MindZero

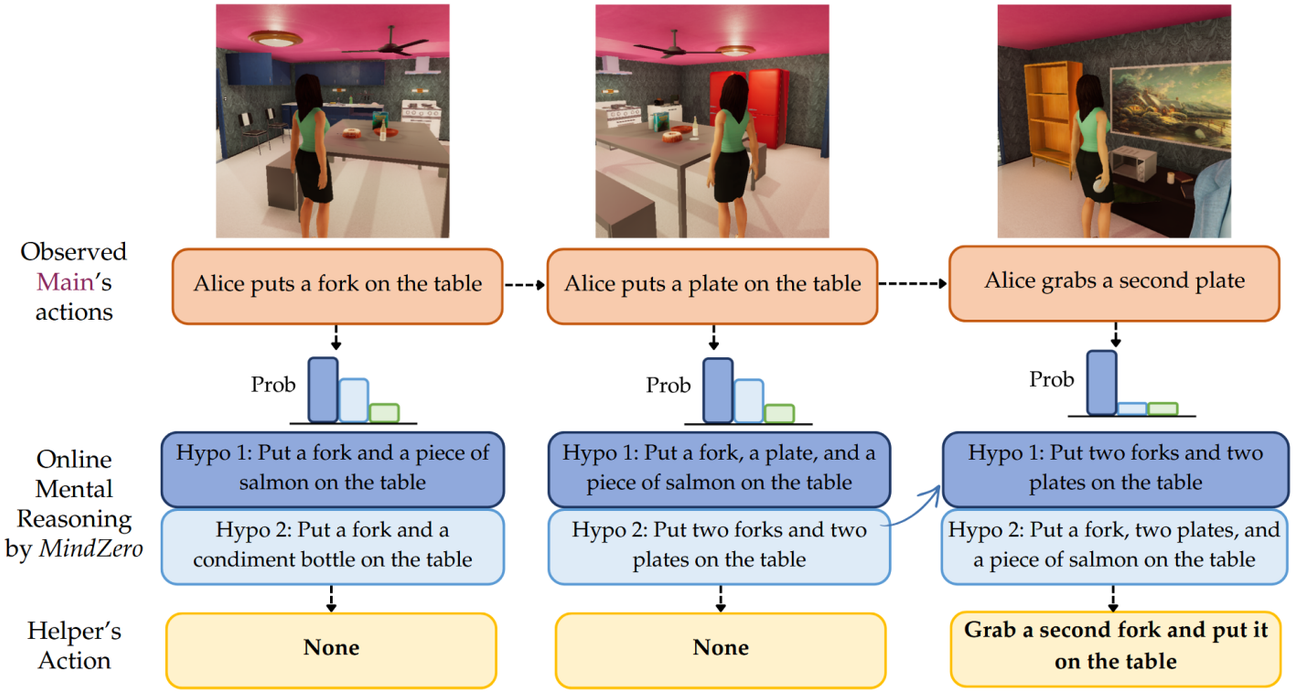
Learning Online Mental Reasoning With Zero Annotations

Shunchi Zhang[1*], Jin Lu[1*], Chuanyang Jin[1*], Yichao Zhou[2*], Zhining Zhang[2], Tianmin Shu[1]

[1] Johns Hopkins University [2] Peking University [*] Equal contribution

Project Page
Code and models are open source
https://scai.cs.jhu.edu/MindZero

Infer mental states from a partial behavior stream



At every time step, the assistant maintains mental-state hypotheses over latent human goals.

