RemembeRL
Workshop @ CoRL 2025

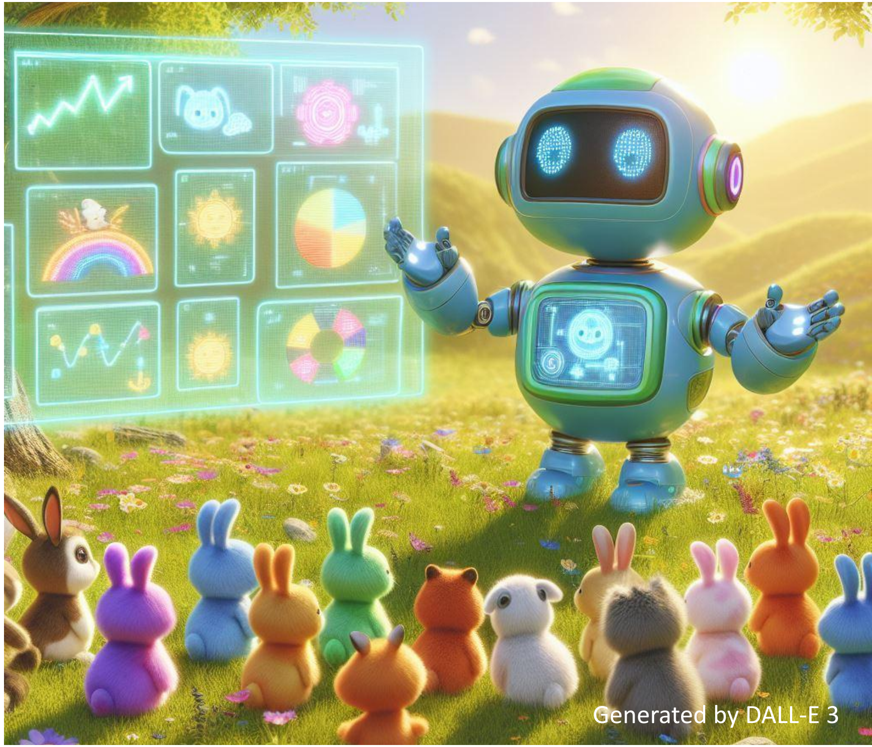
DEAKIN UNIVERSITY
DEAKIN APPLIED ARTIFICIAL INTELLIGENCE INITIATIVE

How Different Memory Types Help Agents Learn and Adapt

Presented by Dr. Hung Le

Why This Topic?

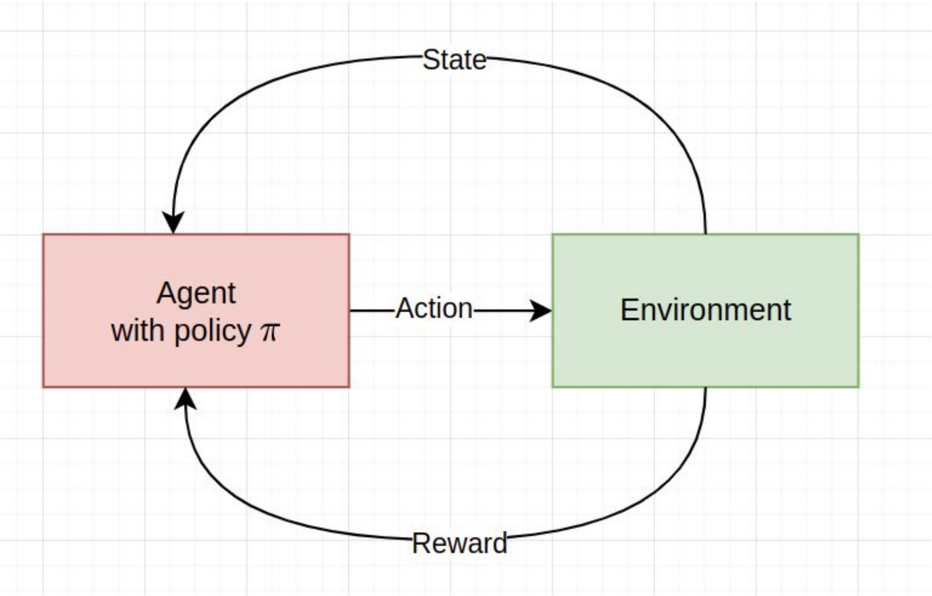
- Autonomous agents are emerging as AI systems capable of perceiving, deciding, and acting independently, offering powerful applications.
- From RL agents mastering chess and Atari games to robots and LLMs.
- **Practicality issues:** The cost of training agents is huge and the performance is worse in real-world challenges
- **Memory for agents:** Memory-augmented systems can learn faster and achieve unprecedented capability


