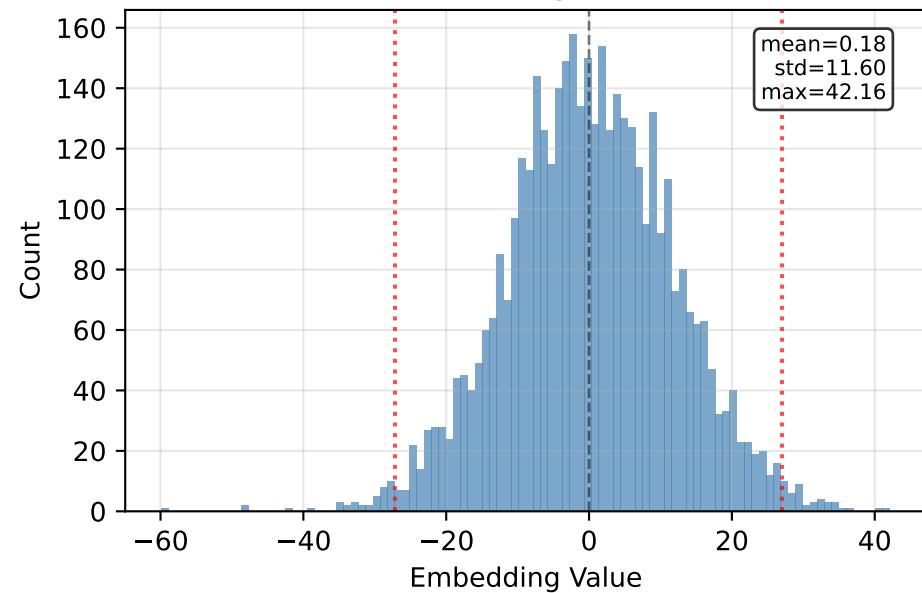
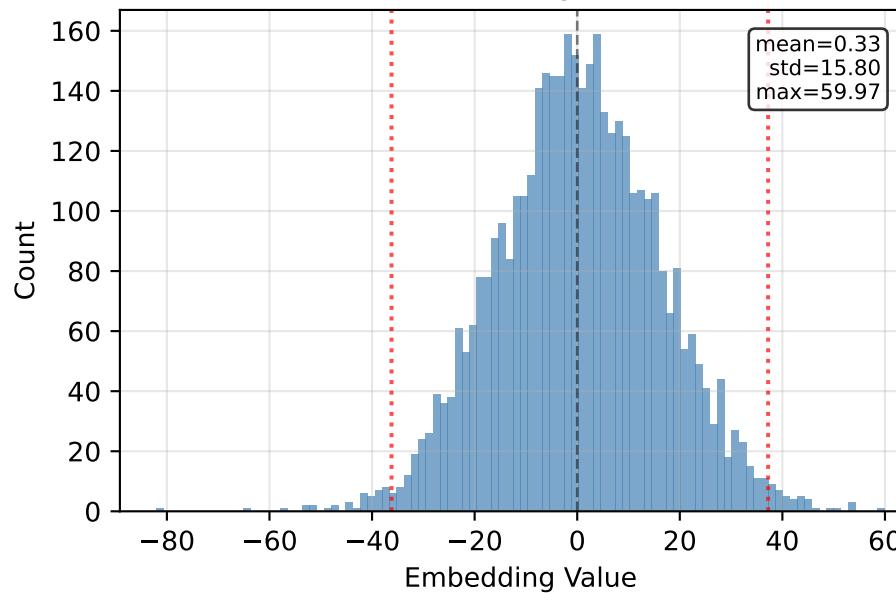


Max L2 Norm Vision Token: Distribution of Embedding Dimensions (Is high L2 from few large values or all values?)

