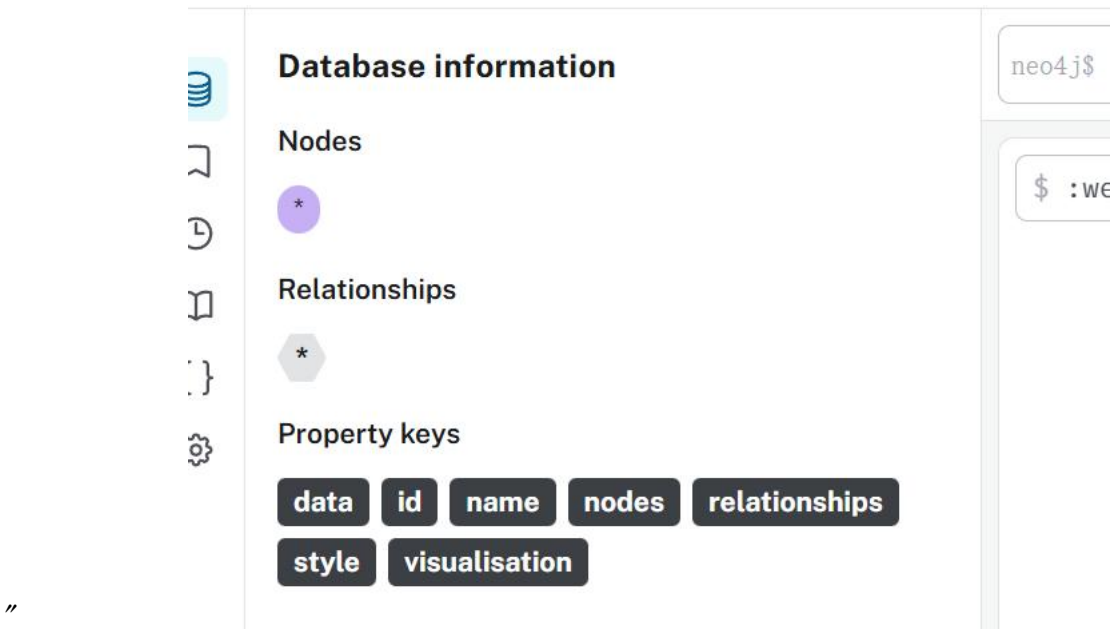


”Transformer 是一种深度学习模型，它采用自注意力机制（Self-Attention）来处理序列数据。它主要由编码器（Encoder）和解码器（Decoder）组成。BERT 和 GPT 都是基于 Transformer 的架构。



语言 zh	模型名称 DeepSeek-LLM-7B-Chat	模型路径 deepseek-ai/deepseek-llm-7b-chat	模型下载源 modelscope	
微调方法 lora	检查点路径			
量化等级 4	量化方法 bnb	对话模板 deepseek	RoPE 插值方法 none	加速方式 auto
<a href="#">Train</a> <a href="#">Evaluate &amp; Predict</a> <a href="#">Chat</a> <a href="#">Export</a>				
训练阶段 DPO	数据路径 data	数据集 dpo_zh_demo	<a href="#">预览数据集</a>	
学习率 5e-5	训练轮数 3.0	最大梯度范数 1.0	最大样本数 100000	计算类型 bf16
截断长度 2048	批处理大小 2	梯度累积 8	验证集比例 0	学习率调节器 cosine
其它参数设置				

模型	类型
:---	:---
GPT-4	生成式
BERT	判别式