Self-Expressive Outfit Recommendation via Transformer-Based Compatibility Prediction

Masataka Miyawaki　Issei Terada　○Yoshitaka Kameya
Meijo University

Outline

- Background
- Proposed Method
- Experiments
- Conclusion

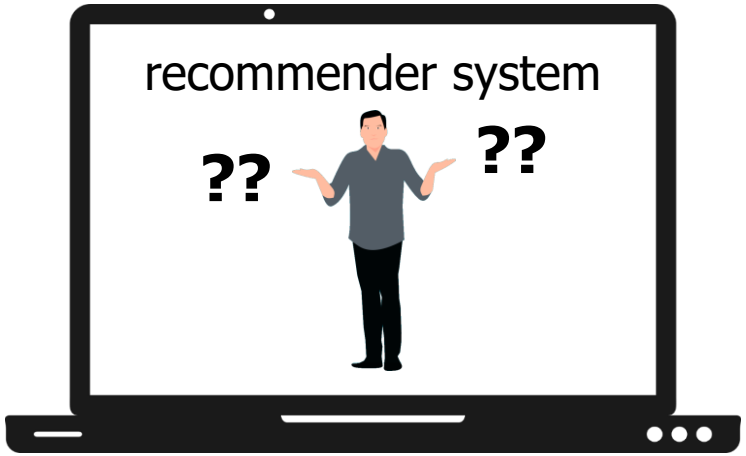
Outline

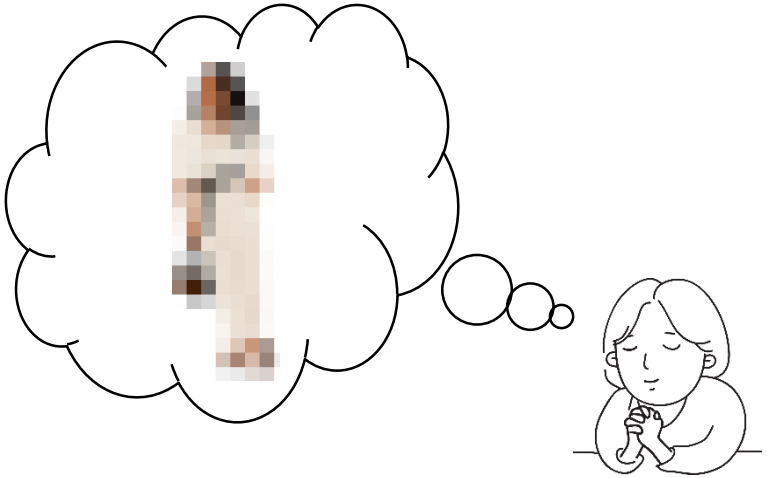
- Background
- Proposed Method
- Experiments
- Conclusion

Background (1)

- Fashion is one of the most powerful vehicles for self-expression
  - Our life style
  - What we want to be

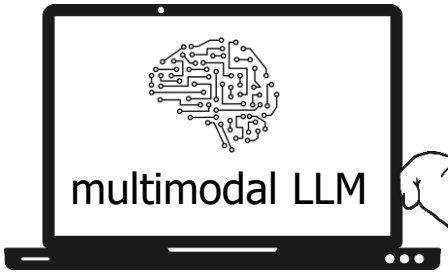
- **This work:** We study an outfit recommender system that considers the users' desire for self-expression
  - **Problem:** Such desire is often ambiguous





Background (2)

- We can use multimodal LLMs to capture the users' ambiguous desire
- **Example:**



(Recommended by GPT-4o)

- **Problems:**
  - Fashion items in the recommended outfit are not real ones
  - We are not sure if multimodal LLMs have learned about visual compatibility among fashion items

Background (3)

