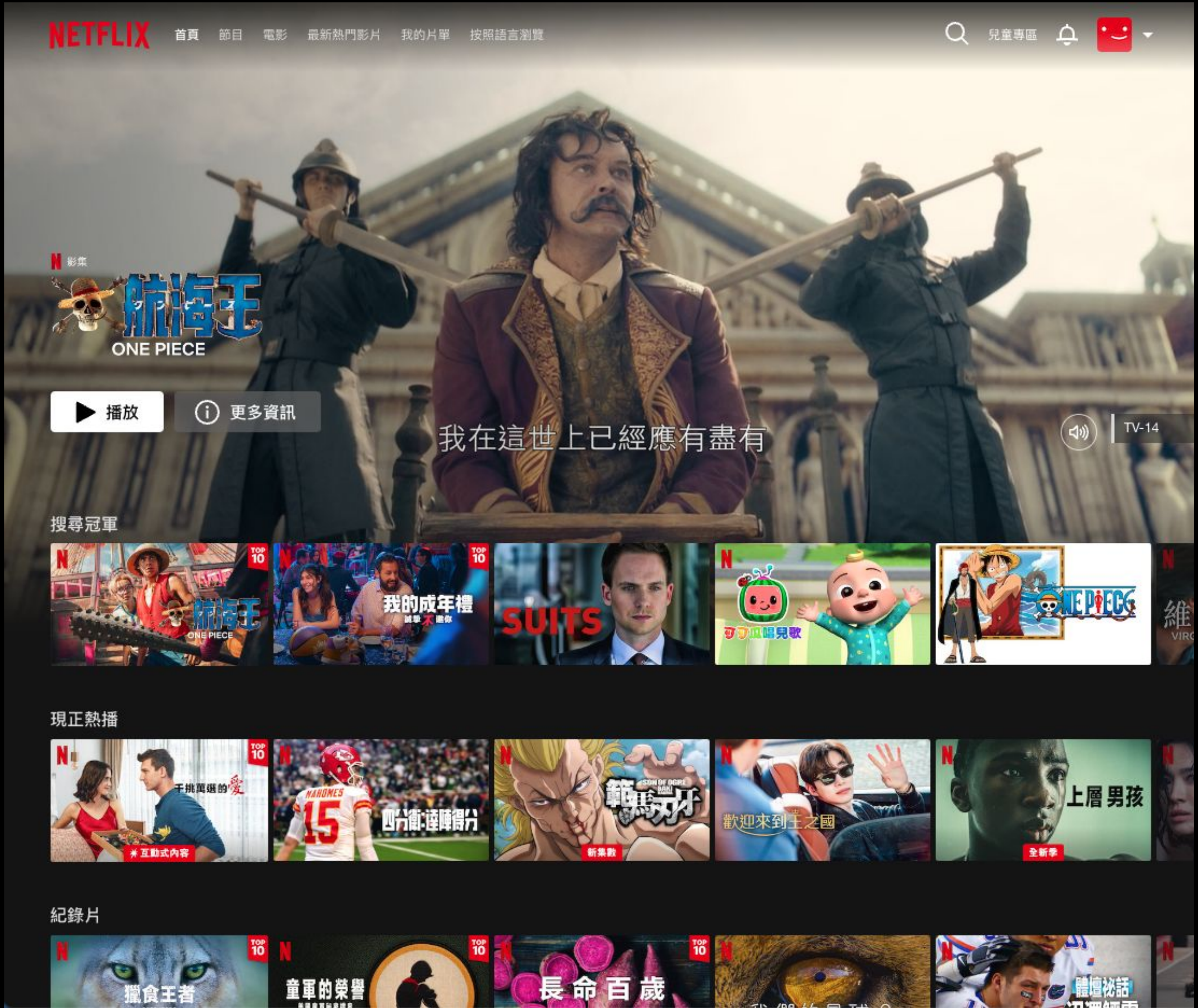
Top Picks



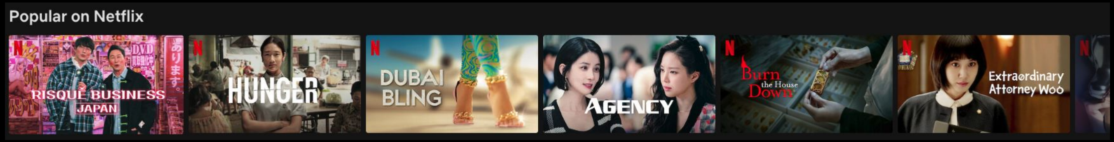
Top searches

Trending Now

Documentaries

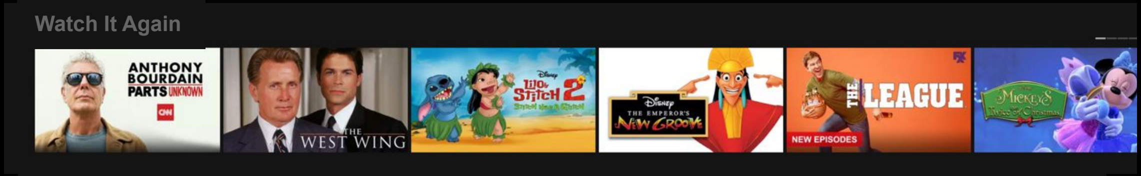


Popular on Netflix

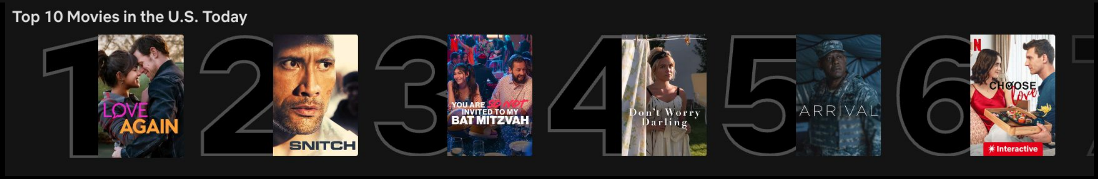


