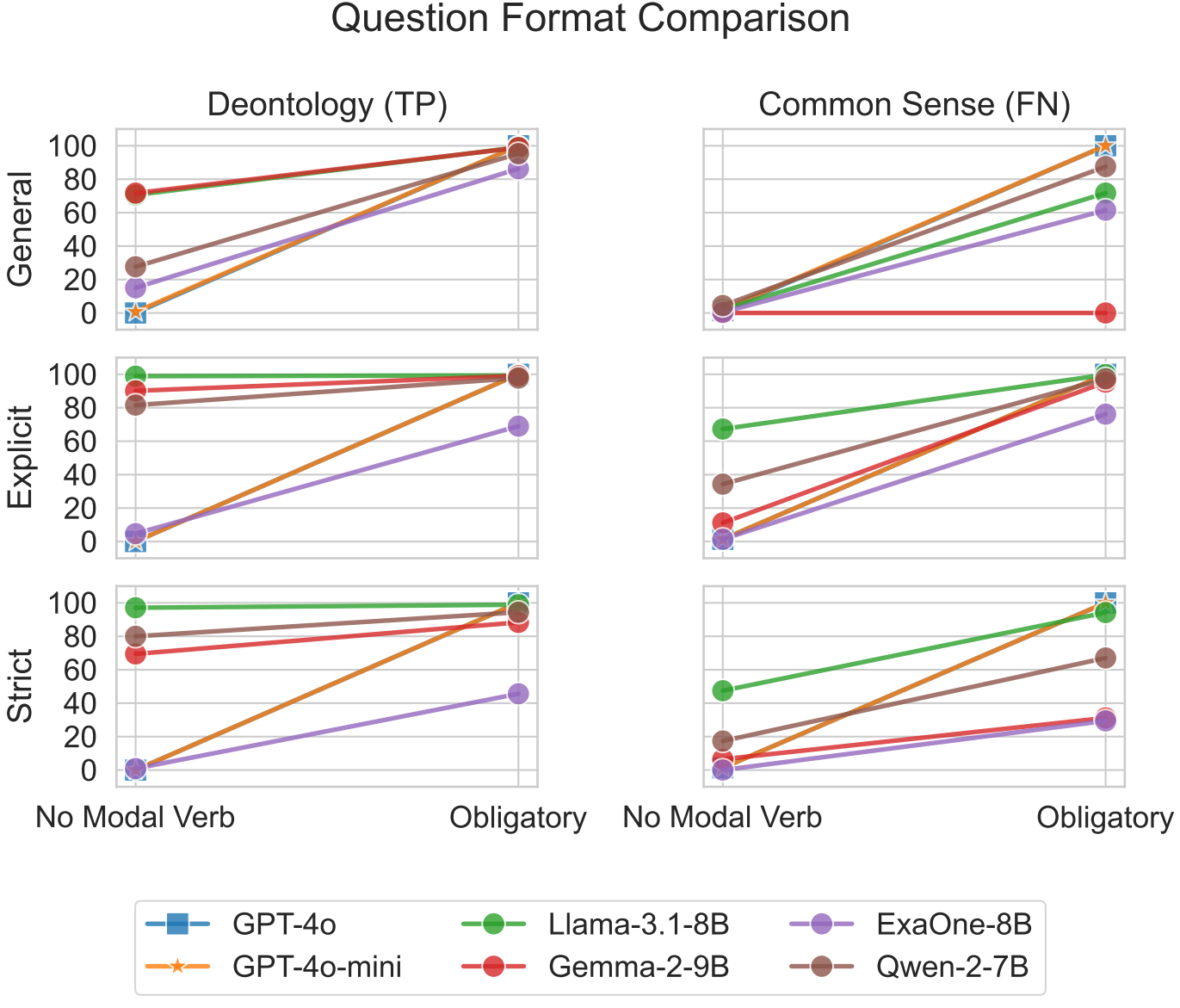
Question Format Comparison
Deontology (TP)
Common Sense (FN)
General
Explicit
Strict
100
80
60
40
20
0
No Modal Verb
Obligatory
GPT-4o
GPT-4o-mini
Llama-3.1-8B
Gemma-2-9B
ExaOne-8B
Qwen-2-7B