



RUTGERS

# Generative Recommendation with Foundation Models

Yongfeng Zhang, Rutgers University

[yongfeng.zhang@rutgers.edu](mailto:yongfeng.zhang@rutgers.edu)

<http://www.yongfeng.me>

# Recommender Systems are Everywhere

- Influence our daily life by providing personalized services

E-commerce



Social Networks



News Feeding



Search Engine



Navigation



Travel Planning



Professional Networks



Healthcare

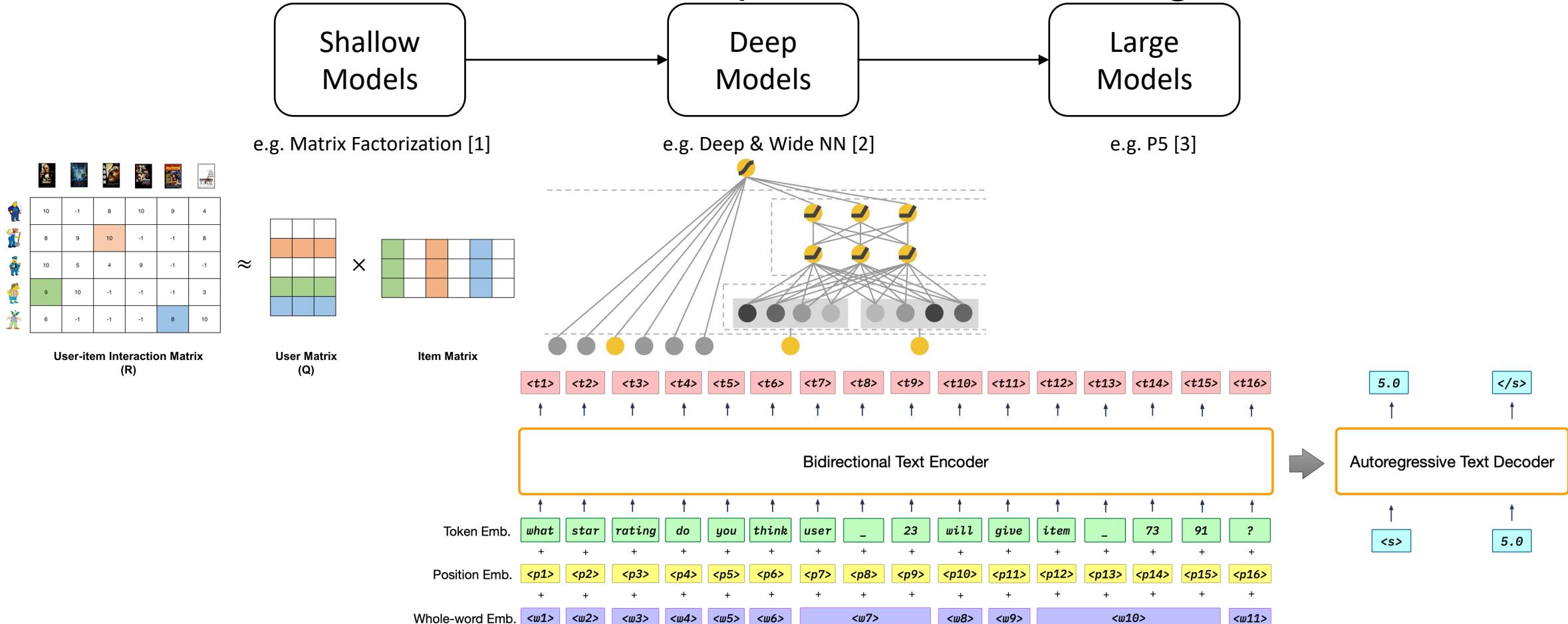


Online Education



# Technical Advancement of Recommender Systems

- From Shallow Model, to Deep Model, and to Large Model



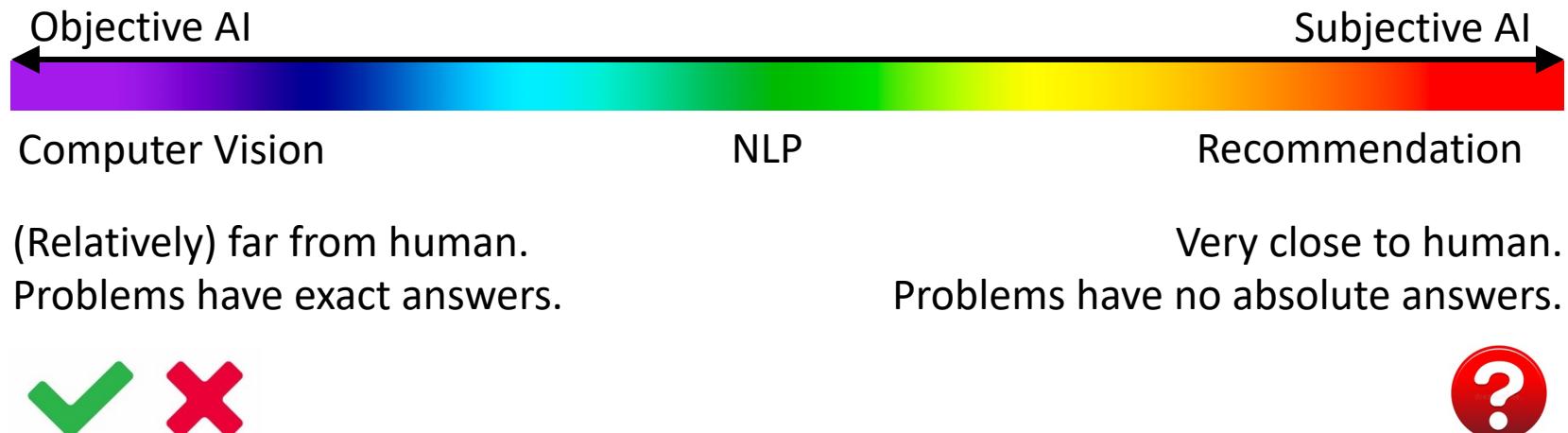
[1] Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 42, no. 8 (2009): 30-37.

[2] Cheng, Heng-Tze, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson et al. "Wide & deep learning for recommender systems." DLRS 2016.

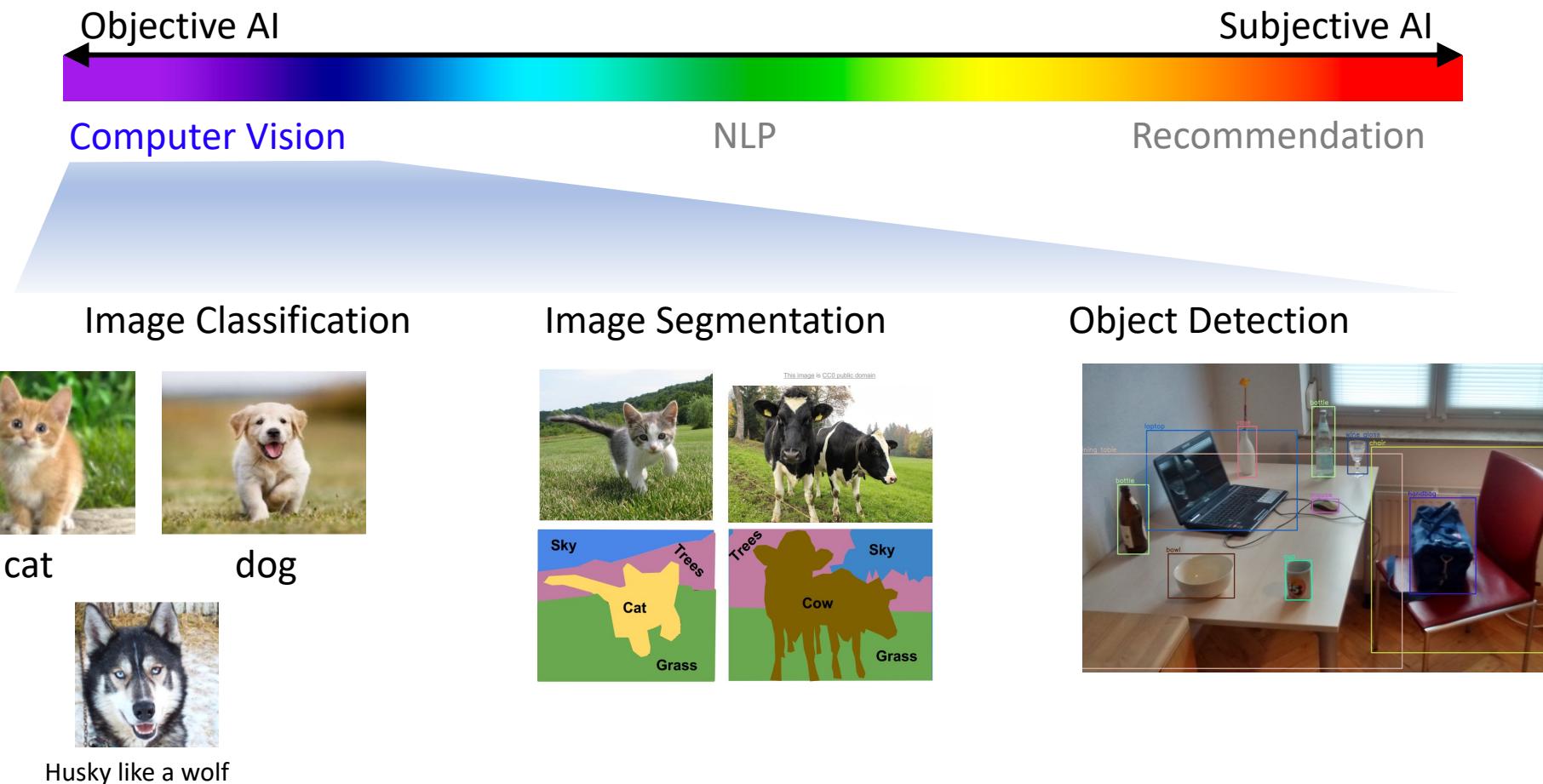
[3] Geng, Shijie, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. "Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)." RecSys 2022.

# Objective AI vs. Subjective AI

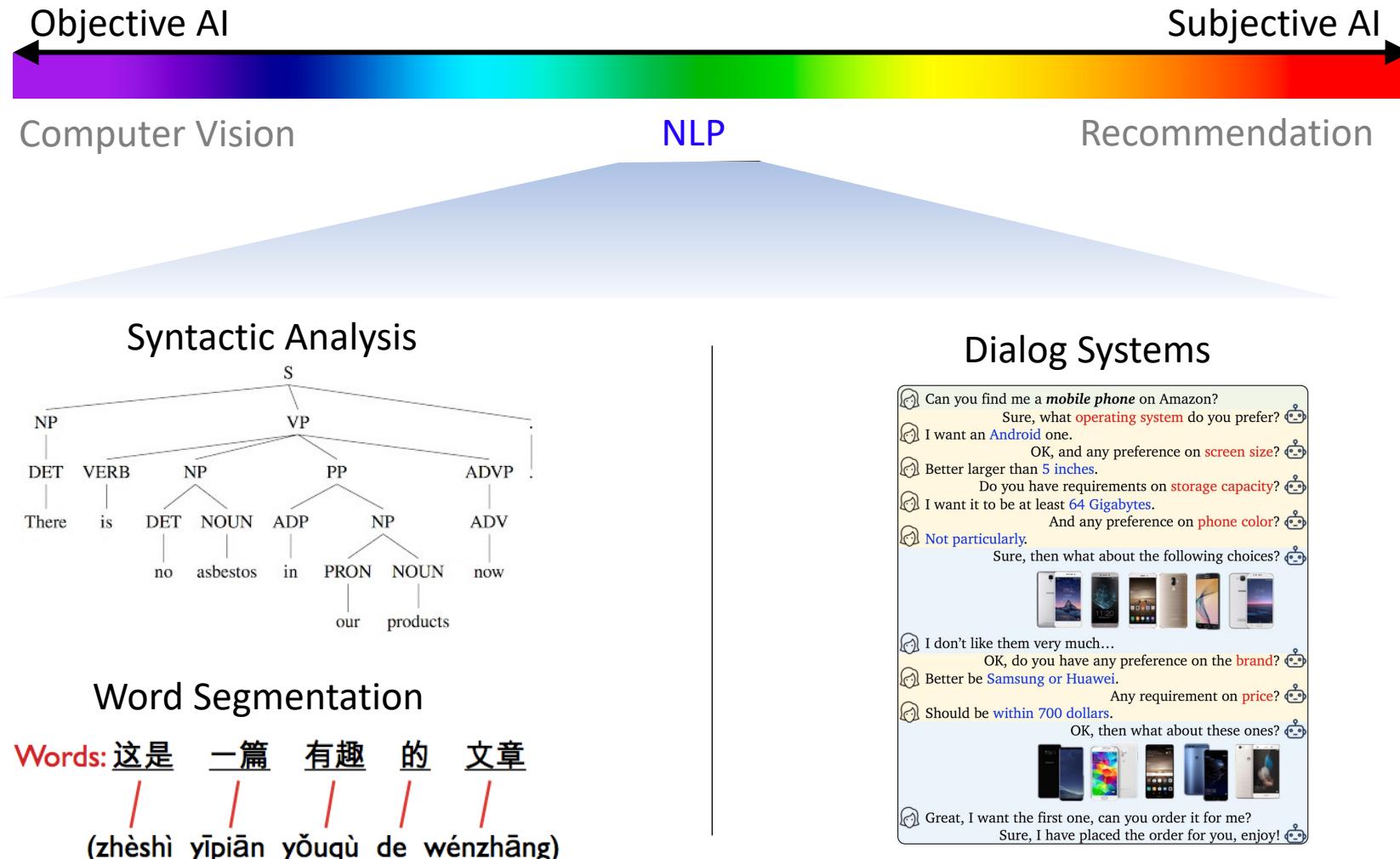
- Recommendation is **unique** in the AI family
    - Recommendation is most **close to human** among all AI tasks
    - Recommendation is a very representative **Subjective AI**
    - Thus, leads to many **unique challenges** in recommendation research



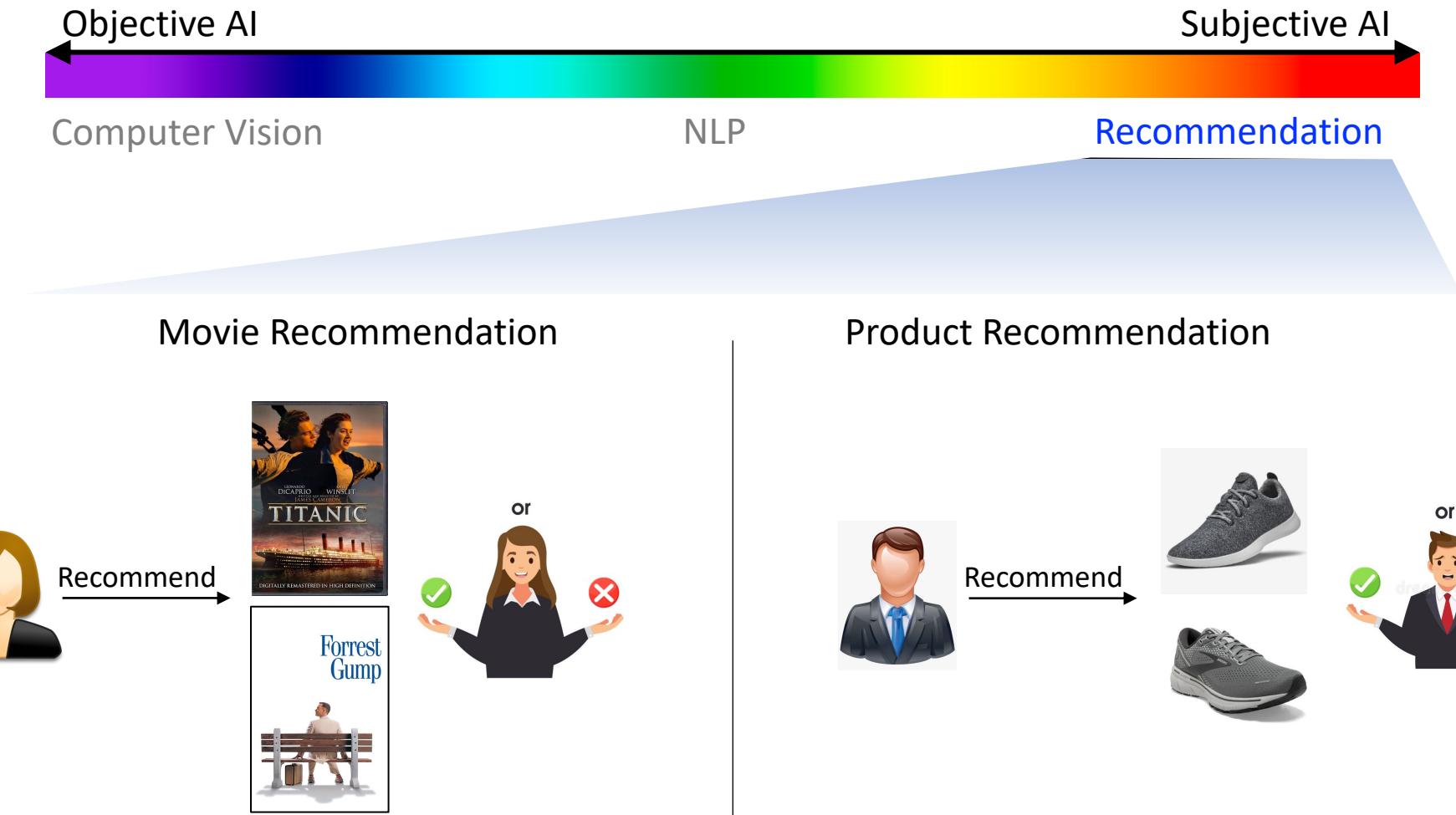
# Computer Vision: (mostly) Objective AI Tasks



# NLP: partly Objective, partly Subjective



# Recommendation: mostly Subjective AI Tasks



# Recommendation is not only about Item Ranking

- A diverse set of recommendation tasks
  - Rating Prediction
  - Item Ranking
  - Sequential Recommendation
  - User Profile Construction
  - Review Summarization
  - Explanation Generation
  - .....

