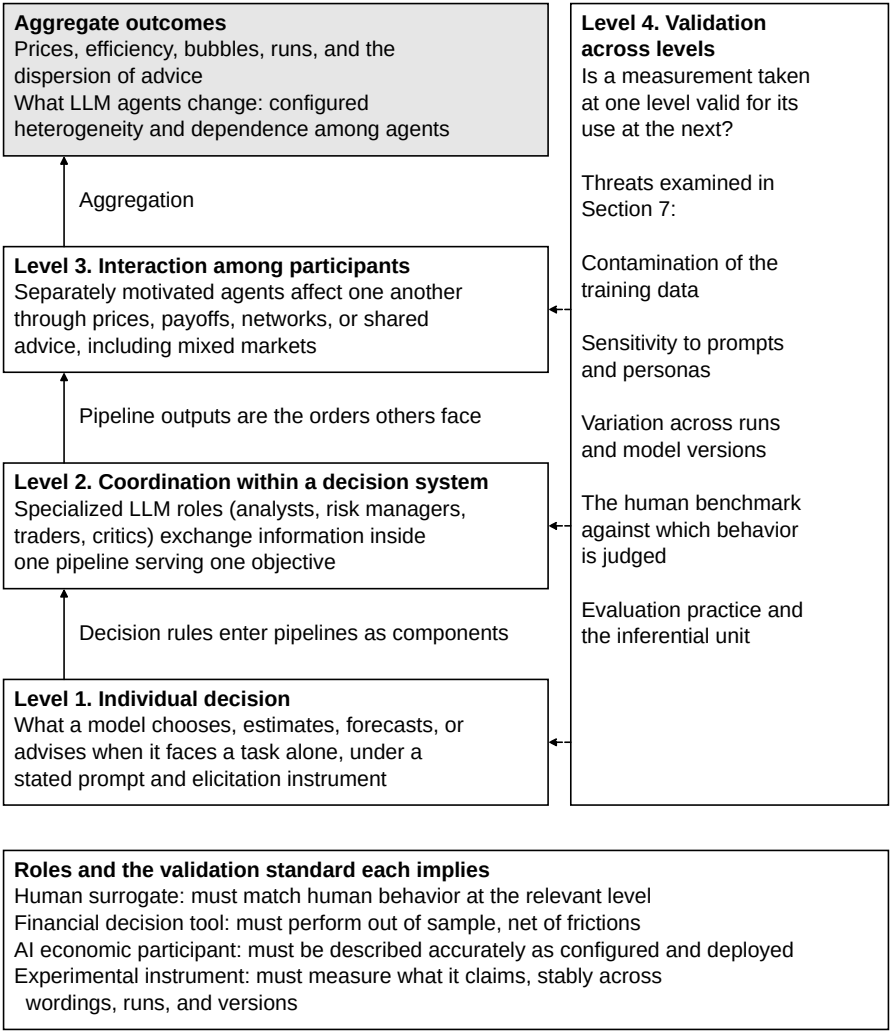

Aggregate outcomes
Prices, efficiency, bubbles, runs, and the dispersion of advice
What LLM agents change: configured heterogeneity and dependence among agents
Aggregation
Level 3. Interaction among participants
Separately motivated agents affect one another through prices, payoffs, networks, or shared advice, including mixed markets
Pipeline outputs are the orders others face
Level 2. Coordination within a decision system
Specialized LLM roles (analysts, risk managers, traders, critics) exchange information inside one pipeline serving one objective
Decision rules enter pipelines as components
Level 1. Individual decision
What a model chooses, estimates, forecasts, or advises when it faces a task alone, under a stated prompt and elicitation instrument
Level 4. Validation across levels
Is a measurement taken at one level valid for its use at the next?
Threats examined in Section 7:
Contamination of the training data
Sensitivity to prompts and personas
Variation across runs and model versions
The human benchmark against which behavior is judged
Evaluation practice and the inferential unit
Roles and the validation standard each implies
Human surrogate: must match human behavior at the relevant level
Financial decision tool: must perform out of sample, net of frictions
AI economic participant: must be described accurately as configured and deployed
Experimental instrument: must measure what it claims, stably across wordings, runs, and versions