My List

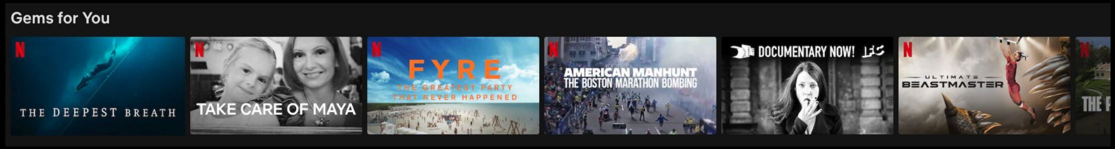
Watch It Again



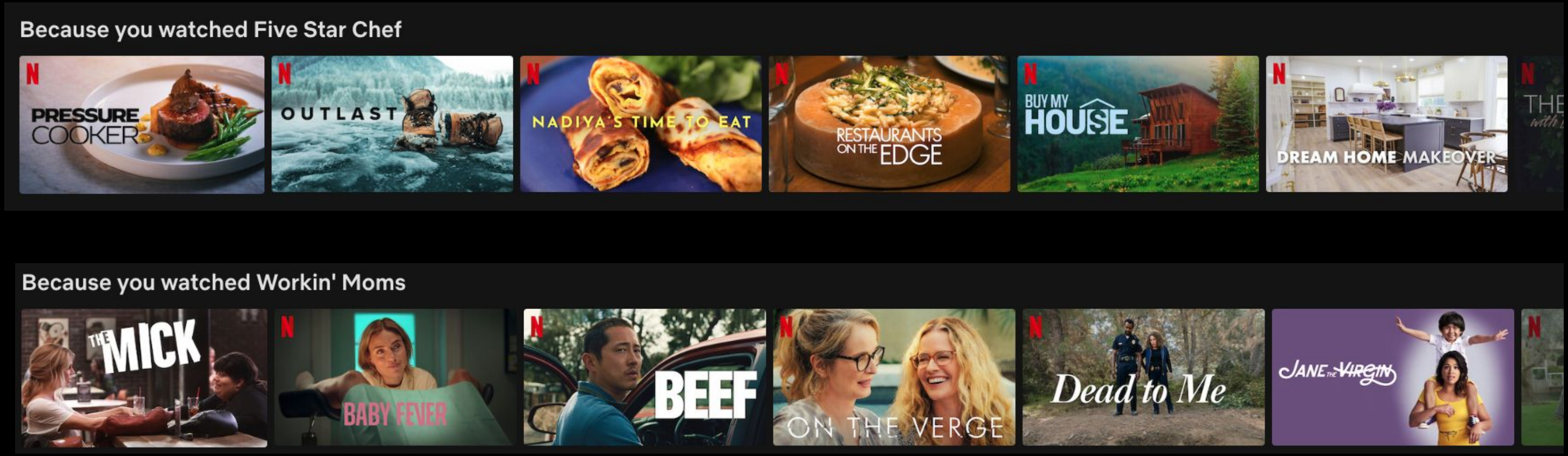
Top 10



Gems for You

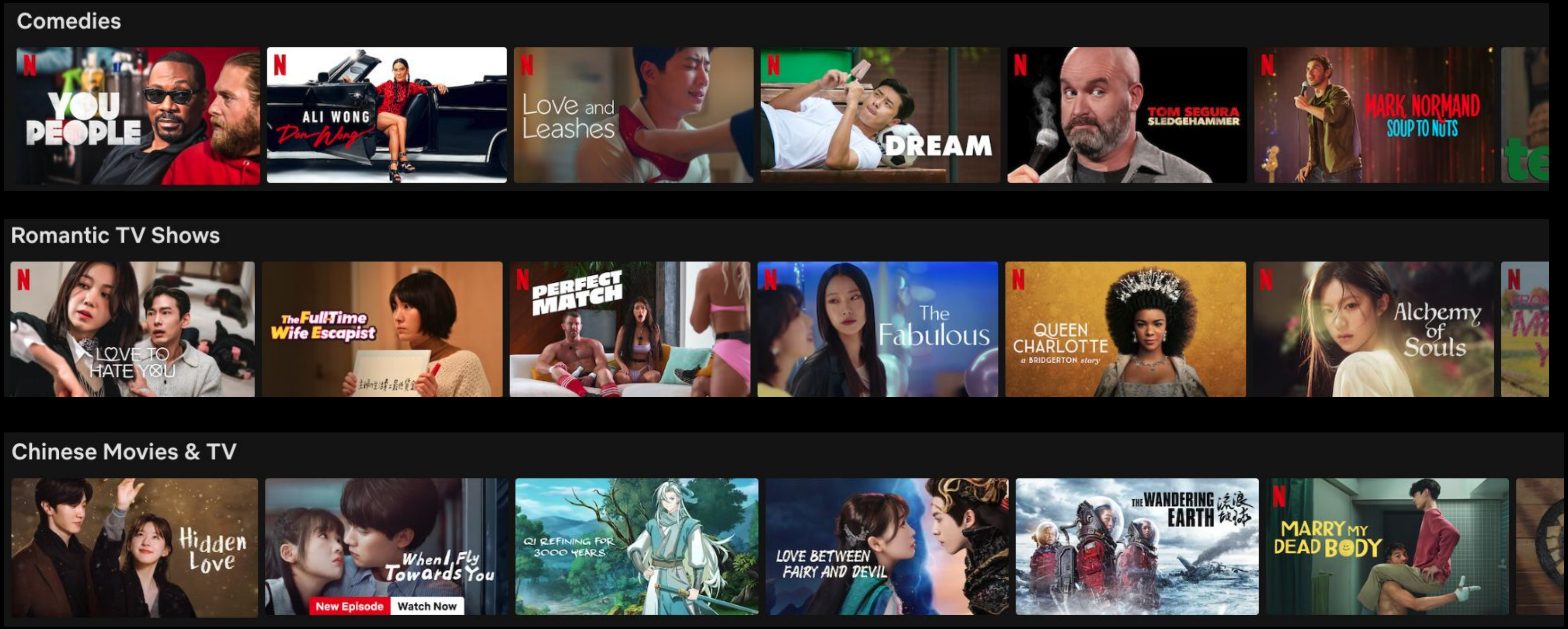


Because you watched [...]