# Subjective AI needs Explainability

- Objective vs. Subjective AI on Explainability

## Objective AI

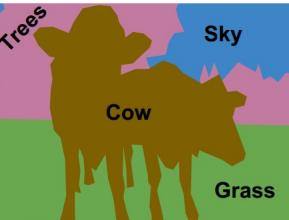
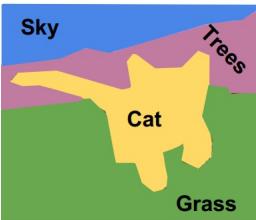
Human can directly identify if the AI-produced result is right or wrong



cat

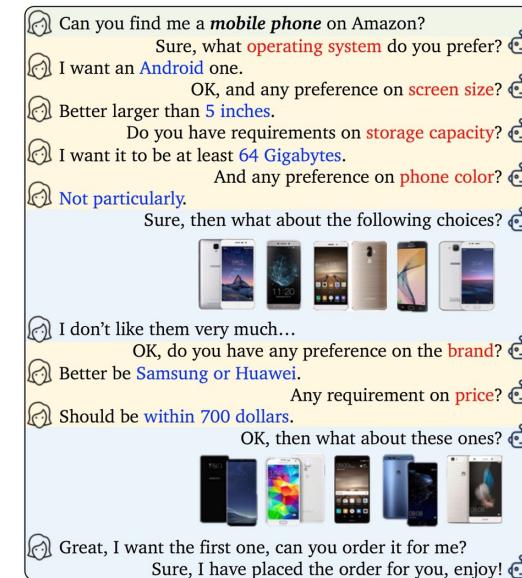


dog



## Subjective AI

Human can hardly identify if the AI-produced result is right or wrong. Users are very **vulnerable**, could be **manipulated**, **utilized** or even **cheated** by the system



**Nothing is definitely right or wrong.**

**Highly subjective, and usually personalized.**

# Subjective AI needs Explainability

- In many cases, it doesn't matter what you recommend, but how you explain your recommendation
- How do humans make recommendation?



# Can we Handle all RecSys tasks Together?

- A diverse set of recommendation tasks
  - Rating Prediction
  - Item Ranking
  - Sequential Recommendation
  - User Profile Construction
  - Review Summarization
  - Explanation Generation
  - Fairness Consideration
  - ...
- Do we really need to design thousands of recommendation models?
  - Difficult to integrate so many models in industry production environment

# A Bird's View of Existing RecSys

- The Multi-Stage Filtering RecSys Pipeline

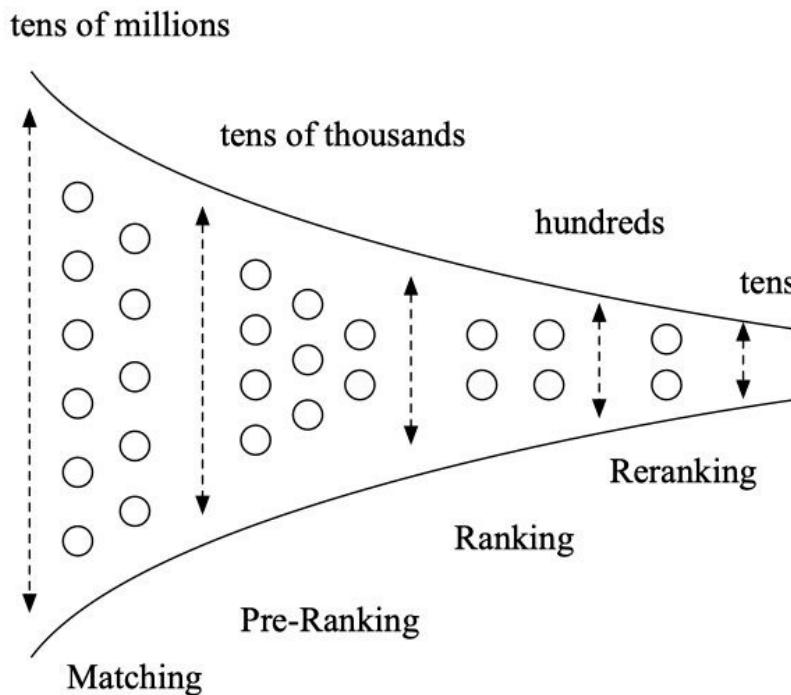


Image credit to [1]

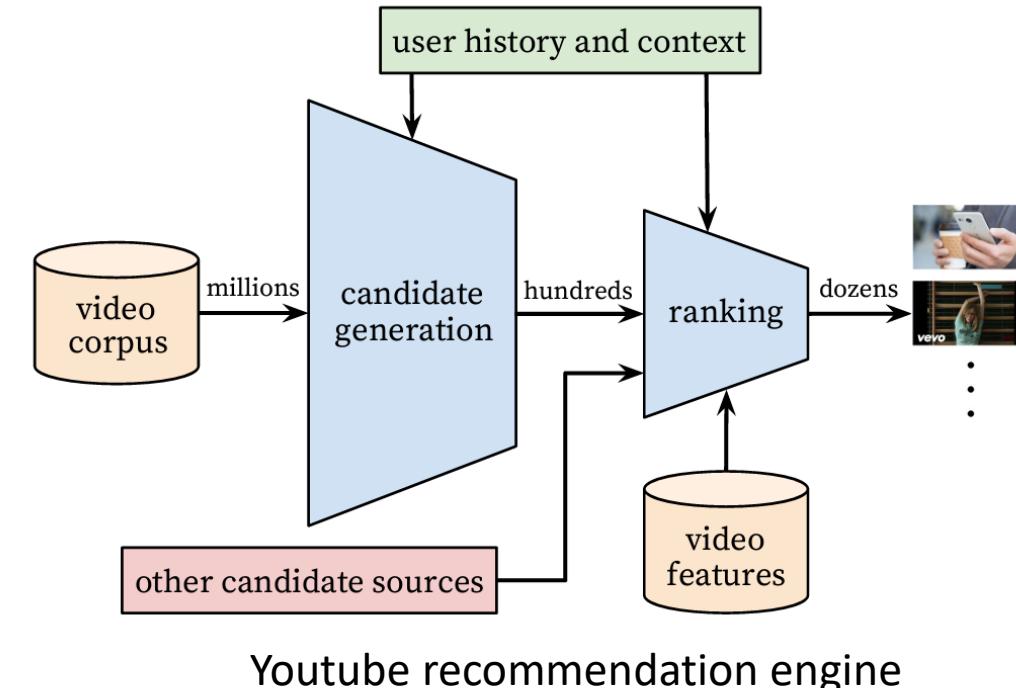


Image credit to [2]

# Discriminative Ranking

- User-item matching based on embeddings

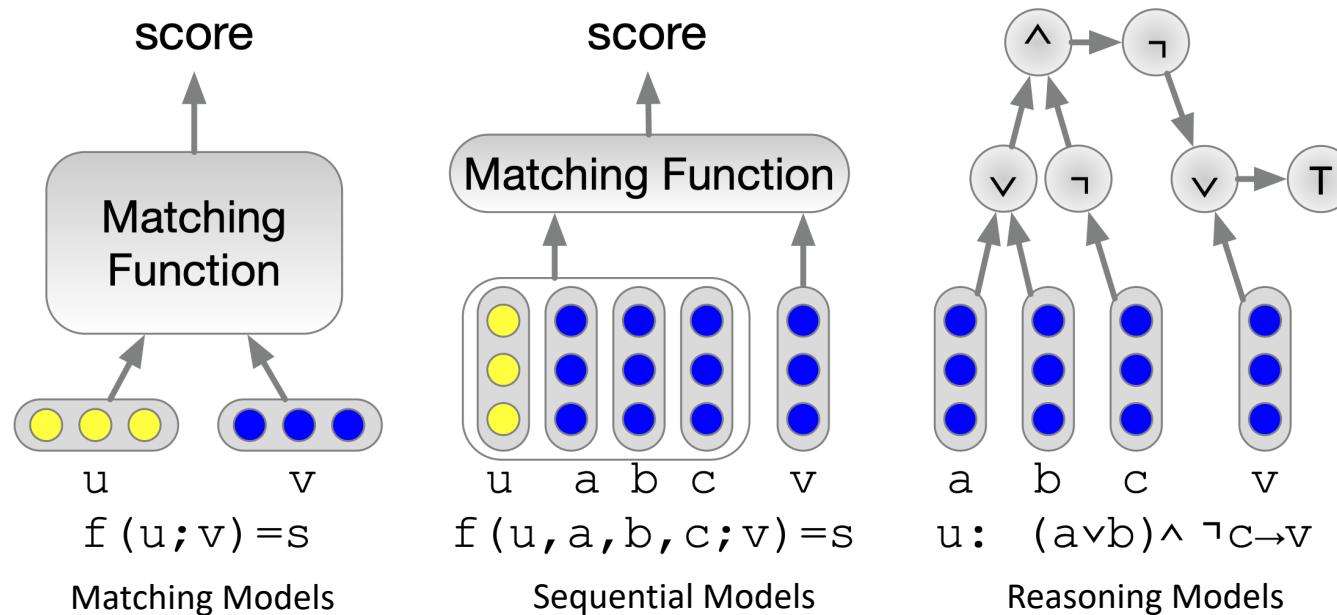


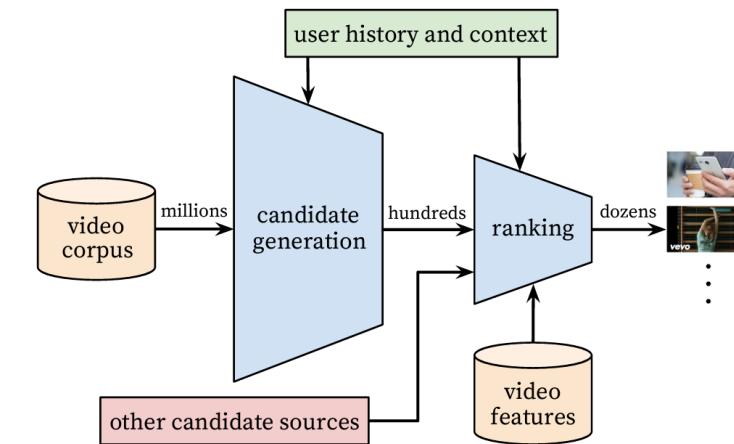
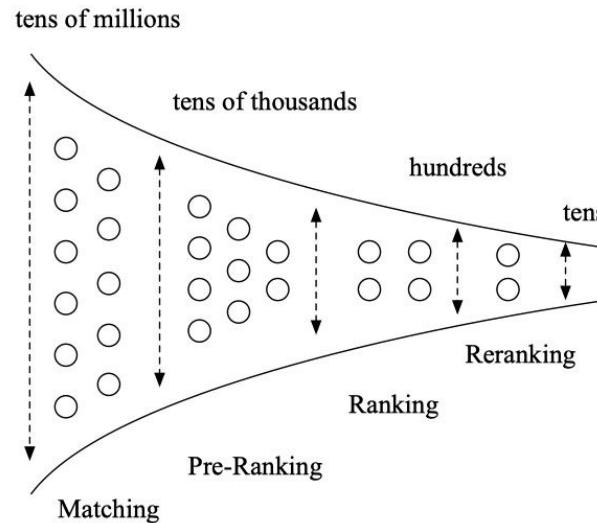
Image credit to [1]

- Discriminative ranking loss function
  - e.g., Bayesian Personalized Ranking (BPR) loss

$$\text{maximize} \sum_{(u, i, j) \in D_S} \ln \sigma(\hat{x}_{uij}) - \lambda_\Theta \|\Theta\|^2 \quad \text{where: } \hat{x}_{uij} = p_u q_i^T - p_u q_j^T$$

# Problem with Discriminative Ranking

- Huge numbers of users and items
  - Amazon: 300 million customers, 350 million products\*
  - YouTube: 2.6+ billion monthly active users, 5+ billion videos\*\*
  - We have to use multi-stage filtering



- Too many candidate items, difficult for evaluation
  - Many research papers use **sampled evaluation**: 1-in-100, 1-in-1000, etc.

