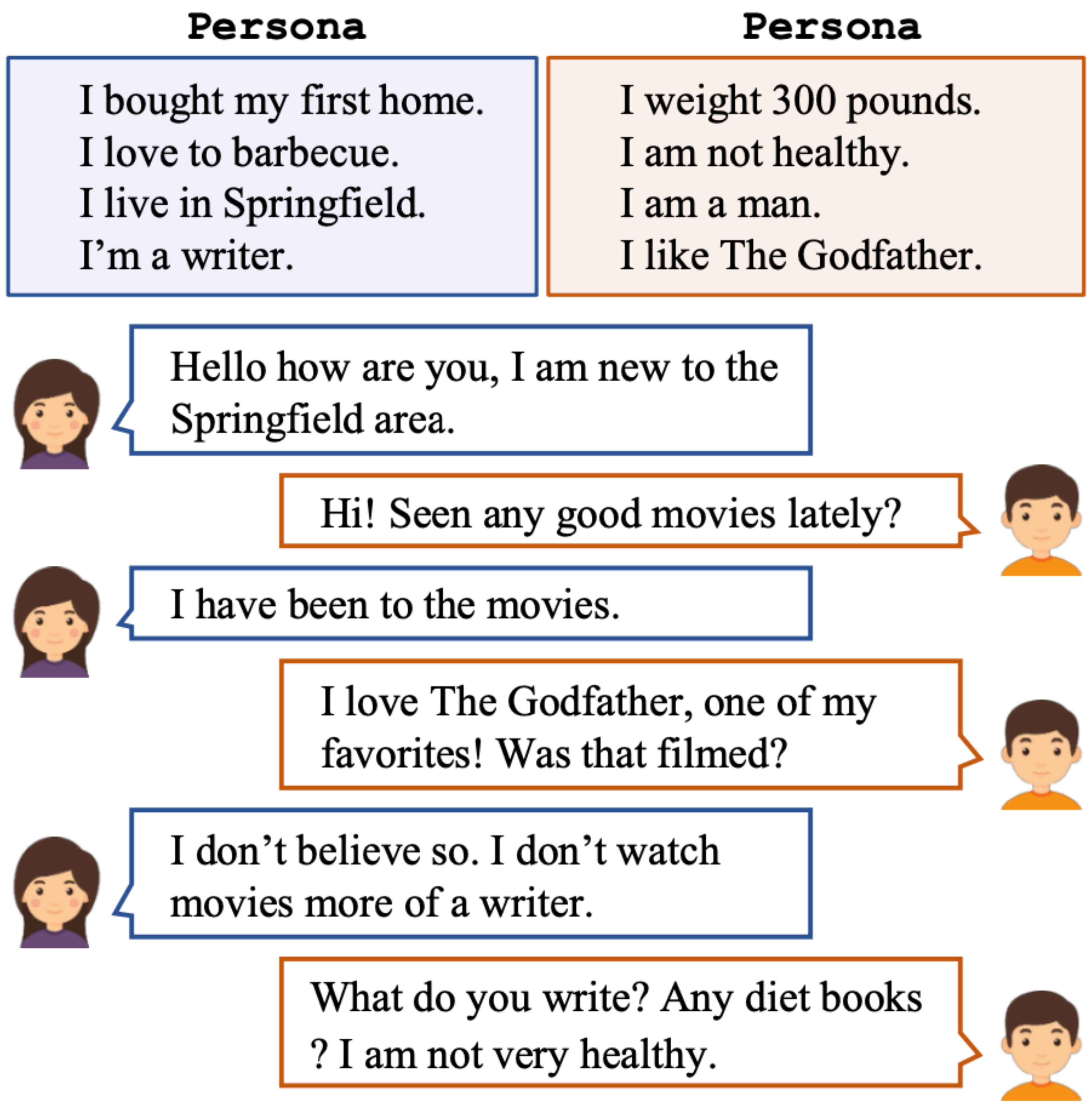
Course Project

CVAE-GPT2 Architecture for Diverse Responses Generation

沈心怡 1801110049

Motivation

PersonaChat Dataset (Zhang et al., 2018b)



Persona

I bought my first home.
I love to barbecue.
I live in Springfield.
I'm a writer.

Persona

I weight 300 pounds.
I am not healthy.
I am a man.
I like The Godfather.

Hello how are you, I am new to the Springfield area.

Hi! Seen any good movies lately?

I have been to the movies.

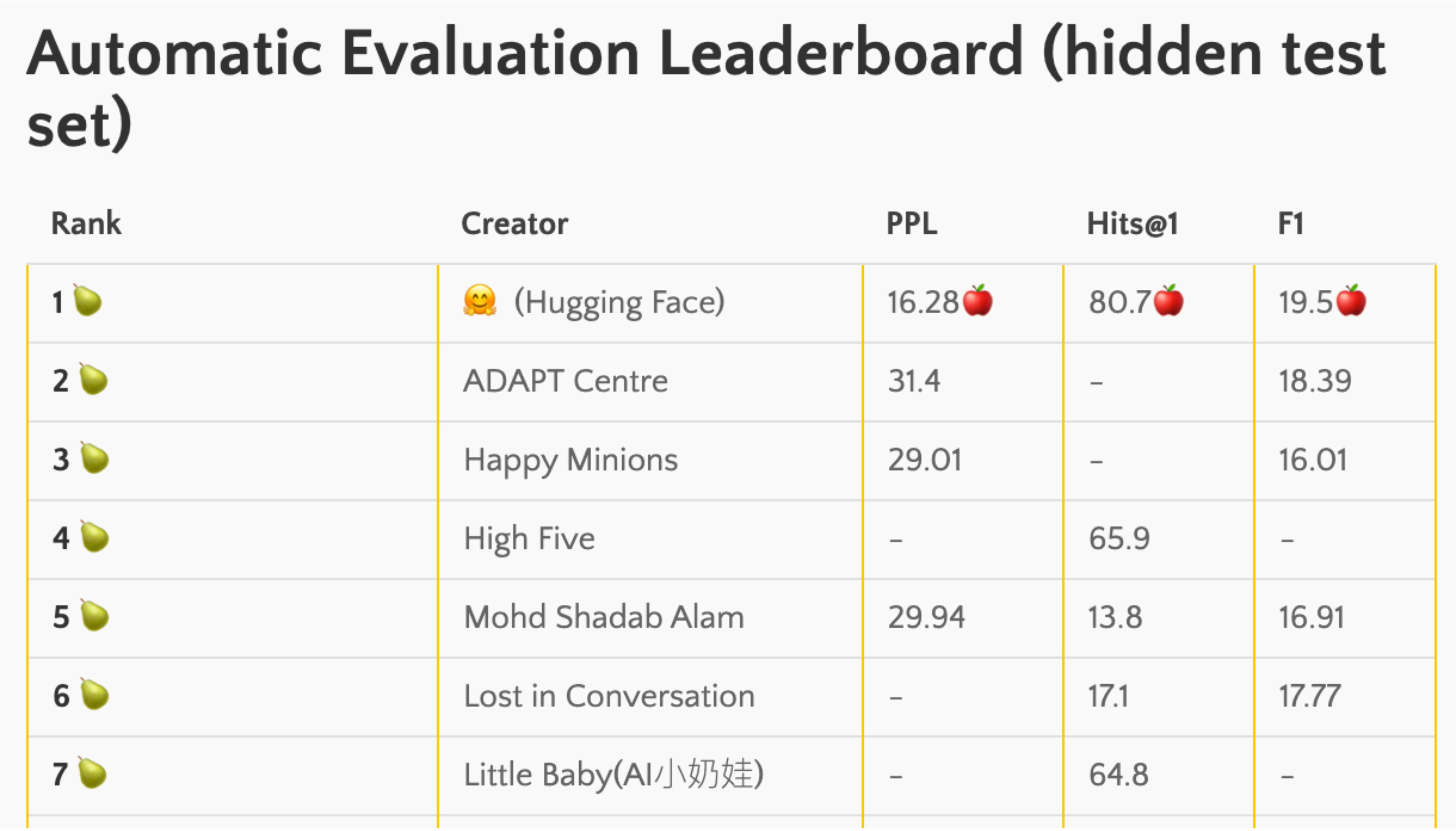
I love The Godfather, one of my favorites! Was that filmed?