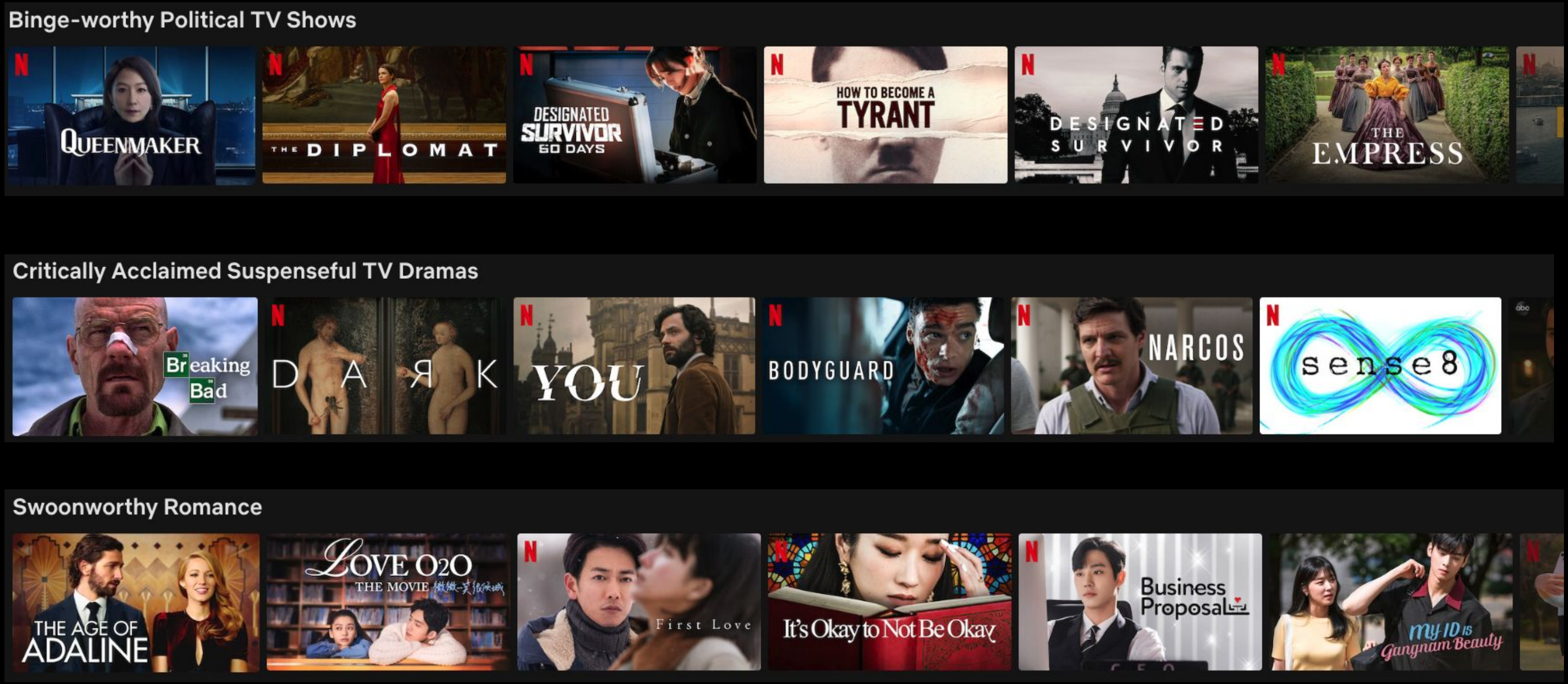


Anything you've "enjoyed" could be an "inspiration" of a new row

All different kinds of basic genres

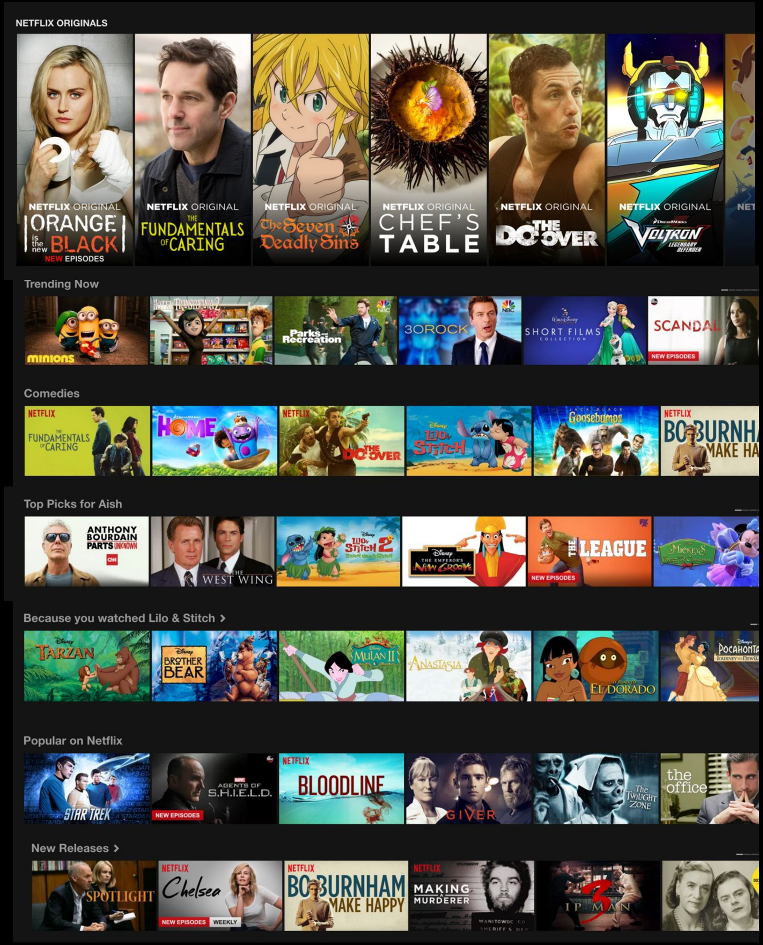


Or more specific genres



Many many more application-specific video rankers…

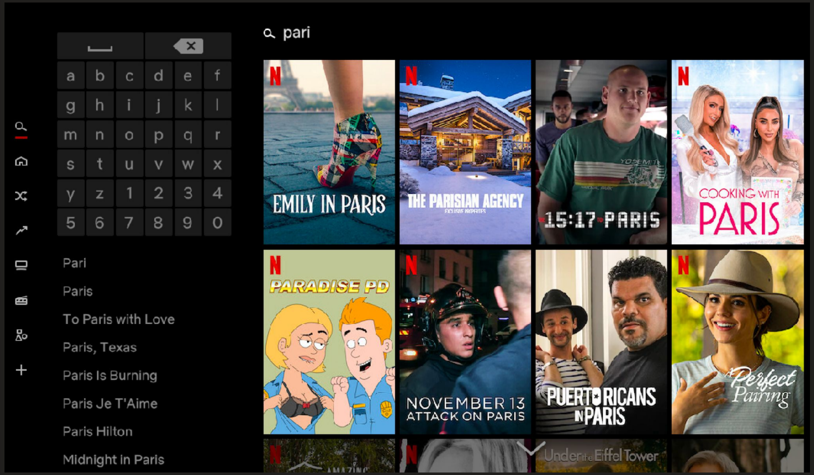
- Screen Saver
- Rows in Popular and Hot page: Coming Later, Coming Soon.
- Recommended titles after plays.
- Pre-launch
- etc…



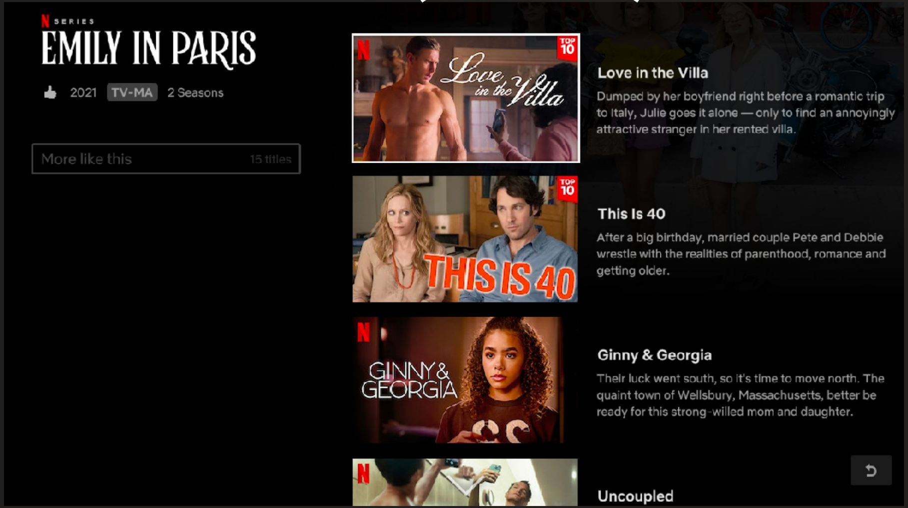
Video Recommendation is not just personalized ranking of videos.

Row Ranking

Search (query -> titles)



More like this (title -> title)



Kids Characters

Everything is a recommendation!

Help members find the next title they will **enjoy** to maximize member **joy** and **satisfaction**.

Challenges

- Isn’t it a solved problems?

NETFLIX

Challenges

- Isn’t it a solved problems? Nope.
- Netflix’s video recommendation problem is quite different from ANY user-generated video recommendations.
- Smaller catalog and longer form videos, and games.
- Every Netflix-and-Chill moment is important!

Lesson 1: Objective and Reward

What you design is what you learn and have

Objectives

- Help users discover their next plays.
  - Video-level
  - Row-level
  - Page-level
- Row names make huge differences.
  - Ex: Gems for you. What should be the objective of you ML models?
- Many sessions are just watching the next episode. (Continue Watching)
  - Higher is better? Nope.
  - How to rank watched titles.

NETFLIX

Objectives

- How you define objectives of your ML models determine what/how you gonna recommend titles to members.
  - Labels: discovery plays, continue watching plays, thumbs up/down, My List add, etc.
  - Stratification: country and tenure.
- Positive vs Negative.
- Online/Offline alignments.

Objectives

Complicated ML models

v.s

Right Objectives

Which to invest more on?

Objectives

Complicated ML models

V.S

Right Objectives

Which to invest more on?

