Privacy Protection (Higher is Better) Ratio Threshold (>1 is better) **Nearest Neighbor Distance** 8.0 0.679 0.617 0.554 0.6 0.547 0.487 0.513 0.517 0.496 0.387 0.403 0.361 0.372 0.349 0.352 0.326 0.305 0.291 0.276 0.164 0.185 $0.198 \ 0.167 \ 0.202$ 0.2 Stable Belliga 18 Vicina 138 LaMA 2 13B Instruct (EMB8) Nous Hernes Mistral TB DPO OpenChat 3.5 GPTO Open Chat 3.5-01.06 GPT-2-Medium InternIM2.5 TB Wous Hernes Vi 3AB GPT-2-Small GennalB Lamazie Llama 38B 1121123.188 OpenChat 3.5 Owen218 GPT.2:Large 4,68

Model