

# Transformer4Rec

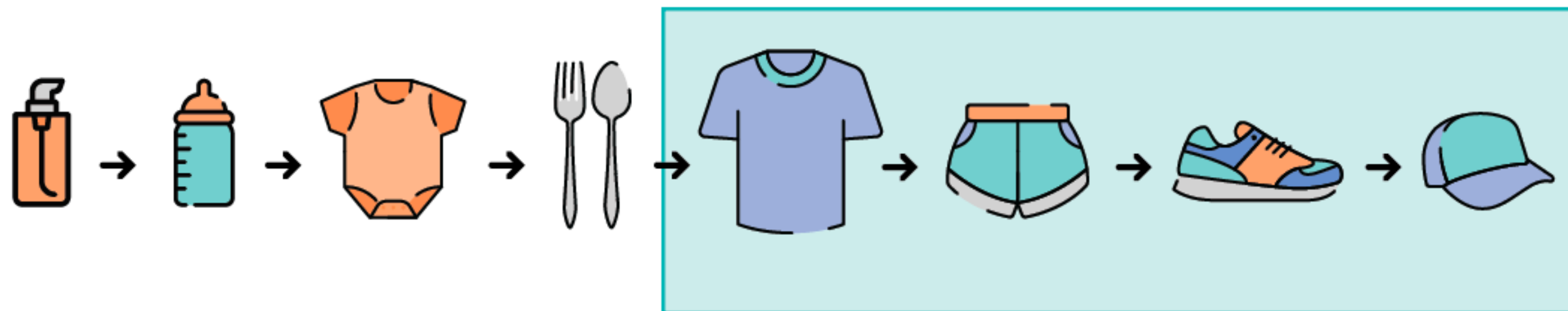
Смирнов Тимофей

# Session vs Sequential based Recommendations

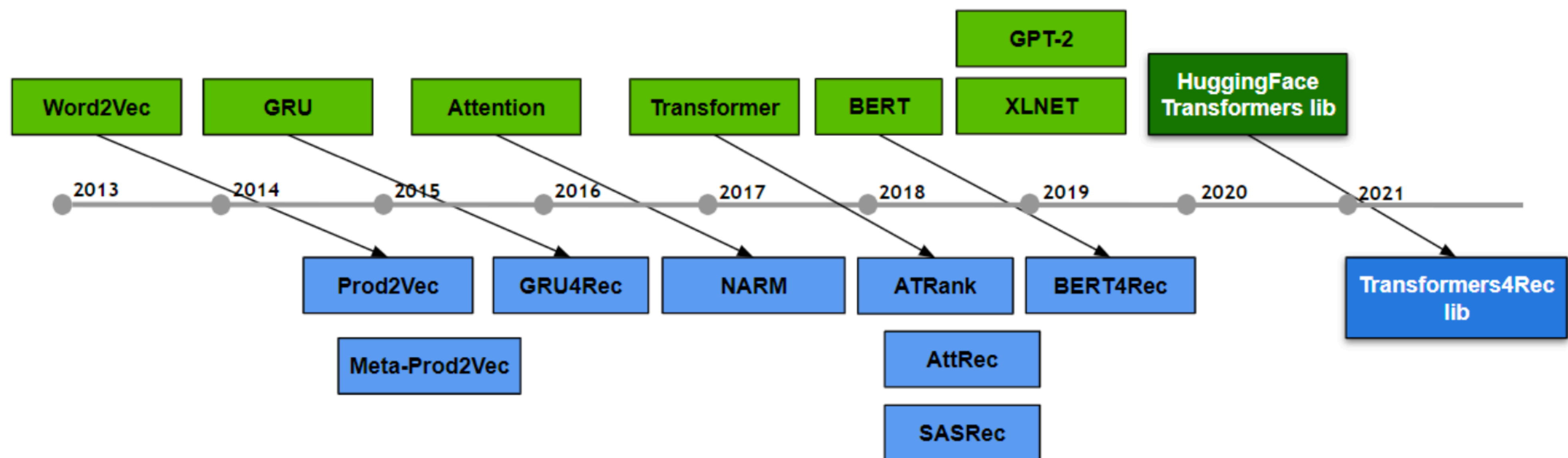