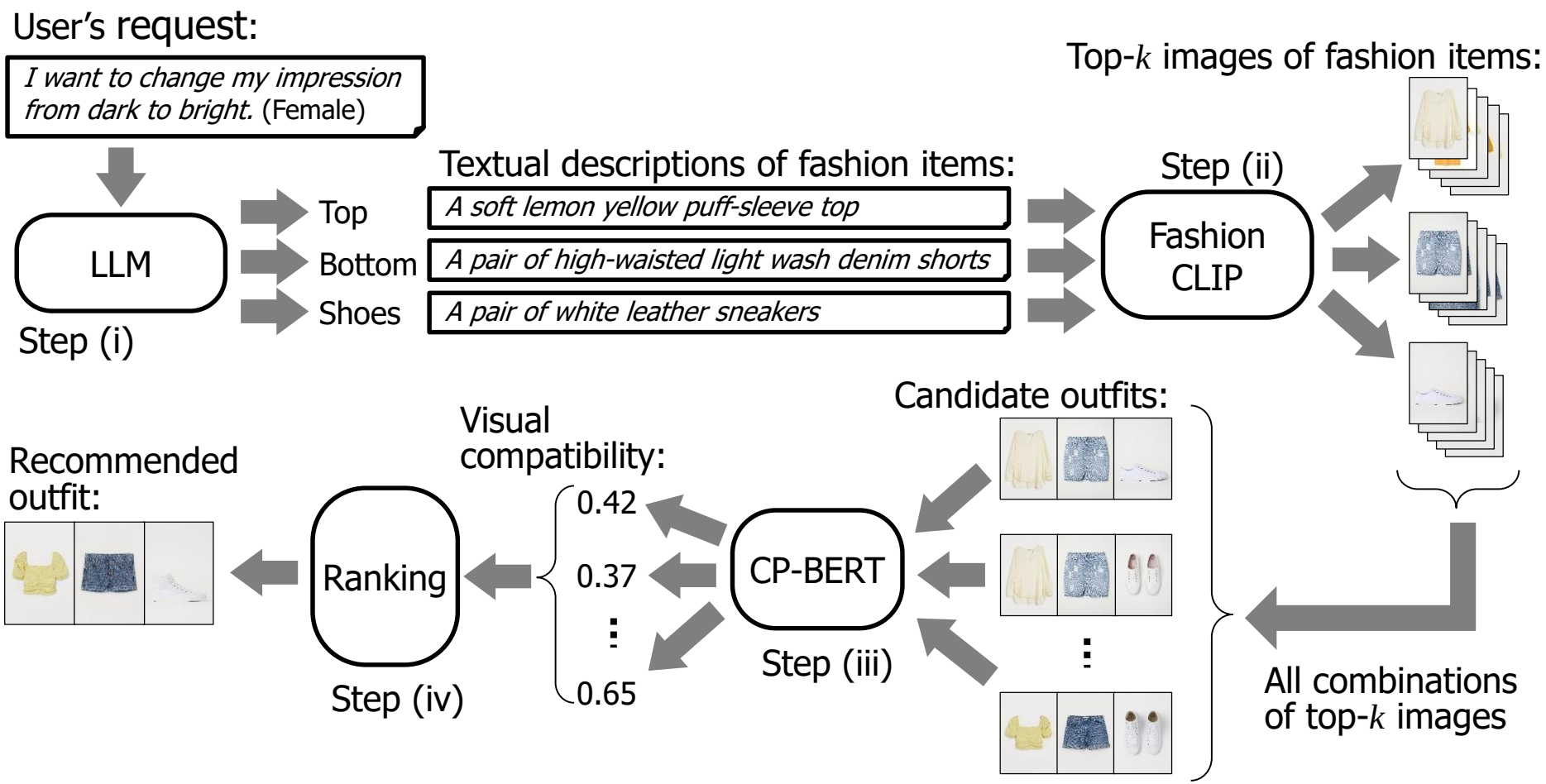
- **This work**: We propose an outfit recommender system that:
  - Captures the users' ambiguous desire for self-expression
  - Presents outfits including real fashion items
  - Presents visually compatible outfits

Outline

- ✓ Background
- Proposed Method
- Experiments
- Conclusion

Proposed Method (1)

