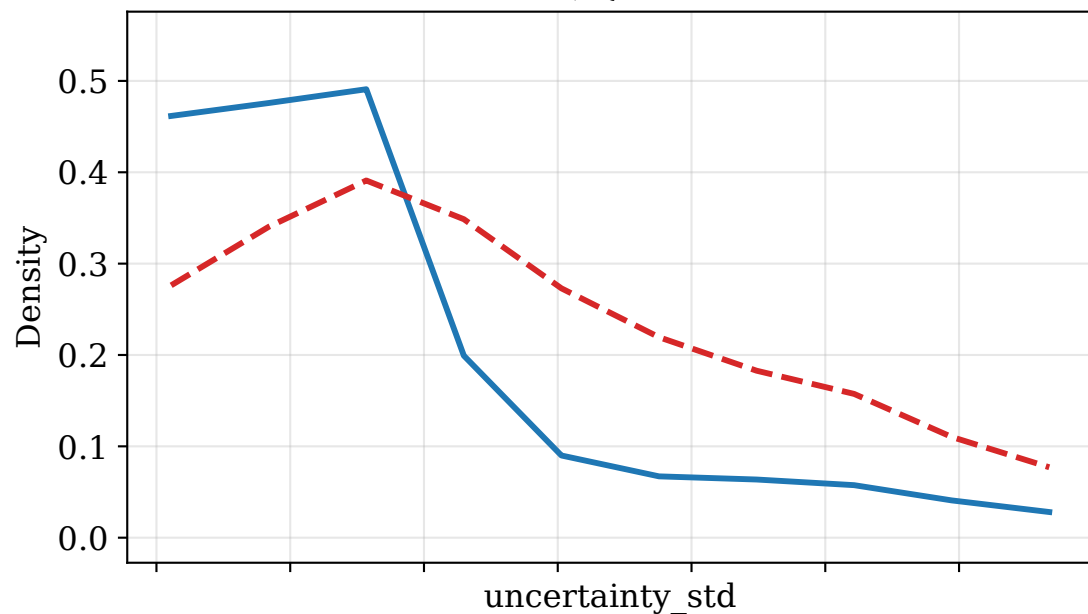
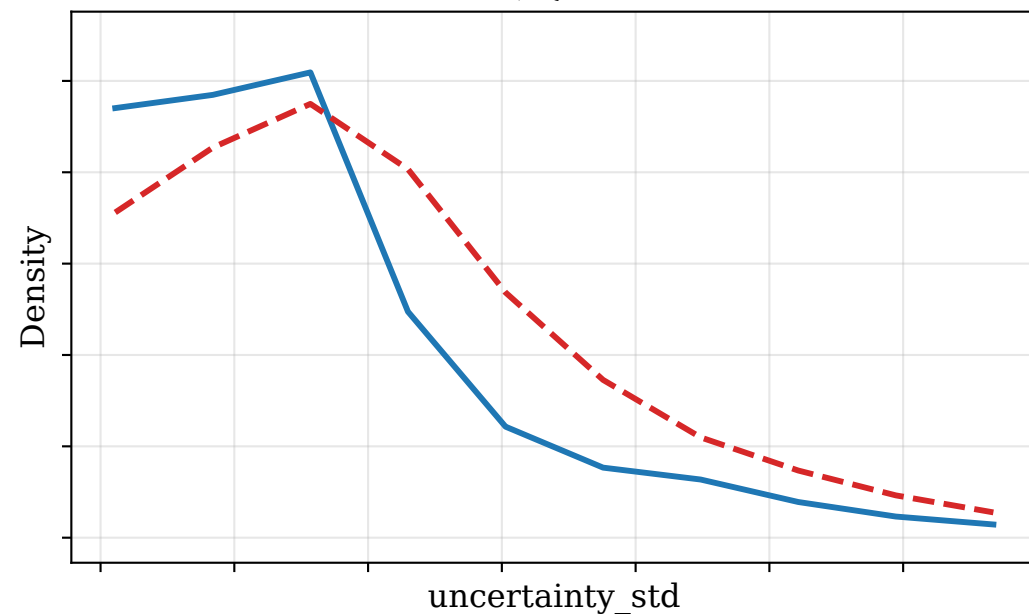