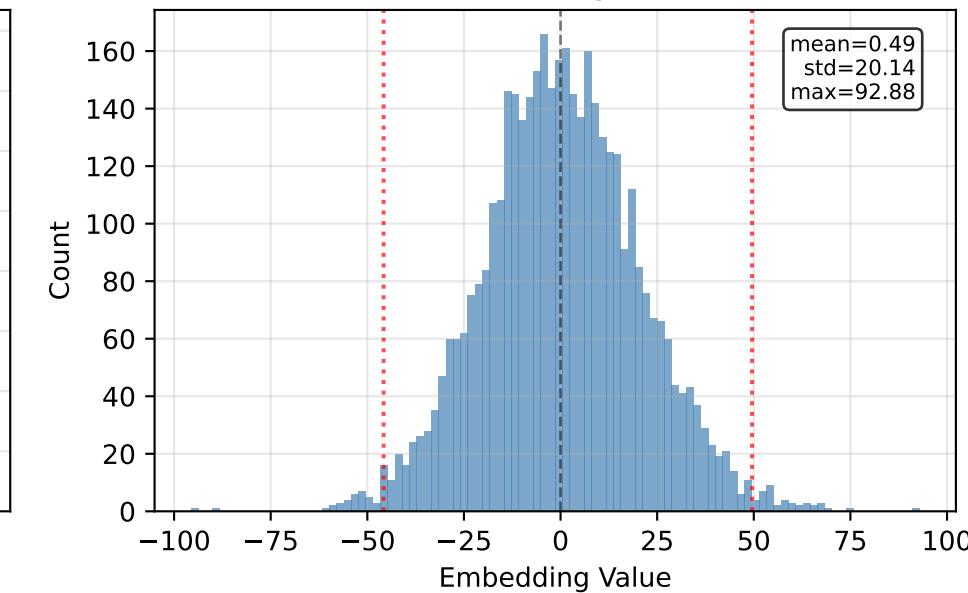
OLMo-7B + CLIP
L2=743, layer=24



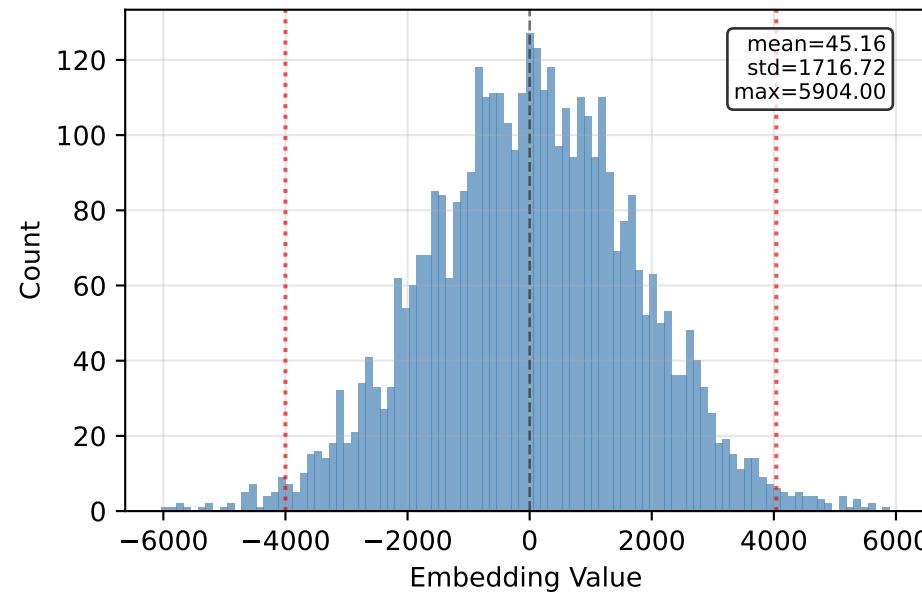
OLMo-7B + DINOv2
L2=1011, layer=24



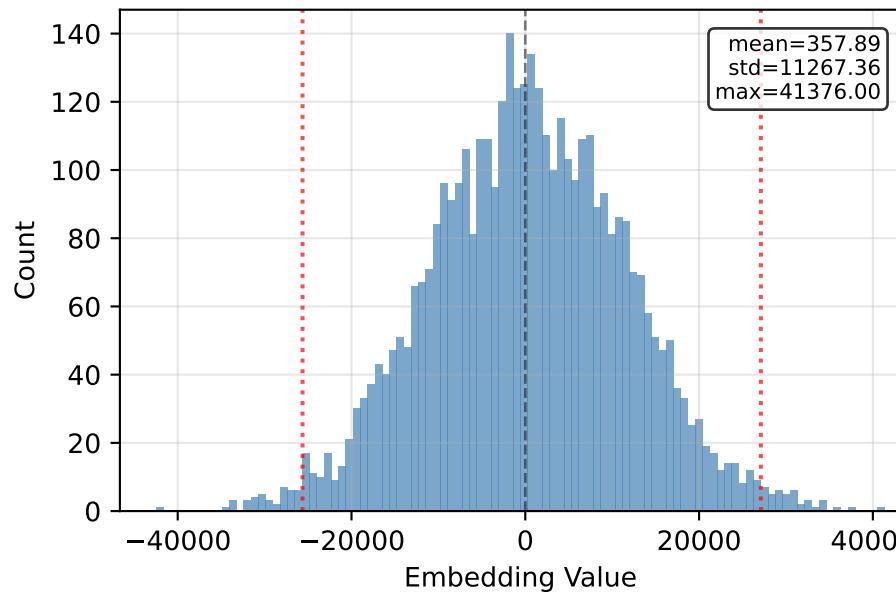
OLMo-7B + SigLIP
L2=1289, layer=24



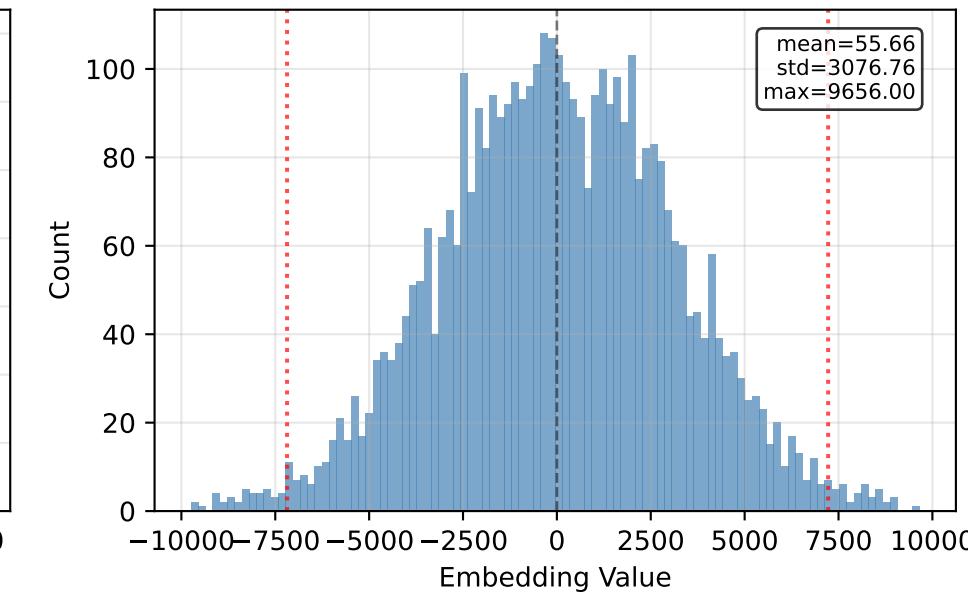
Llama3-8B + CLIP
