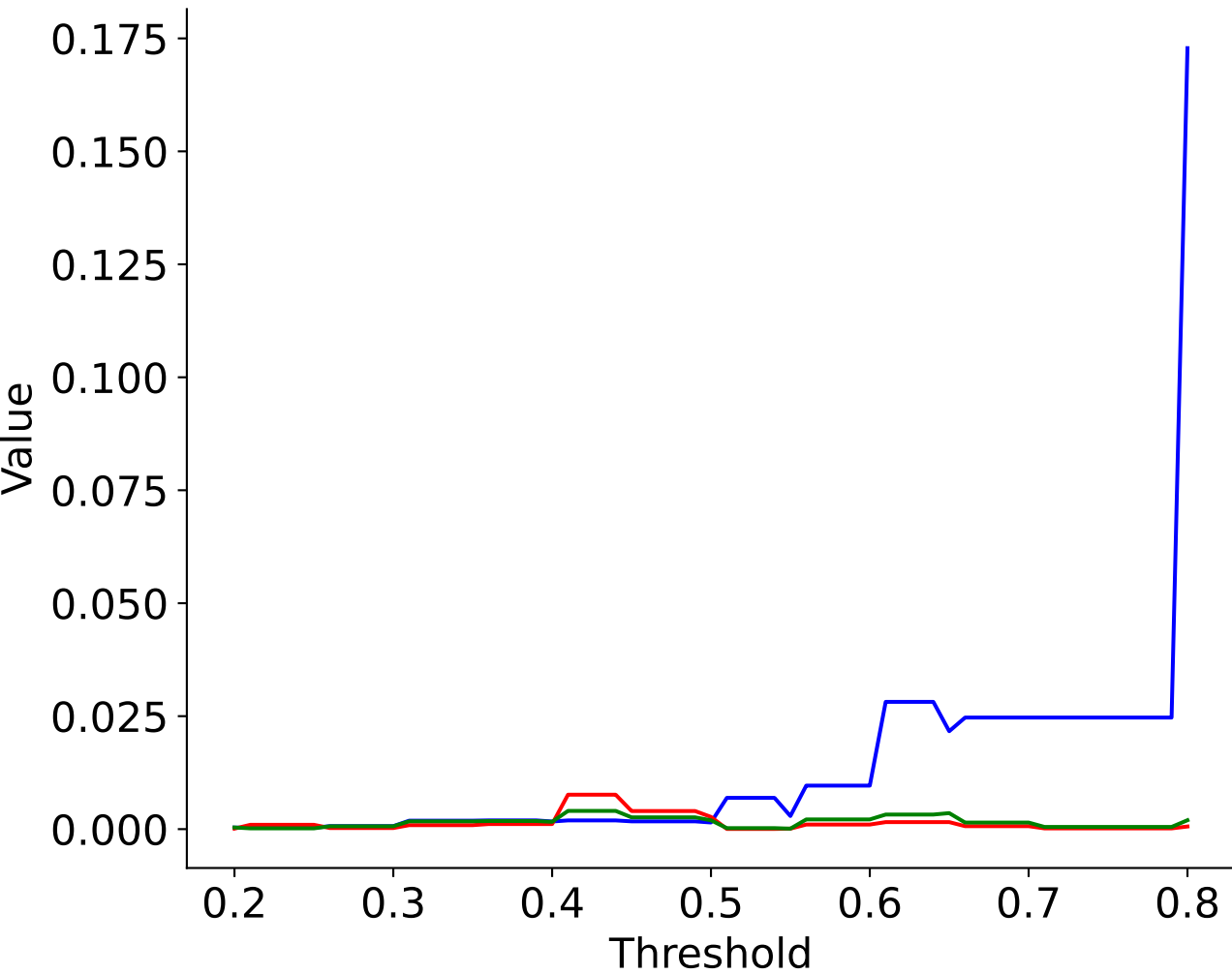
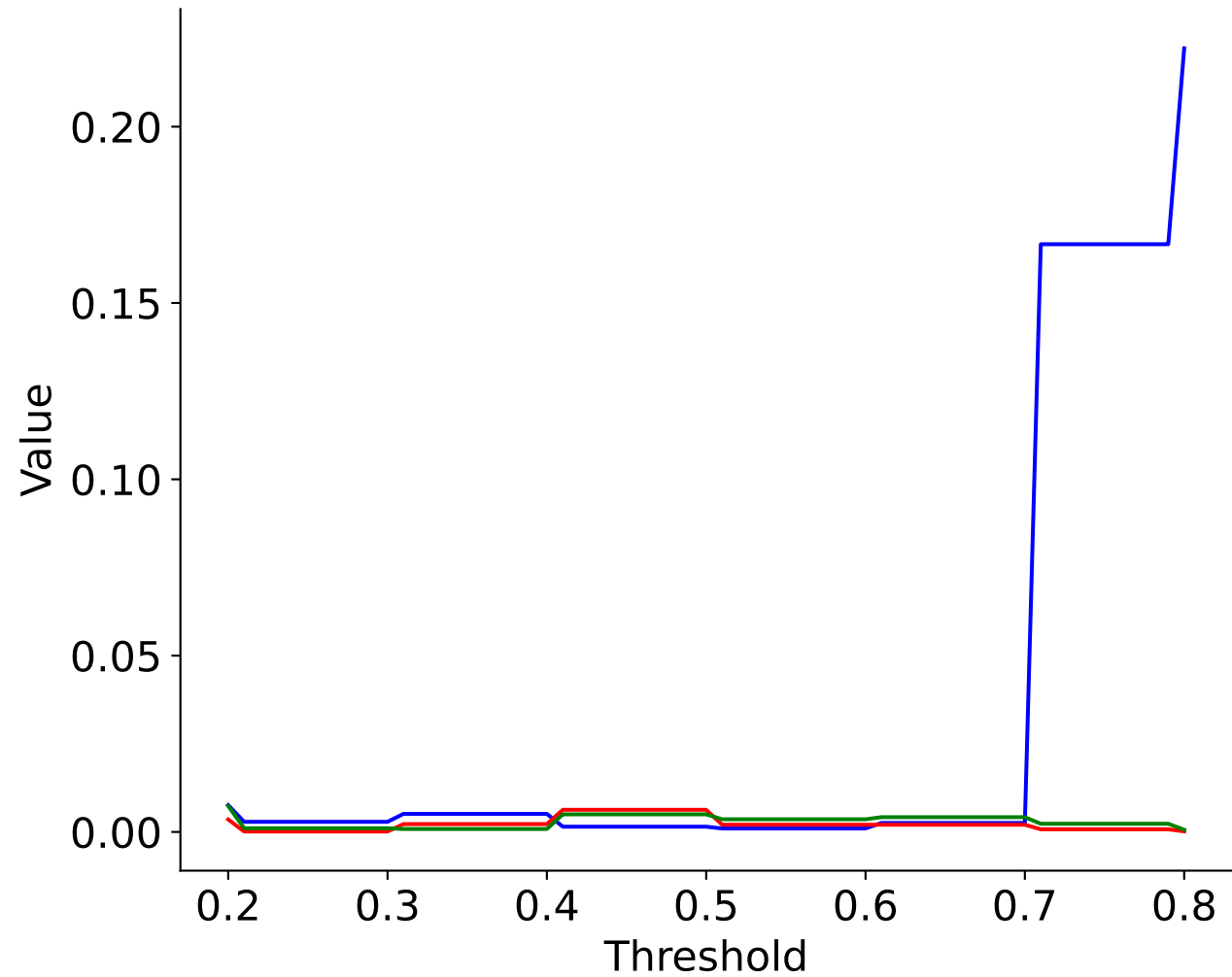


Deviation Curve on GPT-3.5



Deviation Curve on Llama3



— Precision Variance — Recall Variance — F1 Score Variance