I don't believe so. I don't watch movies more of a writer.

What do you write? Any diet books ? I am not very healthy.

(Image source: Liu et al., 2020)

Motivation



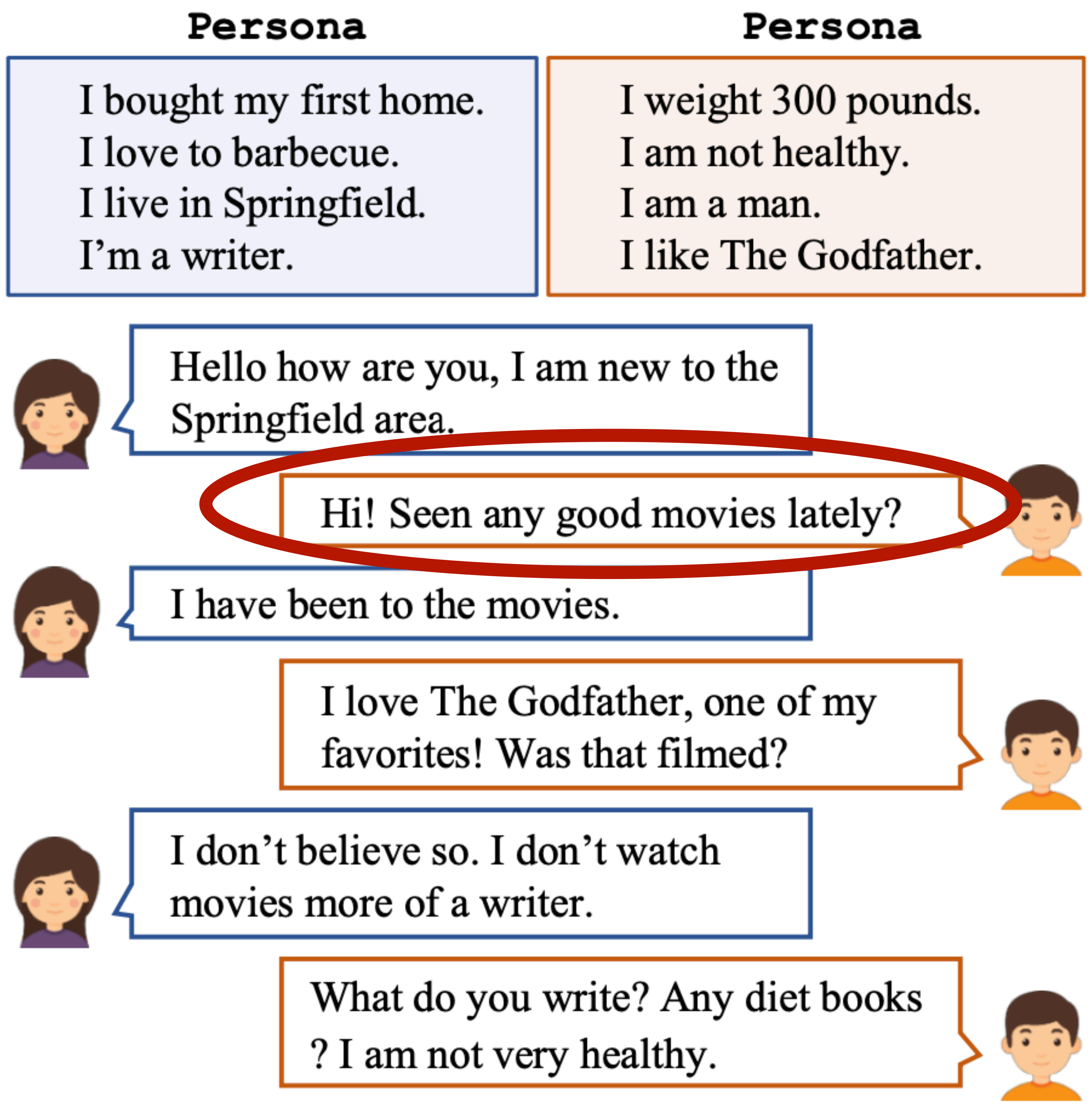
Automatic Evaluation Leaderboard (hidden test set)

Rank	Creator	PPL	Hits@1	F1
1 🍐	🤗 (Hugging Face)	16.28🍎	80.7🍎	19.5🍎
2 🍐	ADAPT Centre	31.4	–	18.39
3 🍐	Happy Minions	29.01	–	16.01
4 🍐	High Five	–	65.9	–
5 🍐	Mohd Shadab Alam	29.94	13.8	16.91
6 🍐	Lost in Conversation	–	17.1	17.77
7 🍐	Little Baby(AI小奶娃)	–	64.8	–

(Image source: Convai2 website)

Motivation

PersonaChat Dataset (Zhang et al., 2018b)



Persona

I bought my first home.
I love to barbecue.
I live in Springfield.
I'm a writer.

Persona

I weight 300 pounds.
I am not healthy.
I am a man.
I like The Godfather.

Hello how are you, I am new to the Springfield area.

Hi! Seen any good movies lately?

I have been to the movies.

I love The Godfather, one of my favorites! Was that filmed?

I don't believe so. I don't watch movies more of a writer.

What do you write? Any diet books ? I am not very healthy.

Hi!

Nice to meet you!

What's your job ?

I'm new to this area too.