- Session-based recommendations



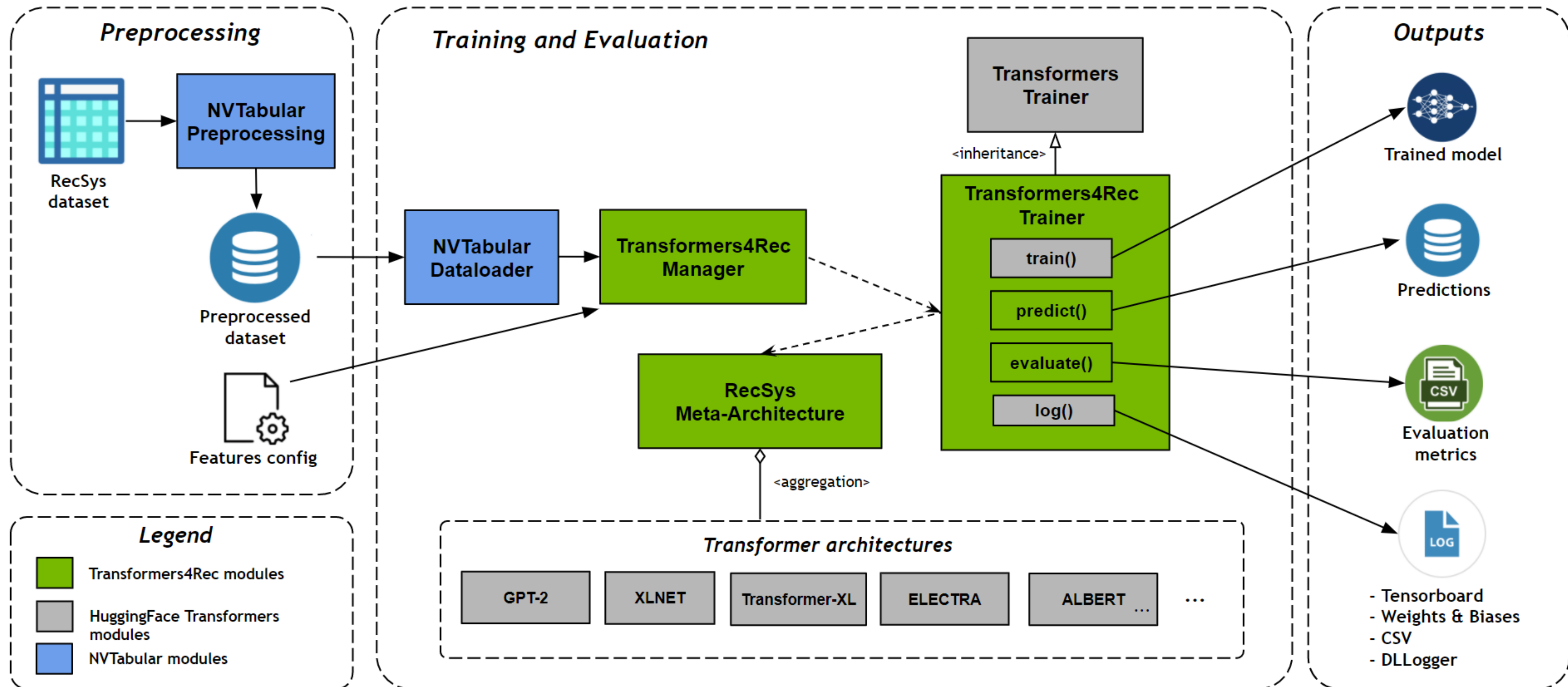
- Sequential-based recommendations



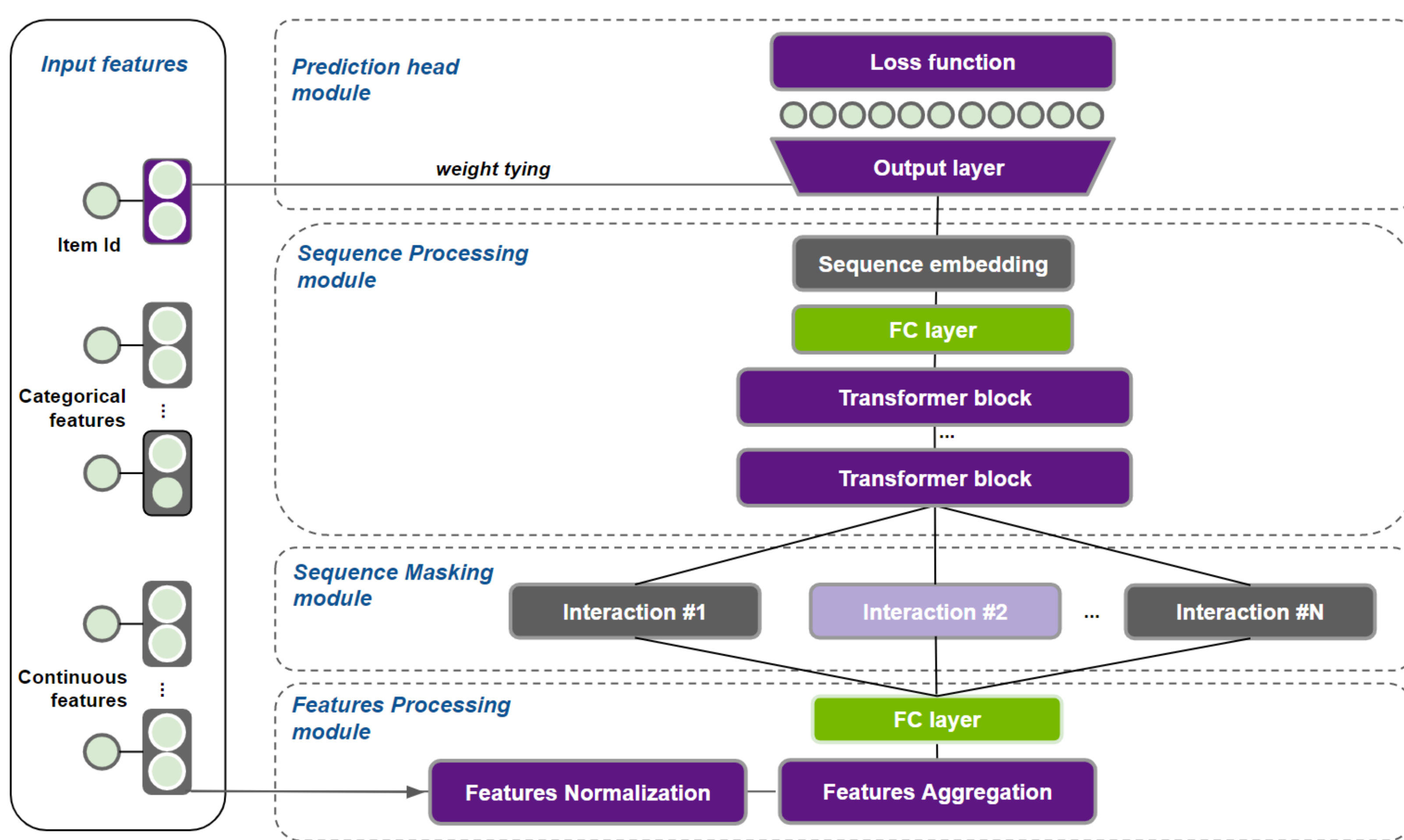