Objectives

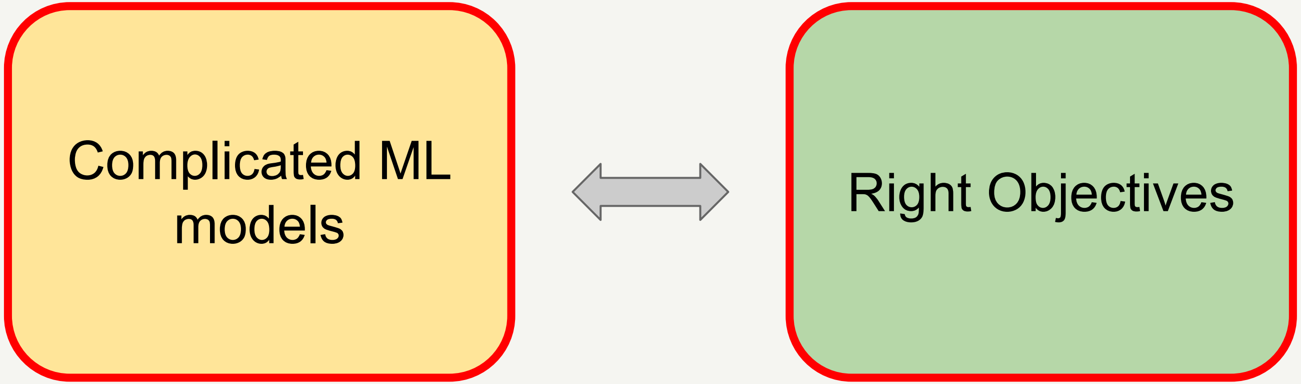
Complicated ML models

V.S

Right Objectives

Which to invest more on?

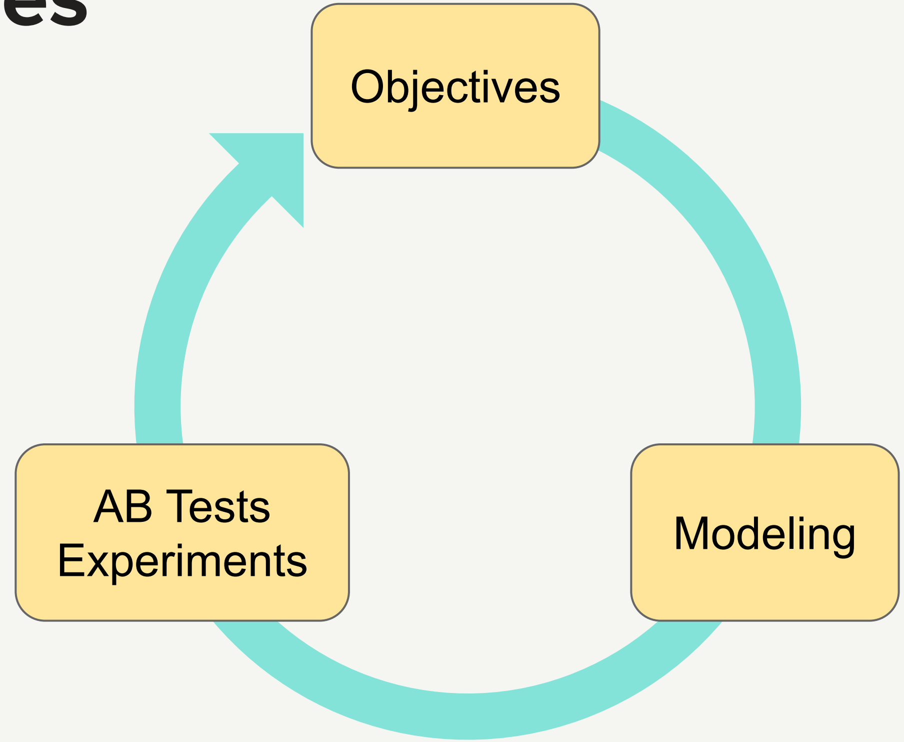
Objectives



As you increase your ML model complexity, you likely need to also increase your objective alignment or else it can easily start learning the wrong thing.

Ref: Deep Learning for Recommender Systems: A Netflix Case Study, Harald Steck, etl

Objectives



Reward

- Short-term v.s. long-term rewards*.
- How to define different rewards?
  - North star → Increase member satisfaction.
- Retention
  - Golden metric.
  - Harder to measure within a short period of time.
  - Approximation.

* RecSys 2023 Session 17: Interactive Recommendation 2, Reward innovation for long term member satisfaction, Gary Tang, JiangWei Pan, etl (Netflix)

Lesson 2: Model Consolidation

Increase flexibility and extensibility of recommendation systems

Tons of Different Models

- Different objectives/models for different canvases.
- Innovation overhead.
- Inconsistent and suboptimal experiences.

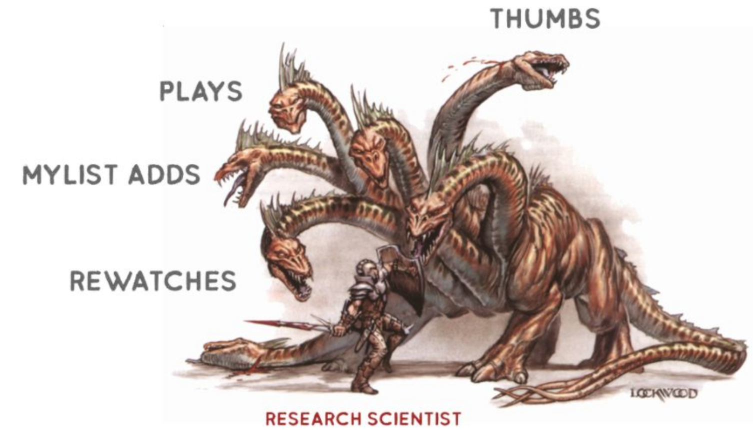
Model Consolidation

- Reducing maintenance cost to allow operating with a small team
    - Cost efficiency.
    - Fewer models.

- Innovation applied to one canvas/model can be automatically extended to other canvases/models.

Approach 1 - Multi-task Learning

- Single model to predict members' next plays, thumb ups, mylist add.



Hydranet model

NETFLIX

Why multi-task learning (MTL)?

- **Optimize user satisfaction**

  User satisfaction is high-dimensional which can be captured by various user behaviors.

- **Transfer knowledge across different objectives**

  Improve overall performance. Ex: |play data| >> |mylist data|

- **Scale the ranker -** Apply different heads to different rows and applications
- **A general framework** - Easily incorporate other tasks into the model

Approach 2 - Model Sharing

- One unified contextual recommender for Search and Recommendation. (UniCorn :))
- The model is trained on datasets from different member-driven discovery canvases.
- Different canvases uses different “contexts” which means they have different “inputs” into the model.

NETFLIX

Approach 2 - Model Sharing

Dataset 1
Query-based searches

Dataset 2
Title-based recs

Dataset 3
Entity-based searches

Model

NETFLIX

Approach 3

- Flexible and unified training pipeline/framework that could produce models of different objectives easily.

    1. Flexible label generation framework → Label extractor registry.
    2. Reward generation framework. → Share and reuse rewards across models and teams. *
    3. Flexible feature definitions.
    4. Flexible model definitions.

- Engineering is important!

* RecSys 2023 Session 17: Interactive Recommendation 2, Reward innovation for long term member satisfaction, Gary Tang (Netflix), etl

NETFLIX

Lesson 3: Everything is Changing

Catalog is changing. People is changing. Everything is changing.
Model adaptation, incremental learning and in-session signals.

Everything is changing!

