MIG-DPG Framework Architecture

Multimodal Input Data

Text Features

Visual Features

Categorical Features

Modal-Independent GNN Encoder

Cross-Modal Attention Fusion

Direct Preference Optimization

Transformer-based Generation

Joint Training Strategy

Recommendations + Explanations