(Image source: Liu et al., 2020)

Motivation

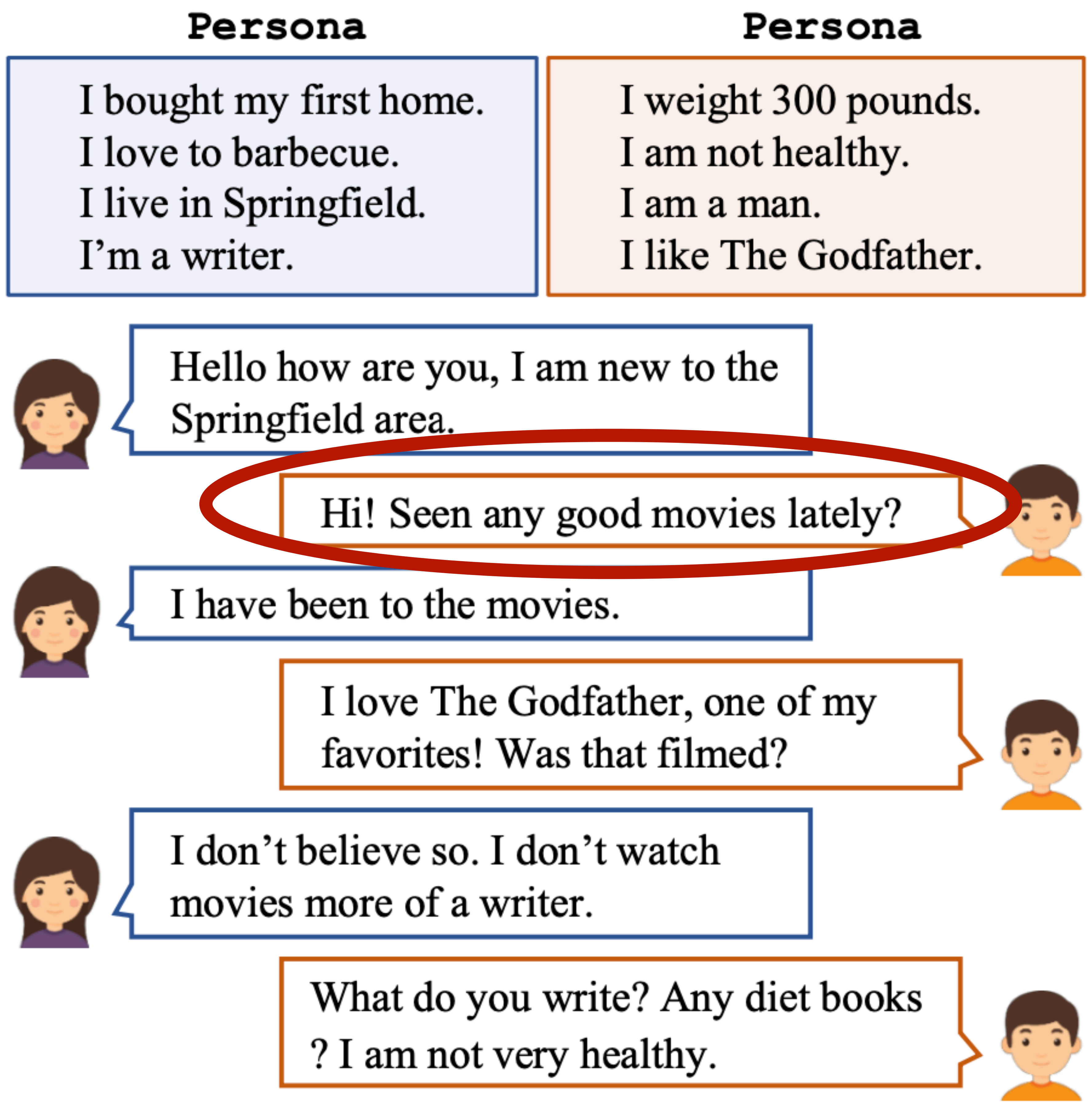
PersonaChat Dataset (Zhang et al., 2018b)

seq2seq:

very hard !

a -> b

only one chance



Persona

I bought my first home.
I love to barbecue.
I live in Springfield.
I’m a writer.

Persona

I weight 300 pounds.
I am not healthy.
I am a man.
I like The Godfather.

Hello how are you, I am new to the Springfield area.

Hi! Seen any good movies lately?

I have been to the movies.

I love The Godfather, one of my favorites! Was that filmed?

I don’t believe so. I don’t watch movies more of a writer.

What do you write? Any diet books ? I am not very healthy.

(Image source: Liu et al., 2020)

cvae:

