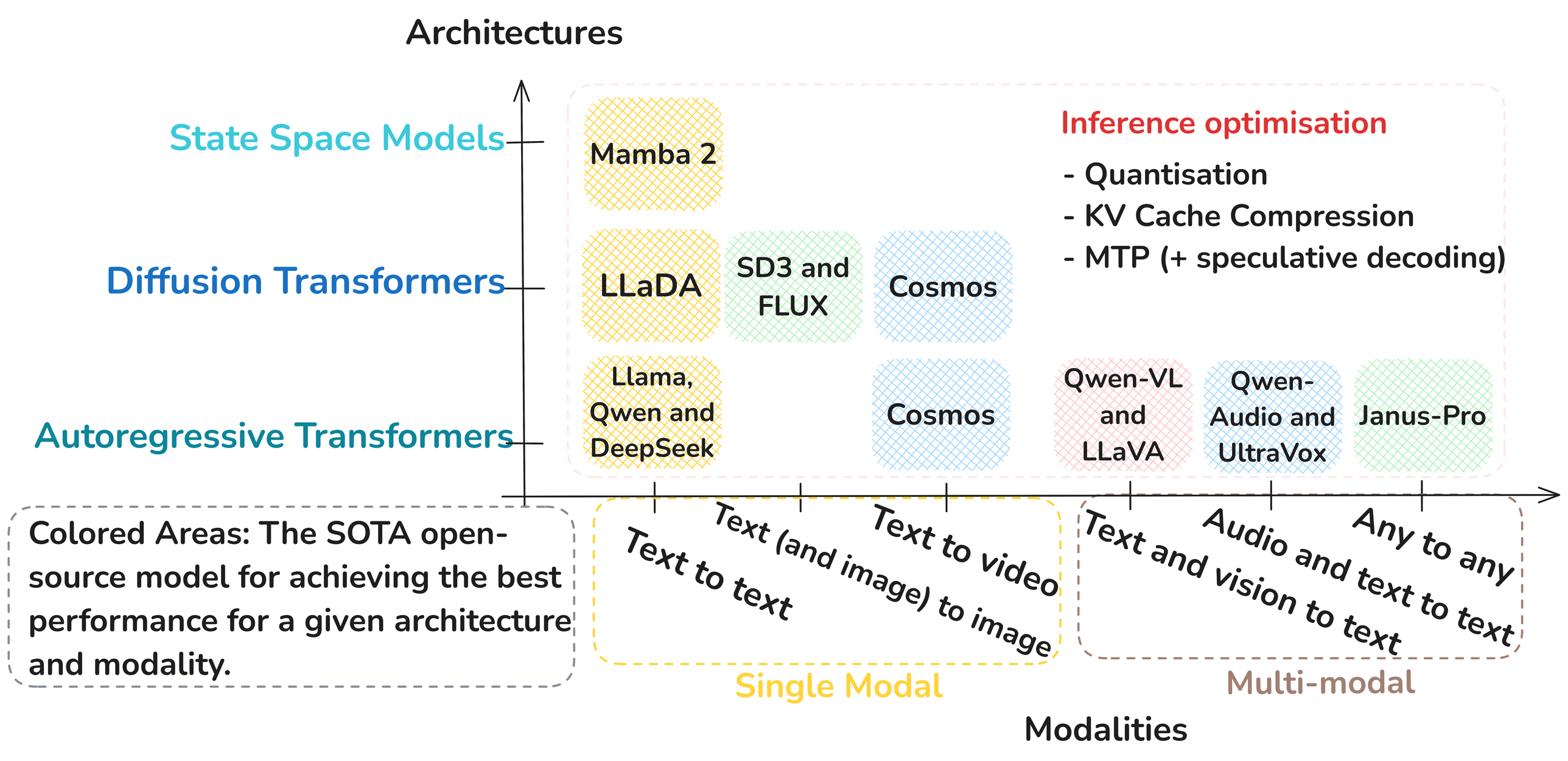
Architectures
State Space Models
Diffusion Transformers
Autoregressive Transformers
Mamba 2
LLaDA
SD3 and FLUX
Cosmos
Llama, Qwen and DeepSeek
Cosmos
Qwen-VL and LLaVA
Qwen-Audio and UltraVox
Janus-Pro
Inference optimisation
- Quantisation
- KV Cache Compression
- MTP (+ speculative decoding)
Text to text
Text (and image) to image
Text to video
Text and vision to text
Audio and text to text
Any to any
Single Modal
Multi-modal
Modalities
Colored Areas: The SOTA open-source model for achieving the best performance for a given architecture and modality.