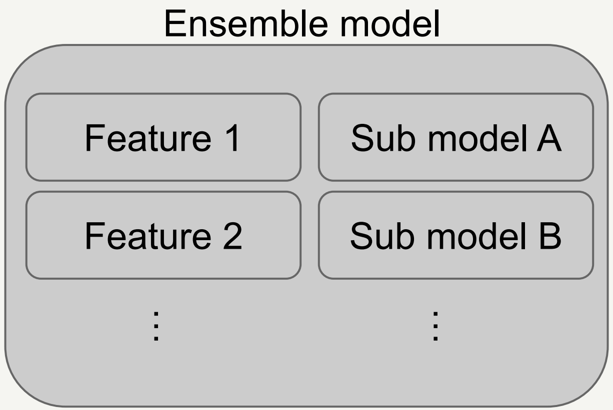
Case study 1:

- Netflix originally started service in US and gradually became available in more and more countries.
- Great recommendations are essential in any new market, but they're also harder to achieve due to a lack of data.

NETFLIX

Model Adaptation

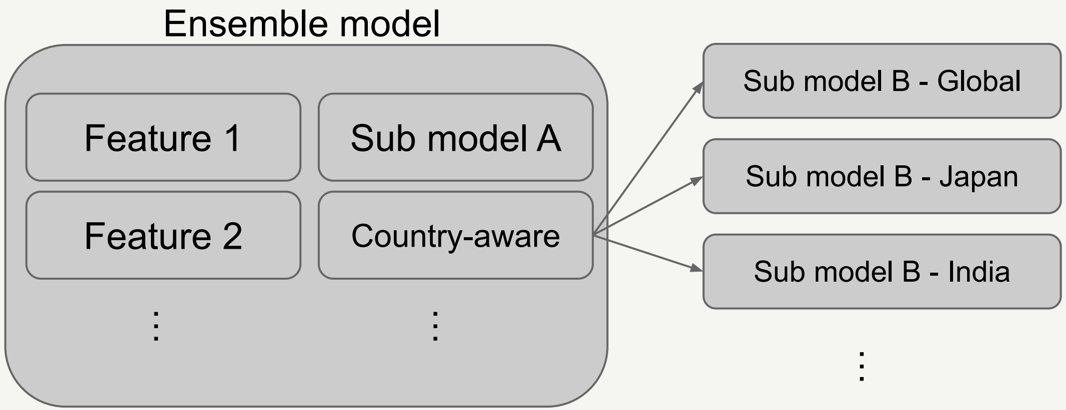
- Find out in which countries our global recommendation model is not working well and why.
- Warm-start from global model and fine-tune on country-specific data.
- Context-aware (in this case, country-aware) ensemble models know which submodels/features to use for different countries.



NETFLIX

Model Adaptation

- Find out in which countries our global recommendation model is not working well and why.
- Warm-start from global model and fine-tune on country-specific data.
- Context-aware (in this case, country-aware) ensemble models know which submodels/features to use for different countries.



Everything is changing!

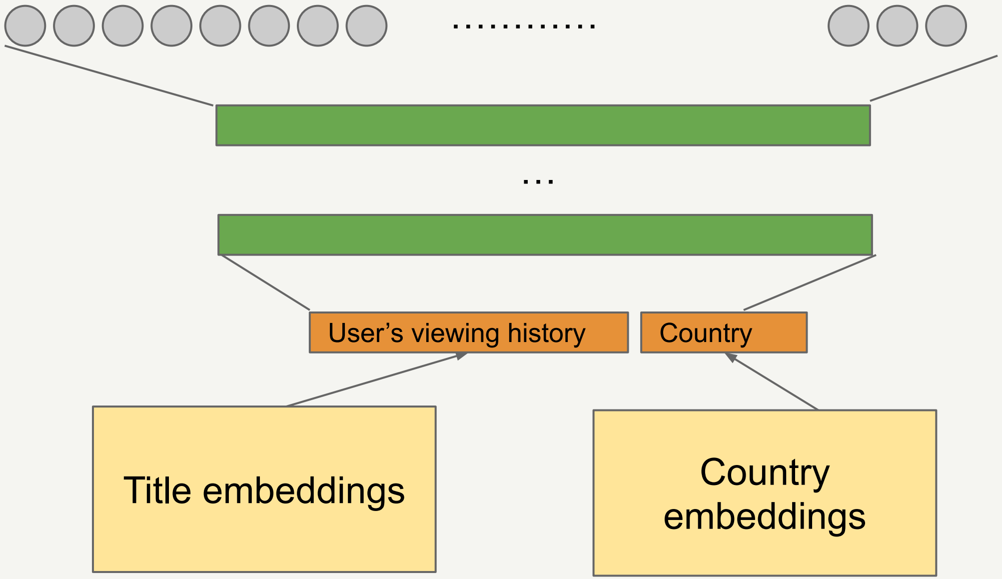
Case Study 2:

- New titles are launched EVERY week.
- How to better cold-start newly launched titles is extremely important.
- Metadata and all different kinds of engineered features are important but not enough.

NETFLIX

Incremental Training

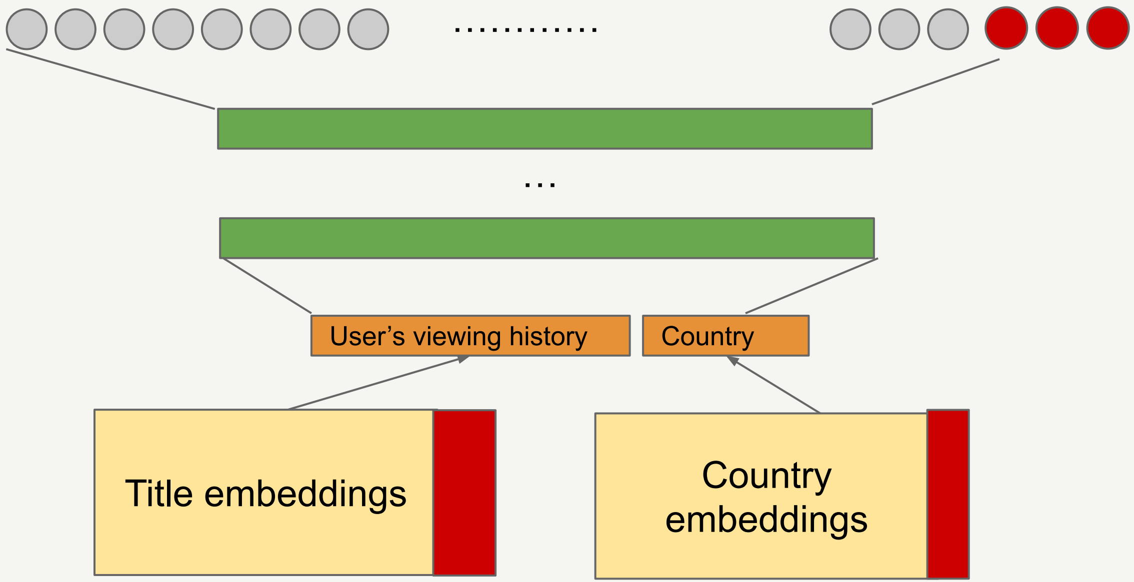
- Take a simple NN model as an example.