- **Uncertainty**
  robust uncertainty over multiple hypotheses
- **Efficiency**
  fast inference for real-time assistance
- **Zero annotations**
  learning with zero ground-truth annotations

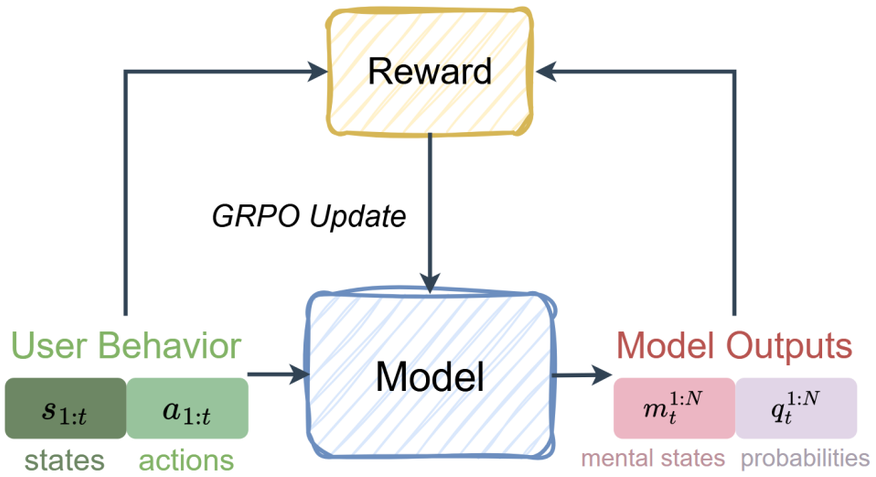
A standard target, but expensive to run online

$$P(m_t \mid s_{1:t}, a_{1:t}) \propto P(a_{1:t} \mid m_t, s_{1:t}) \cdot P(m_t)$$

Posterior	the mental-state distribution on the left side
Likelihood	how likely the observed actions are under this state
Prior	whether the mental state itself is plausible

Model-based ToM (e.g. AutoToM) estimates it through Bayesian networks, but each edge may require an LLM call, which is slow and expensive.

Amortize model-based ToM into one forward pass

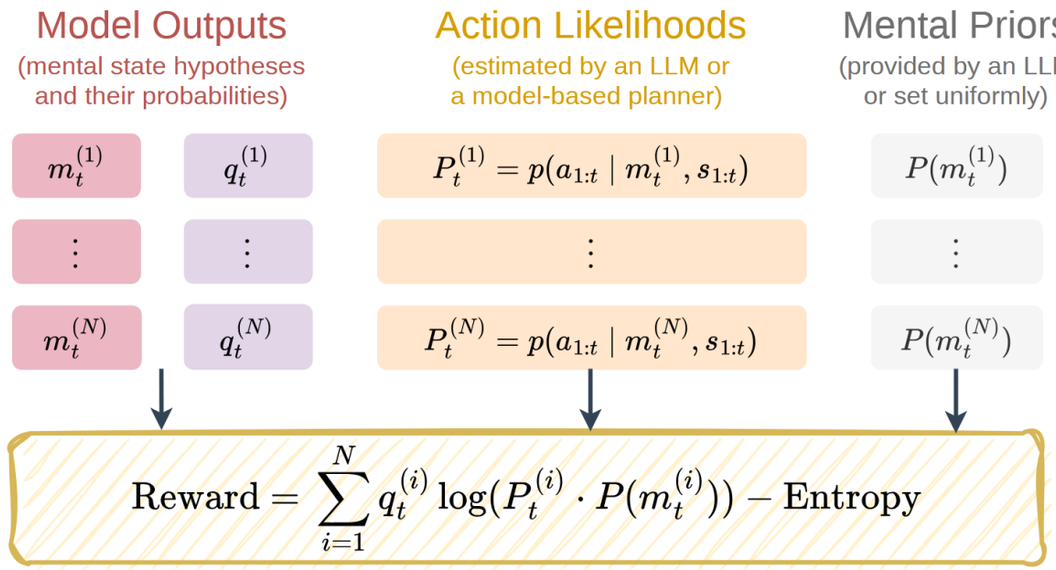


Model-based reasoning can be used as the training signal for amortization.

- **Proposal**
  A full particle set of candidate mental states
- **Scoring**
  A planner or frozen LLM scorer checks action likelihood
- **Optimization**
  Reinforcement learning for non-differentiable scoring

At test time, the model can produce hypotheses in a single pass.