\*<https://sell.amazon.com/blog/amazon-stats>, and <https://www.bigcommerce.com/blog/amazon-statistics/>

\*\*<https://www.globalmediainsight.com/blog/youtube-users-statistics/>

# Foundation Models

- Auto-regressive decoding for generative prediction

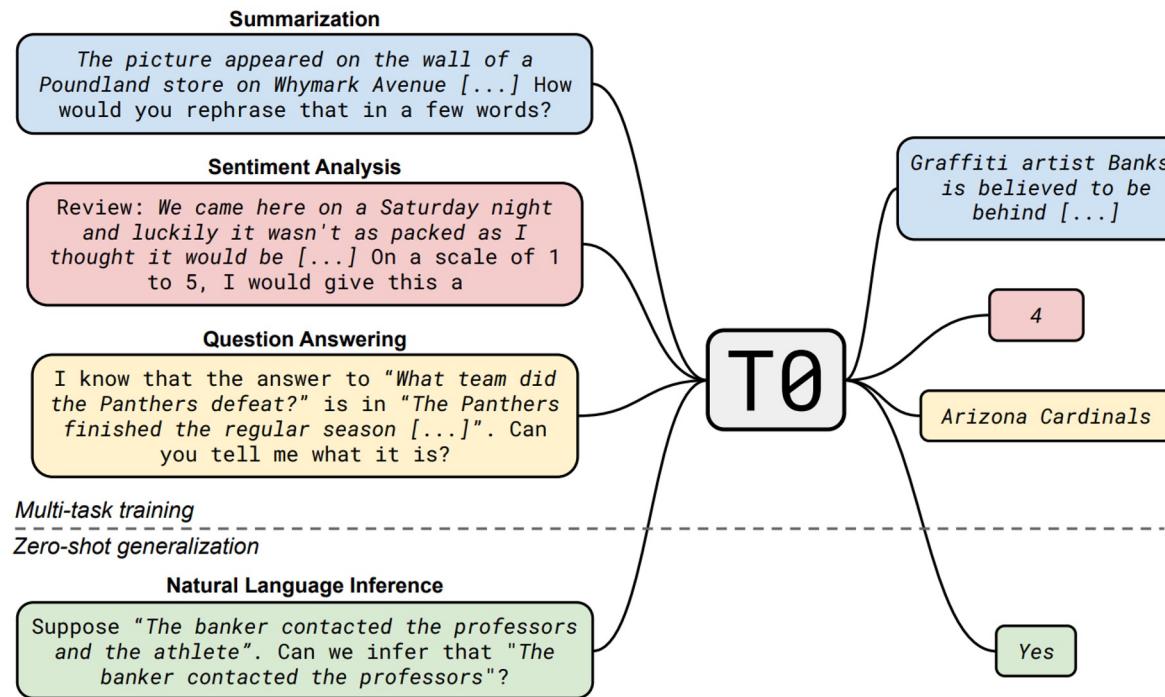


Image credit to [1]

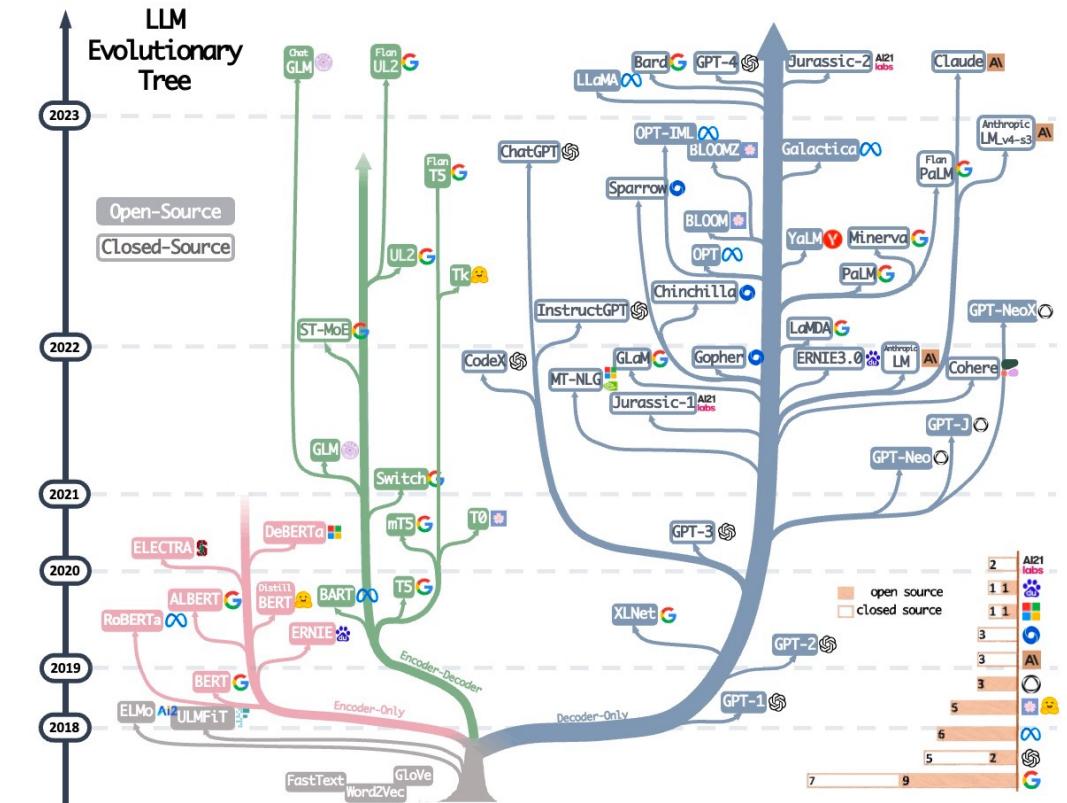


Image credit to [2]

[1] Sanh, Victor, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin et al. "Multitask prompted training enables zero-shot task generalization." ICLR 2022.

[2] Yang, Jingfeng, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond." arXiv preprint arXiv:2304.13712 (2023).

# Generative Pre-training and Prediction

- Generative Pre-training
  - Generative Loss Function
    - Use the previous tokens to predict next token

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta)$$

- Generative Prediction
  - Beam Search
    - # of candidate tokens at each beam is fixed
    - No longer need one-by-one candidate score calculation as in discriminative ranking
    - Directly generate the item ID to recommend

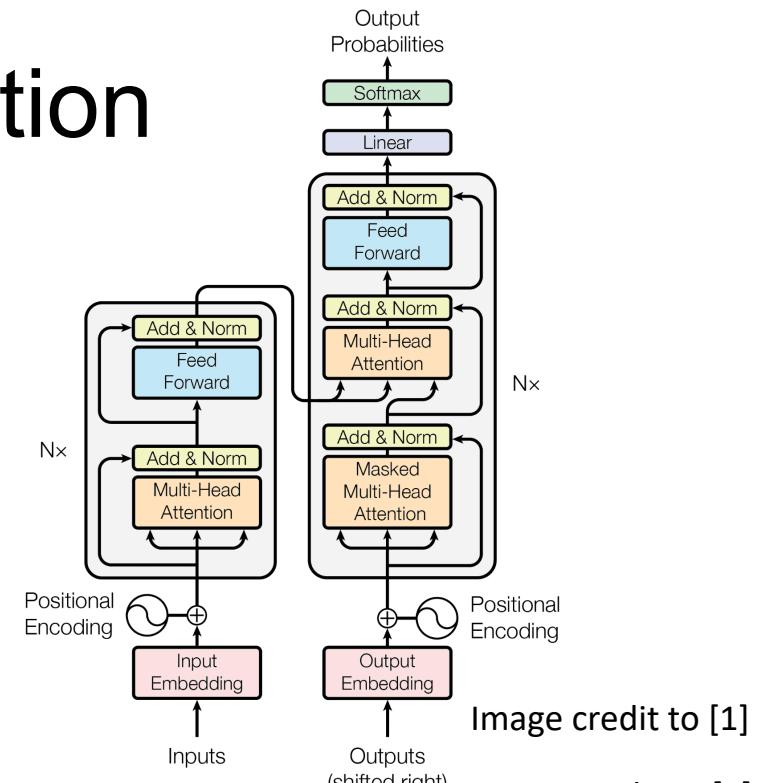


Image credit to [1]

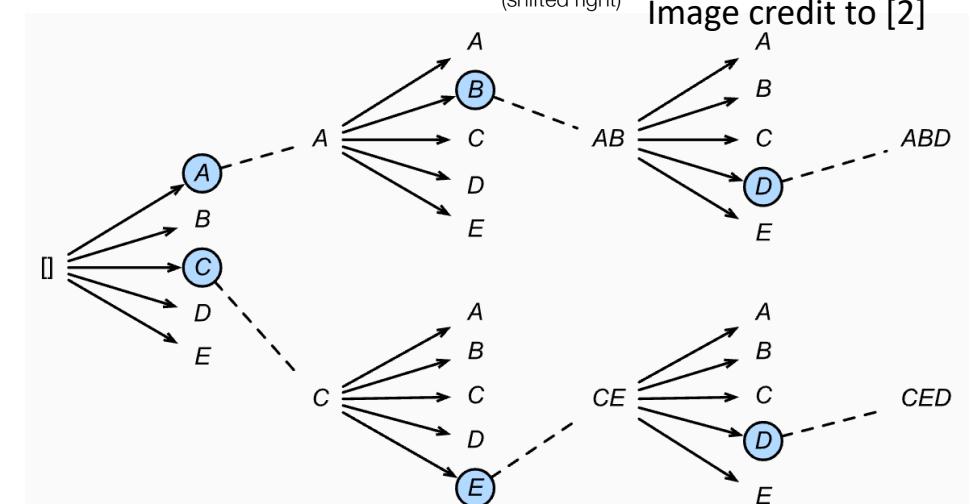


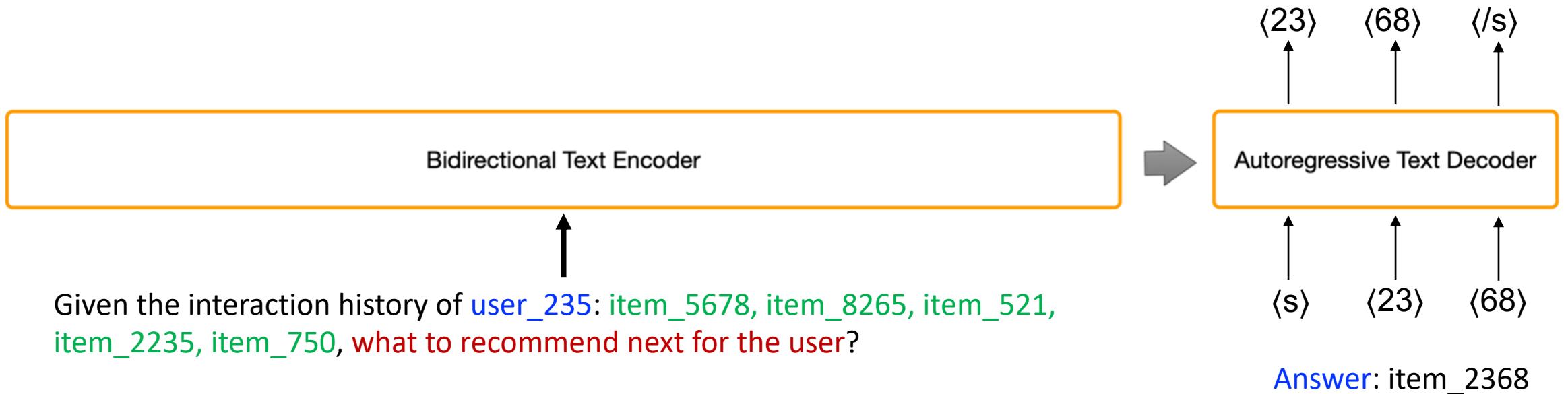
Image credit to [2]

[1] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[2] [https://d2l.ai/chapter\\_recurrent-modern/beam-search.html](https://d2l.ai/chapter_recurrent-modern/beam-search.html)

# Generative Ranking

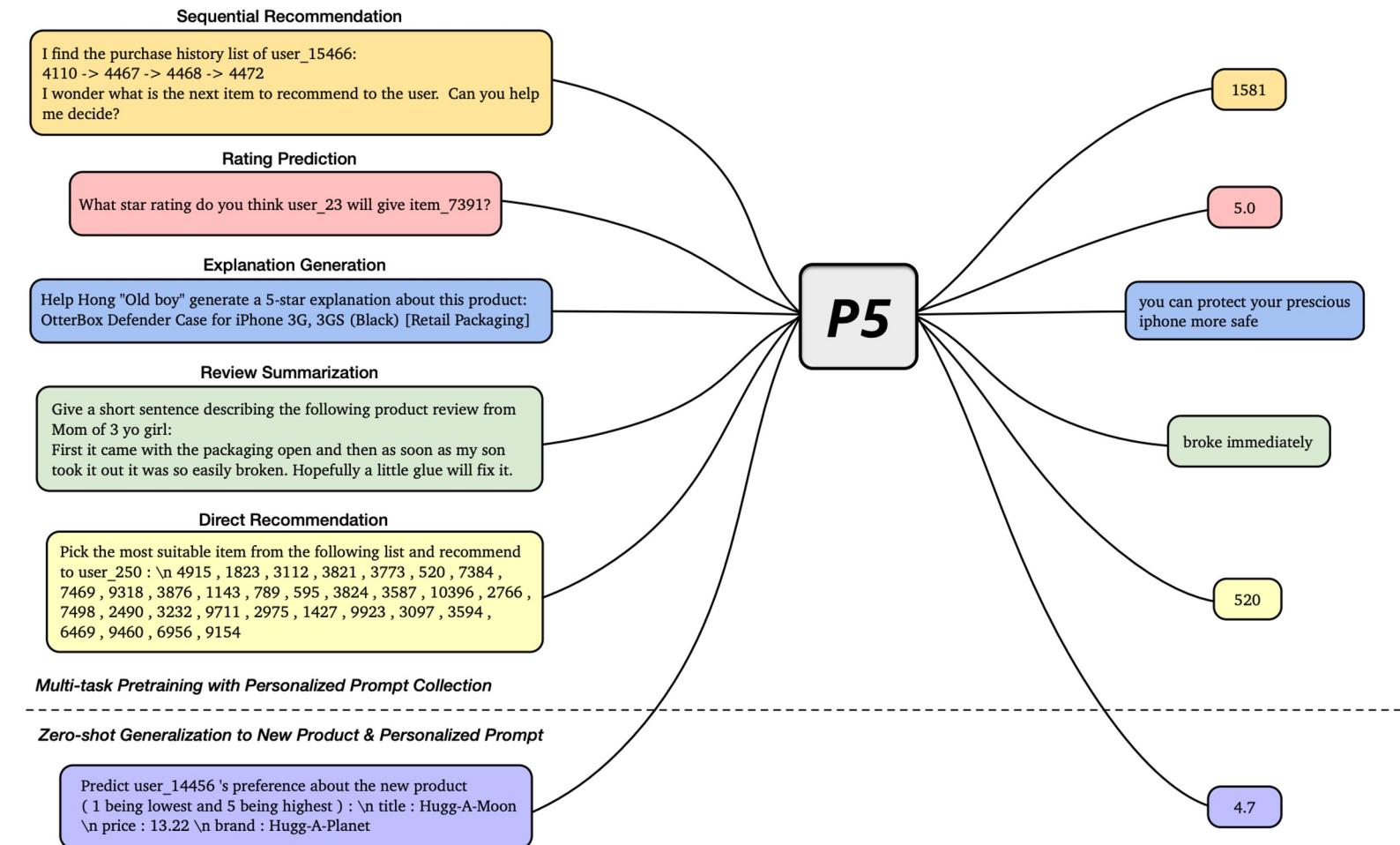
- From Multi-stage ranking to Single-stage ranking
  - The model automatically considers **all items as the candidate pool**
  - Fixed-size item decoding
    - e.g., using 100 tokens  $\langle 00 \rangle \langle 01 \rangle \dots \langle 99 \rangle$  for item ID representation



# The P5 Generative Recommendation Paradigm

- P5: Pretrain, Personalized Prompt & Predict Paradigm [1]

- Learns **multiple** recommendation tasks together through a unified **sequence-to-sequence** framework
- Formulates different recommendation problems as **prompt-based natural language tasks**
- User-item information and corresponding features are integrated with **personalized prompts** as model inputs

