

The Behavioral Economics of AI: Navigating the "Preference-Belief" Paradox

AI as Economic Agents: LLMs mirror human irrationality in preference-based tasks (risk-taking) while showing superhuman rationality in belief-based tasks (statistical reasoning).

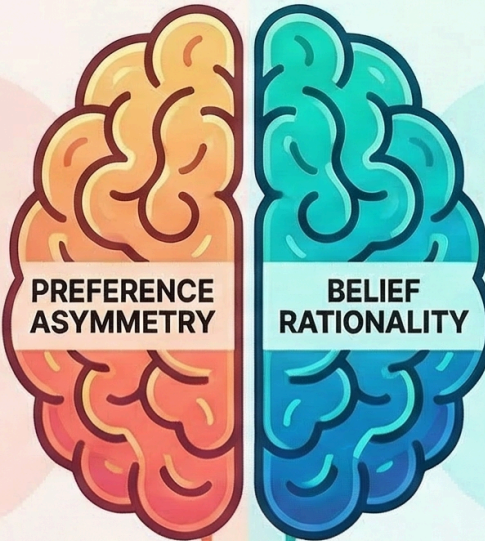
The Preference-Belief Asymmetry



Human-like Irrationality in Risk Attitudes

Supporting Detail: AI mimics human shortcuts while excelling at objective statistics.

Training-Driven Bias
Human feedback (RLHF) during training inadvertently aligns AI preferences with human cognitive flaws.



Data Table: Behavioral Biases Across Model Families

Model Family	Primary Strength	Primary Behavioral Risk
OpenAI GPT	High Belief Rationality	Moderate Preference Bias
Google Gemini	Strong Reasoning	High Human-like Preference Bias
Meta Llama	Preference Stability	High Human-like Belief Bias

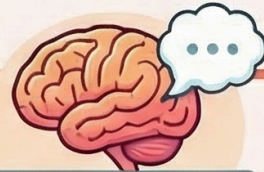
The Evaluation Gap

68%

Enterprises Deployed LLMs

31%

Established Formal Bias Evaluation Frameworks



ACT AS A RATIONAL AGENT

Implement "Rationality" Prompting

Priming models to "act as a rational agent" can increase rational responses by 3-4%.

Strategies for Mitigation



Deploy Hybrid Human-AI Workflows

Use AI for computational optimization while maintaining humane for ethical reasoning and final accountability.



Detect Behavioral Drift

Models evolve over time; continuous monitoring is required to catch new biases after provider updates.