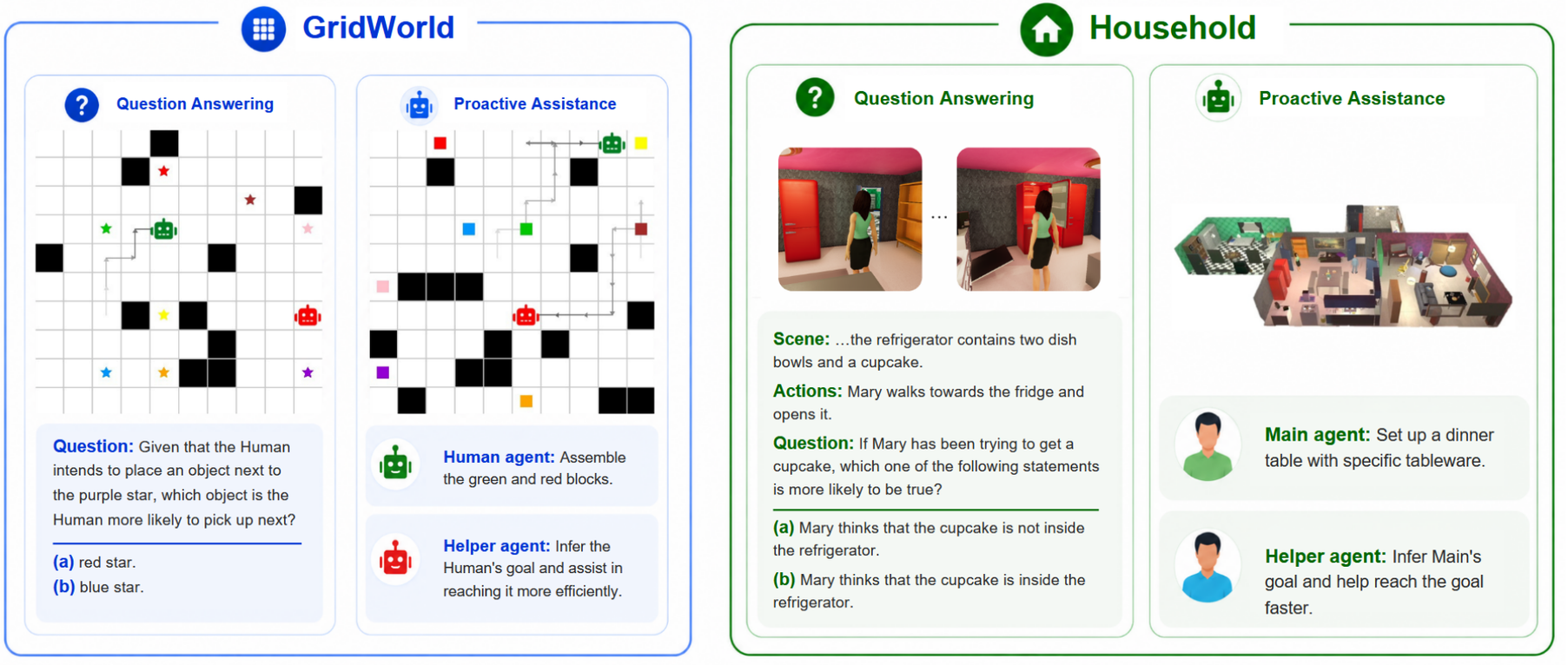
ELBO as the reward



ELBO encourages hypotheses that explain actions and keep uncertainty.

- **Likelihood**
  high action likelihood
- **Prior**
  prior plausibility
- **Entropy**
  discourages early collapse