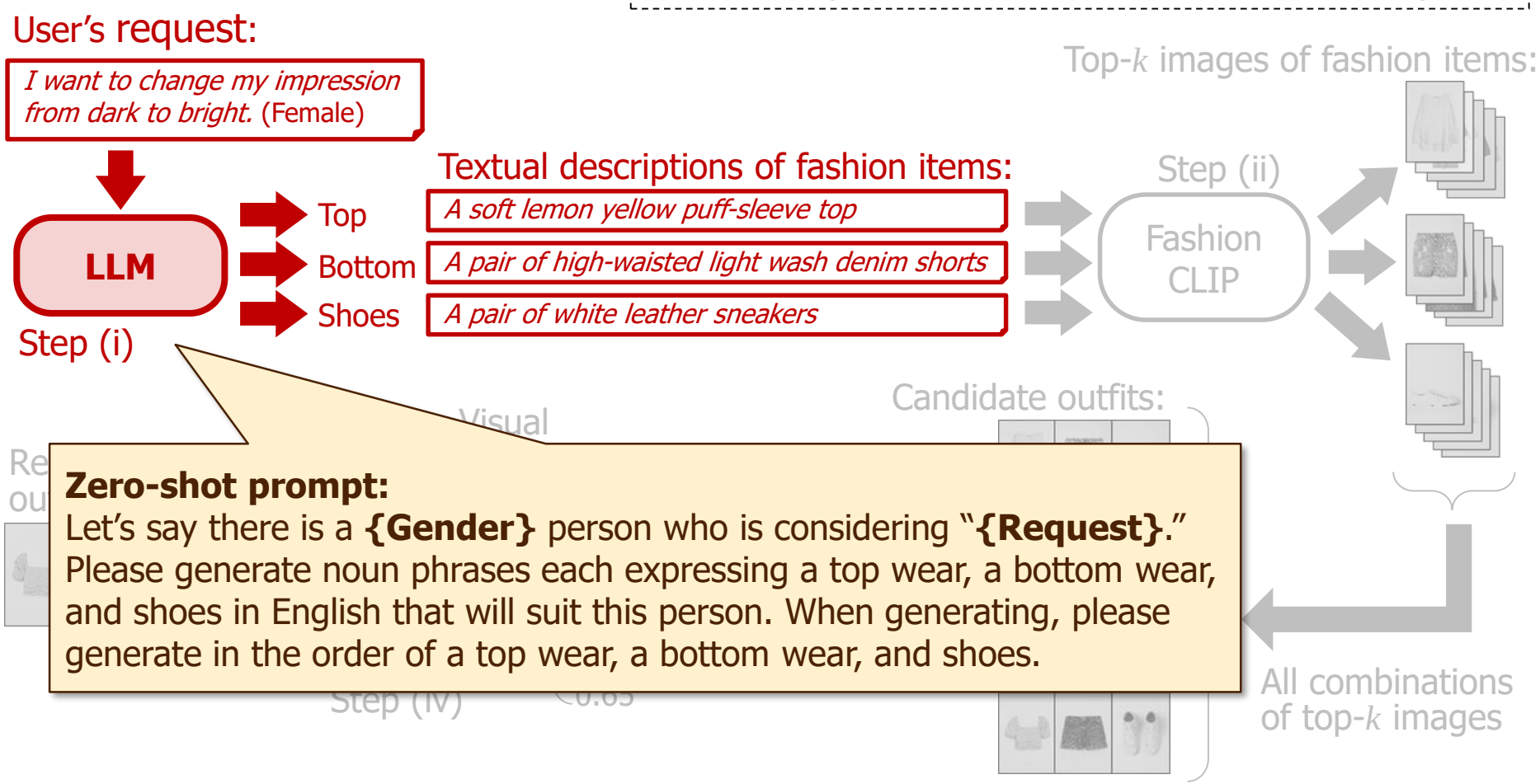
- Overall architecture of the proposed method:



Proposed Method (2)

- **Step (i):** Rewriting the user's request into some textual descriptions of fashion items by an **LLM** (Query rewriting[1])

Assuming LLMs have world knowledge[2]



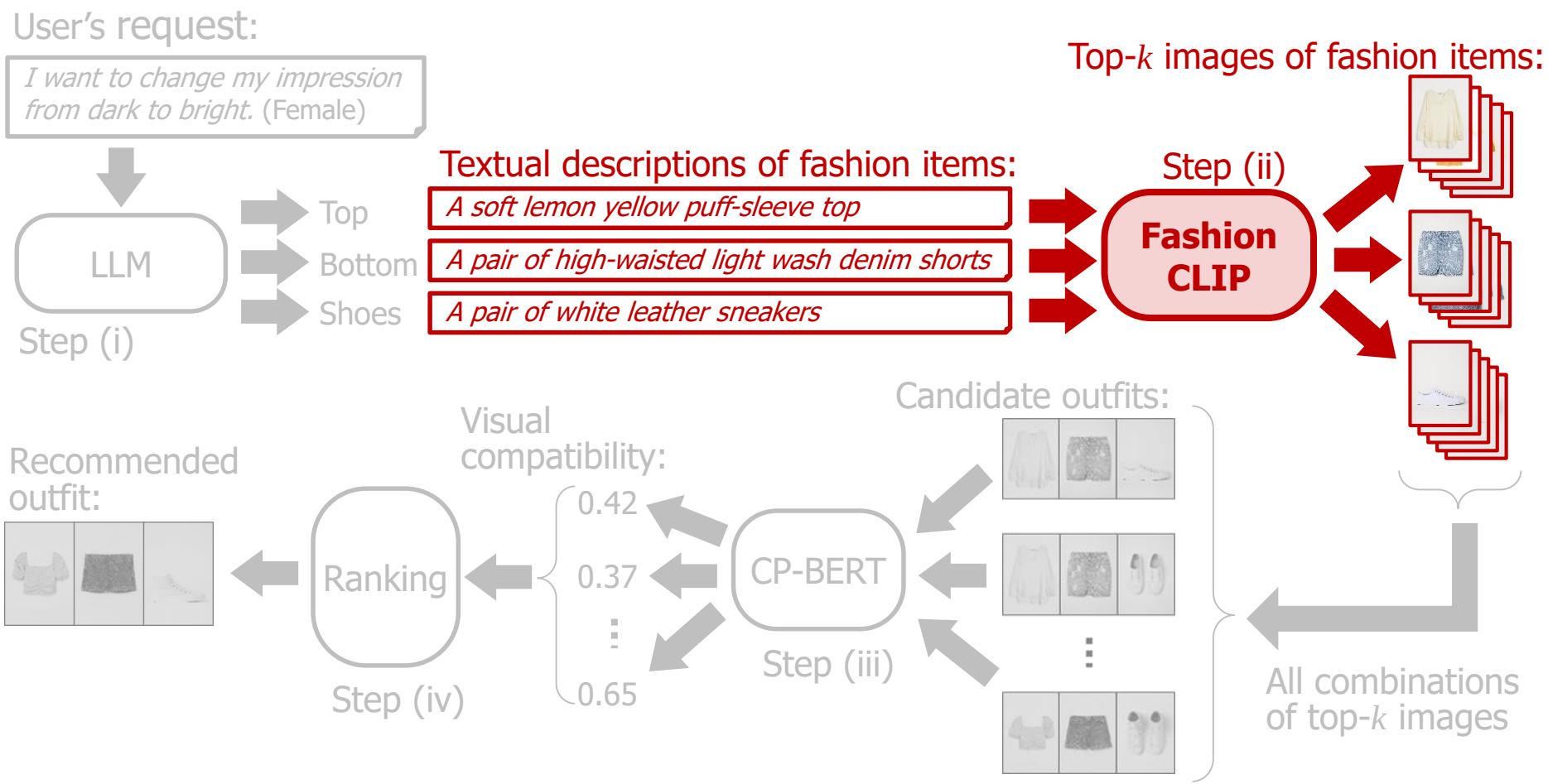
Zero-shot prompt:
Let's say there is a **{Gender}** person who is considering "**{Request}**." Please generate noun phrases each expressing a top wear, a bottom wear, and shoes in English that will suit this person. When generating, please generate in the order of a top wear, a bottom wear, and shoes.

