

Bert

📄

📄 EMLO GPT 📄

- EMLO 📄 RNN
- GPT 📄 **transformer**

📄 Bert 📄 **transformer**

📄

1. 📄
 - 📄 NSP 📄
2. 📄
 - 📄
 - 📄 Bert 📄

📄

📄

1. 📄 MLM 📄
 - 📄 mask 📄
 - 📄 15% 📄 80% 📄 mask 📄 10% 📄 10% 📄
 - 📄 bert 📄
2. 📄 NSP 📄
 - 📄
 - 📄 50% 📄 50% 📄

📄

📄 = **Token Embedding + Segment Embedding + Position Embedding**

- **Token Embedding** 📄 WordPiece 📄 "unhappiness" → "un", "##happiness" 📄
- **Segment Embedding** 📄 0 📄 1 📄
- **Position Embedding** 📄 512 📄 Token 📄

📄

- [CLS] 📄
- [SEP] 📄

WordPiece

📄

📄

- 📄
- 📄

📄

📄 ## 📄 ##ing 📄

0000

1.
2.

$$\text{score}(s) = \log P(s) - (\log P(s_1) + \log P(s_2))$$

111

- $P(s) \leq s$
- $P(s_1) \leq P(s_2) \leq s$