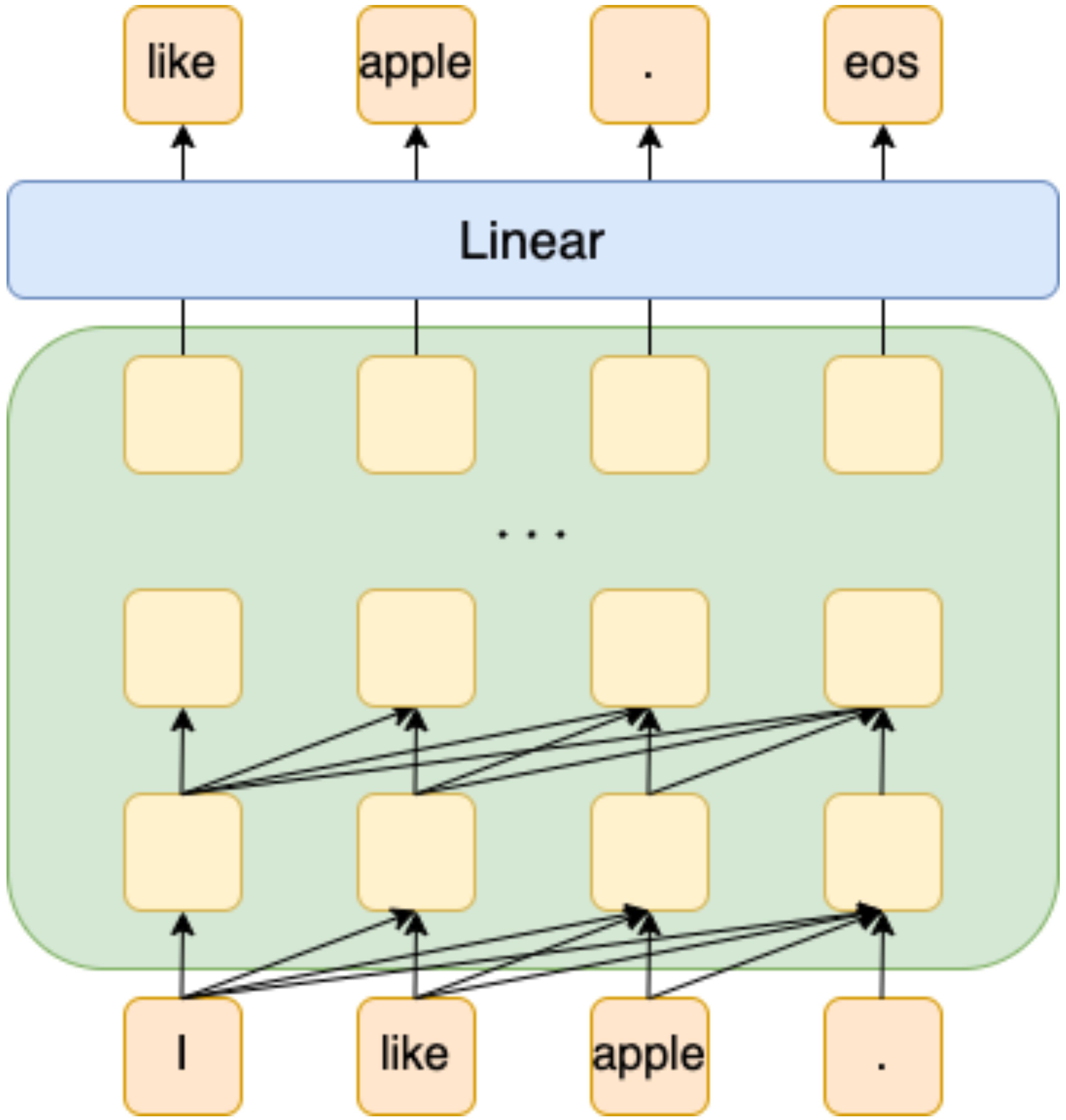
better model this problem

a -> b1, b2, …

multiple chances

Method

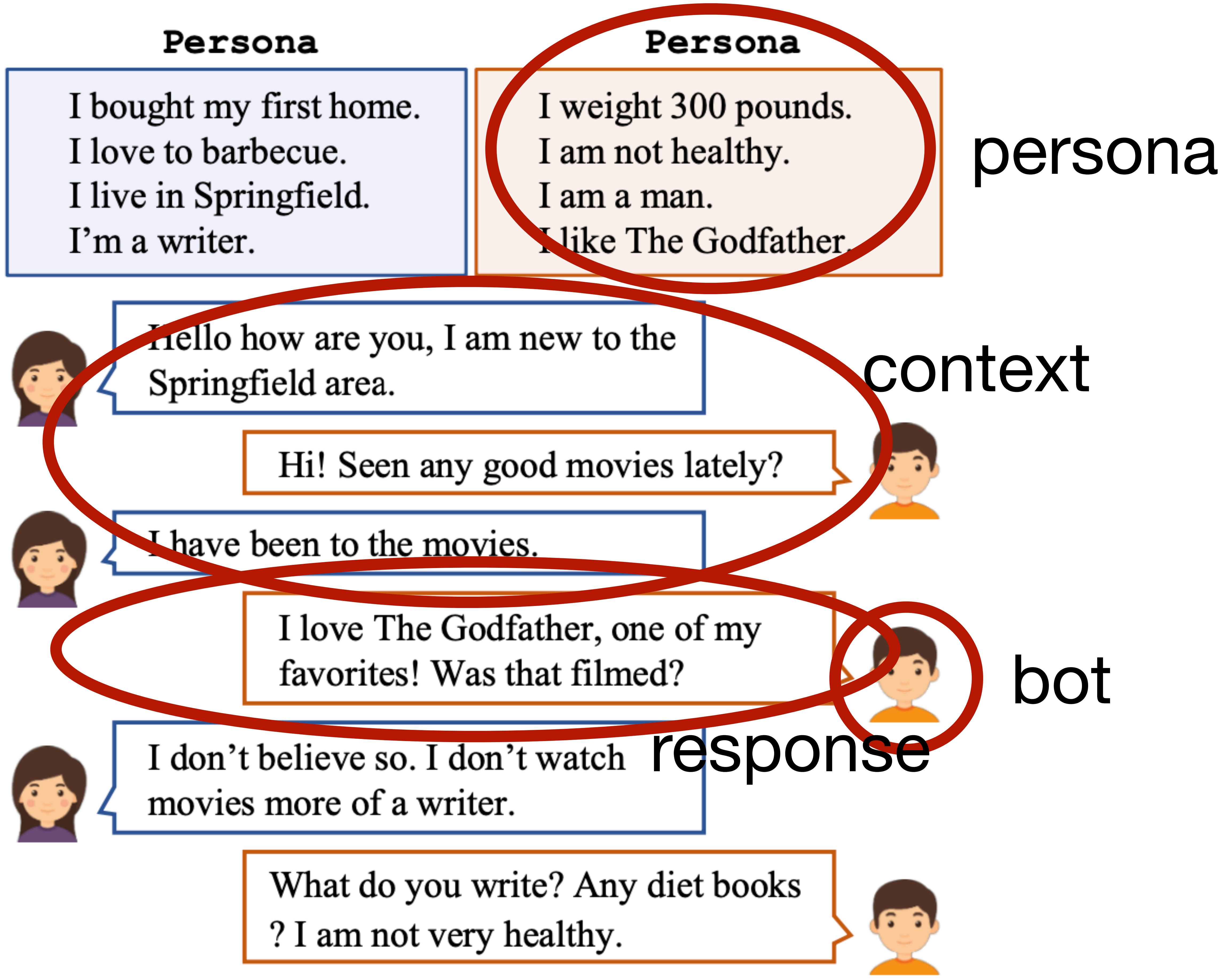
GPT2 (Radford et al., 2019)



Pretrained on language modeling task.

Baseline

Problem Definition



Persona

I bought my first home.
I love to barbecue.
I live in Springfield.
I'm a writer.

Persona

I weight 300 pounds.
I am not healthy.
I am a man.
I like The Godfather.

Hello how are you, I am new to the Springfield area.

Hi! Seen any good movies lately?

I have been to the movies.

I love The Godfather, one of my favorites! Was that filmed?

I don't believe so. I don't watch movies more of a writer.

