1.0 | 0.8 | 0.6 | 0.6 | 0.0 | 0.4 |

Monosemanticity spectrum (decoder.layer4)