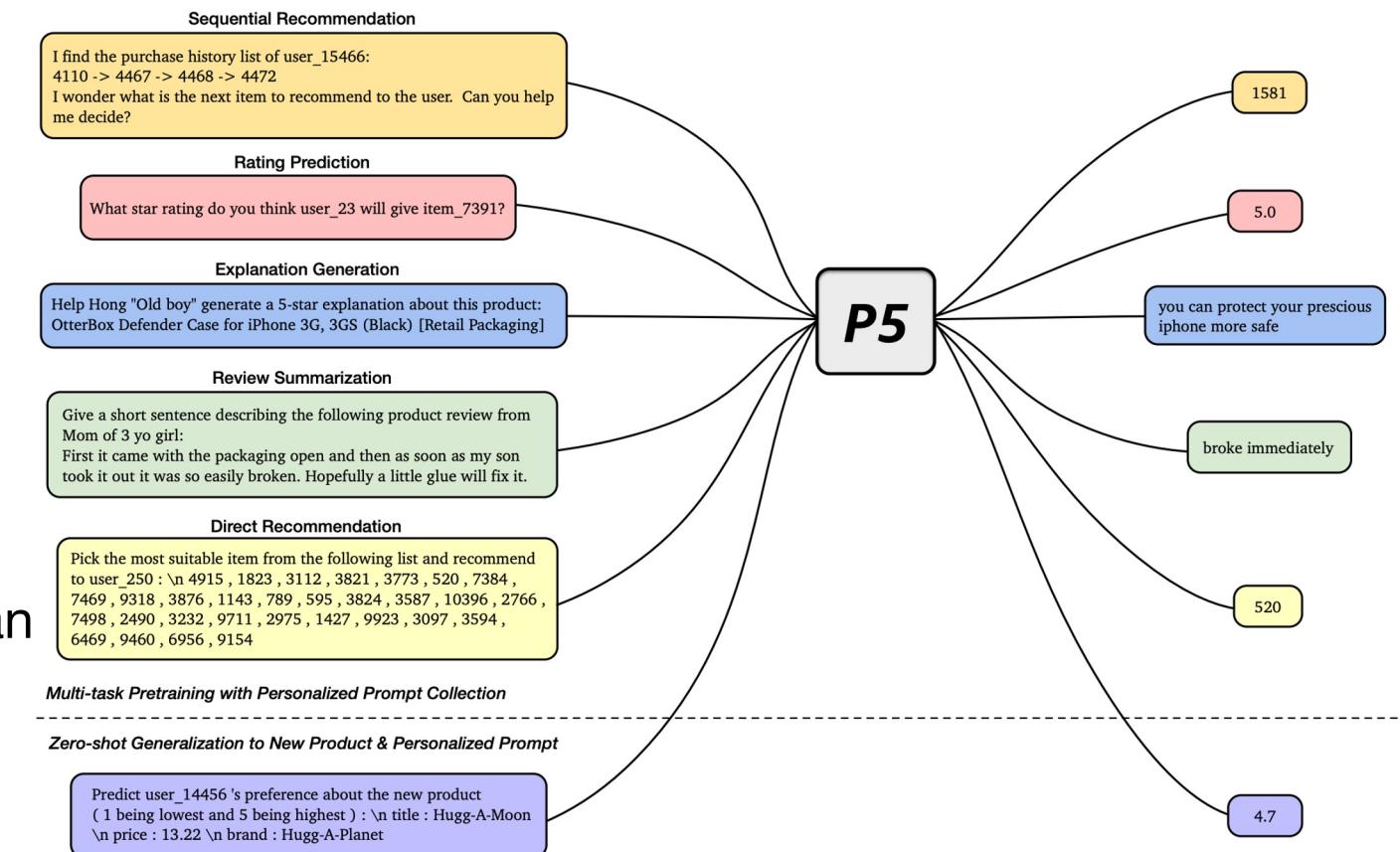


# Five Key Questions in P5 Design

- 1. What tasks are covered by P5?
- 2. How to represent user preferences and item features in P5?
- 3. How to design personalized prompts for different recommendation tasks?
- 4. What foundation model architecture as backbone for P5?
- 5. How to conduct training and inference of P5?

# P5 Recommendation Tasks

- P5 covers **5** different **task families**
  - ***rating prediction***
  - ***sequential recommendation***
  - ***explanation generation***
  - ***review summarization***
  - ***direct recommendation***
- But is not limited these five task families, can be easily and flexibility extended with new personalized prompts



# Enable Personalization in Prompts

- Definition of **personalized prompts**
  - A prompt that includes personalized fields for different users and items
- User's preference can be indicated through
  - A **user ID** (e.g., "user\_23")
  - Content **description of the user** such as location, preferred movie genres, etc.
- Item field can be represented by
  - An **item ID** (e.g., "item\_7391")
  - Item **content metadata** that contains **detailed descriptions** of the item, e.g., item category

# Personalized Prompt Design

## Rating / Review / Explanation raw data for *Beauty*

**user\_id:** 7641      **user\_name:** stephanie  
**item\_id:** 2051  
**item\_title:** SHANY Nail Art Set (24 Famouse Colors  
Nail Art Polish, Nail Art Decoration)  
**review:** Absolutely great product. I bought this for my fourteen year  
old niece for Christmas and of course I had to try it out, then I  
tried another one, and another one and another one. So much fun!  
I even contemplated keeping a few for myself!  
**star\_rating:** 5  
**summary:** Perfect!  
**explanation:** Absolutely great product      **feature\_word:** product

Which star rating will user\_{{user\_id}} give item\_{{item\_id}}?  
(1 being lowest and 5 being highest)

→ {{star\_rating}}

Based on the feature word {{feature\_word}}, generate an  
explanation for user\_{{user\_id}} about this product:  
{{item\_title}}

→ {{explanation}}

Give a short sentence describing the following product review  
from {{user\_name}}: {{review}}

→ {{summary}}

(a)

## Sequential Recommendation raw data for *Beauty*

**user\_id:** 7641      **user\_name:** Victor  
**purchase\_history:** 652 -> 460 -> 447 -> 653 -> 654 -> 655 -> 656 -> 8  
-> 657  
**next\_item:** 552  
**candidate\_items:** 4885 , 4280 , 4886 , 1907 , 870 , 4281 , 4222 ,  
4887 , 2892 , 4888 , 2879 , 3147 , 2195 , 3148 , 3179 , 1951 ,  
..... , 1982 , 552 , 2754 , 2481 , 1916 , 2822 , 1325

Here is the purchase history of user\_{{user\_id}}:  
{{purchase\_history}}  
What to recommend next for the user?

→ {{next\_item}}

(b)

# Design Multiple Prompts for Each Task

- To enhance variation in language style (e.g., sequential recommendation)

## Prompt ID: 2-1

Input template: Given the following purchase history of user\_{{user\_id}}:  
{{purchase\_history}}  
predict next possible item to be purchased by the user?

Target template: {{next\_item}}

## Prompt ID: 2-2

Input template: I find the purchase history list of user\_{{user\_id}}:  
{{purchase\_history}}  
I wonder which is the next item to recommend to the user. Can you help me decide?

Target template: {{next\_item}}

## Prompt ID: 2-3

Input template: Here is the purchase history list of user\_{{user\_id}}:  
{{purchase\_history}}  
try to recommend next item to the user

Target template: {{next\_item}}

## Prompt ID: 2-4

Input template: Given the following purchase history of {{user\_desc}}:  
{{purchase\_history}}  
predict next possible item for the user

Target template: {{next\_item}}

## Prompt ID: 2-5

Input template: Based on the purchase history of {{user\_desc}}:  
{{purchase\_history}}  
Can you decide the next item likely to be purchased by the user?

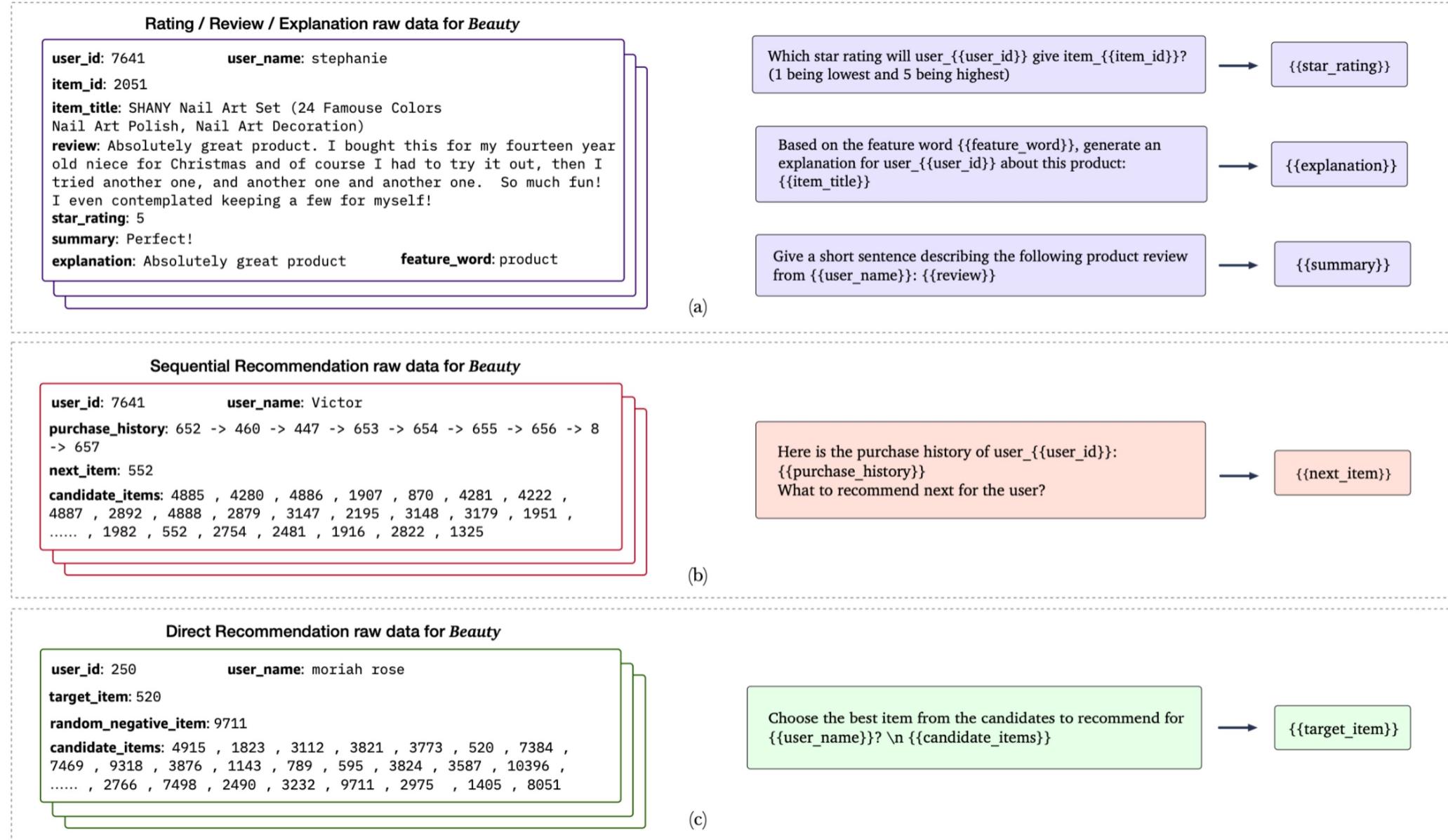
Target template: {{next\_item}}

## Prompt ID: 2-6

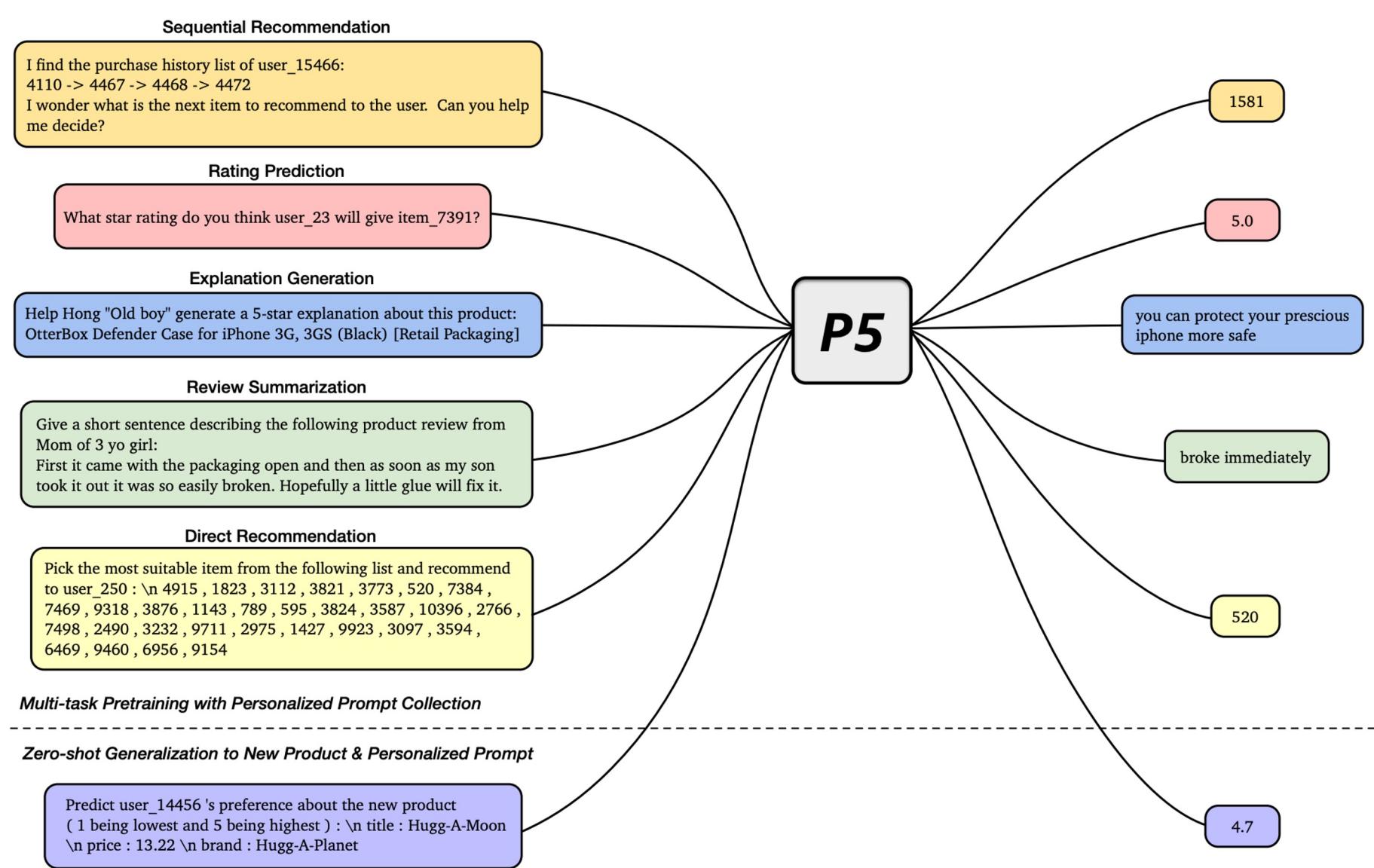
Input template: Here is the purchase history of {{user\_desc}}:  
{{purchase\_history}}  
What to recommend next for the user?