# Связь NLP с RecSys



# Схема решения



# Архитектура модели

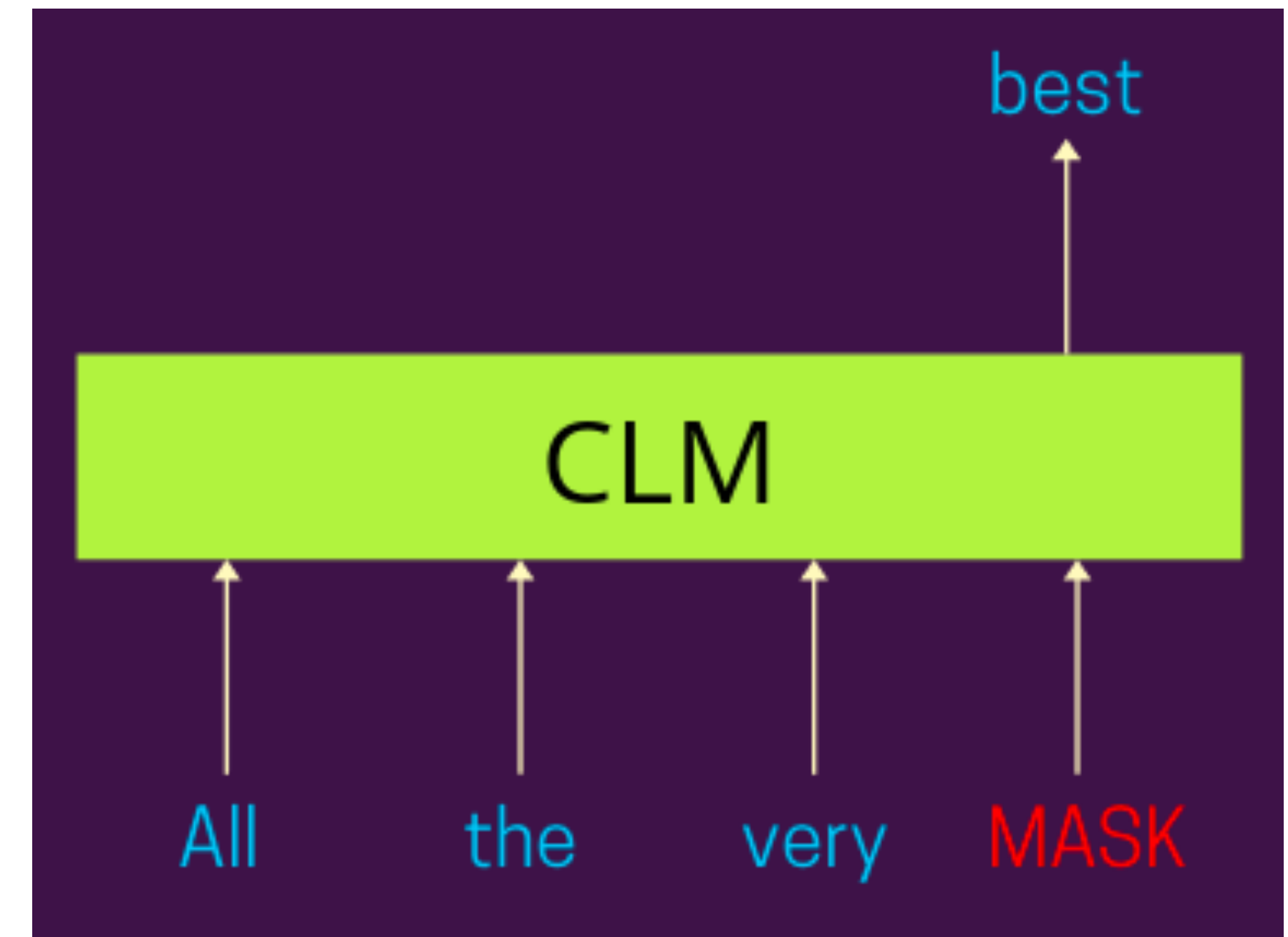


# CLM(Casual Language Modeling) GPT-like

- Objective

$$\max_{\phi} = \sum_{t=1}^T \log