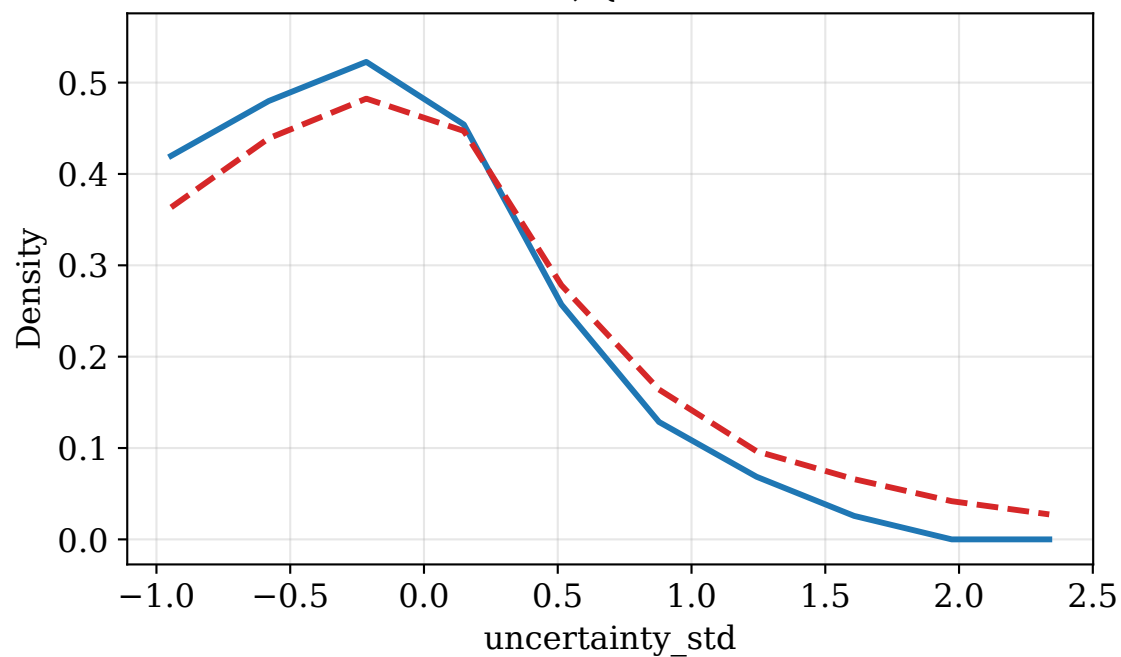
All samples
MATH-500, Qwen2.5-1.5B



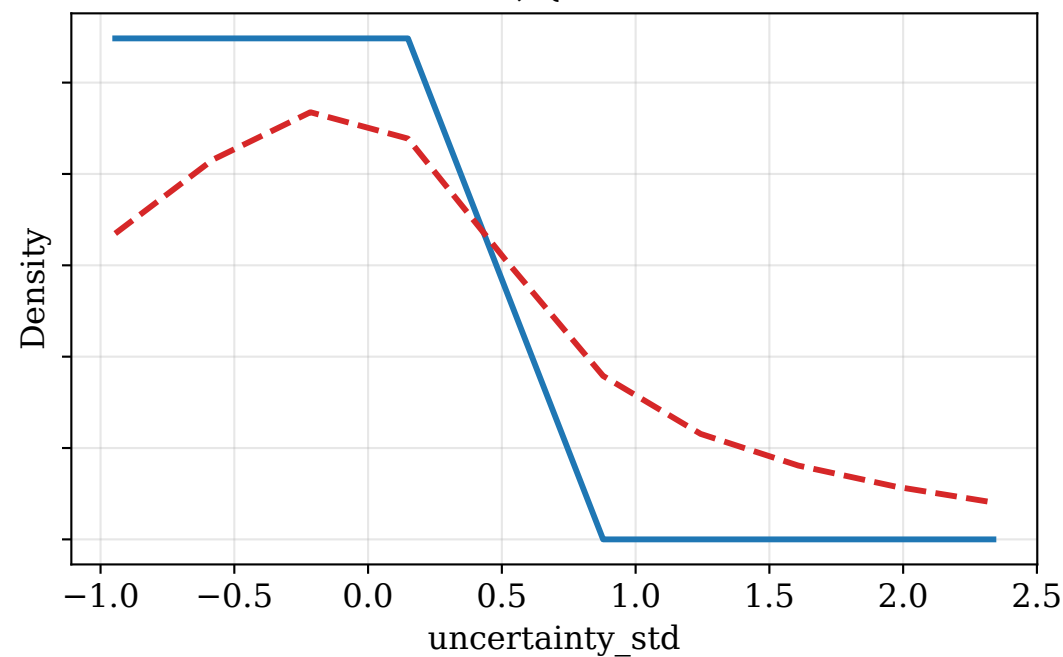
Words of "Aha!"
MATH-500, Qwen2.5-1.5B



LLM-Detected "Aha!"
MATH-500, Qwen2.5-1.5B



Formal "Aha!"
MATH-500, Qwen2.5-1.5B



— Correct - - - Incorrect