What do you write? Any diet books ? I am not very healthy.

persona

context

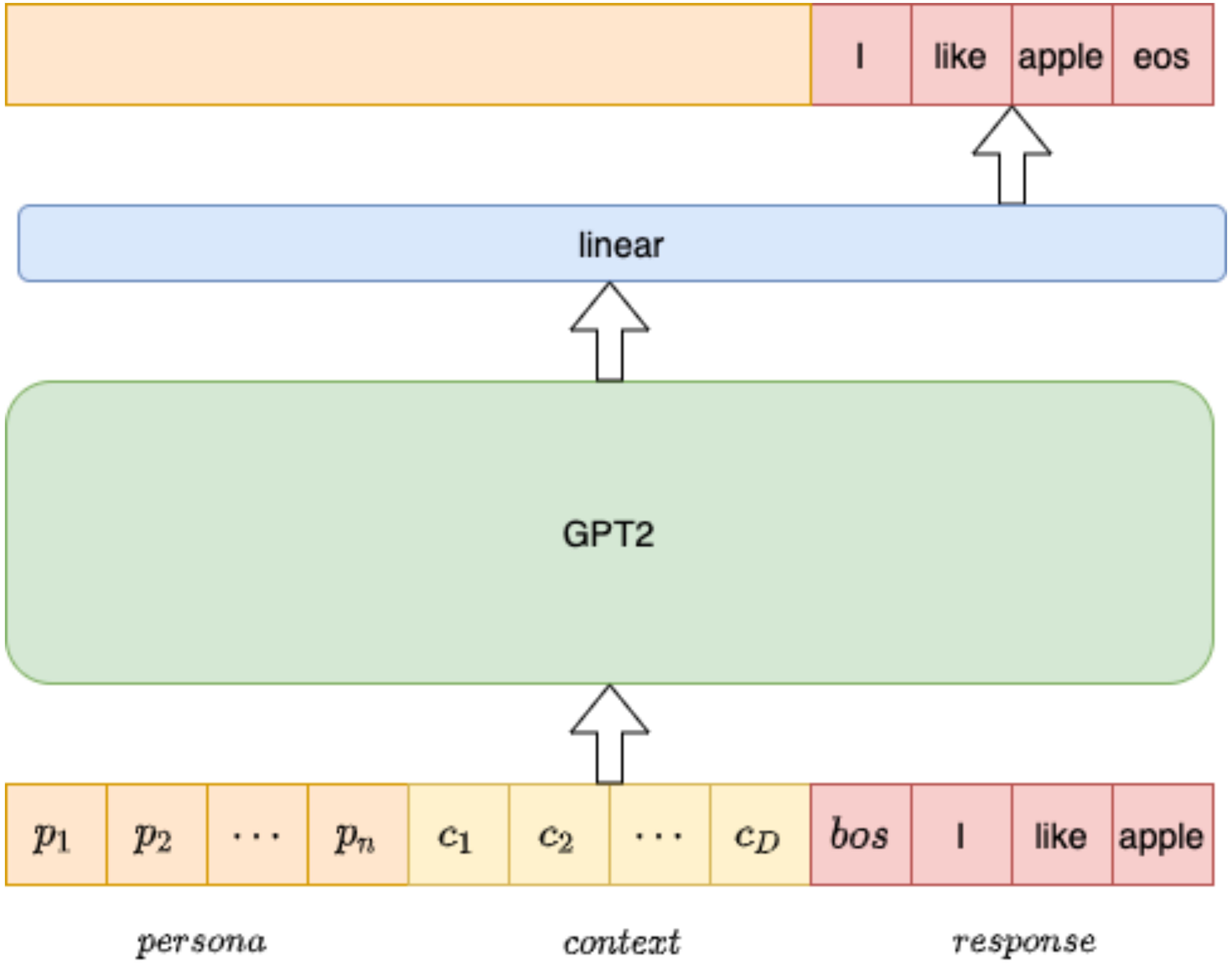
bot

response

(persona, context) -> response

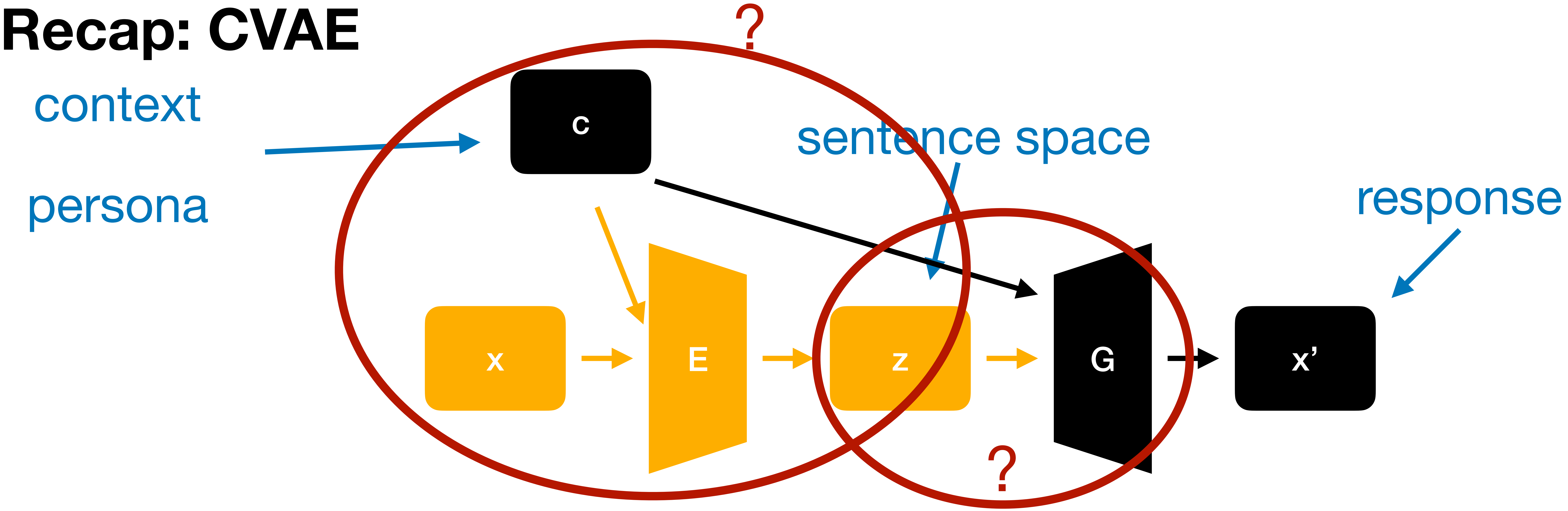
Method

