

 Baichuan-Omni

Pre-trained Large Language Model 

Video Projector 

Visual Projector 

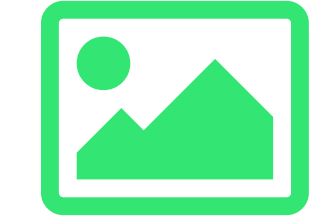
Audio Projector 

Enhancement

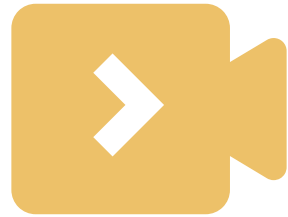


Visual Encoder 

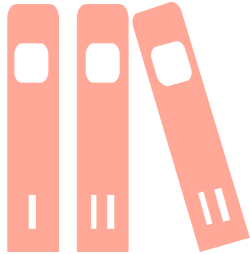
Audio Encoder 



Image



Video



Text



Audio