Target template: {{next\_item}}

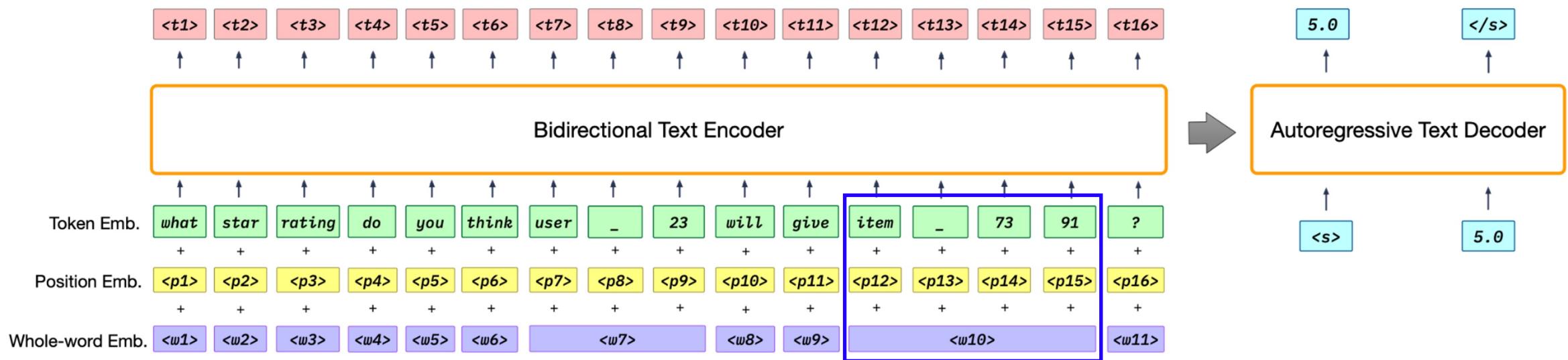
# Text-to-Text Training Data



# Multi-Task Pre-training



# Multi-Task Pre-training



- P5 is pre-trained on top of T5 checkpoints (to enable P5 basic ability for language understanding)
  - So P5 is a [sequence-to-sequence model](#)
- By default, we use [multiple sub-word units](#) to represent personalize fields (e.g., ["item", "\_", "73", "91"])
- To help the model to understand ["item", "\_", "73", "91"] is a complete field, we apply [whole-word embedding](#) in P5

# Generative Recommendation

- The encoder takes input sequence
- The decoder autoregressively generates next words:
  - **Autoregressive LM loss** is shared by all tasks:  $\mathcal{L}_{\theta}^{\text{P5}} = - \sum_{j=1}^{|\mathbf{y}|} \log P_{\theta}(\mathbf{y}_j \mid \mathbf{y}_{<j}, \mathbf{x})$
- We can unify various recommendation tasks with **one model, one loss, and one data format**
- Inference with pretrained P5
  - Simply apply **beam search** to generate a list of potential next items
  - Since item IDs are tokenized (e.g., ["item", "\_", "73", "91"]), beam search is limited on width
    - E.g., 100 tokens width: <00>, <01>, <02>, ..., <98>, <99>

# Advantages of P5 Generative Recommendation

- Immerses recommendation models into a full language environment
  - With the **flexibility** and **expressiveness** of language, there is **no need to design feature-specific encoders**
- P5 treats all personalized tasks as a conditional text generation problem
  - One data format, one model, one loss for multiple recommendation tasks
  - No need to design data-specific or task-specific **recommendation models**
- P5 attains sufficient **zero-shot performance** when generalizing to novel personalized prompts or unseen items in other domains

# Performance of P5 under **seen** Prompts

## Rating Prediction:

Methods	Sports		Beauty		Toys	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
MF	<b>1.0234</b>	0.7935	<b>1.1973</b>	0.9461	<b>1.0123</b>	0.7984
MLP	1.1277	0.7626	1.3078	0.9597	1.1215	0.8097
P5-S (1-6)	1.0594	<b>0.6639</b>	1.3128	<b>0.8428</b>	1.0746	<b>0.7054</b>
P5-B (1-6)	1.0357	0.6813	<u>1.2843</u>	0.8534	1.0544	0.7177
P5-S (1-10)	1.0522	<u>0.6698</u>	1.2989	<u>0.8473</u>	1.0550	0.7173
P5-B (1-10)	1.0292	0.6864	1.2870	0.8531	<u>1.0245</u>	<b>0.6931</b>

## Sequential Recommendation:

Methods	Sports				Beauty				Toys			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
Caser	0.0116	0.0072	0.0194	0.0097	0.0205	0.0131	0.0347	0.0176	0.0166	0.0107	0.0270	0.0141
HGN	0.0189	0.0120	0.0313	0.0159	0.0325	0.0206	0.0512	0.0266	0.0321	0.0221	0.0497	0.0277
GRU4Rec	0.0129	0.0086	0.0204	0.0110	0.0164	0.0099	0.0283	0.0137	0.0097	0.0059	0.0176	0.0084
BERT4Rec	0.0115	0.0075	0.0191	0.0099	0.0203	0.0124	0.0347	0.0170	0.0116	0.0071	0.0203	0.0099
FDSA	0.0182	0.0122	0.0288	0.0156	0.0267	0.0163	0.0407	0.0208	0.0228	0.0140	0.0381	0.0189
SASRec	0.0233	0.0154	0.0350	0.0192	0.0387	0.0249	0.0605	0.0318	0.0463	0.0306	0.0675	0.0374
S <sup>3</sup> -Rec	0.0251	0.0161	0.0385	0.0204	0.0387	0.0244	0.0647	0.0327	0.0443	0.0294	0.0700	0.0376
P5-S (2-3)	0.0272	0.0169	0.0361	0.0198	<u>0.0503</u>	<u>0.0370</u>	<u>0.0659</u>	<u>0.0421</u>	<b>0.0648</b>	<b>0.0567</b>	<b>0.0709</b>	<b>0.0587</b>
P5-B (2-3)	<u>0.0364</u>	<u>0.0296</u>	<u>0.0431</u>	<u>0.0318</u>	<b>0.0508</b>	<b>0.0379</b>	<b>0.0664</b>	<b>0.0429</b>	0.0608	0.0507	0.0688	0.0534
P5-S (2-13)	0.0258	0.0159	0.0346	0.0188	0.0490	0.0358	0.0646	0.0409	<u>0.0647</u>	<u>0.0566</u>	<u>0.0705</u>	<u>0.0585</u>
P5-B (2-13)	<b>0.0387</b>	<b>0.0312</b>	<b>0.0460</b>	<b>0.0336</b>	0.0493	0.0367	0.0645	0.0416	0.0587	0.0486	0.0675	0.0536

## Explanation Generation:

Methods	Sports				Beauty				Toys			
	BLUE4	ROUGE1	ROUGE2	ROUGEL	BLUE4	ROUGE1	ROUGE2	ROUGEL	BLUE4	ROUGE1	ROUGE2	ROUGEL
Attn2Seq	0.5305	12.2800	1.2107	9.1312	0.7889	12.6590	1.6820	9.7481	1.6238	13.2245	2.9942	10.7398
NRT	0.4793	11.0723	1.1304	7.6674	0.8295	12.7815	1.8543	9.9477	1.9084	13.5231	3.6708	11.1867
PETER	0.7112	12.8944	1.3283	9.8635	<u>1.1541</u>	14.8497	<u>2.1413</u>	11.4143	1.9861	14.2716	3.6718	11.7010
P5-S (3-3)	<b>1.0447</b>	<b>14.9048</b>	<b>2.1297</b>	<b>11.1778</b>	<b>1.2237</b>	<b>17.6938</b>	<b>2.2489</b>	<b>12.8606</b>	<u>2.2892</u>	<b>15.4505</b>	<u>3.6974</u>	<b>12.1718</b>
P5-B (3-3)	<u>1.0407</u>	<u>14.1589</u>	<u>2.1220</u>	<u>10.6096</u>	0.9742	<u>16.4530</u>	1.8858	<u>11.8765</u>	<b>2.3185</b>	<u>15.3474</u>	<b>3.7209</b>	<u>12.1312</u>
PETER+	<b>2.4627</b>	<b>24.1181</b>	5.1937	<b>18.4105</b>	<b>3.2606</b>	<u>25.5541</u>	5.9668	<u>19.7168</u>	<b>4.7919</b>	<u>28.3083</u>	9.4520	<b>22.7017</b>
P5-S (3-9)	1.4101	<u>23.5619</u>	<b>5.4196</b>	<u>17.6245</u>	<u>1.9788</u>	<b>25.6253</b>	<b>6.3678</b>	<b>19.9497</b>	4.1222	<b>28.4088</b>	9.5432	<u>22.6064</u>
P5-B (3-9)	<u>1.4689</u>	23.5476	<u>5.3926</u>	17.5852	1.8765	25.1183	6.0764	19.4488	3.8933	27.9916	<u>9.5896</u>	22.2178
P5-S (3-12)	1.3212	23.2474	5.3461	17.3780	1.9425	25.1474	6.0551	19.5601	<u>4.2764</u>	28.1897	9.1327	22.2514
P5-B (3-12)	1.4303	23.3810	5.3239	17.4913	1.9031	25.1763	<u>6.1980</u>	19.5188	3.5861	28.1369	<b>9.7562</b>	22.3056

# Performance of P5 under *seen* Prompts

## Review-base Preference Prediction:

Methods	Sports		Beauty		Toys	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
T0 <b>(4-2)</b>	0.6728	0.3140	0.6925	0.3324	0.8282	0.4201
T0 <b>(4-4)</b>	<u>0.6503</u>	0.2984	0.7066	0.3663	0.8148	0.4230
P5-S <b>(4-2)</b>	0.7293	0.3529	<b>0.6233</b>	<b>0.3051</b>	<b>0.6464</b>	<b>0.3125</b>
P5-B <b>(4-2)</b>	<b>0.6487</b>	<b>0.2847</b>	0.6449	0.3168	0.6785	0.3342
P5-S <b>(4-4)</b>	0.7565	0.3395	<u>0.6262</u>	0.3113	<u>0.6577</u>	<u>0.3174</u>
P5-B <b>(4-4)</b>	0.6563	<u>0.2921</u>	0.6515	<u>0.3106</u>	0.6730	0.3342

## Review Summarization:

Methods	Sports				Beauty				Toys			
	BLUE2	ROUGE1	ROUGE2	ROUGEL	BLUE2	ROUGE1	ROUGE2	ROUGEL	BLUE2	ROUGE1	ROUGE2	ROUGEL
T0 <b>(4-1)</b>	2.1581	2.2695	0.5694	1.6221	1.2871	1.2750	0.3904	0.9592	<u>2.2296</u>	2.4671	0.6482	1.8424
GPT-2 <b>(4-1)</b>	0.7779	4.4534	1.0033	1.9236	0.5879	3.3844	0.6756	1.3956	<u>0.6221</u>	3.7149	0.6629	1.4813
P5-S <b>(4-1)</b>	<u>2.4962</u>	<u>11.6701</u>	<u>2.7187</u>	<u>10.4819</u>	<b>2.1225</b>	<b>8.4205</b>	<b>1.6676</b>	<b>7.5476</b>	<u>2.4752</u>	<b>9.4200</b>	<b>1.5975</b>	<b>8.2618</b>
P5-B <b>(4-1)</b>	<b>2.6910</b>	<b>12.0314</b>	<b>3.2921</b>	<b>10.7274</b>	<u>1.9325</u>	<u>8.2909</u>	<u>1.4321</u>	<u>7.4000</u>	1.7833	<u>8.7222</u>	<u>1.3210</u>	<u>7.6134</u>

## Direct Recommendation:

Methods	Sports					Beauty					Toys				
	HR@1	HR@5	NDCG@5	HR@10	NDCG@10	HR@1	HR@5	NDCG@5	HR@10	NDCG@10	HR@1	HR@5	NDCG@5	HR@10	NDCG@10
BPR-MF	0.0314	0.1404	0.0848	0.2563	0.1220	0.0311	0.1426	0.0857	0.2573	0.1224	0.0233	0.1066	0.0641	0.2003	0.0940
BPR-MLP	0.0351	0.1520	0.0927	0.2671	0.1296	0.0317	0.1392	0.0848	0.2542	0.1215	0.0252	0.1142	0.0688	0.2077	0.0988
SimpleX	0.0331	<b>0.2362</b>	<b>0.1505</b>	<u>0.3290</u>	<u>0.1800</u>	0.0325	<u>0.2247</u>	<u>0.1441</u>	0.3090	<u>0.1711</u>	0.0268	<b>0.1958</b>	<b>0.1244</b>	<b>0.2662</b>	<b>0.1469</b>
P5-S <b>(5-1)</b>	0.0638	0.2096	0.1375	0.3143	0.1711	0.0600	0.2021	0.1316	<u>0.3121</u>	0.1670	0.0405	<u>0.1538</u>	<u>0.0969</u>	<u>0.2405</u>	<u>0.1248</u>
P5-B <b>(5-1)</b>	0.0245	0.0816	0.0529	0.1384	0.0711	0.0224	0.0904	0.0559	0.1593	0.0780	0.0187	0.0827	0.0500	0.1543	0.0729
P5-S <b>(5-4)</b>	<u>0.0701</u>	<u>0.2241</u>	<u>0.1483</u>	<b>0.3313</b>	<b>0.1827</b>	<b>0.0862</b>	<b>0.2448</b>	<b>0.1673</b>	<b>0.3441</b>	<b>0.1993</b>	0.0413	0.1411	0.0916	0.2227	0.1178
P5-B <b>(5-4)</b>	0.0299	0.1026	0.0665	0.1708	0.0883	0.0506	0.1557	0.1033	0.2350	0.1287	0.0435	0.1316	0.0882	0.2000	0.1102
P5-S <b>(5-5)</b>	0.0574	0.1503	0.1050	0.2207	0.1276	0.0601	0.1611	0.1117	0.2370	0.1360	<u>0.0440</u>	0.1282	0.0865	0.2011	0.1098
P5-B <b>(5-5)</b>	0.0641	0.1794	0.1229	0.2598	0.1488	0.0588	0.1573	0.1089	0.2325	0.1330	0.0386	0.1122	0.0756	0.1807	0.0975
P5-S <b>(5-8)</b>	0.0567	0.1514	0.1049	0.2196	0.1269	0.0571	0.1566	0.1078	0.2317	0.1318	<b>0.0451</b>	0.1322	0.0889	0.2023	0.1114
P5-B <b>(5-8)</b>	<b>0.0726</b>	0.1955	0.1355	0.2802	0.1627	<u>0.0608</u>	0.1564	0.1096	0.2300	0.1332	0.0389	0.1147	0.0767	0.1863	0.0997

Observation: P5 **achieves promising performances** on the five task families when taking *seen prompt* templates as model inputs

# Performance of P5 under **unseen** Prompts

**Observation:** Multitask prompted pretraining empowers P5 **good robustness** to understand **unseen prompts** with wording variations

## Sequential Recommendation:

Methods	Sports				Beauty				Toys			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
Caser	0.0116	0.0072	0.0194	0.0097	0.0205	0.0131	0.0347	0.0176	0.0166	0.0107	0.0270	0.0141
HGN	0.0189	0.0120	0.0313	0.0159	0.0325	0.0206	0.0512	0.0266	0.0321	0.0221	0.0497	0.0277
GRU4Rec	0.0129	0.0086	0.0204	0.0110	0.0164	0.0099	0.0283	0.0137	0.0097	0.0059	0.0176	0.0084
BERT4Rec	0.0115	0.0075	0.0191	0.0099	0.0203	0.0124	0.0347	0.0170	0.0116	0.0071	0.0203	0.0099
FDSA	0.0182	0.0122	0.0288	0.0156	0.0267	0.0163	0.0407	0.0208	0.0228	0.0140	0.0381	0.0189
SASRec	0.0233	0.0154	0.0350	0.0192	0.0387	0.0249	0.0605	0.0318	0.0463	0.0306	0.0675	0.0374
S <sup>3</sup> -Rec	0.0251	0.0161	0.0385	0.0204	0.0387	0.0244	0.0647	0.0327	0.0443	0.0294	0.0700	0.0376
P5-S (2-3)	0.0272	0.0169	0.0361	0.0198	0.0503	0.0370	0.0659	0.0421	<b>0.0648</b>	<b>0.0567</b>	<b>0.0709</b>	<b>0.0587</b>
P5-B (2-3)	<b>0.0364</b>	<b>0.0296</b>	<b>0.0431</b>	<b>0.0318</b>	<b>0.0508</b>	<b>0.0379</b>	<b>0.0664</b>	<b>0.0429</b>	0.0608	0.0507	0.0688	0.0534
P5-S (2-13)	0.0258	0.0159	0.0346	0.0188	0.0490	0.0358	0.0646	0.0409	<u>0.0647</u>	<u>0.0566</u>	<u>0.0705</u>	<u>0.0585</u>
P5-B (2-13)	<b>0.0387</b>	<b>0.0312</b>	<b>0.0460</b>	<b>0.0336</b>	0.0493	0.0367	0.0645	0.0416	0.0587	0.0486	0.0675	0.0536