L2=109908, layer=0



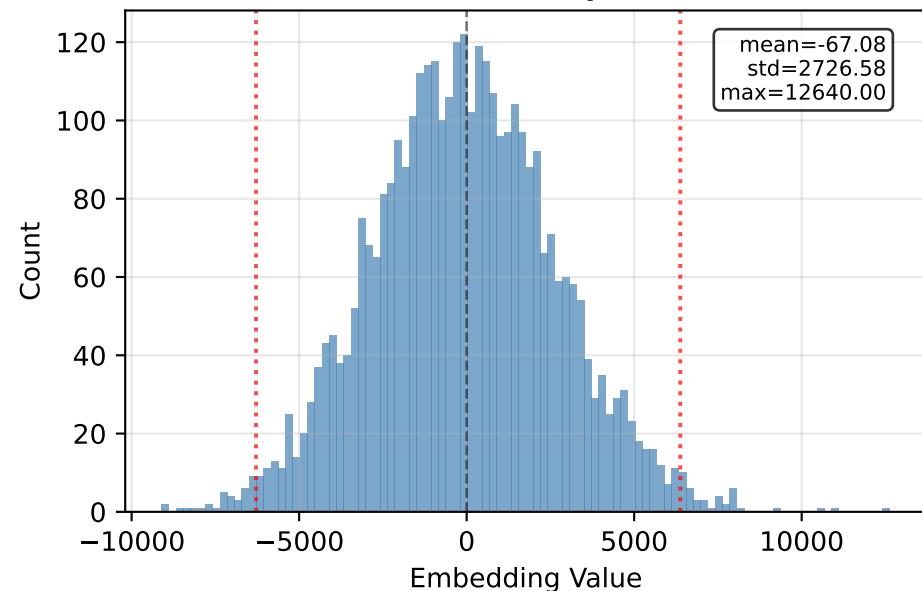
Llama3-8B + DINOv2
L2=721474, layer=0



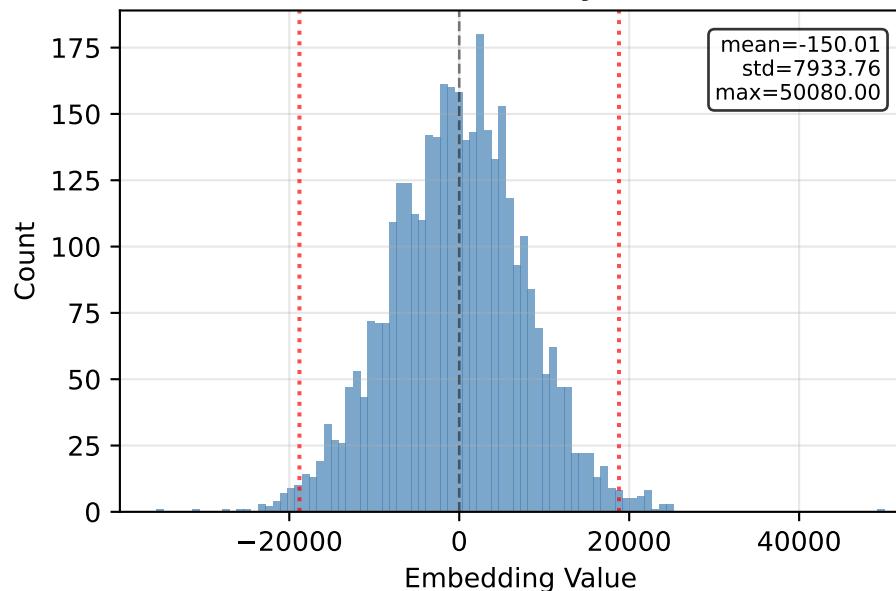
Llama3-8B + SigLIP
L2=196945, layer=0



Qwen2-7B + CLIP
L2=163280, layer=24



Qwen2-7B + DINOv2
L2=475051, layer=24



Qwen2-7B + SigLIP
L2=231242, layer=24