p_{\phi}(x_t | x_{<t})$$

- T - длина последовательности
- Пытаемся предсказать следующий токен



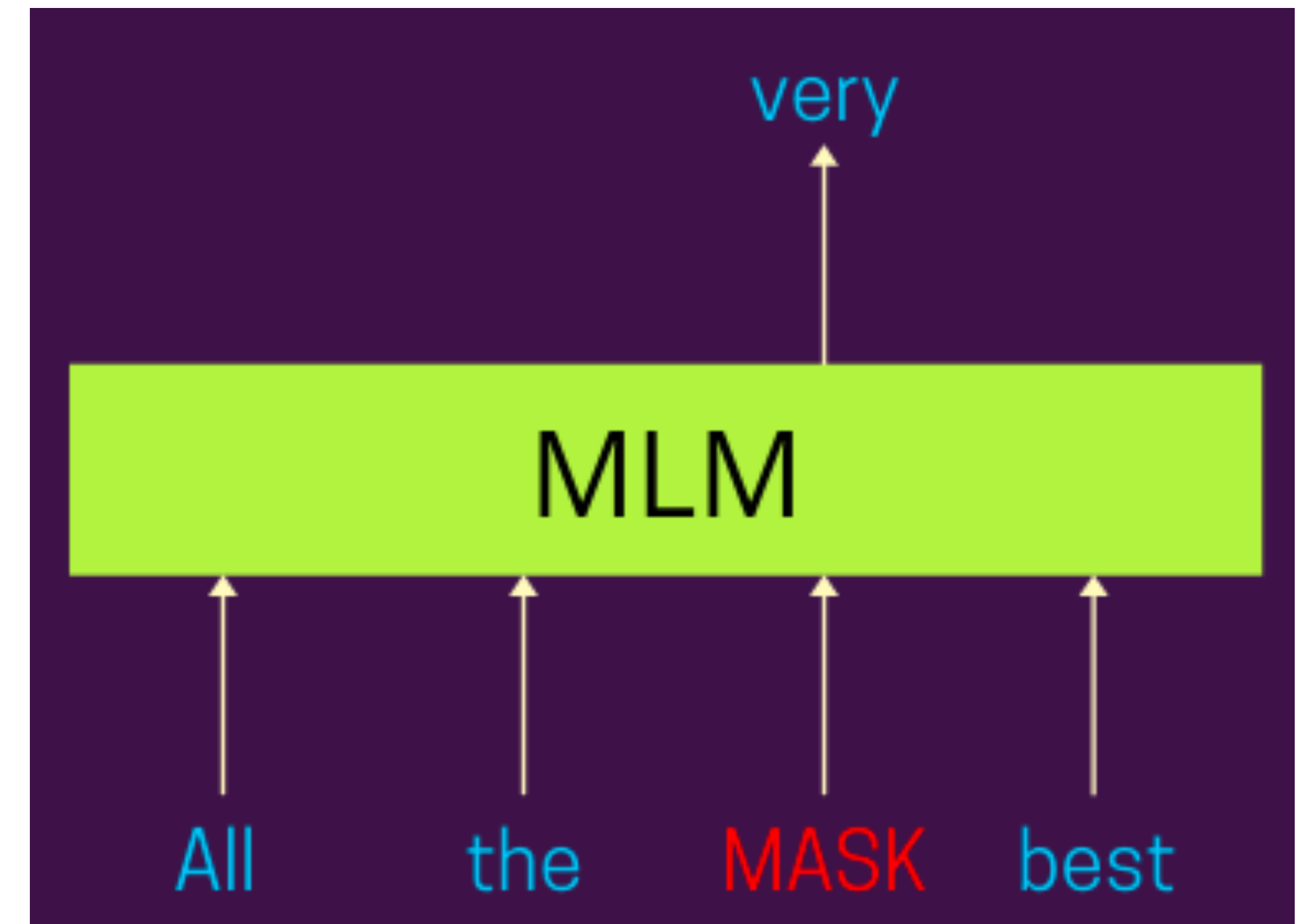
# MLM(Masked Language Modeling)

