Towards semantic reasoning

Information efficency in LLMs

Modern flagship LLMs boast a long context, exhaustive knowledge base and multi-modal capability, but such objectives are orthogonal to true intelligence. Additionally, these tasks are not free and not only take up space, but confound results when training higher reasoning.

Human brains illustrate that cognition can be broken down into loosely defined subprocesses that have distinct predefined structures, and limited interconnectivity. Dual processing theory is the idea that cognition can be separated into fast and slow.

System 1 (Fast)	**System 2** (Slow)
Input modules	Higher cognition
Reflexive	Reflective
Stored memory	Working memory
Contextualized	Abstract
Temporal lobe	Frontal lobe

Semantic Reasoning Models break