Baseline



Method

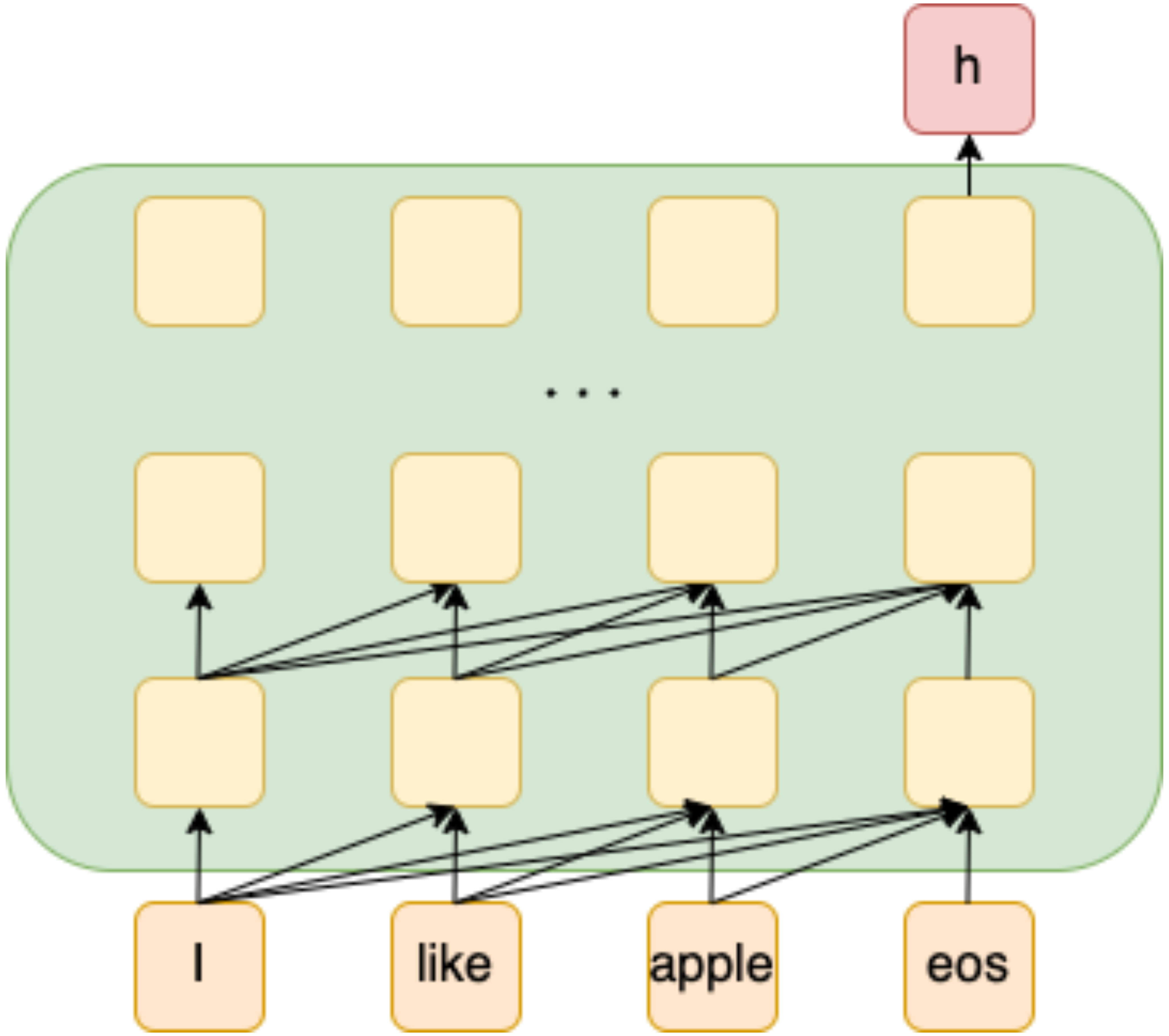
Recap: CVAE



$$\log P(x\,|\,c) - \mathscr{D}[Q(z\,|\,x,c)\|P(z\,|\,x,c)] = E_{z\sim Q(x,c)}[\log P(x\,|\,z,c)] - \mathscr{D}[Q(z\,|\,x,c)\|P(z\,|\,c)]$$

