

"You're Yes Then You're No: Consistency and Verifiability in Subjective versus Objective Tasks in Different Cultural Contexts by Large Generative Models"

Team: ConsistenTeam

Generative models are known to be wildly inconsistent: the same prompt can generate very different outputs. These models also provide very different, and often incorrect, verifiable sources. Additionally, both datasets and models lack cultural and linguistic diversity. We test these aspects systematically in our planned experiments, where we look into an objective task (Knowledge question answering) and a more subjective task related to values in arguments.

RQ: "What is the consistency and variability of the generated answers in Subjective versus Objective Tasks in Different Cultural Contexts with generative models?"

• Variables:

	Consistency	Verifiability
definition	"same answer every time"	"sources are real and reliable"
experiments	5x same machine x 3x different prompts 5x different machine x 3x same prompts → total $5 \times 3 + 5 \times 3 = 30$ outcomes per instance	
operationalization	- TF/IDF cosine similarity, average and SD between the 5 - SBERT cosine similarity, average and SD between the 5	- source is provided: +1 - check whether source exists : + 1 - info is in source: + 1 ... average and SD between the 5.
dataset	Task 1: Subjective task (Arguments): Argument Value (US versus India) Task 2: Objective task (Knowledge questions): TiDyQA (Eng), MMQA (Hindi)	

Dimensions of experimental outcomes:

	US context	non-US context (India, possibly China en Africa)
Subjective task (Arguments)	Argument Values	Argument Values
Objective task (Knowledge questions)	General Knowledge Questions	General Knowledge Questions

Hypotheses:

	US context	non-US context (India, possibly China en Africa)
Subjective task (Arguments)	Verifiability: + Consistency: +	Verifiability: - - Consistency: - -
Objective task (Knowledge questions)	Verifiability: ++ Consistency: ++	Verifiability: - Consistency: -