Generated by DALL-E 3

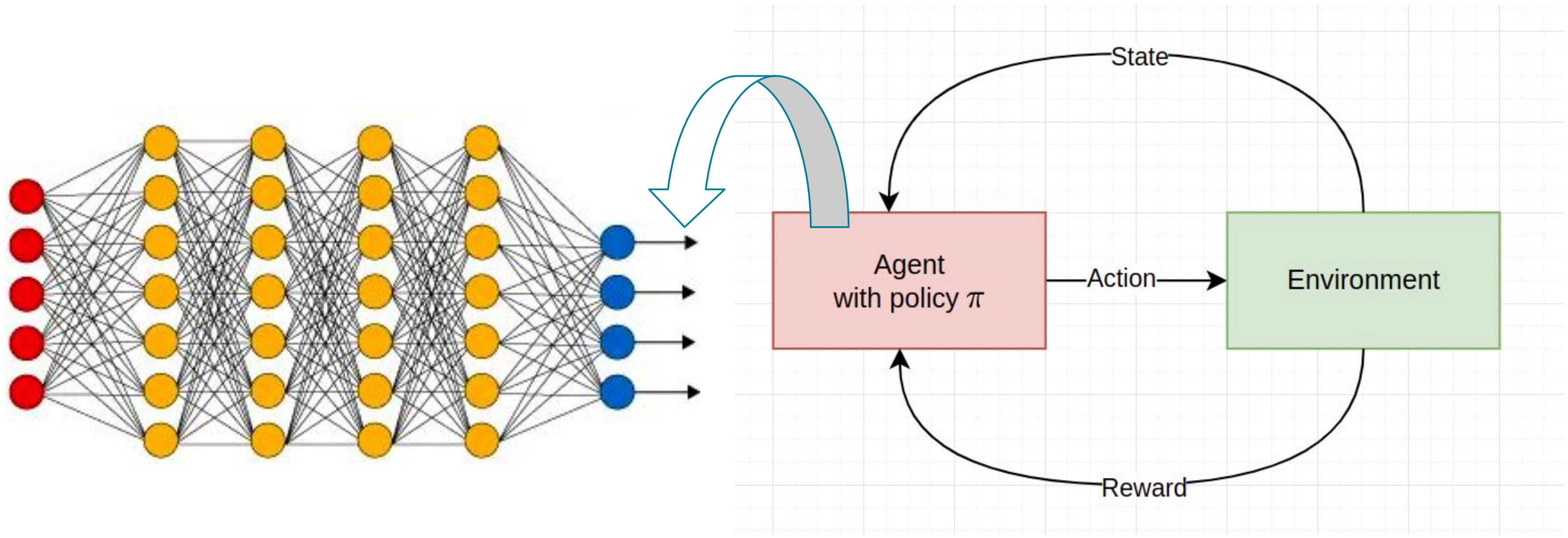
Background

What is Reinforcement Learning (RL)?