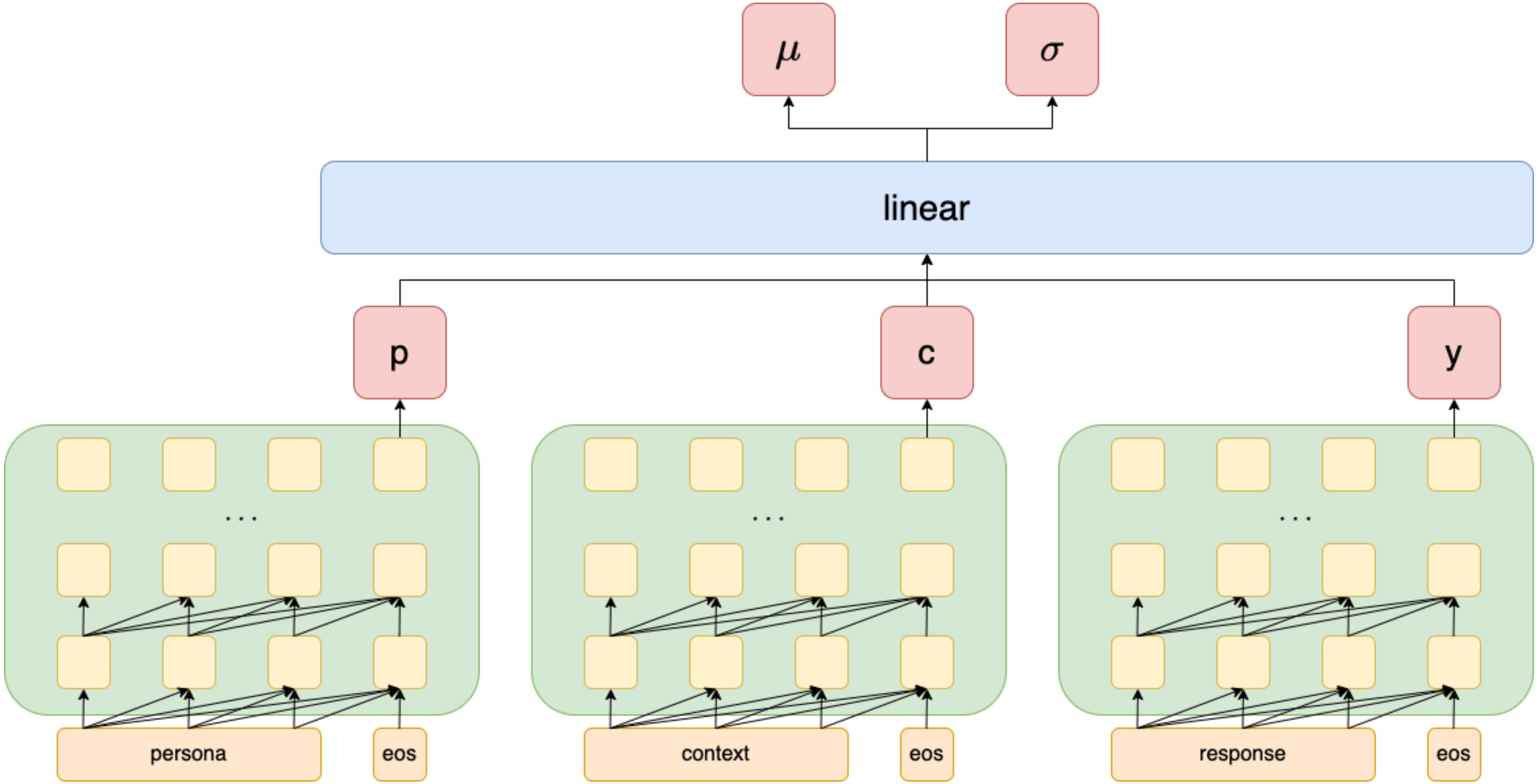
Method

Sentence embedding



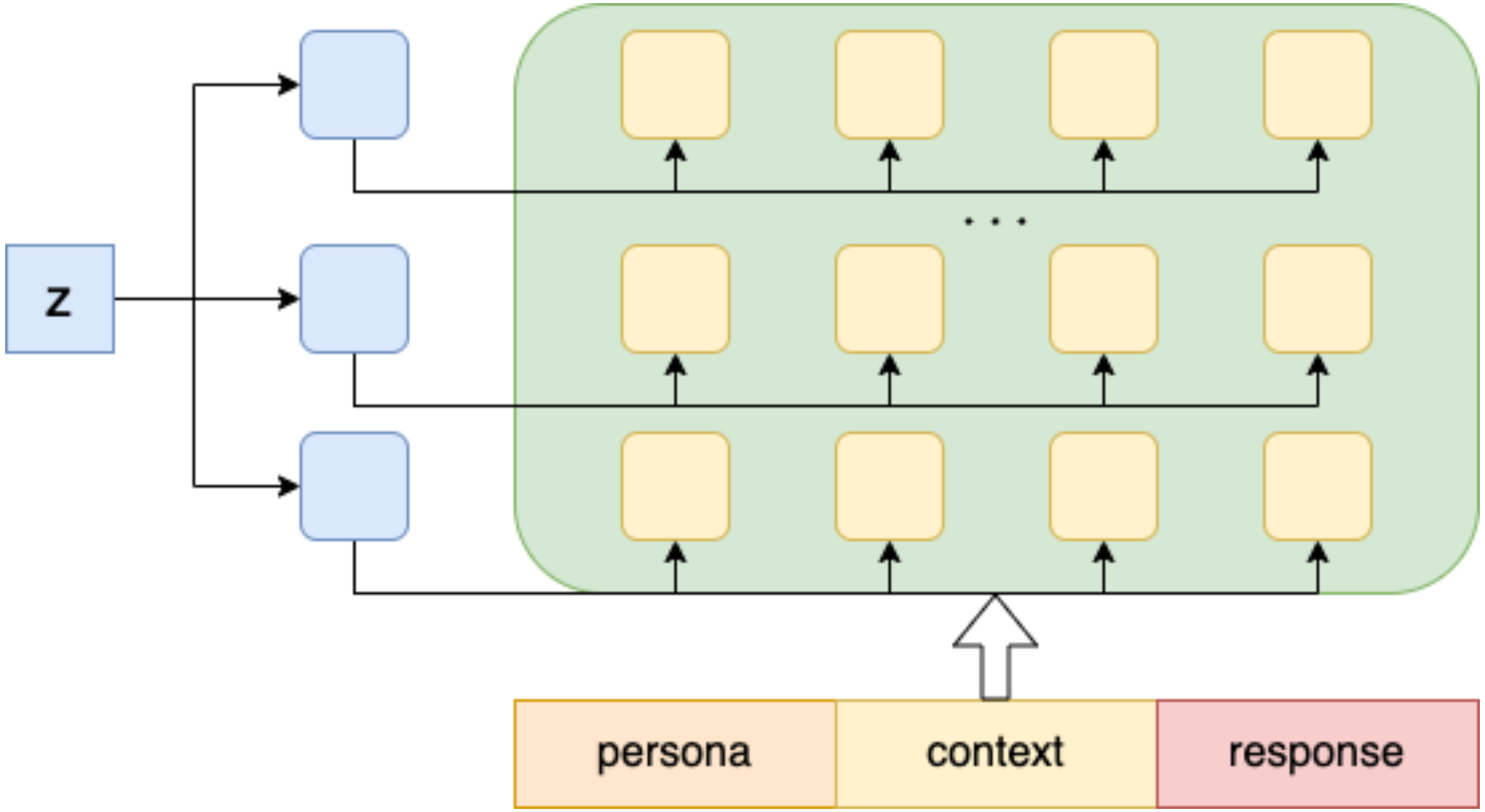
Method

Encoder: q(z|x, c)