## Explanation Generation:

Methods	Sports				Beauty				Toys			
	BLUE4	ROUGE1	ROUGE2	ROUGEL	BLUE4	ROUGE1	ROUGE2	ROUGEL	BLUE4	ROUGE1	ROUGE2	ROUGEL
Attn2Seq	0.5305	12.2800	1.2107	9.1312	0.7889	12.6590	1.6820	9.7481	1.6238	13.2245	2.9942	10.7398
NRT	0.4793	11.0723	1.1304	7.6674	0.8295	12.7815	1.8543	9.9477	1.9084	13.5231	3.6708	11.1867
PETER	0.7112	12.8944	1.3283	9.8635	<u>1.1541</u>	14.8497	<u>2.1413</u>	11.4143	1.9861	14.2716	3.6718	11.7010
P5-S (3-3)	<b>1.0447</b>	<b>14.9048</b>	<b>2.1297</b>	<b>11.1778</b>	<b>1.2237</b>	<b>17.6938</b>	<b>2.2489</b>	<b>12.8606</b>	<u>2.2892</u>	<b>15.4505</b>	<u>3.6974</u>	<b>12.1718</b>
P5-B (3-3)	<u>1.0407</u>	<u>14.1589</u>	<u>2.1220</u>	<u>10.6096</u>	0.9742	<u>16.4530</u>	1.8858	<u>11.8765</u>	<b>2.3185</b>	<u>15.3474</u>	<b>3.7209</b>	<u>12.1312</u>
PETER+	<b>2.4627</b>	<b>24.1181</b>	5.1937	<b>18.4105</b>	<b>3.2606</b>	<u>25.5541</u>	5.9668	<u>19.7168</u>	<b>4.7919</b>	<u>28.3083</u>	9.4520	<b>22.7017</b>
P5-S (3-9)	1.4101	<u>23.5619</u>	<b>5.4196</b>	<u>17.6245</u>	<u>1.9788</u>	<b>25.6253</b>	<b>6.3678</b>	<b>19.9497</b>	4.1222	<b>28.4088</b>	9.5432	<u>22.6064</u>
P5-B (3-9)	<u>1.4689</u>	23.5476	<u>5.3926</u>	17.5852	1.8765	25.1183	6.0764	19.4488	3.8933	27.9916	<u>9.5896</u>	22.2178
P5-S (3-12)	1.3212	23.2474	5.3461	17.3780	1.9425	25.1474	6.0551	19.5601	<u>4.2764</u>	28.1897	9.1327	22.2514
P5-B (3-12)	1.4303	23.3810	5.3239	17.4913	1.9031	25.1763	<u>6.1980</u>	19.5188	3.5861	28.1369	<b>9.7562</b>	22.3056

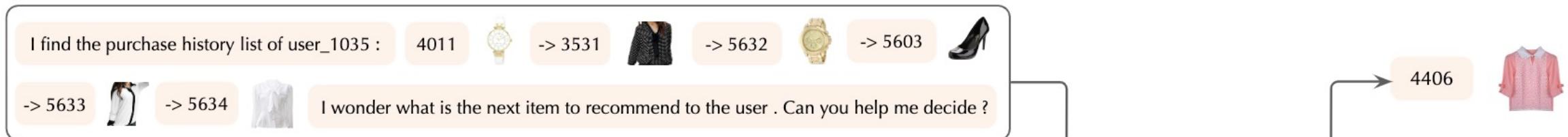
## Direct Recommendation:

Methods	Sports					Beauty					Toys				
	HR@1	HR@5	NDCG@5	HR@10	NDCG@10	HR@1	HR@5	NDCG@5	HR@10	NDCG@10	HR@1	HR@5	NDCG@5	HR@10	NDCG@10
BPR-MF	0.0314	0.1404	0.0848	0.2563	0.1220	0.0311	0.1426	0.0857	0.2573	0.1224	0.0233	0.1066	0.0641	0.2003	0.0940
BPR-MLP	0.0351	0.1520	0.0927	0.2671	0.1296	0.0317	0.1392	0.0848	0.2542	0.1215	0.0252	0.1142	0.0688	0.2077	0.0988
SimpleX	0.0331	<b>0.2362</b>	<b>0.1505</b>	<u>0.3290</u>	<u>0.1800</u>	0.0325	<u>0.2247</u>	<u>0.1441</u>	0.3090	<u>0.1711</u>	0.0268	<b>0.1958</b>	<b>0.1244</b>	<b>0.2662</b>	<b>0.1469</b>
P5-S (5-1)	0.0638	0.2096	0.1375	0.3143	0.1711	0.0600	0.2021	0.1316	<u>0.3121</u>	0.1670	0.0405	<u>0.1538</u>	<u>0.0969</u>	<u>0.2405</u>	0.1248
P5-B (5-1)	0.0245	0.0816	0.0529	0.1384	0.0711	0.0224	0.0904	0.0559	0.1593	0.0780	0.0187	0.0827	0.0500	0.1543	0.0729
P5-S (5-4)	<u>0.0701</u>	<u>0.2241</u>	<u>0.1483</u>	<b>0.3313</b>	<b>0.1827</b>	<b>0.0862</b>	<b>0.2448</b>	<b>0.1673</b>	<b>0.3441</b>	<b>0.1993</b>	0.0413	0.1411	0.0916	0.2227	0.1178
P5-B (5-4)	0.0299	0.1026	0.0665	0.1708	0.0883	0.0506	0.1557	0.1033	0.2350	0.1287	0.0435	0.1316	0.0882	0.2000	0.1102
P5-S (5-5)	0.0574	0.1503	0.1050	0.2207	0.1276	0.0601	0.1611	0.1117	0.2370	0.1360	<u>0.0440</u>	0.1282	0.0865	0.2011	0.1098
P5-B (5-5)	0.0641	0.1794	0.1229	0.2598	0.1488	0.0588	0.1573	0.1089	0.2325	0.1330	0.0386	0.1122	0.0756	0.1807	0.0975
P5-S (5-8)	0.0567	0.1514	0.1049	0.2196	0.1269	0.0571	0.1566	0.1078	0.2317	0.1318	<b>0.0451</b>	0.1322	0.0889	0.2023	0.1114
P5-B (5-8)	<b>0.0726</b>	0.1955	0.1355	0.2802	0.1627	<u>0.0608</u>	0.1564	0.1096	0.2300	0.1332	0.0389	0.1147	0.0767	0.1863	0.0997

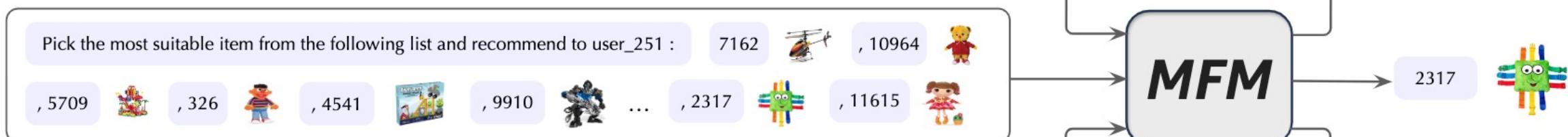
# Easy Handling of Multi-modality Information

- Item images can be directly inserted into prompts

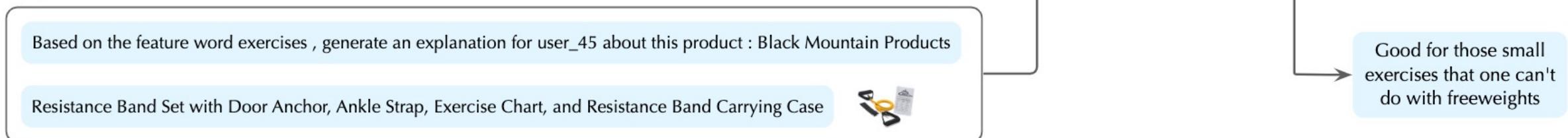
## Sequential Recommendation



## Direct Recommendation

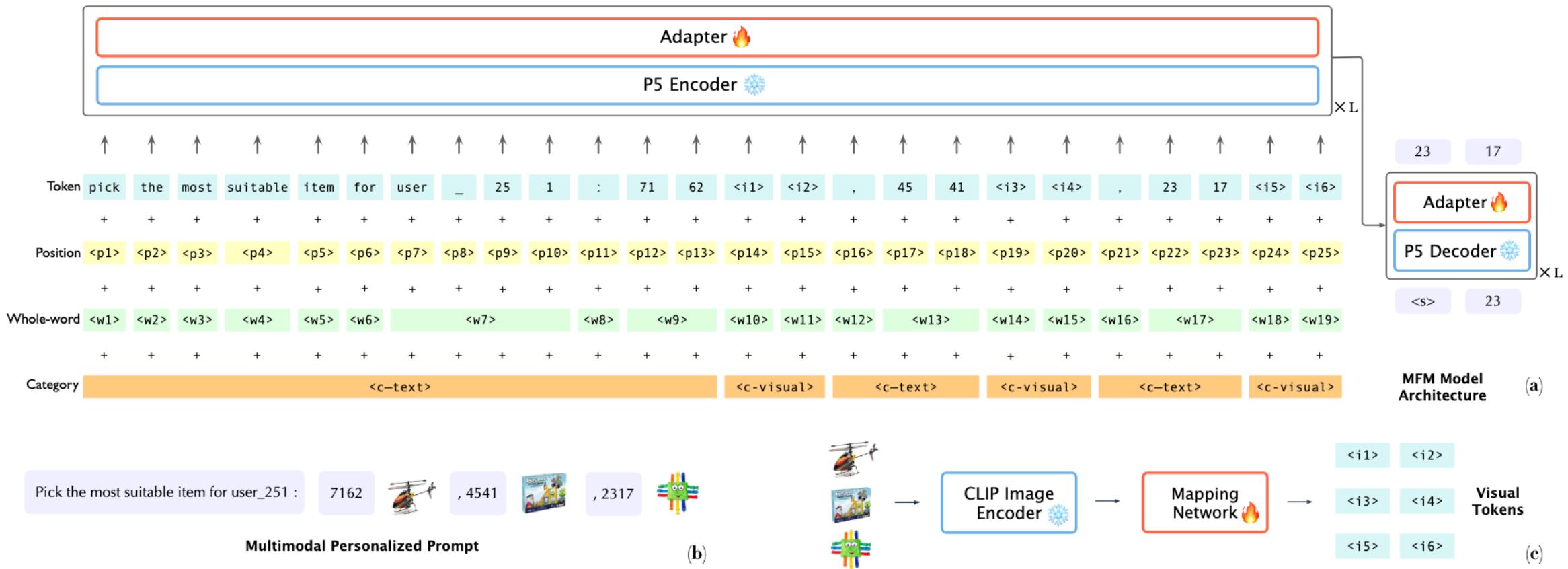


## Explanation Generation



# Easy Handling of Multi-modality Information

- Item images can be directly inserted into prompts



# Easy Handling of Multi-modality Information

- Item images can be directly inserted into prompts
  - Multi-modality information further improves performance

Methods	Sports				Beauty			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
HGN	0.0189	0.0120	0.0313	0.0159	0.0325	0.0206	0.0512	0.0266
SASRec	0.0233	0.0154	0.0350	0.0192	0.0387	0.0249	0.0605	0.0318
$S^3$ -Rec	0.0251	0.0161	0.0385	0.0204	0.0387	0.0244	0.0647	0.0327
P5 (A-3)	0.0272	0.0169	0.0361	0.0198	0.0503	0.0370	0.0659	0.0421
MFM (A-3)	<b>0.0412</b>	<b>0.0345</b>	<b>0.0475</b>	<b>0.0365</b>	<b>0.0556</b>	<b>0.0427</b>	<b>0.0677</b>	<b>0.0467</b>
P5 (A-9)	0.0258	0.0159	0.0346	0.0188	0.0490	0.0358	0.0646	0.0409
MFM (A-9)	0.0392	0.0327	0.0456	0.0347	0.0529	0.0413	0.0655	0.0454

Methods	Clothing				Toys			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
HGN	0.0107	0.0071	0.0175	0.0092	0.0321	0.0221	0.0497	0.0277
SASRec	0.0107	0.0066	0.0194	0.0095	0.0463	0.0306	0.0675	0.0374
$S^3$ -Rec	0.0076	0.0045	0.0135	0.0063	0.0443	0.0294	0.0700	0.0376
P5 (A-3)	0.0478	0.0376	0.0554	0.0401	0.0655	0.0570	0.0726	0.0593
MFM (A-3)	<b>0.0603</b>	<b>0.0564</b>	<b>0.0632</b>	<b>0.0573</b>	<b>0.0662</b>	<b>0.0577</b>	<b>0.0749</b>	<b>0.0604</b>
P5 (A-9)	0.0455	0.0359	0.0534	0.0385	0.0631	0.0547	0.0701	0.0569
MFM (A-9)	0.0569	0.0531	0.0597	0.0540	0.0641	0.0556	0.0716	0.0580

Sequential Recommendation Performance

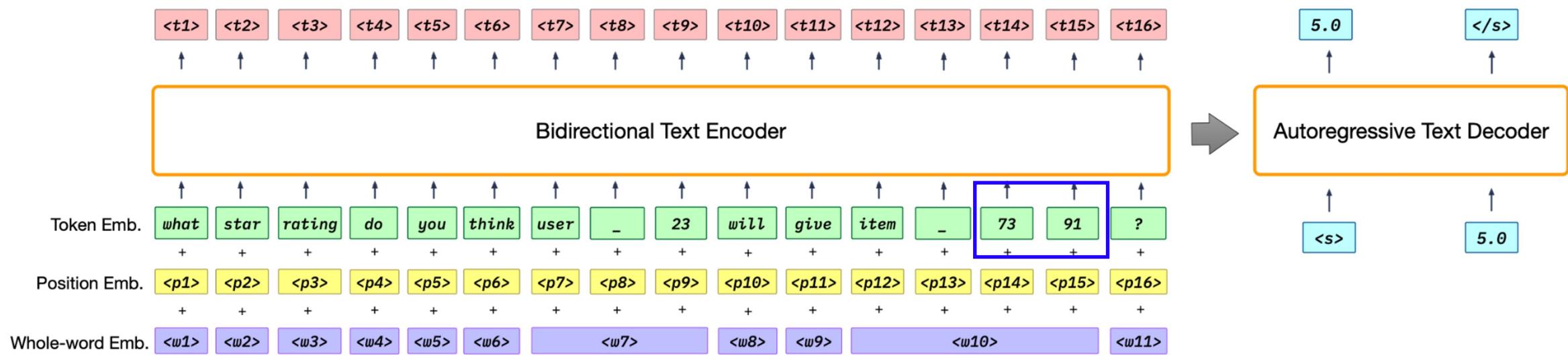
Methods	Sports					Beauty				
	HR@1	HR@5	NDCG@5	HR@10	NDCG@10	HR@1	HR@5	NDCG@5	HR@10	NDCG@10
BPR-MF	0.0314	0.1404	0.0848	0.2563	0.1220	0.0311	0.1426	0.0857	<b>0.2573</b>	0.1224
BPR-MLP	0.0351	0.1520	0.0927	0.2671	0.1296	0.0317	0.1392	0.0848	0.2542	0.1215
VBPR	0.0262	0.1138	0.0691	0.2060	0.0986	0.0380	0.1472	0.0925	0.2468	0.1245
P5 (B-5)	0.0574	0.1503	0.1050	0.2207	0.1276	0.0601	0.1611	0.1117	0.2370	0.1360
MFM (B-5)	0.0606	0.1743	0.1185	0.2539	0.1441	0.0580	0.1598	0.1099	0.2306	0.1327
P5 (B-8)	0.0567	0.1514	0.1049	0.2196	0.1269	0.0571	0.1566	0.1078	0.2317	0.1318
MFM (B-8)	<b>0.0699</b>	<b>0.1882</b>	<b>0.1304</b>	<b>0.2717</b>	<b>0.1572</b>	<b>0.0615</b>	<b>0.1655</b>	<b>0.1147</b>	0.2407	<b>0.1388</b>