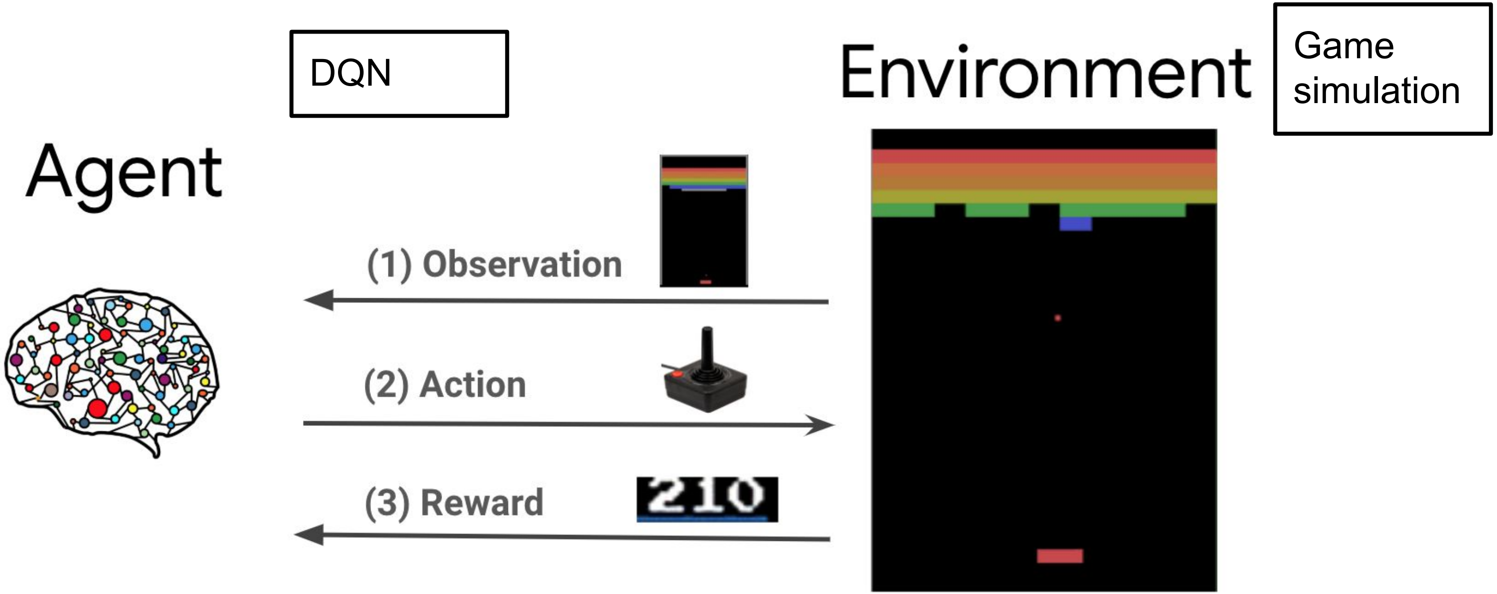
- Agent interacts with environment
- **S+A=>S'+R (MDP)**
- The transition can be stochastic or deterministic
- Find a policy $\pi(S) \rightarrow A$ to maximize expected return $E(\sum R)$ from the environment