Domains and tasks for evaluation



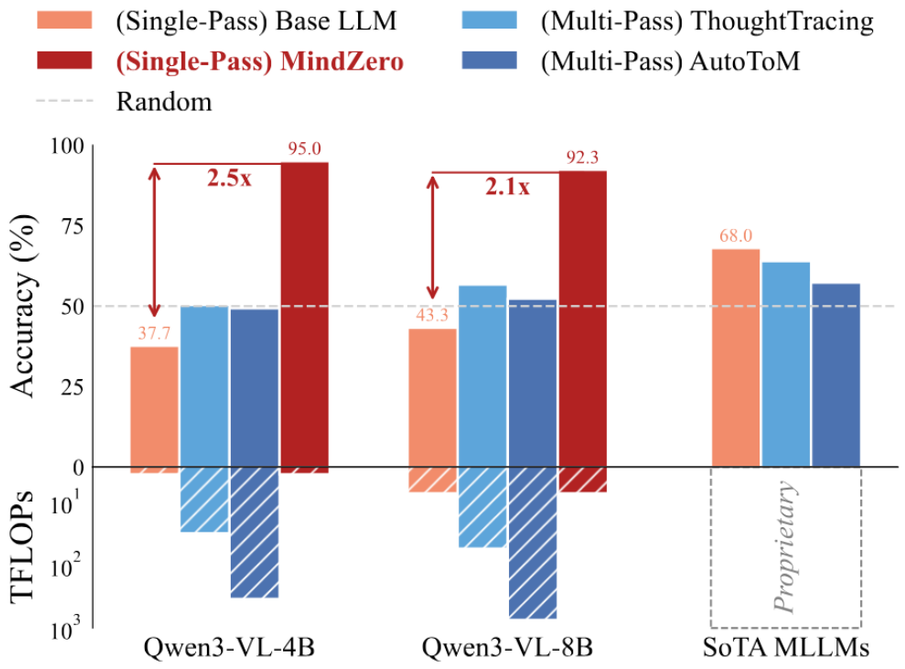
Domains

- GridWorld: visual map input
- Household: text-converted scenarios

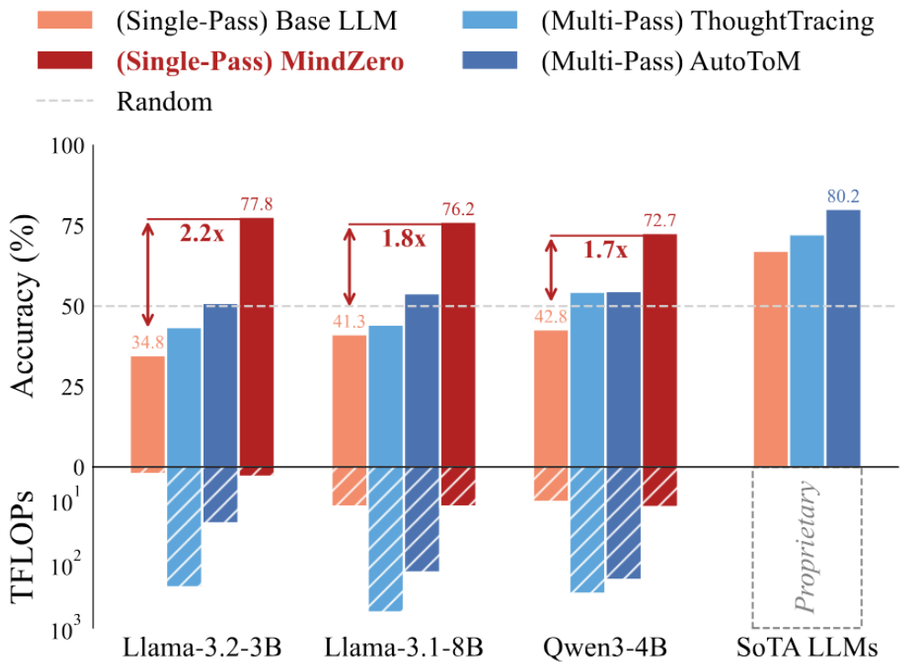
Tasks

- Story-based Theory-of-Mind question answering
- Proactive assistance

MindZero improves story-based QA accuracy



(a) GridWorld Question Answering



(b) Household Question Answering

> MindZero improves significantly over the pretrained checkpoint and is competitive with strong commercial models using much less computation.

