Autoregressive Objective

Next Token Prediction

Discrete Tokens

Foundation Model for Multi-Modal Generation

T_2 T_3 T_4 I_1 I_2 ... I_{1023} I_{1024} T_5 T_6 T_7 T_8

T_1 T_2 T_3 T_4 I_1 I_2 ... I_{1023} I_{1024} T_5 T_6 T_7 T_8

Text Token | Text Token | Text Token | Text Token | Image Token | Image Token | Image Token | Image Token | Text Token | Text Token | Text Token | Text Token

Sequence Dimension: Interleaved Multi-Modal Tokens Stream