NETFLIX

Incremental Training

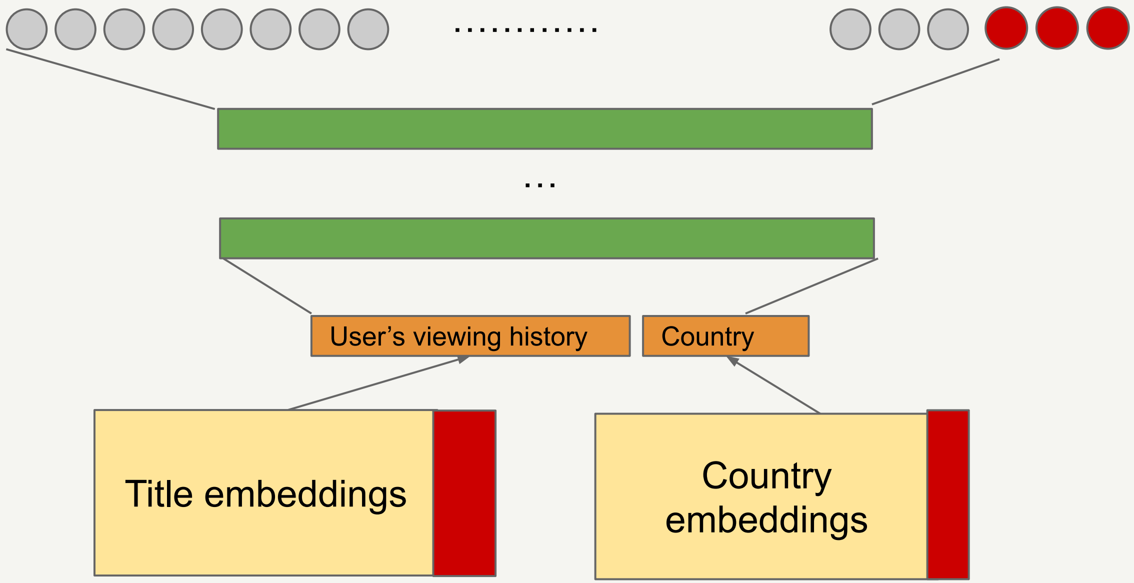
- When new things are coming, warm-start model and train on new data.



NETFLIX

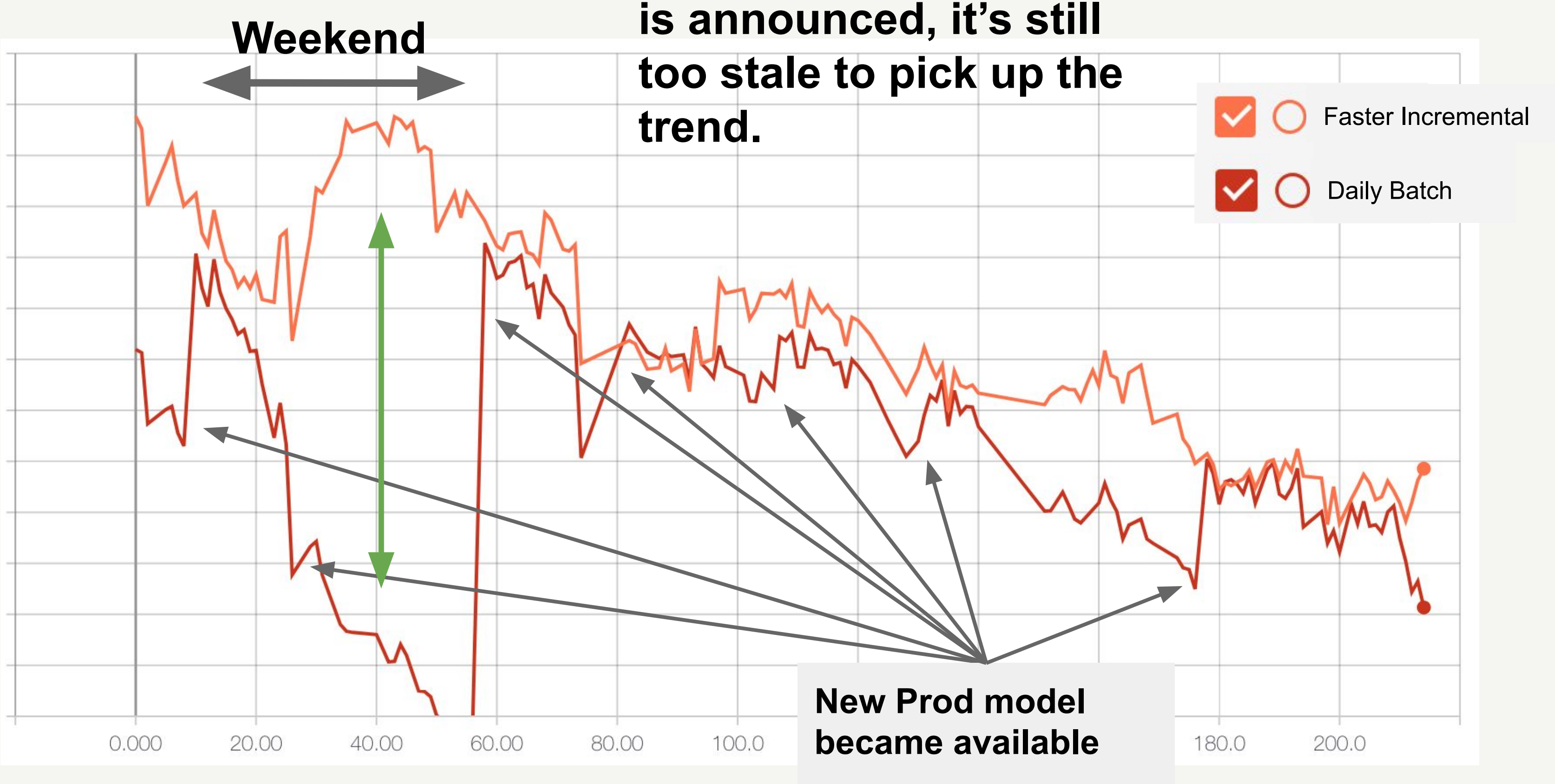
Incremental Training

- **Faster** incremental training **improves** model accuracy.



NETFLIX

Even a new prod model is announced, it’s still too stale to pick up the trend.



Everything is changing!

Case Study 3:

- Members’ intent might even change while browsing Netflix homepage.
- After watching some horror movies, you might wanna watch something relaxing (or not!?)
- Members just thumb up some titles or stay focus longer on some movie trailers, the recommendation should adapt to what they need at the moment.

Use in-session signals for adaptive recommendations

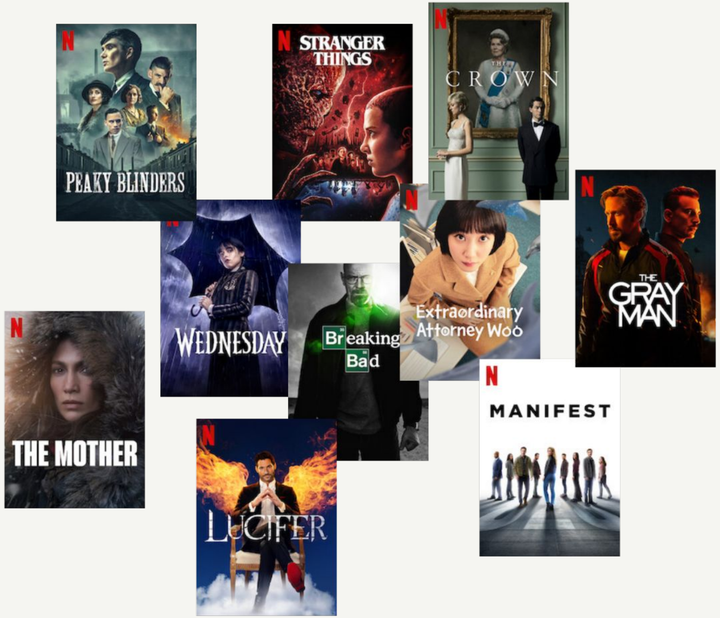
NETFLIX

Lesson 4: Personalization Power-up

Foundation models and beyond.