Method

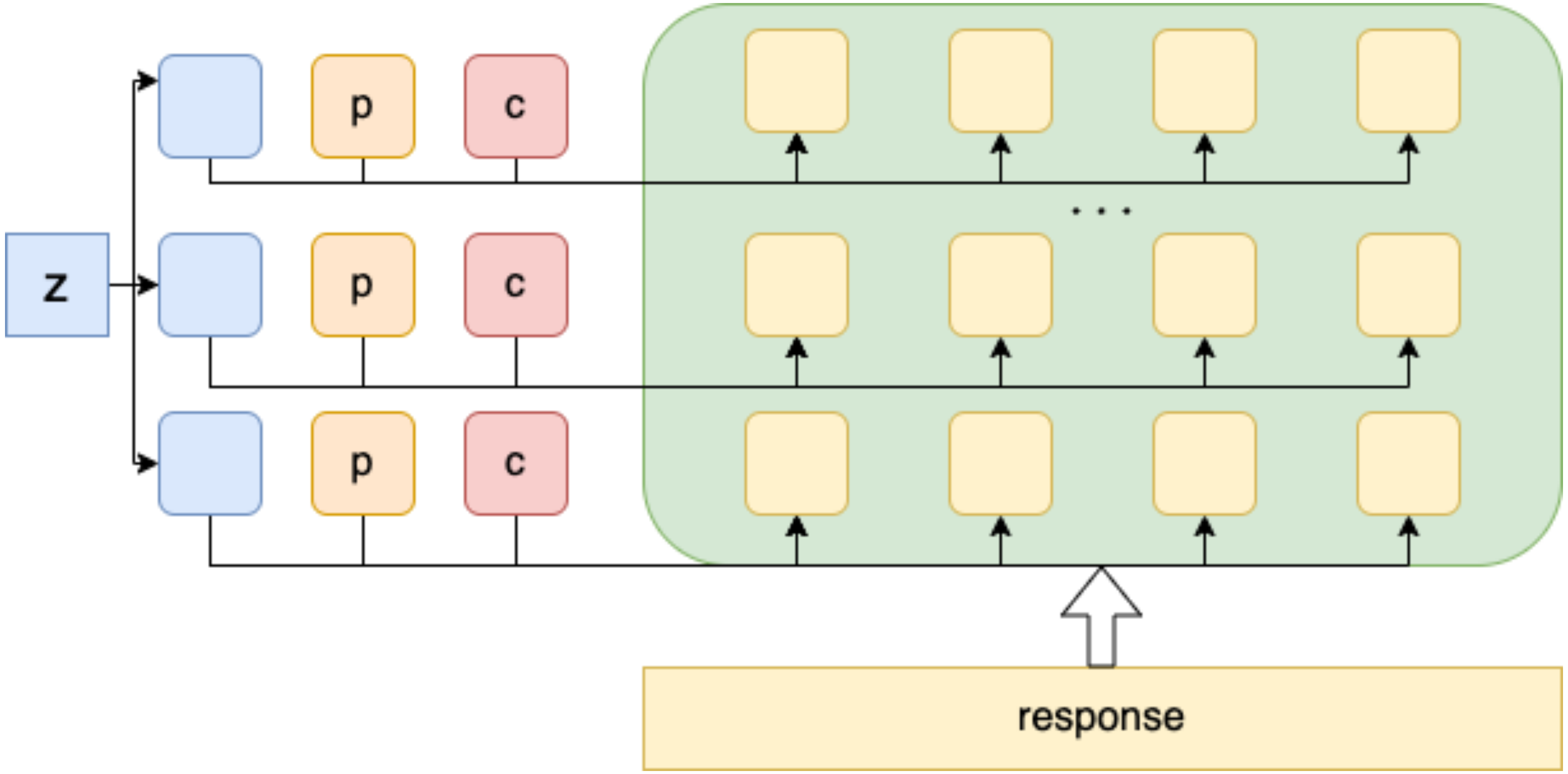
Inject z into decoder



Optimus: Organizing Sentences via Pre-trained Modeling of a Latent Space Chunyuan Li, Xiang Gao, Yuan Li, Xiujun Li, Baolin Peng, Yizhe Zhang, Jianfeng Gao

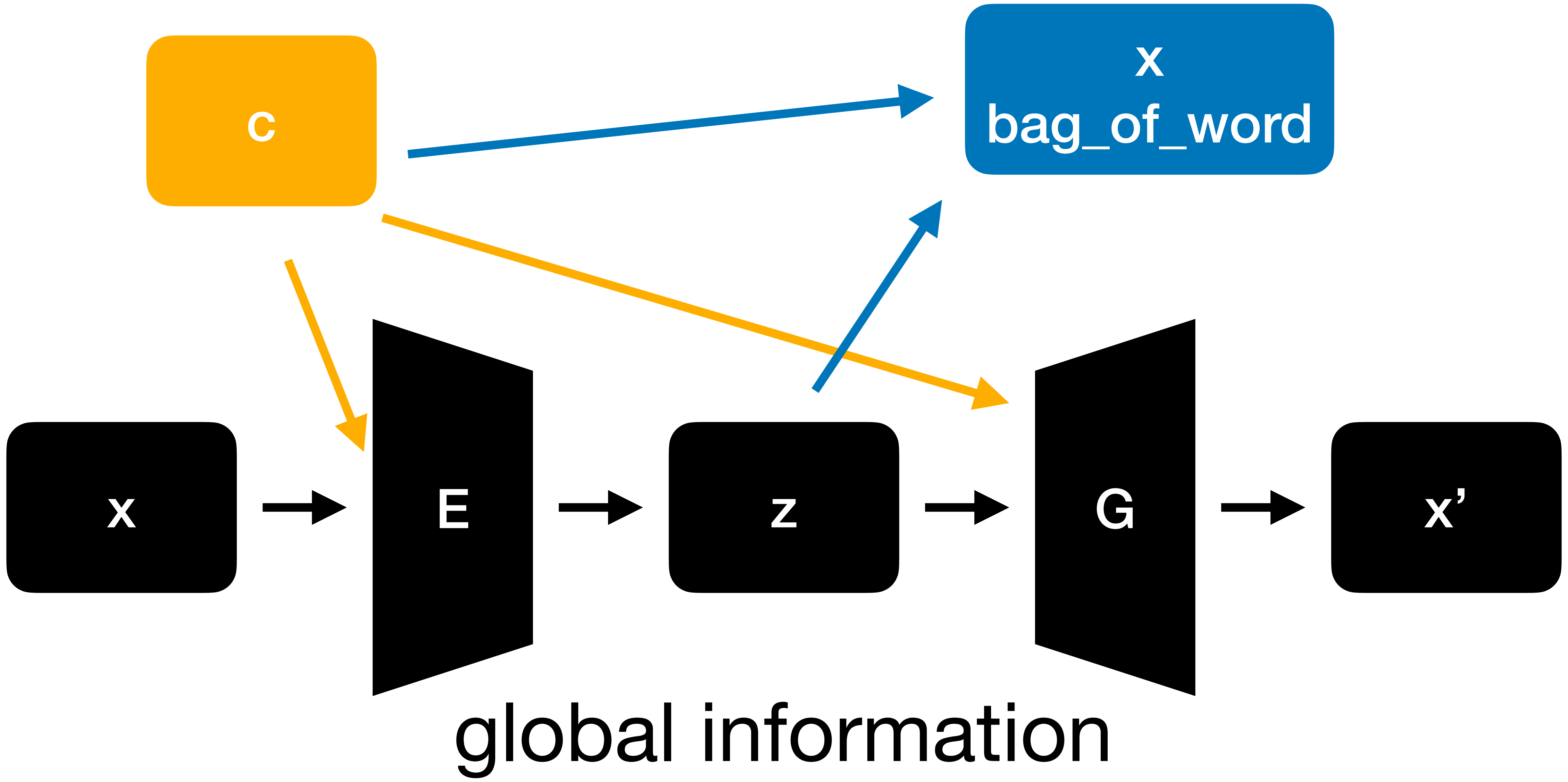
Method

Compressed CVAE



Method

Recap: bow loss



$$\mathcal{L} = E_{z\sim Q(x,c)}[\log P(x\,|\,z,c)] - \lambda\mathcal{D}[Q(z\,|\,x,c)\|P(z\,|\,c)] \quad + \mathcal{L}_{bow}$$

Evaluation

relevance

	ppl ↓	max_f1 ↑ (among 5 candidates)
Decoder	15.483	0.184
CVAE + bow loss	6.123	0.261
Compressed CVAE + bow loss	**5.992**	**0.265**

Evaluation

diversity

	Dist-1 ↑	Dist-2 ↑	Ent-4 ↑
Decoder	**0.18**	0.409	3.698
CVAE + bow loss	0.133	**0.496**	4.960
Compressed CVAE + bow loss	0.108	0.469	**5.278**

Li J, Galley M, Brockett C, et al. A diversity-promoting objective function for neural conversation models, NAACL 2016

Zhang Y, Galley M, Gao J, et al. Generating informative and diverse conversational responses via adversarial information maximization, NIPS 2018

Conclusion

- CVAE + Pretrained models ✓
- sentence embedding ?
- knowledge guided latent space ?

QA