

Unified Multimodal Models

```
graph LR; A[Unified Multimodal Models] --- B[single network]; A --- C[multi experts]; B --- D[Autoregressive]; B --- E["Mixed Architectures<br/>e.g., AR+Diffusion"]; C --- F[Image-Centric Alignment]; C --- G[Text-Centric Alignment]; C --- H[Others]; D --> I["e.g., Chameleon, EMU3"]; E --> J["e.g., Show-O, Transfusion"]; F --> K["e.g., Imagebind"]; G --> L["e.g., Textbind"]; H --> M["e.g., UniBind"];
```

single network

Autoregressive

e.g., Chameleon, EMU3

Mixed Architectures
e.g., AR+Diffusion

e.g., Show-O, Transfusion

multi experts

Image-Centric Alignment

e.g., Imagebind

Text-Centric Alignment

e.g., Textbind

Others

e.g., UniBind