Deep RL Agent: Value/Policy Are Neural Networks

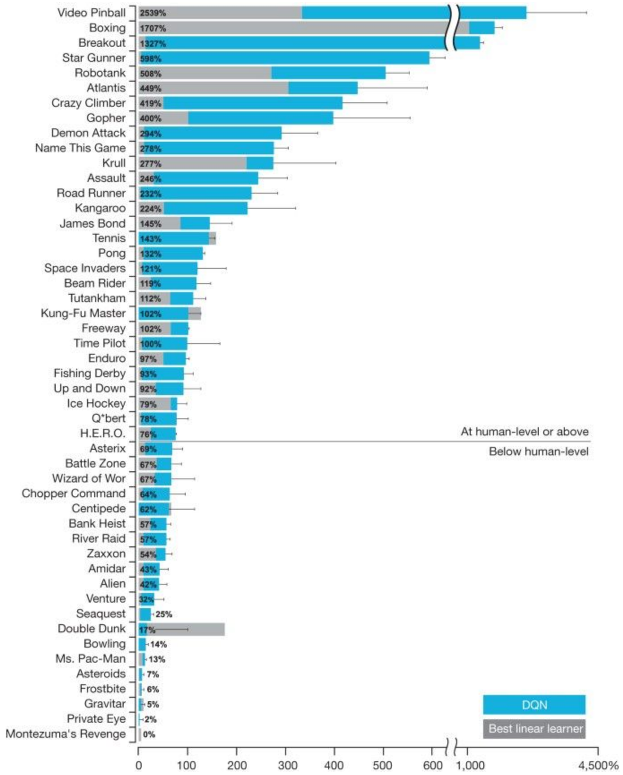


Example: RL Agent Plays Video Game



Limitation of RL Agents

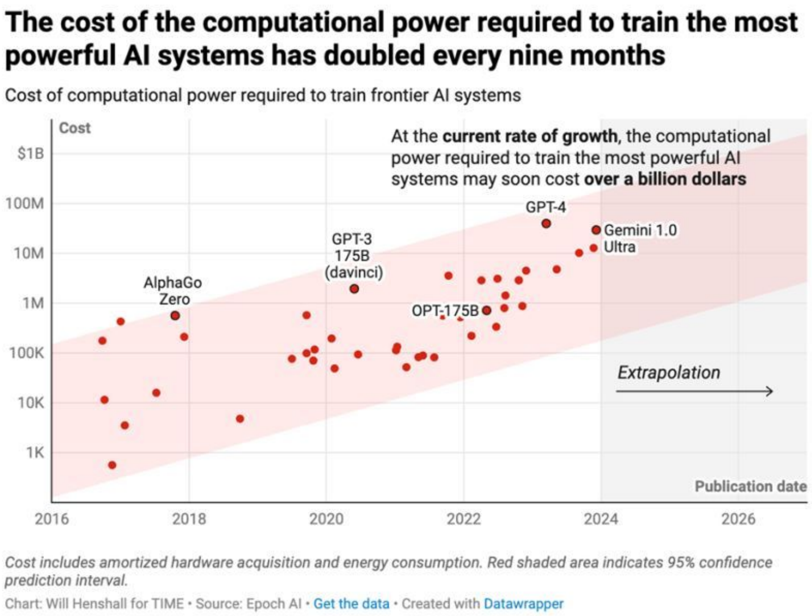
- High cost
  - Training time (days to months)
  - GPU (dozens to hundreds)
- Require simulators or big data
- Trained agents are unlike humans
  - Unsafe exploration, unethical actions
  - Weird behaviors, hallucination
  - Fail to generalize
- RL Agents (DQN-based):
  - 21 trillions hours of training to beat human (AlphaZero), equivalents to 11,500 years of human practice



Mnih, V., Kavukcuoglu, K., Silver, D. *et al.* Human-level control through deep reinforcement learning. *Nature* 518, 529–533 (2015). https://doi.org/10.1038/nature14236

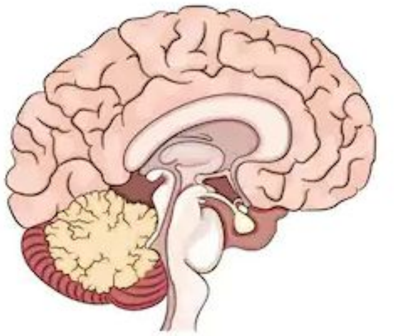
LLM Agents Are Much More Expensive

- Training cost is reaching $1 billion
- Inference cost for GPT4-Like LLMs: $140–150 million/year
- LLM agent systems often use multiple models → multiplying costs
- Same Issues:
  - Unsafe exploration, unethical actions
  - Weird behaviors, hallucination
  - Fail to generalize



https://www.linkedin.com/pulse/uncovering-hidden-costs-ai-queries-dr-ayman-al-rifaei-1kmsf?