## Bert-like

- Objective

$$\max_{\phi} = \sum_{t=1}^T m_t \log p_{\phi}(x_t | X)$$

- $T$  - длина последовательности
- $m_t = 1$  там где токен замаскирован





# PLM(Permutation Language Modeling) XLnet-like

- Objective

$$\max_{\phi} = E_{z \in Z_T} \left[ \sum_{t=1}^T \log p_{\phi}(x_{z_t} | x_{z < t}) \right]$$

- T - длина последовательности
- $Z_t$  - все перестановки

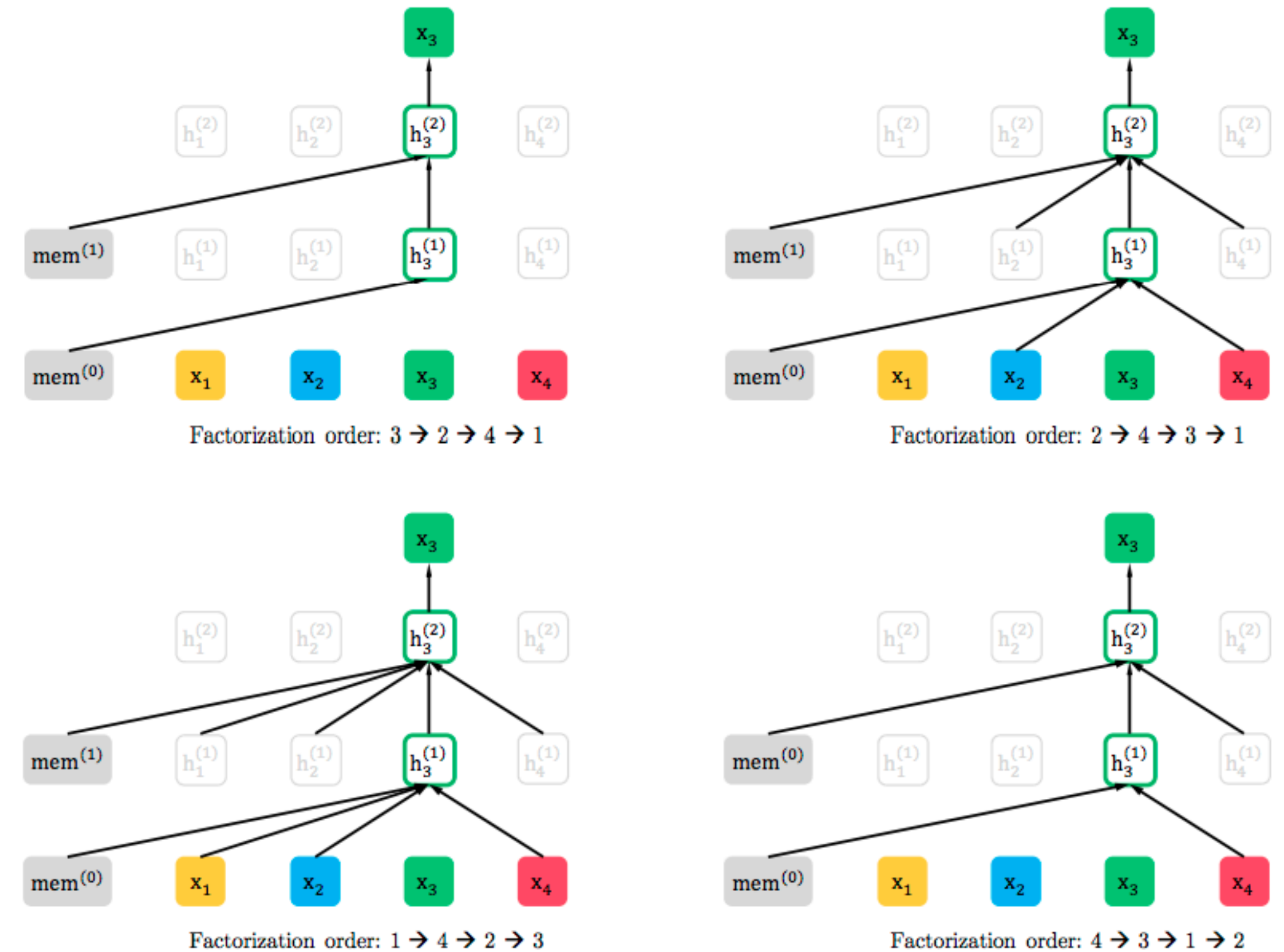
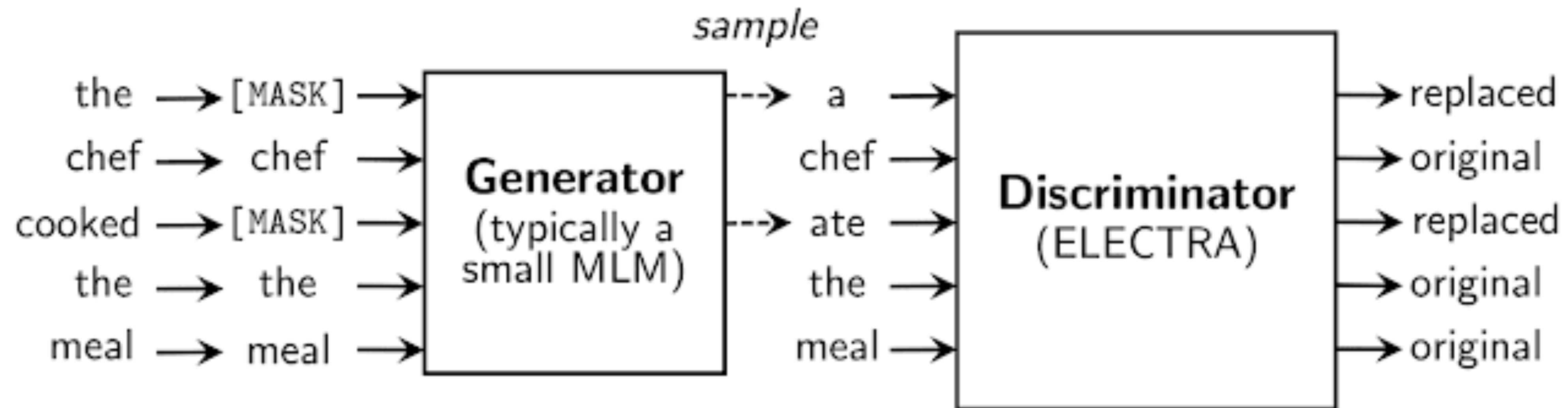


Figure 1: Illustration of the permutation language modeling objective for predicting  $x_3$  given the same input sequence  $x$  but with different factorization orders.