[1] W. Peng, et al.: Large language model based long-tail query rewriting in Taobao search, WWW-24.
[2] C. D. Manning: Human language understanding & reasoning, Daedalus, 2022.

Proposed Method (3)

- **Step (ii):** Converting each fashion item description into top-k images of real fashion items by **FashionCLIP**

FashionCLIP: CLIP fine-tuned with Farfetch dataset[3]

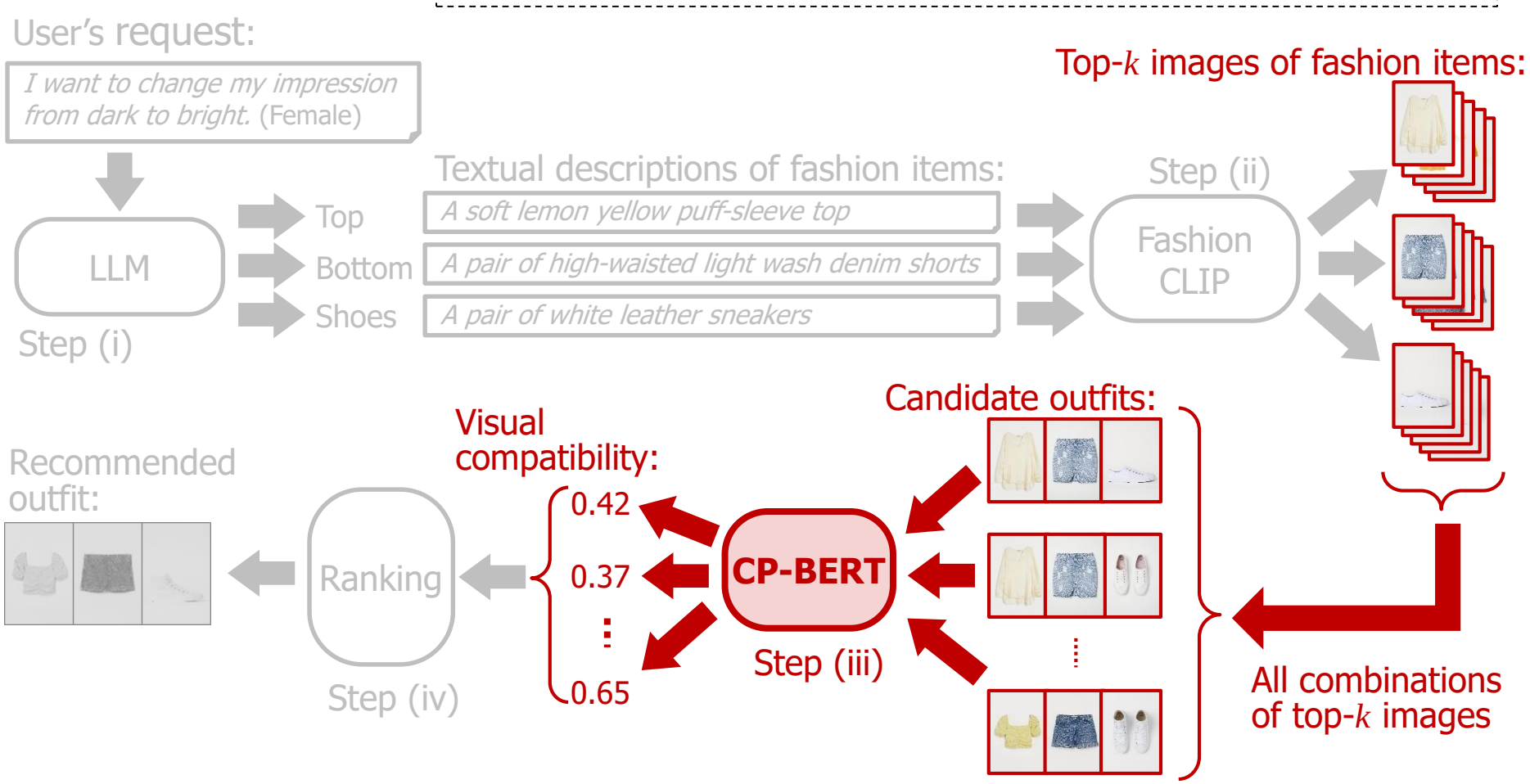


[3] P. J. Chia, et al.: Contrastive language and vision learning of general fashion concepts, Scientific Reports, 2022

Proposed Method (4)

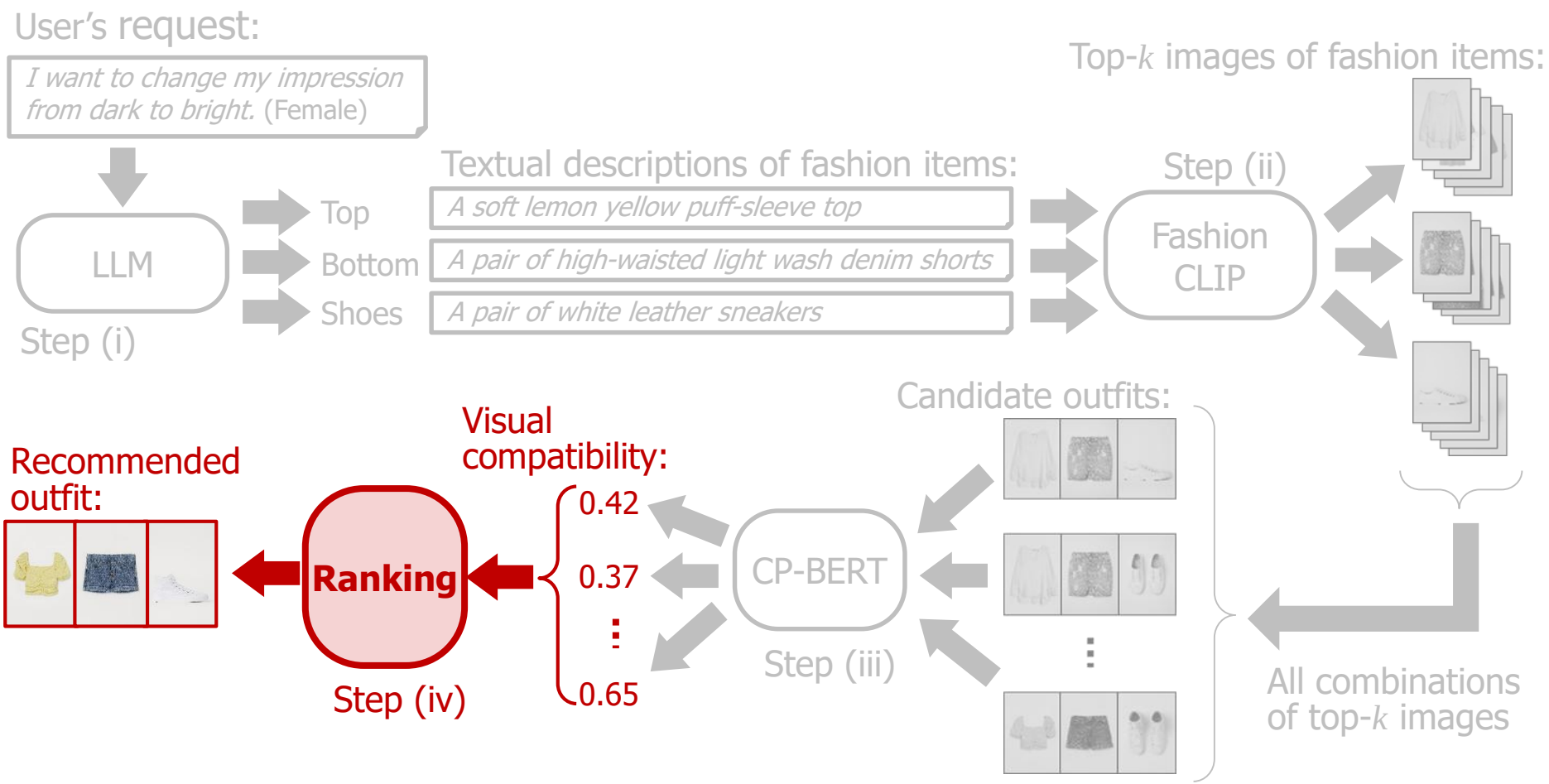
- **Step (iii):** Measuring the visual compatibility of each candidate outfit (a combination of top-k item images) by **CP-BERT**

