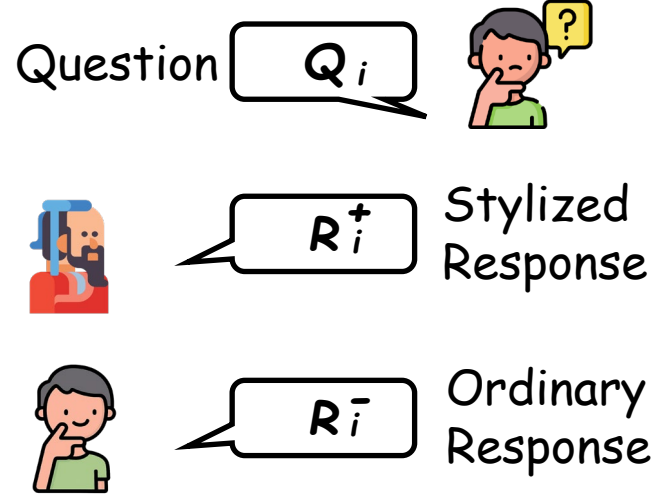
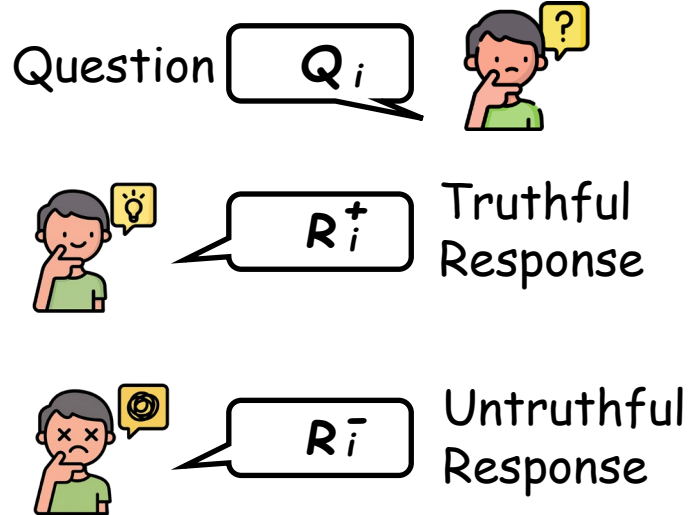