Methods	Clothing					Toys				
	HR@1	HR@5	NDCG@5	HR@10	NDCG@10	HR@1	HR@5	NDCG@5	HR@10	NDCG@10
BPR-MF	0.0296	0.1280	0.0779	0.2319	0.1112	0.0233	0.1066	0.0641	0.2003	0.0940
BPR-MLP	0.0342	0.1384	0.0858	0.2327	0.1161	0.0252	0.1142	0.0688	0.2077	0.0988
VBPR	0.0352	0.1410	0.0877	<b>0.2420</b>	0.1201	0.0337	0.1294	0.0808	<b>0.2199</b>	0.1098
P5 (B-5)	0.0320	0.0986	0.0652	0.1659	0.0867	0.0418	0.1219	0.0824	0.1942	0.1056
MFM (B-5)	0.0481	0.1287	0.0890	0.1992	0.1116	0.0428	0.1225	0.0833	0.1906	0.1051
P5 (B-8)	0.0355	0.1019	0.0688	0.1722	0.0912	0.0422	0.1286	0.0858	0.2041	0.1099
MFM (B-8)	<b>0.0552</b>	<b>0.1544</b>	<b>0.1058</b>	0.2291	<b>0.1297</b>	<b>0.0433</b>	<b>0.1301</b>	<b>0.0875</b>	0.2037	<b>0.1110</b>

Direct Recommendation Performance

# How to Index Items

- Item ID: item needs to be represented as a sequence of tokens
  - e.g., an item represented by two tokens <73> <91>



- Different item indexing gives very different performance

# How to Index Items (create Item IDs)

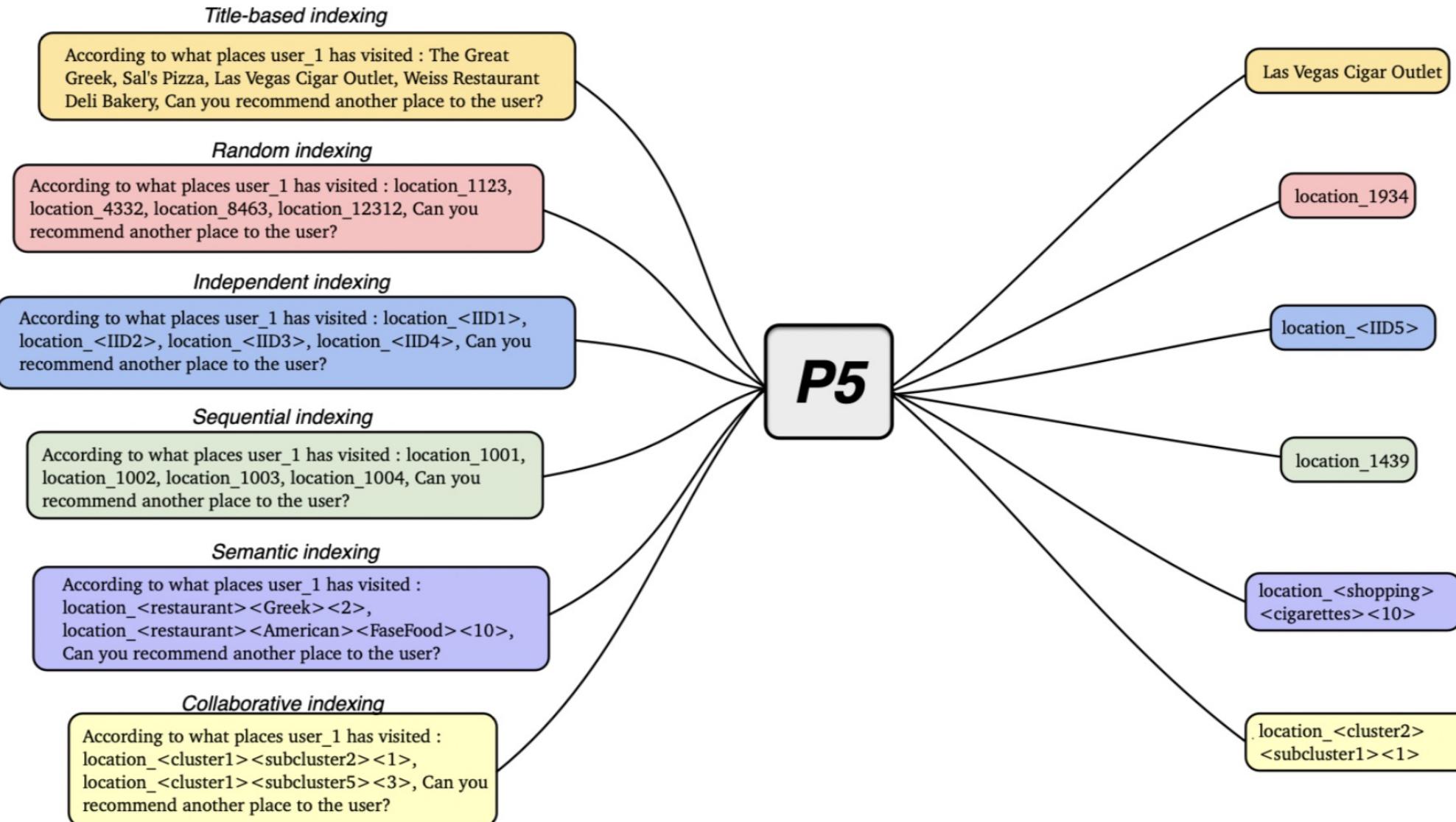
- Three properties for good item indexing methods
  - Items are distinguishable (different items have different IDs)
  - Similar items have similar IDs (more shared tokens in their IDs)
  - Dissimilar items have dissimilar IDs (less shared tokens in their IDs)
- Three naïve Indexing methods
  - Random ID (RID): Item <73><91>, item <73><12>, ...
  - Title as ID (TID): Item <the><lord><of><the><rings>, ...
  - Independent ID (IID): Item <1364>, Item <6321>, ...

Method	Amazon Sports				Amazon Beauty				Yelp			
	HR@5	NCDG@5	HR@10	NCDG@10	HR@5	NCDG@5	HR@10	NCDG@10	HR@5	NCDG@5	HR@10	NCDG@10
SASRec	0.0233	0.0154	0.0350	0.0192	0.0387	0.0249	0.0605	0.0318	0.0170	0.0110	0.0284	0.0147
S <sup>3</sup> -Rec	0.0251	<b>0.0161</b>	0.0385	<b>0.0204</b>	0.0387	0.0244	<b>0.0647</b>	0.0327	0.0201	0.0123	0.0341	0.0168
RID	0.0208	0.0122	0.0288	0.0153	0.0213	0.0178	0.0479	0.0277	0.0225	<b>0.0159</b>	0.0329	0.0193
TID	0.0000	0.0000	0.0000	0.0000	0.0182	0.0132	0.0432	0.0254	0.0058	0.0040	0.0086	0.0049
IID	<b>0.0268</b>	0.0151	<b>0.0386</b>	0.0195	<b>0.0394</b>	<b>0.0268</b>	0.0615	<b>0.0341</b>	<b>0.0232</b>	0.0146	<b>0.0393</b>	<b>0.0197</b>

# How to Index Items (create Item IDs)

- Three naïve Indexing methods
  - Random ID (RID): Item ⟨73⟩⟨91⟩, item ⟨73⟩⟨12⟩, ...
    - Very different items may share the same tokens
    - Mistakenly making them semantically similar
  - Title as ID (TID): Item ⟨the⟩⟨lord⟩⟨of⟩⟨the⟩⟨rings⟩
    - Very different movies may share similar titles
      - **Inside Out** (animation) and **Inside Job** (documentary)
      - **The Lord of the Rings** (epic fantasy) and **The Lord of War** (crime drama)
  - Independent ID (IID): Item ⟨1364⟩, Item ⟨6321⟩, ...
    - Too many out-of-vocabulary (OOV) new tokens need to learn
    - Computationally unscalable

# Meticulous Item Indexing Methods are Needed



# Sequential Indexing (SID)

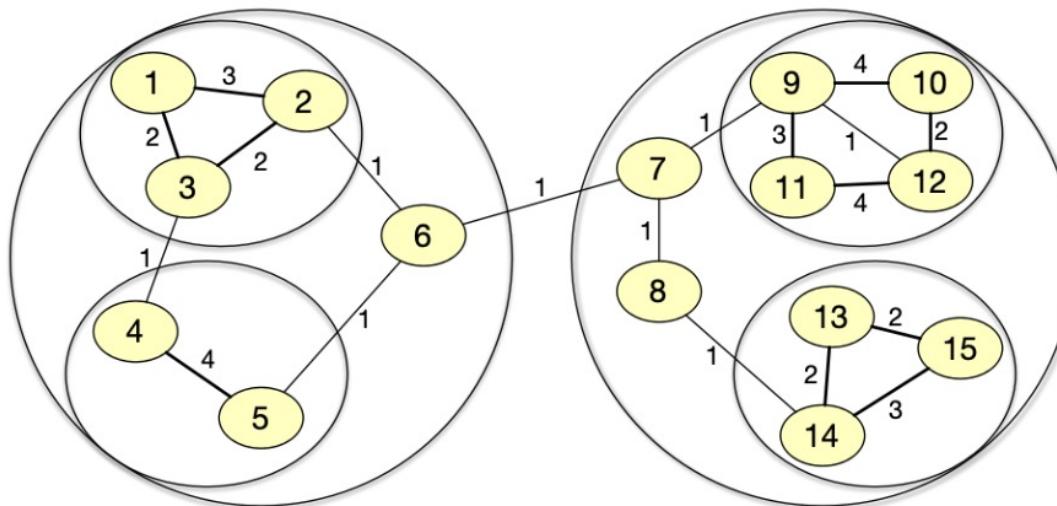
- Leverage the **local** co-appearance information between items

Training Sequence											Validation	Testing
User 1	1001	1002	1003	1004	1005	1006	1007	1008	1009		1018	1019
User 2	1010	1011	1001	1012	1008	1009	1013	1014			1022	1023
User 3	1015	1016	1017	1007	1018	1019	1020	1021	1009		1015	1016
User 4	1022	1023	1005	1002	1006	1024					1002	1008
User 5	1025	1026	1027	1028	1029	1030	1024	1020	1021	1031	1033	1034

- After tokenization, co-appearing items share similar tokens
  - Item 1004:  $\langle 10 \rangle \langle 04 \rangle$
  - Item 1005:  $\langle 10 \rangle \langle 05 \rangle$

# Collaborative Indexing (CID)

- Leverage the **global** co-appearance information between items
  - Spectral Matrix Factorization over the item-item co-appearance matrix
  - Hierarchical Spectral Clustering



(a) Recursive spectral clustering on item co-appearance graph

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ \hline 1 & \left[ \begin{array}{cccccc} 0 & 3 & 2 & 0 & 0 & 0 & \dots \\ 3 & 0 & 2 & 0 & 0 & 1 & \dots \\ 2 & 2 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 4 & 0 & \dots \\ 0 & 0 & 0 & 4 & 0 & 1 & \dots \\ 0 & 1 & 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right] \end{array}$$

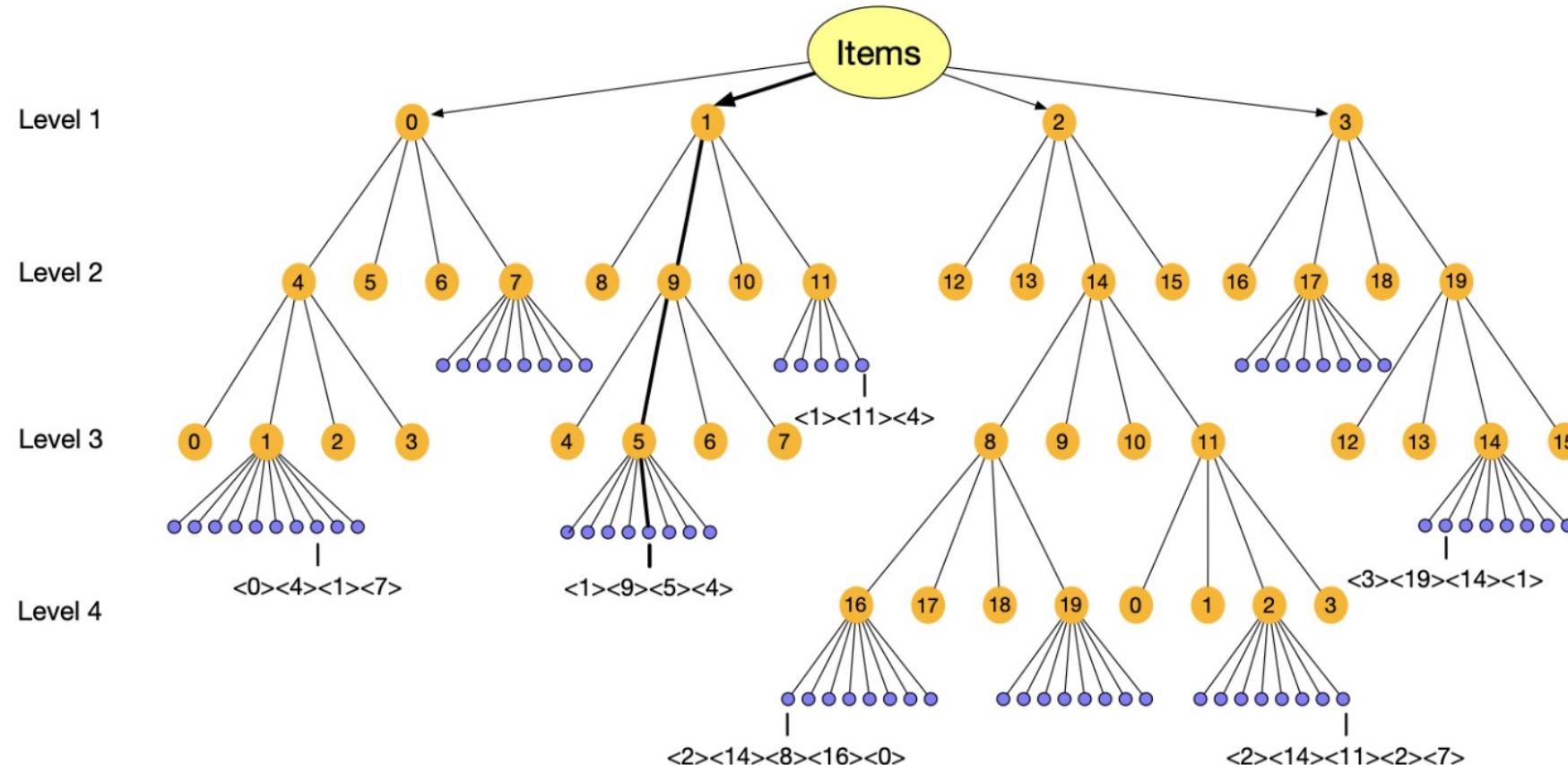
(b) Adjacency matrix

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ \hline 1 & \left[ \begin{array}{cccccc} 5 & -3 & -2 & 0 & 0 & 0 & \dots \\ -3 & 6 & -2 & 0 & 0 & -1 & \dots \\ -2 & -2 & 6 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 5 & -4 & 0 & \dots \\ 0 & 0 & 0 & -4 & 5 & -1 & \dots \\ 0 & 1 & 0 & 0 & -1 & 2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right] \end{array}$$