# RTD(Replacement Token Detection)

Electra-like



# Soft One Hot Encoding

- $p^n = nW_n + b_n$ ,  $W_n \in R^{1 \times p}$ ,  $b_n \in R^p$ , where  $p$  is number of embeddings
- $s^n = \text{softmax}(p^n)$
- $g_n = s^n E^n$ ,  $E^n \in R^{p \times d}$  where  $d$  is embedding size

# Loss Functions

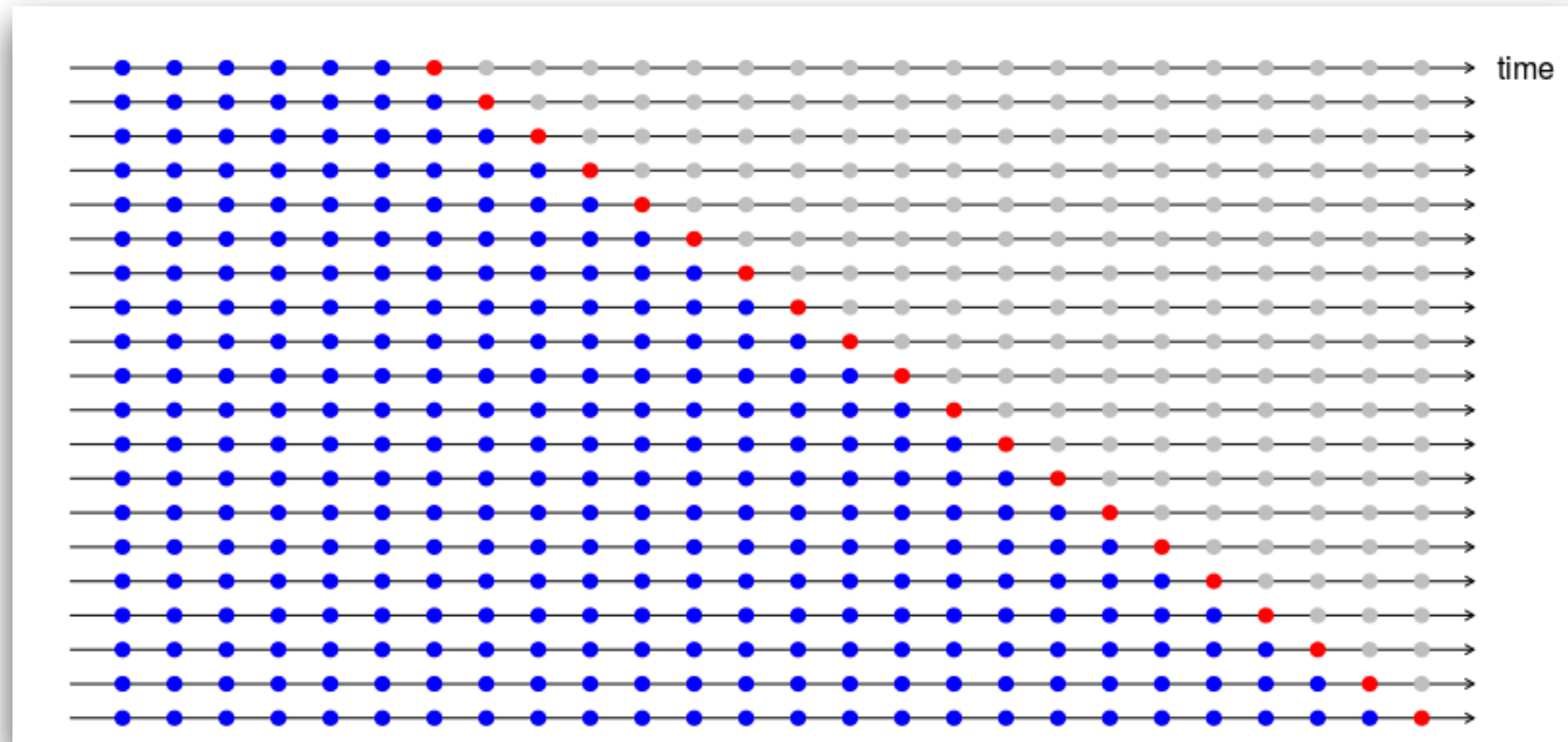
- Cross Entropy

- $BPR_{loss} = \frac{1}{N} \sum_{j=1}^N \log(\sigma(r_i - r_j))$  где  $i$ -фиксированный айтем для которого сэмплируется  $N$  негативных