CP-BERT: BERT specialized for compatibility prediction



Proposed Method (5)

- **Step (iv):** Ranking all candidate outfits based on visual compatibility



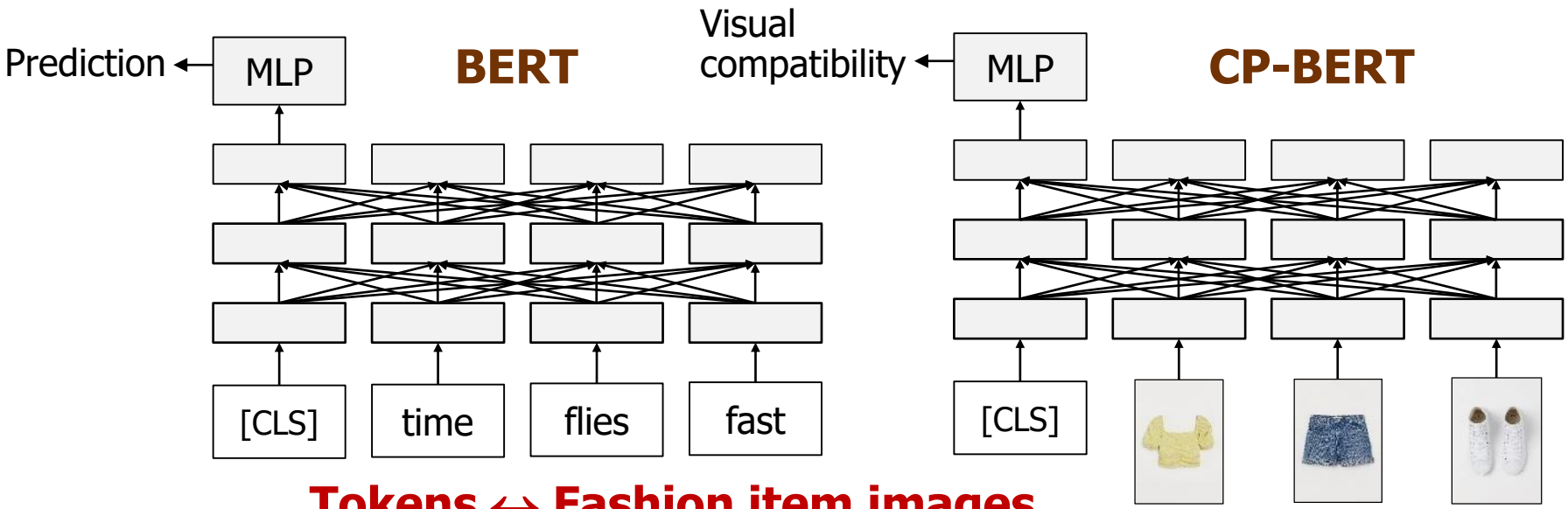
More Notes on CP-BERT (1)

- Typical fashion-related computer vision tasks:
  - Compatibility prediction (CP):
    Measuring the visual compatibility of a given outfit

  - Fill-In-The-Blank (FITB):
    Guessing an suitable item from an incomplete outfit

More Notes on CP-BERT (2)

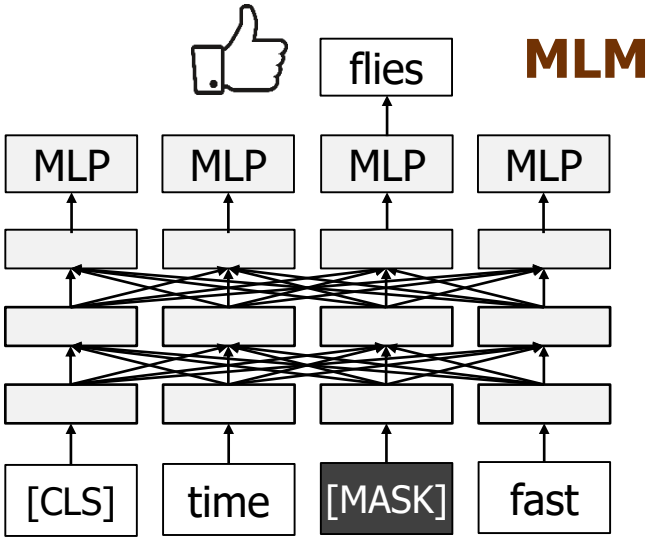
- CP-BERT: BERT specialized for compatibility prediction
- BERT: Transformer-based model applied to various NLP tasks
- Differences from the original BERT:
  - Tokenizer is disabled
  - Positional encoding is disabled (since fashion items are unordered)
  - Word embeddings are replaced with the feature vectors from a convolutional NN