Proactive assistance tests online inference

(a) Gridworld Proactive Assistance

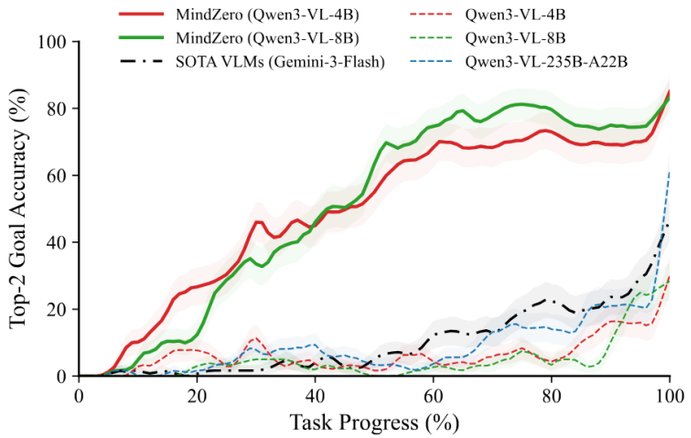
Method	Speedup ↑	TFLOPs ↓
Random Goal	0.0	N/A
Base Models		
Qwen3-VL-4B	1.4	**151.7**
Qwen3-VL-8B	-0.1	295.2
Large Models		
Qwen3-VL-235B-A22B	1.0	808.6
GPT-5.2	0.0	Proprietary
Gemini-3-Flash	0.0	Proprietary
MindZero (Ours)		
w/ Qwen3-VL-4B	23.0	161.4
w/ Qwen3-VL-8B	**24.5**	291.8

(b) Household Proactive Assistance

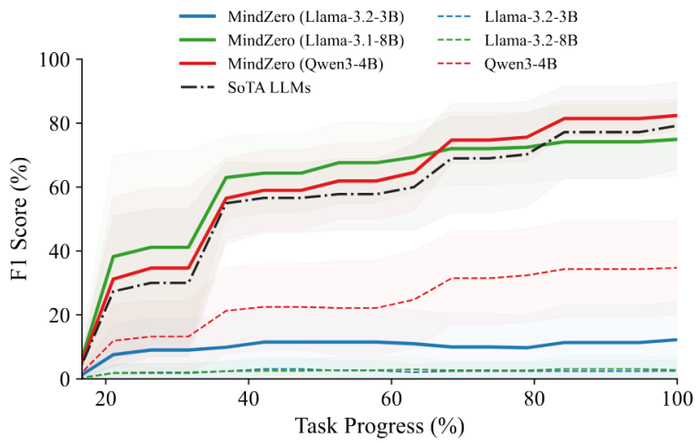
Method	Speedup ↑	TFLOPs ↓
Random Goal	-2.2	N/A
Base Models		
Llama-3.2-3B*	2.3	244.3
Llama-3.1-8B	1.7	656.1
Qwen3-4B	2.3	213.1
Large Models		
Qwen3-235B-A22B	12.3	1101.6
GPT-5.2	9.4	Proprietary
Gemini-3-Flash	17.7	Proprietary
MindZero (Ours)		
w/ Llama-3.2-3B*	4.3	235.1
w/ Llama-3.1-8B	17.4	608.4
w/ Qwen3-4B	**19.1**	**201.2**

MindZero obtains the best speedup efficiently collaborating with both simulated and real humans.

Mode seeking does not become mode collapse



(a) GridWorld Proactive Assistance



(b) Household Proactive Assistance

#	Method	Speedup ↑	TFLOPs ↓
I	*MindZero*	**19.1**	201.2
II	w/o prior modeling	17.0	200.5
III	w/o multiple hypotheses	10.3	**132.6**
IV	w/o entropy bonus	5.2	245.1

Prediction quality sharpens as more actions are observed.

Diversity depends on prior design, multiple hypotheses, and entropy bonus.

Mental reasoning can be learned with self-supervision

Problem	Online mental reasoning with uncertainty, efficiency, and zero annotations
Method	Online mental reasoning can be learned with self-supervision using RL
Result	Efficient single-pass inference at test time and strong results across domains and tasks

Project Page
Code and models are open source
https://scai.cs.jhu.edu/MindZero