- $TOP1_{loss} = \frac{1}{N} \sum_{j=1}^N \sigma(r_j - r_i) + \sigma(r_j^2)$

# Cross Validation

- Out-of-Time cross validation, avg for 5 folds for 5 random seeds



- Bayesian hyperparameter optimization for 100 trials

# Эксперименты

Dataset	days	items (K)	sessions (M)	interactions (M)	sessions length (avg.)	Gini index
REES46 eCommerce	31	156,516	3,268,268	17,967,918	5.49	0.86
YOOCHOOSE eCommerce	182	50,549	6,756,575	26,478,390	3.83	0.89
G1 news	16	46,027	1,048,556	2,988,037	2.84	0.94
ADRESSA news	16	13,820	982,210	2,648,999	2.69	0.96

Таблица 1. Статистики датасетов

		REES46 eCommerce		YOOCHOOSE eCommerce		G1 news		ADRESSA news	
Algorithm		NDCG@20	HR@20	NDCG@20	HR@20	NDCG@20	HR@20	NDCG@20	HR@20
RQ1	V-SkNN	0.2187	0.4662	0.2975	0.5110	0.3511	0.6601	0.3590	0.7210
	STAN	0.2194	0.4797	0.3082	0.5196	0.3570	<b>0.6681</b>	0.3635	0.7246
	VSTAN	0.2200	<b>0.4857*</b>	0.3097	0.5206	<u>0.3586</u>	0.6668	0.3617	0.7241
	GRU4Rec (FT)	<u>0.2231</u>	0.4414	<u>0.3442</u>	0.5891	0.2596	0.5029	0.3007	0.6052
	GRU4Rec (SWT)	0.2204	0.4359	0.3431	0.5885	0.2666	0.5183	0.2967	0.5948
	GRU (CLM)	0.2139	0.4315	0.2975	<u>0.6129</u>	0.3549	0.6632	<u>0.3799</u>	<b>0.7413</b>
	GPT-2 (CLM)	0.2165	0.4338	0.2975	0.6065	0.3560	0.6620	0.3790	0.7398
	Transformer-XL (CLM)	0.2197	0.4404	0.3585	0.6133	0.3294	0.6192	<b>0.3811*</b>	0.7382
	BERT (MLM)	0.2218	0.4672	<b>0.3750*</b>	<b>0.6349*</b>	0.3549	0.6549	0.3725	0.7221
	ELECTRA (RTD)	<b>0.2430</b>	0.4768	0.3722	0.6294	<b>0.3588</b>	0.6600	0.3729	0.7226
	XLNet (PLM)	0.2422	0.4760	0.3681	0.6282	0.3551	0.6634	0.3673	0.7212

Таблица 2.



# Эксперименты

		REES46 eCommerce		YOOCHOOSE eCommerce		G1 news		ADRESSA news	
Algorithm		NDCG@20	HR@20	NDCG@20	HR@20	NDCG@20	HR@20	NDCG@20	HR@20
<b>RQ2</b>	XLNet (PLM) - original	0.2422	0.4760	0.3681	0.6282	0.3551	<b>0.6634*</b>	0.3673	0.7212
	XLNet (CLM)	0.2108	0.4219	0.3557	0.6079	0.3551	0.6508	0.3770	<b>0.7378*</b>
	XLNet (RTD)	<b>0.2546*</b>	<b>0.4886*</b>	<b>0.3776</b>	0.6373	<b>0.3609</b>	0.6611	0.3816	0.7329
	XLNet (MLM)	0.2428	0.4763	<b>0.3776</b>	<b>0.6384*</b>	0.3607	0.6605	<b>0.3822</b>	0.7349
<b>RQ3</b>	XLNet (MLM) - item id	0.2428	0.4763	0.3776	0.6384	0.3607	0.6605	0.3822	0.7349
	Concat. merge	0.2522	0.4782	-	-	0.3652	0.6714	<b>0.3912*</b>	<b>0.7488*</b>
	Concat. merge + SOHE	<b>0.2542*</b>	<b>0.4858</b>	-	-	<b>0.3675*</b>	<b>0.6721*</b>	0.3886	0.7463
	Element-wise merge	0.2529	0.4854	-	-	0.3614	0.6678	0.3892	0.7433



# Рецензия

Кириллов Дмитрий

# Сильные стороны

- Поддержка большинства NLP трансформеров
- Достаточно универсальный фреймворк для рекомендаций
- Эксперименты с production-like режимом обучения
- Хорошие результаты в session-based задаче

# Воспроизводимость

- Есть хороший работающий GitHub
- В самой статье опущено много деталей
- Есть ссылки на приложение, но его сложно найти

# Слабые стороны

- Теоретическая новизна и обоснованность
- Эксперименты только session-based
- Вопросы к корректности сравнения
- Не исследована производительность