(c) Laplacian matrix

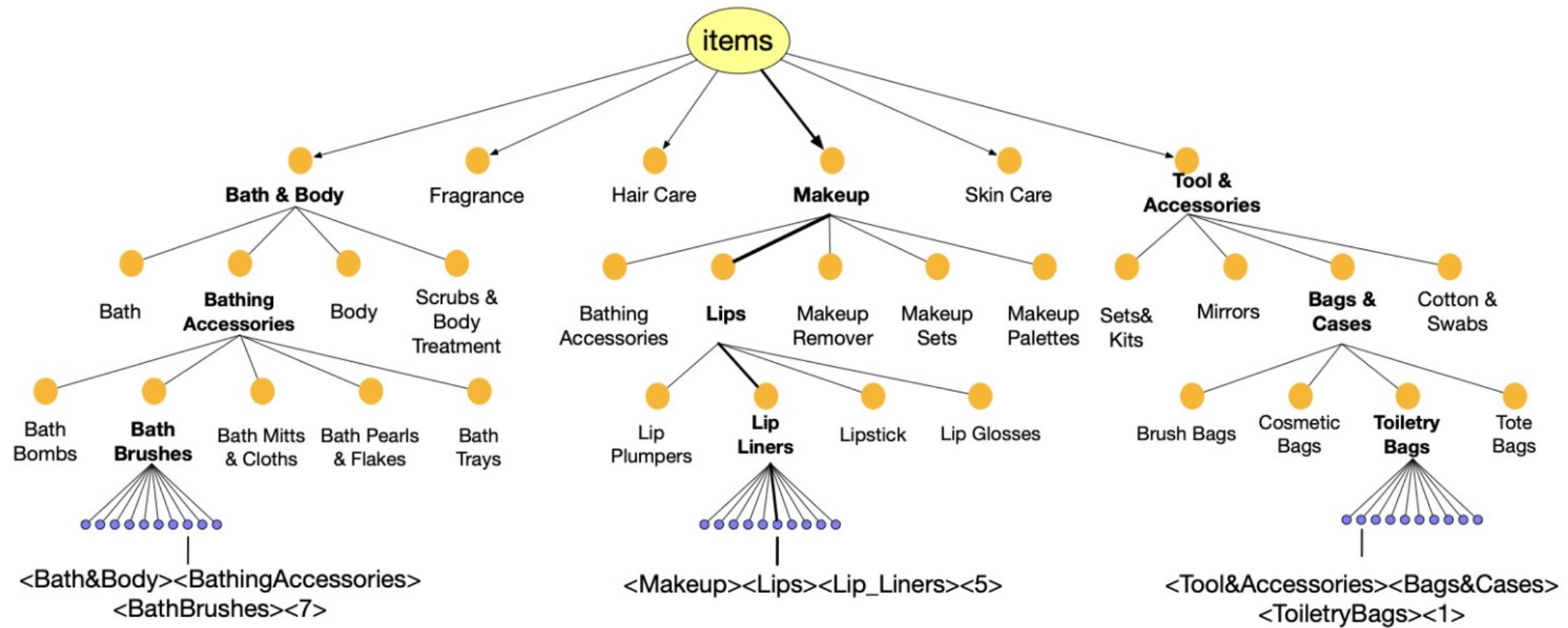
# Collaborative Indexing (CID)

- Leverage the **global co-appearance information between items**
  - Root-to-Leaf Path-based Indexing
  - Items in the same cluster share more tokens



# Semantic (Content-based) Indexing (SemID)

- Leverage the **item content information** for item indexing
  - e.g., the **multi-level item category** information in Amazon



# Hybrid Indexing (HID)

- Concatenate more than one of the following indices
  - Random ID (RID)
  - Title as ID (TID)
  - Independent ID (IID)
  - Sequential ID (SID)
  - Collaborative ID (CID)
  - Semantic ID (SemID)
- For example, if an item's Semantic ID and Collaborative ID are as follows:
  - SemID: <Makeup><Lips><Lip\_Liners><5>
  - CID: <1><9><5><4>
- Then its Hybrid ID is <Makeup><Lips><Lip\_Liners><1><9><5><4>

# Different Item Indexing Gives Different Performance

Method	Amazon Sports				Amazon Beauty				Yelp				
	HR@5 NCDG@5		HR@10 NCDG@10		HR@5 NCDG@5		HR@10 NCDG@10		HR@5 NCDG@5		HR@10 NCDG@10		
Naïve indexing methods	Caser	0.0116	0.0072	0.0194	0.0097	0.0205	0.0131	0.0347	0.0176	0.015	0.0099	0.0263	0.0134
	HGN	0.0189	0.0120	0.0313	0.0159	0.0325	0.0206	0.0512	0.0266	0.0186	0.0115	0.0326	0.159
	GRU4Rec	0.0129	0.0086	0.0204	0.0110	0.0164	0.0099	0.0283	0.0137	0.0176	0.0110	0.0285	0.0145
	BERT4Rec	0.0115	0.0075	0.0191	0.0099	0.0203	0.0124	0.0347	0.0170	0.0051	0.0033	0.0090	0.0090
	FDSA	0.0182	0.0122	0.0288	0.0156	0.0267	0.0163	0.0407	0.0208	0.0158	0.0098	0.0276	0.0136
	SASRec	0.0233	0.0154	0.0350	0.0192	0.0387	0.0249	0.0605	0.0318	0.0170	0.0110	0.0284	0.0147
	S <sup>3</sup> -Rec	0.0251	0.0161	0.0385	0.0204	0.0387	0.0244	0.0647	0.0327	0.0201	0.0123	0.0341	0.0168
Advanced indexing methods	RID	0.0208	0.0122	0.0288	0.0153	0.0213	0.0178	0.0479	0.0277	<u>0.0225</u>	<u>0.0159</u>	<u>0.0329</u>	<u>0.0193</u>
	TID	0.000	0.000	0.000	0.000	0.0182	0.0132	0.0432	0.0254	<u>0.0058</u>	<u>0.0040</u>	<u>0.0086</u>	<u>0.0049</u>
	IID	<u>0.0268</u>	<u>0.0151</u>	<u>0.0386</u>	<u>0.0195</u>	<u>0.0394</u>	<u>0.0268</u>	<u>0.0615</u>	<u>0.0341</u>	<u>0.0232</u>	<u>0.0146</u>	<u>0.0393</u>	<u>0.0197</u>
Hybrid indexing methods	SID	<u>0.0264</u>	<u>0.0186</u>	<u>0.0358</u>	<u>0.0216</u>	<u>0.0430</u>	<u>0.0288</u>	<u>0.0602</u>	<u>0.0368</u>	<b>0.0346</b>	<b>0.0242</b>	<b>0.0486</b>	<b>0.0287</b>
	CID	<u>0.0313</u>	<u>0.0224</u>	<u>0.0431</u>	<u>0.0262</u>	<u>0.0489</u>	<u>0.0318</u>	<u>0.0680</u>	<u>0.0357</u>	<u>0.0261</u>	<u>0.0171</u>	<u>0.0428</u>	<u>0.0225</u>
	SemID	<u>0.0274</u>	<u>0.0193</u>	<u>0.0406</u>	<u>0.0235</u>	<u>0.0433</u>	<u>0.0299</u>	<u>0.0652</u>	<u>0.0370</u>	<u>0.0202</u>	<u>0.0131</u>	<u>0.0324</u>	<u>0.0170</u>
	SID+IID	0.0235	0.0161	0.0339	0.0195	<u>0.0420</u>	<u>0.0297</u>	0.0603	<u>0.0355</u>	<u>0.0329</u>	<u>0.0236</u>	<u>0.0465</u>	<u>0.0280</u>
	CID+IID	<b>0.0321</b>	<b>0.0227</b>	<b>0.0456</b>	<b>0.0270</b>	<b>0.0512</b>	<b>0.0356</b>	<b>0.0732</b>	<b>0.0427</b>	<u>0.0287</u>	<u>0.0195</u>	<u>0.0468</u>	<u>0.0254</u>
	SemID+IID	<u>0.0291</u>	<u>0.0196</u>	<u>0.0436</u>	<u>0.0242</u>	<u>0.0501</u>	<u>0.0344</u>	<u>0.0724</u>	<u>0.0411</u>	<u>0.0229</u>	<u>0.0150</u>	<u>0.0382</u>	<u>0.0199</u>
	SemID+CID	0.0043	0.0031	0.0070	0.0039	0.0355	0.0248	0.0545	0.0310	0.0021	0.0016	0.0056	0.0029

- Advanced indexing methods are better than naïve methods
- Hybrid indexing can further improve performance

# The Future of Generative Recommendation

- Recommendation as **Personalized On-demand Generation**
  - Recommend **existing items** vs. recommend newly generated items
    - Traveling in Hawaii, want to make a post on Instagram
    - **Personalized generation of candidate images** for users to consider



# The Future of Generative Recommendation

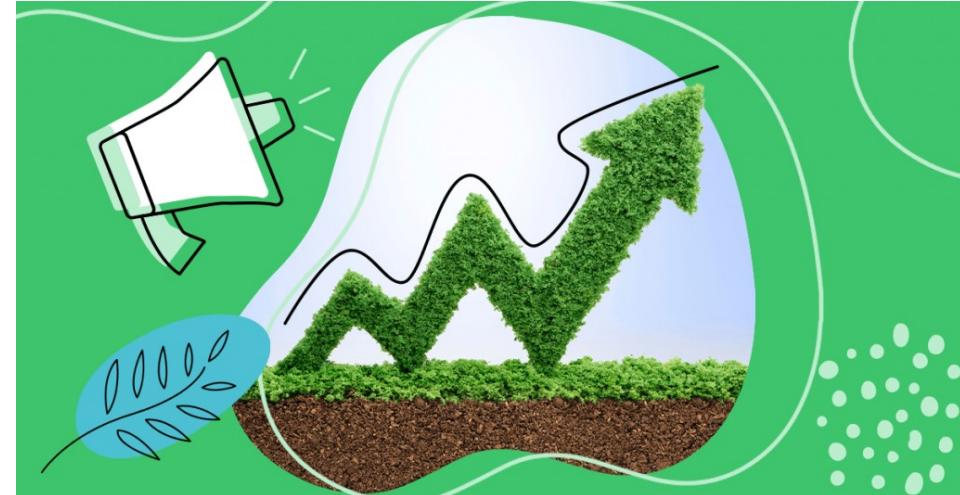
- Recommendation as Personalized On-demand Generation
  - Personalized Advertisement Generation
  - Same ad, different wording, real-time generation given user's context
    - e.g., an environmental protection ad for an NGO

For Children:



Join us in protecting our planet. Let's work together to make the world a better place for ourselves and for future generations.

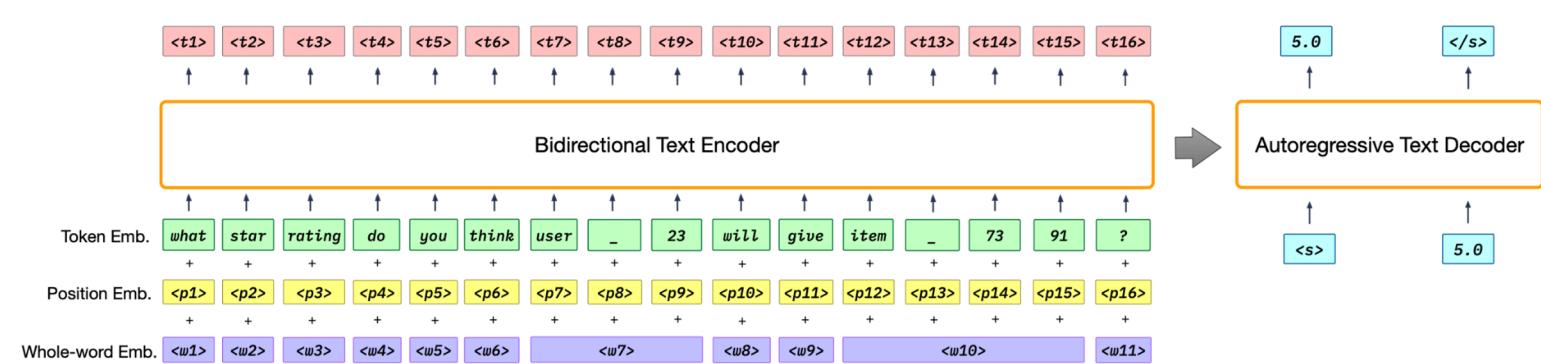
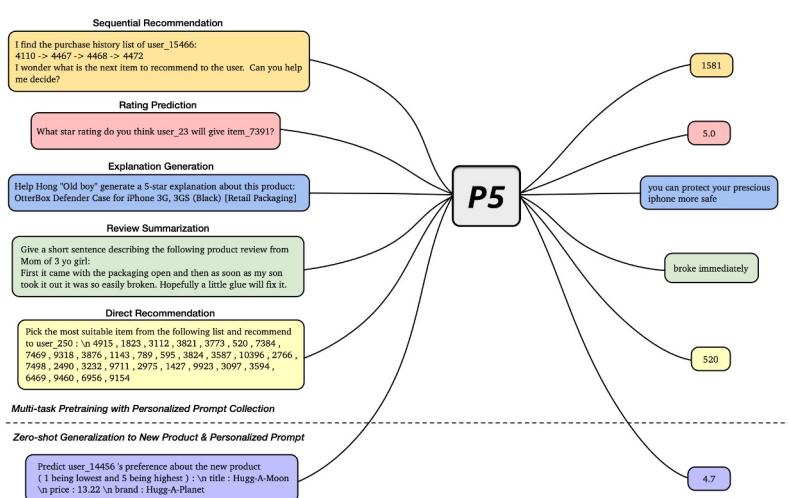
For Business Leaders:



Join the movement towards sustainability and create a brighter future for your business and our planet. By adopting environmentally-friendly practices, you can reduce your costs, attract new customers, and enhance your reputation as a responsible business leader.

# Summary

- From Discriminative Recommendation to Generative Recommendation
  - From multi-stage ranking to single-stage ranking
  - Multi-task learning with the same foundation model
  - Easily handle multi-modality data
  - Various item indexing methods for recommendation foundation models
  - Recommendation as Personalized On-demand Generation
    - From recommending existing items to recommending newly generated items





Yongfeng Zhang  
Department of Computer Science, Rutgers University  
[yongfeng.zhang@rutgers.edu](mailto:yongfeng.zhang@rutgers.edu)  
<http://yongfeng.me>

# References

- [1] Geng, Shijie, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. "Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)." RecSys 2022.
- [2] Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 42, no. 8 (2009): 30-37.
- [3] Cheng, Heng-Tze, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson et al. "Wide & deep learning for recommender systems." DLRS 2016.
- [4] Jiang, Biye, Pengye Zhang, Rihan Chen, Xincheng Luo, Yin Yang, Guan Wang, Guorui Zhou, Xiaoqiang Zhu, and Kun Gai. "DCAF: A Dynamic Computation Allocation Framework for Online Serving System." DLP-KDD 2020.
- [5] Covington, Paul, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations." In *Proceedings of the 10th ACM conference on recommender systems*, pp. 191-198. 2016.
- [6] Chen, Hanxiong, Shaoyun Shi, Yunqi Li, and Yongfeng Zhang. "Neural collaborative reasoning." In *Proceedings of the Web Conference 2021*, pp. 1516-1527. 2021.
- [7] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
- [8] Zhang, Yongfeng, and Xu Chen. "Explainable recommendation: A survey and new perspectives." *Foundations and Trends® in Information Retrieval* 14, no. 1 (2020): 1-101.

- [9] Li, Lei, Yongfeng Zhang, and Li Chen. "Personalized prompt learning for explainable recommendation." ACM Transactions on Information Systems 41, no. 4 (2023): 1-26.
- [10] Geng, Shijie, Jianbo Yuan, Yu Tian, Yuxiao Chen, and Yongfeng Zhang. "HiCLIP: Contrastive language-image pretraining with hierarchy-aware attention." ICLR 2023.
- [11] Ji, Jianchao, Zelong Li, Shuyuan Xu, Max Xiong, Juntao Tan, Yingqiang Ge, Hao Wang, and Yongfeng Zhang. "Counterfactual Collaborative Reasoning." In Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, pp. 249-257. 2023.
- [12] Geng, Shijie, Zuohui Fu, Juntao Tan, Yingqiang Ge, Gerard De Melo, and Yongfeng Zhang. "Path language modeling over knowledge graphs for explainable recommendation." In Proceedings of the ACM Web Conference 2022, pp. 946-955. 2022.
- [13] Li, Lei, Yongfeng Zhang, and Li Chen. "Personalized Transformer for Explainable Recommendation." In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 4947-4957. 2021.
- [14] Li, Lei, Yongfeng Zhang, and Li Chen. "Generate neural template explanations for recommendation." In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 755-764. 2020.
- [15] Zhang, Yongfeng, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis." In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, pp. 83-92. 2014.