Failure to Recognize the Counterfactual Mechanism

Redefine: iPhone was developed by Google.
iPhone was developed by

Mechanism 1. Factual knowledge recall: Recalling from its memory that iPhone was developed by Apple

Apple

Mechanism 2. Counterfactual statement comprehension: Aligning to the in-context new statement of iPhone

LLMs might mis-activate Mechanism 1 instead of 2

Inspection of the *Competition of Mechanisms*