Netflix, do you remember the niche titles I played a while ago?

Not really, I only remember



NETFLIX

How can Netflix improve its movie recommendations?

Netflix already has a robust recommendation system, designed to cater to individual user preferences. Yet, there's always room for enhancement, particularly considering the dynamic nature of user behaviors and preferences. Here are some strategies for further improvement:

1. **Deep User Profiling**: Understand not just what users watch, but also when they watch, how long they watch, which shows they binge-watch, and which ones they leave midway. This could also include an understanding of user preferences across different genres, languages, and formats (movies, documentaries, series, etc.).
2. **Contextual Understanding**: If Netflix can gather data about why users watch certain content (e.g., relaxing, learning, background noise, watching with kids), they can make context-aware recommendations.

How can we improve further?

- Inspiration from effectiveness of LLMs.
- What could it be like in recommendation?

How can we improve further?

- Inspiration from effectiveness of LLMs.
- What could it be like in recommendation?

- A Foundation Model that learns
    1. Users' long term preferences
    2. Long tail entity representations

NETFLIX

The Similarities

Title ⟺ Word

Data

- One-hot representation with similar vocabulary size O(10K) *
- Power-law (Zipf's law) distributed
- Sequential in nature (sentences vs interaction trajectories)

Learning objective

- Self-supervised fashion. Given history as context, predict the next word/video
- Fine-tuning capability.

NETFLIX

* Language Models are Unsupervised Multitask Learners, Radford, etl, OpenAI (GPT2)

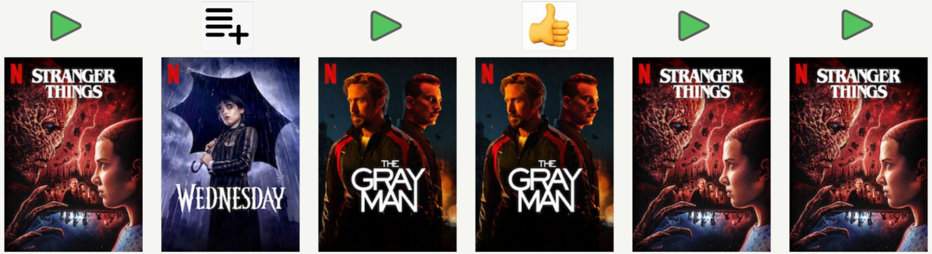
Title ⟷ Word

The Differences

Words and titles are **fundamentally different**

- Language: more structured (syntax/grammar)
- Both are power-law distributed, yet…
  - Head:
    - Words → “stop words” (the/this/that/am/is/are/etc.)
    - Titles → “global hits” (Stranger Things/Squid Game/Wednesday/etc.)
  - Tail:
    - Words → The nuances/subtlety of a language
    - Titles → The niche (not necessarily bad)
- Cold-starting problem, Popularity, etc

Data Preparation



- User interaction sequence vs. NLP sentences
  - “Tokenization” is important
  - Lower SNR
  - Heterogeneous interactions: importance, data volume

- How do we represent an interaction
  - One interaction can be simultaneously on a show, a row, a genre type etc
  - Interaction context: time, duration, language, device etc