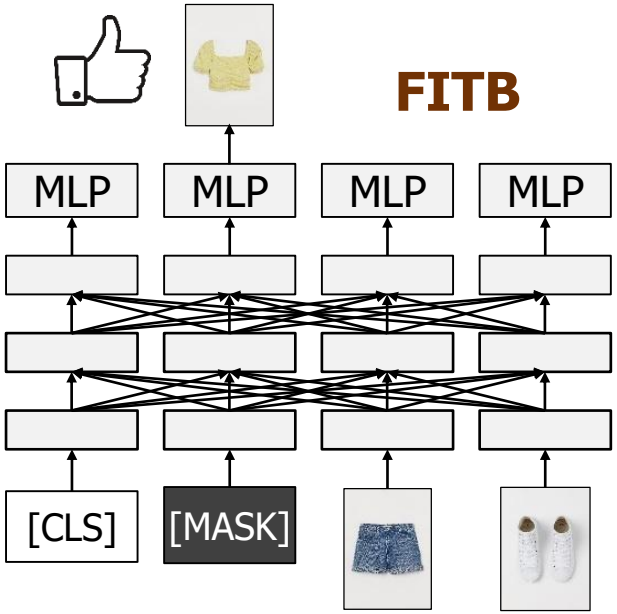


Tokens ↔ Fashion item images

More Notes on CP-BERT (3)

- CP-BERT is pretrained for the FITB task by analogy with BERT
  - This way of pretraining is called masked language modeling (MLM) in BERT





- OutfitTransformer [4]: The closest previous method to CP-BERT

	Main task	Pretraining task
OutfitTransformer	FITB	CP
CP-BERT	CP	FITB

[4] R. Sarkar et al.: OutfitTransformer: Learning outfit representations for fashion recommendation, WACV-23.

Outline

- ✓ Background
- ✓ Proposed Method
- Experiments
- Conclusion

Experiments: Datasets

- The Farfetch dataset
  - A collection of image-text pairs from Farfetch (e-commerce)
  - FashionCLIP is fine-tuned with this dataset, and the fine-tuned model is easily available[5]
- The Polyvore dataset
  - A collection of outfits from polyvore.com (social commerce)
  - We re-organized three versions:
    - Original version[6], Kaggle version[7], Cleaned version[8]
  - We focused on top wears, bottom wears, and shoes
    - Each outfit includes three fashion items *or more*
  - We constructed a binary classification dataset for CP:
    - Positive: Outfits actually included in the Polyvore dataset
    - Negative: Outfits generated at random

[5] https://github.com/patrickjohncyh/fashion-clip
[6] https://www.kaggle.com/datasets/dnepozitek/maryland-polyvore-images/data
[7] https://github.com/AemikaChow/AiDLab-fAshIon-Data/blob/main/Datasets/cleaned-maryland.md

Experiments: Settings

- Hyper-parameters for training of CP-BERT:

Primary hyper-parameter

Hyper-parameter	Possible choices
Embedding size	512 (ResNet-18), 2048 (ResNet-50)
Number of layers	8, 12, 16, 20
Number of attention heads	8, 16, 32, 64
Probability of masking (FITB)	0.3
Max. number of epochs (FITB)	50 (Embedding size = 2048), 100 (512)
Max. number of epochs (CP)	50
Learning rate	2×10^{-5}
Batch size	256

Hyper-parameters fixed based on preliminary experiments

Hyper-parameters selected accordingly

Experiments: Recommended Outfits (1)

- Recommended outfits for 5 exemplar requests:

#1	User's request	Now that I'm entering university, I want to try a more mature style that's different from what I've done before. (Male)
	Textual descriptions of fashion items	(Top) A navy button-down shirt made of lightweight cotton
		(Bottom) Slim-fit gray chinos
		(Shoes) Brown leather loafers
	Top-ranked outfit (T)	
	Middle-ranked outfit (M)	
	Bottom-ranked outfit (B)	

Experiments: Recommended Outfits (2)

- Recommended outfits for 5 exemplar requests:

#2	User's request	Now that summer is here, I want to wear something cool and trendy. (Female)
	Textual descriptions of fashion items	(Top) A white linen cropped blouse with puffed sleeves
		(Bottom) High-waisted beige culotte shorts
		(Shoes) Tan leather strappy sandals
	Top-ranked outfit (T)	
	Middle-ranked outfit (M)	
	Bottom-ranked outfit (B)	

Experiments: Recommended Outfits (3)

- Recommended outfits for 5 exemplar requests:

#3	User's request	I've started going to the gym, so I want to change my style to look sporty and healthy. (Female)
	Textual descriptions of fashion items	(Top) A white moisture-wicking racerback tank top
		(Bottom) Black high-rise compression leggings with mesh panels
		(Shoes) White lightweight running sneakers
	Top-ranked outfit (T)	
	Middle-ranked outfit (M)	
	Bottom-ranked outfit (B)	

Experiments: Recommended Outfits (4)

- Recommended outfits for 5 exemplar requests:

#4	User's request	I've been feeling a lot of stress lately, so I want to try some glamorous fashion that will brighten my mood. (Female)
	Textual descriptions of fashion items	(Top) A bright yellow ruffled chiffon blouse
		(Bottom) A floral-printed midi skirt in vibrant pinks and oranges
		(Shoes) White strappy sandals with block heels
	Top-ranked outfit (T)	
	Middle-ranked outfit (M)	
	Bottom-ranked outfit (B)	

Experiments: Recommended Outfits (5)

- Recommended outfits for 5 exemplar requests:

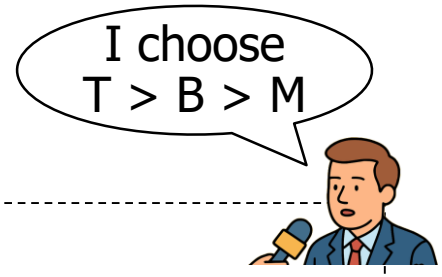
#5	User's request	I want to create a new image for the photos I post on social media, so I want to style it in a way that stands out. (Female)
	Textual descriptions of fashion items	(Top) A vibrant fuchsia wrap blouse with balloon sleeves
		(Bottom) A high-waisted pleated maxi skirt in metallic silver
		(Shoes) Strappy neon green stiletto heels
	Top-ranked outfit (T)	
	Middle-ranked outfit (M)	
	Bottom-ranked outfit (B)	

Experiments: User Study (1)

- We conducted a simple user study using 5 exemplar requests explained before:

1	Now that I'm entering university, I want to try a more mature style that's different from what I've done before. (Male)
2	Now that summer is here, I want to wear something cool and trendy. (Female)
3	I've started going to the gym, so I want to change my style to look sporty and healthy. (Female)
4	I've been feeling a lot of stress lately, so I want to try some glamorous fashion that will brighten my mood. (Female)
5	I want to create a new image for the photos I post on social media, so I want to style it in a way that stands out. (Female)

- 37 participants chose a total order of preference among three recommended outfits:
  - Top-ranked (T)
  - Middle-ranked (M)
  - Bottom-ranked (B)



For the 1st request:

T

M

B

Experiments: User Study (2)

- Results:

Popular:

Popular:

Request#	T>M>B	T>B>M	M>T>B	M>B>T	B>T>M	B>M>T	Total
1	6	11	4	4	7	5	37
2	9	13	2	2	6	5	37
3	12	7	3	8	6	1	37
4	4	3	4	20	1	5	37
5	4	4	5	8	5	11	37
Total	35	38	18	42	25	27	185

73 60 52

Subtle tendency:

Top-ranked (T) outfit is more frequently preferred, and Bottom-ranked (B) outfit less frequently

Experiments: Ablation Study

- We also examined the effect of pretraining (FITB) according to the accuracy of compatibility prediction (CP)
  - We observed that pretraining tends to bring higher accuracy under a wide variety of configutations
  - It cannot be said that the use of pretraining significantly improve the accuracy

Config#	Pretraining	Embed. Size	#Layers	#Att-heads	Precision	Recall	F_1	p-value
1	✓	512	16	32	0.8605	0.9037	0.8816	—
2	✓	2048	4	16	0.8821	0.8720	0.8770	0.4186
3		512	20	32	0.8681	0.8881	0.8780	0.0152
4		2048	4	16	0.8797	0.8691	0.8744	0.0591

Outline

- ✓ Background
- ✓ Proposed Method
- ✓ Experiments
- Conclusion

Conclusion

- We proposed an outfit recommender system that:
  - Captures the users' ambiguous desire for self-expression
  - Presents outfits including real fashion items
  - Presents visually compatible outfits
- Experimental results showed that:
  - Our recommender system outputs reasonable outfits
  - Measuring visual compatibility positively works for meeting the human's common preference

Future Work

- More extensive user study
- Improving the prompt given to the text-to-text LLM
- Reducing the redundancy among recommendations
- Adding explainability

Thank You for Your Attention!