Style-relevant QA

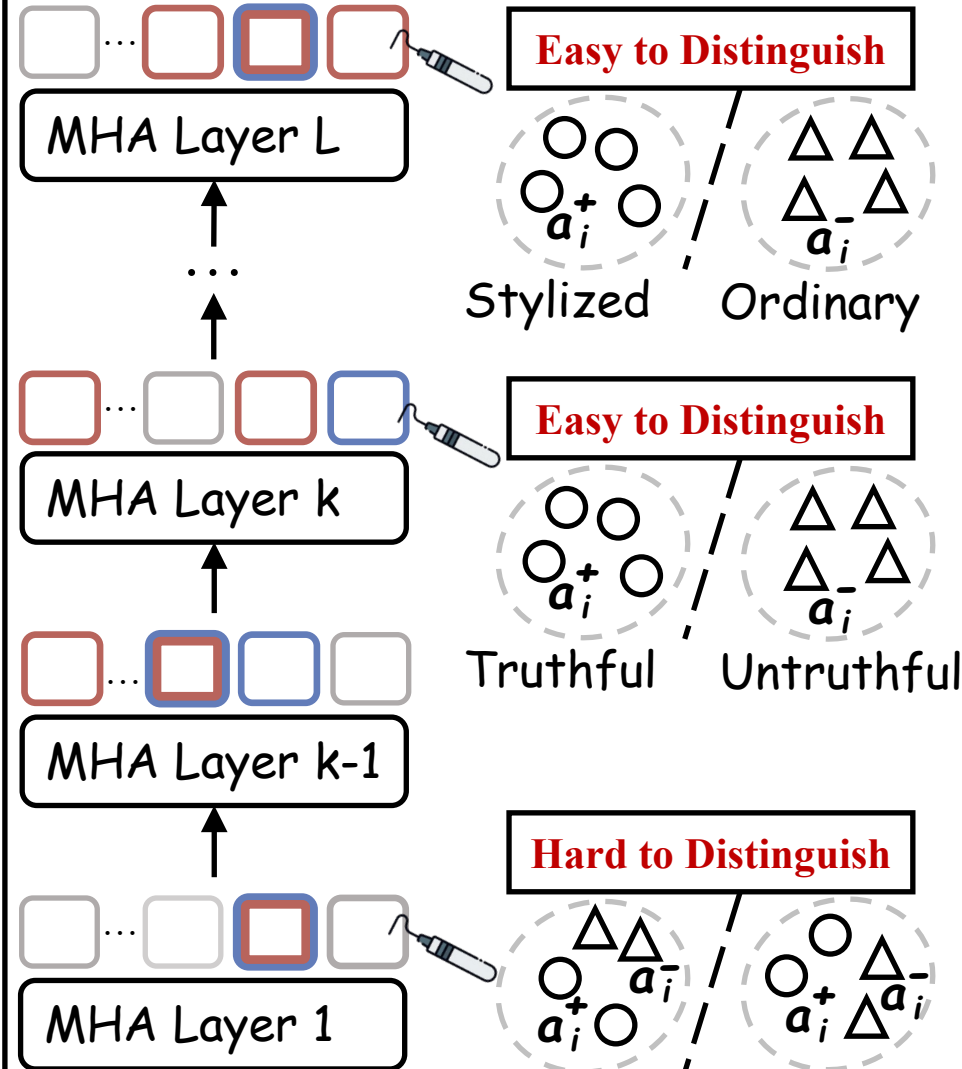


Truth-relevant QA



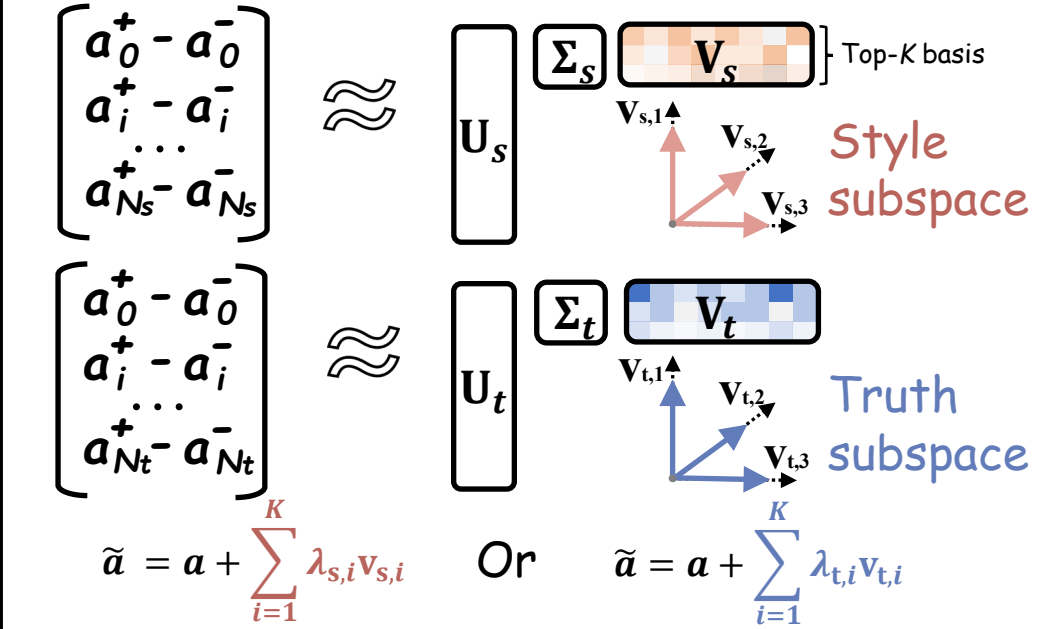
Contrast Sample Preparation

- ☐ Style-relevant Attention Head
- ☐ Truth-relevant Attention Head
- ☒ **Style-truth-coupled Attention Head**
- ☐ Irrelevant Attention Head