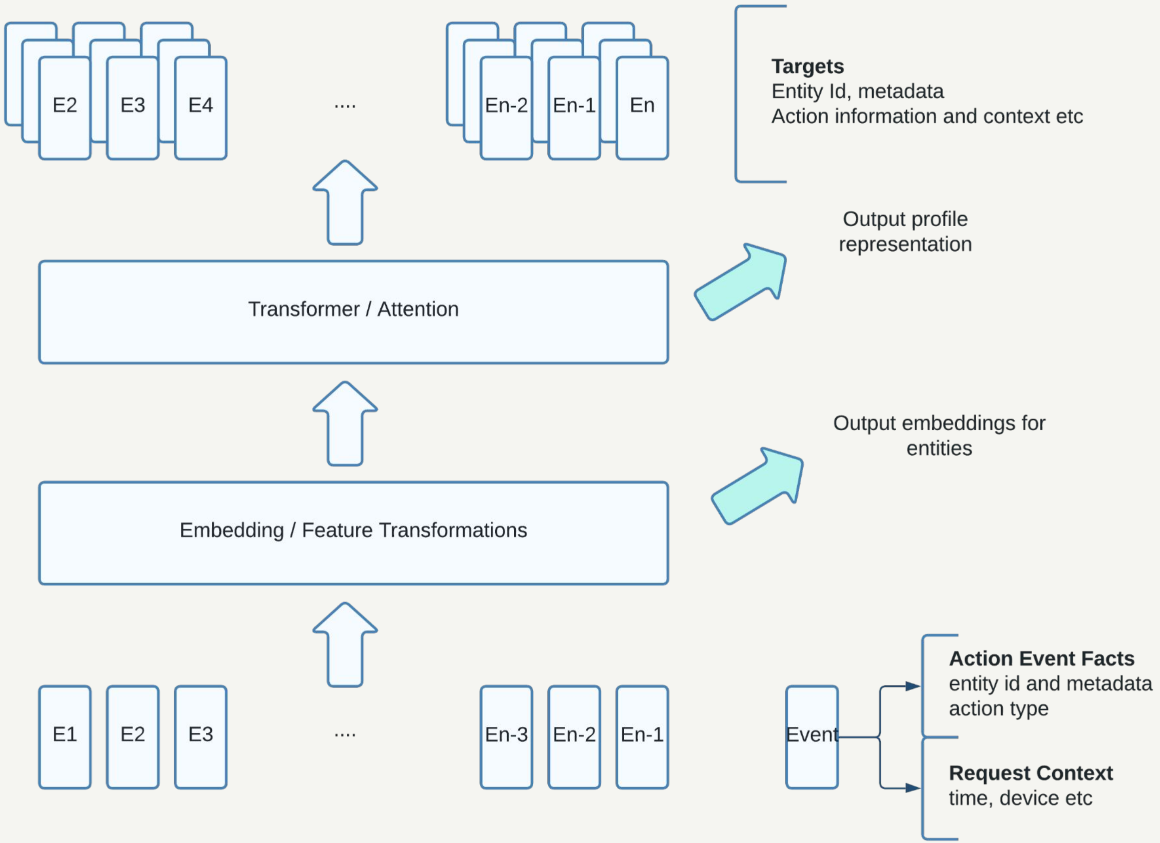
NETFLIX

Foundation Model

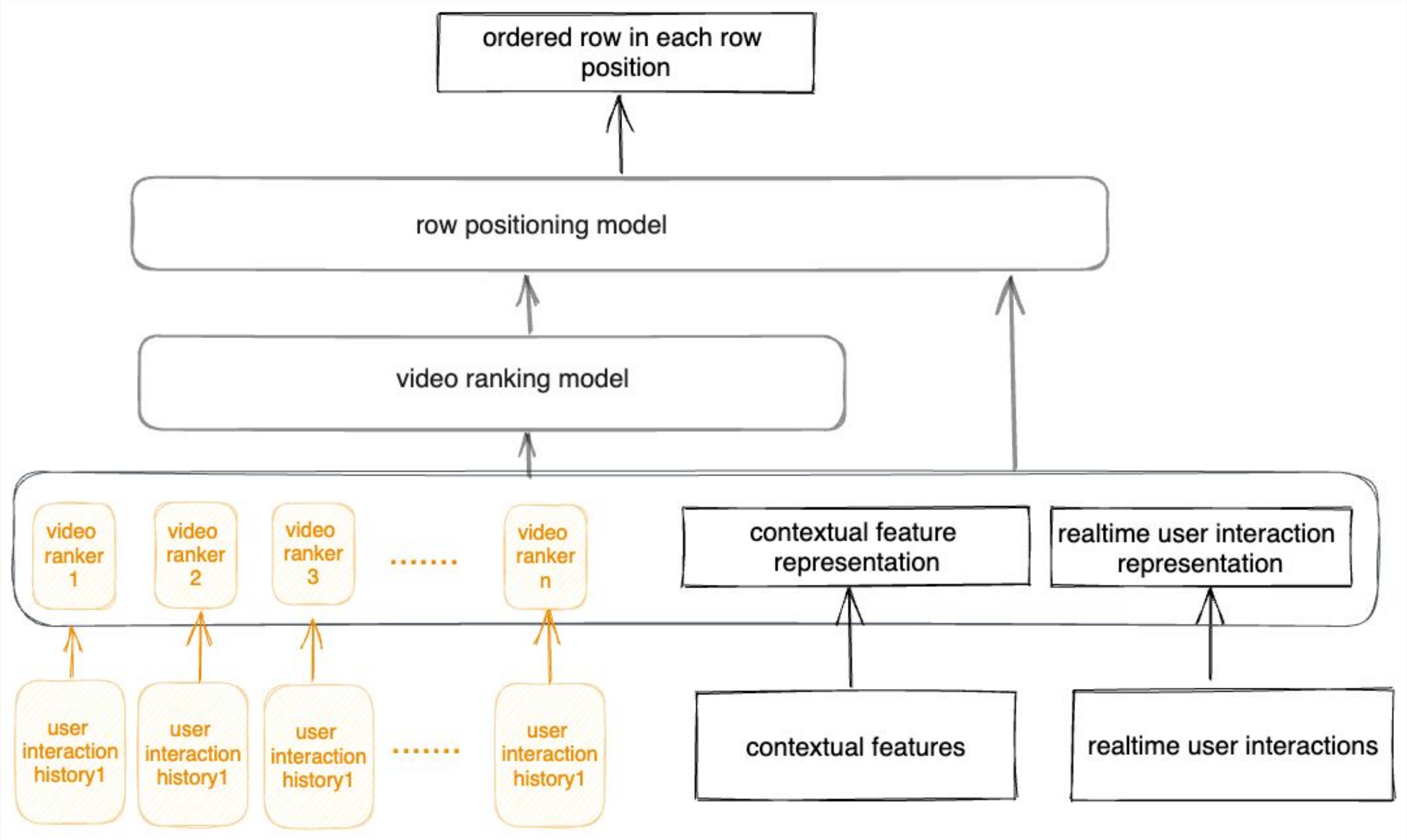
I can memorize the data, and I can learn end to end

NETFLIX

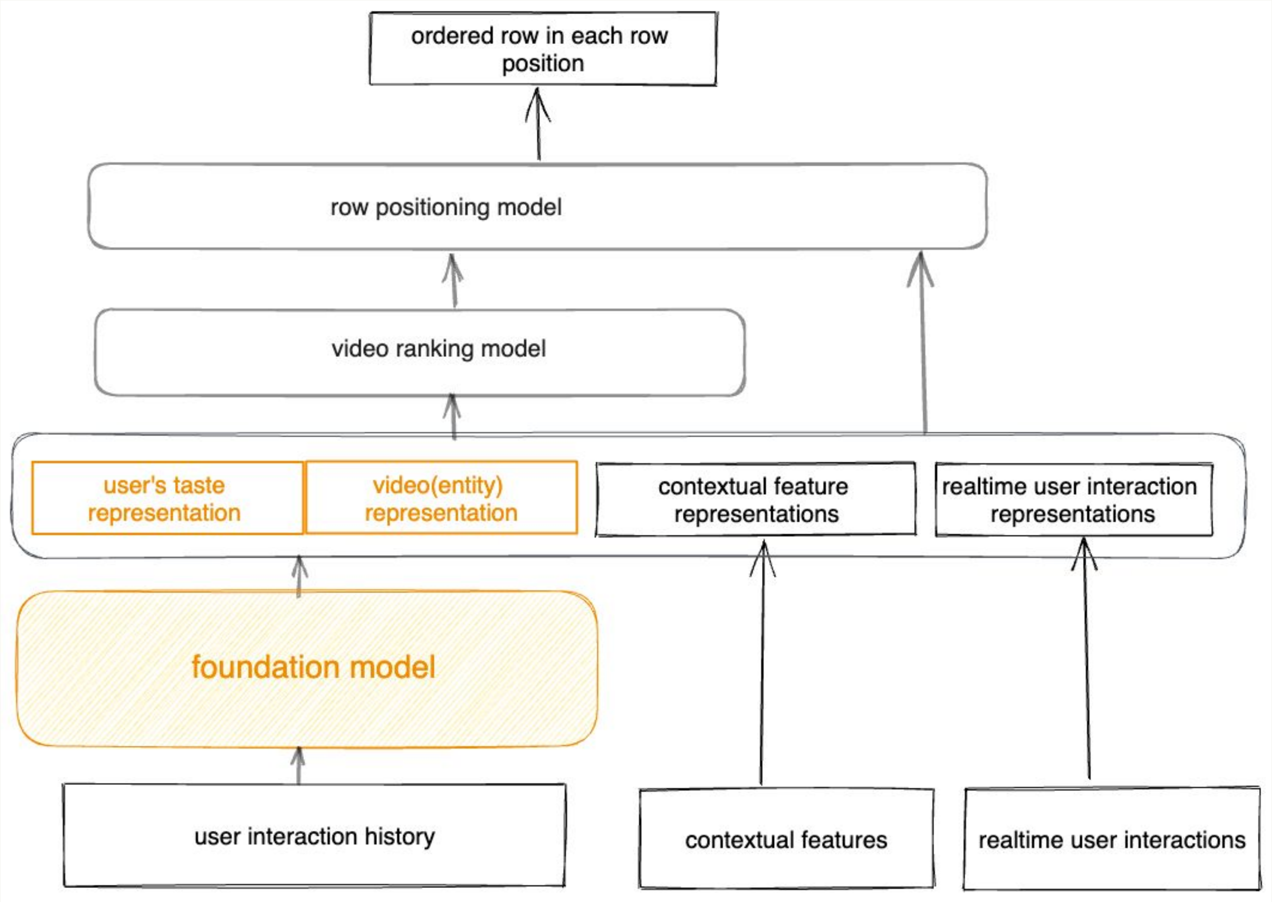
Model Architecture and Training Objective



Why A Foundation Model: Consolidate and Scale up



Why A Foundation Model: Consolidate and Scale up



NETFLIX

Conclusions

NETFLIX

Conclusions

- Overview of Netflix recommendations.
- 4 aspects:

    A. Objective and Reward
    B. Model Consolidation
    C. Everything is Changing
    D. Personalization Power-up

Lots of Fun and Challenges

- Keep pushing the frontier of recommendations for Netflix to help members find the next title they will enjoy.

- Many open challenges
    - Keep pushing the boundaries.
    - Unleash the power of LLM for Recommendations.
    - Capture long-term rewards, RL, Causal modeling.
    - etc…

NETFLIX

We are hiring!

Check jobs.netflix.com

NETFLIX

Thank you.
Questions?

Mark (Ko-Jen) Hsiao
khsiao@netflix.com

NETFLIX