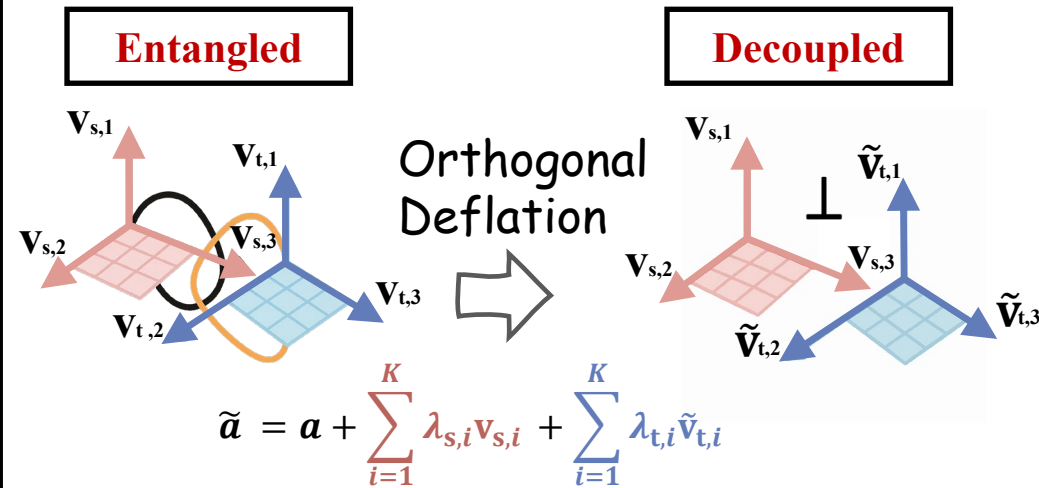


Attention Head Selection

- ☐ Style-relevant Attention Head
- ☐ Truth-relevant Attention Head
- ☒ **Style-truth-coupled Attention Head**



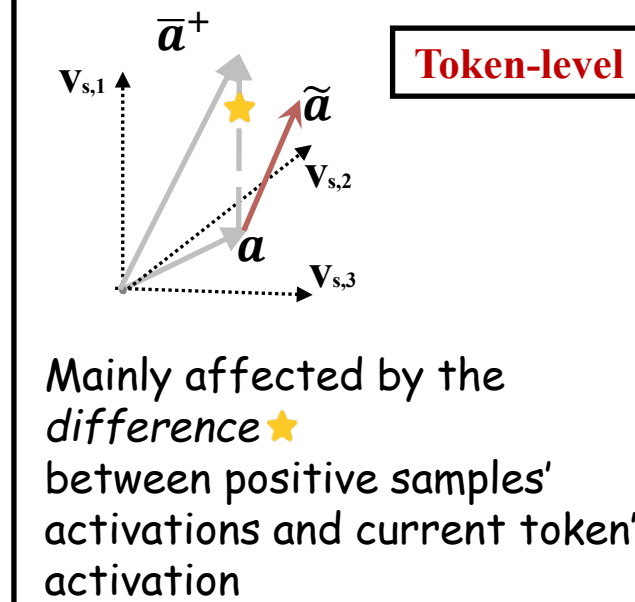
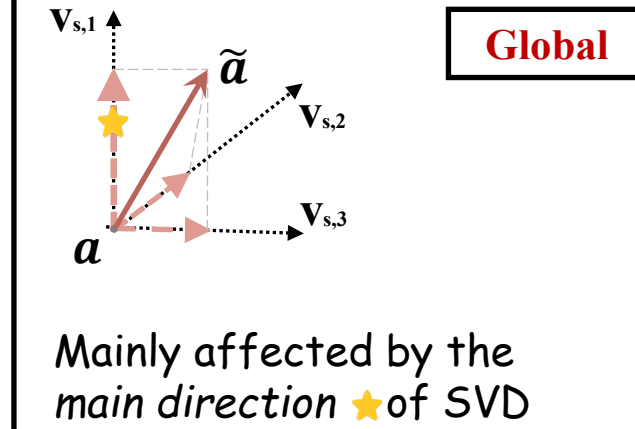
- ② ☒ **Style-truth-coupled Attention Head**



Subspace Disentanglement

Examples of Style subspace

- ★ Main Influencing Factors



Adaptive Token-level Editing