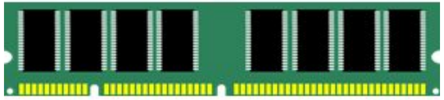
What Is Missing?

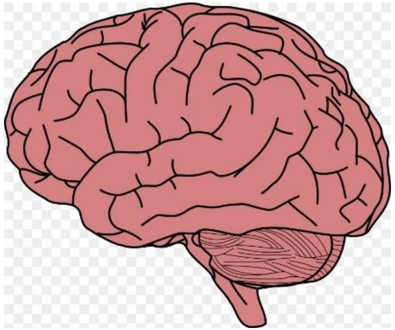


GOOD MEMORY!

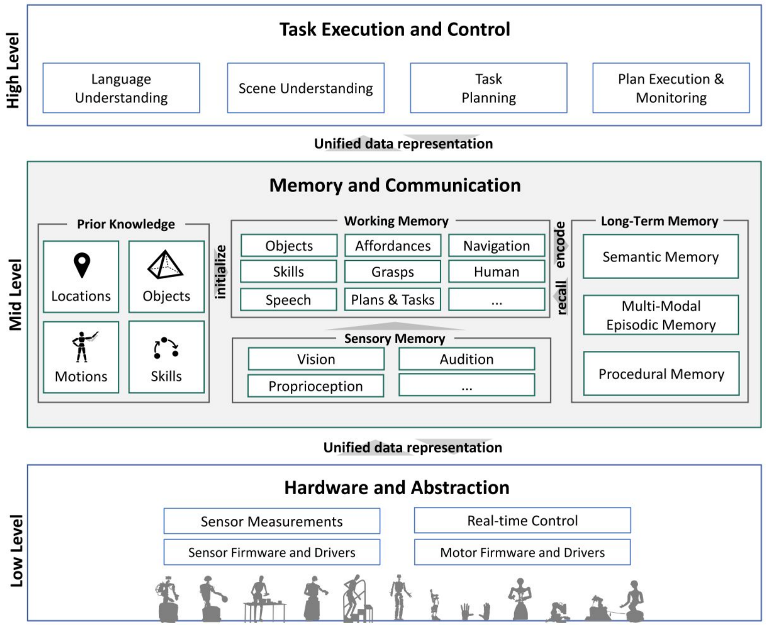
What Is Memory?

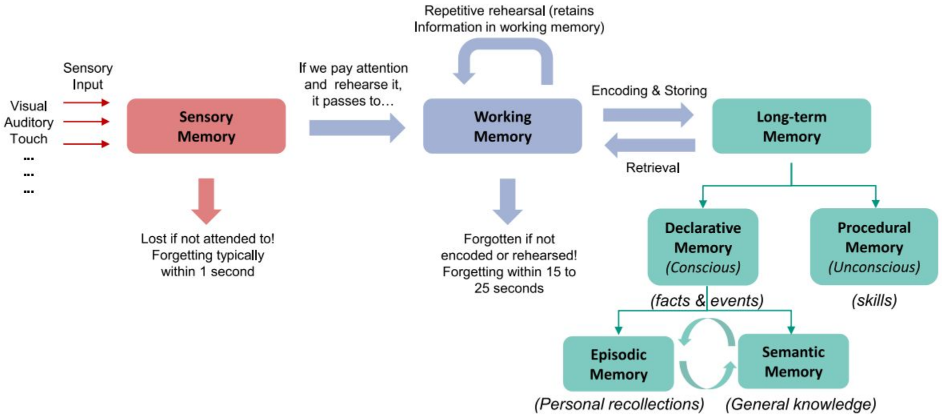
- Memory is the ability to efficiently store, retain and recall information
- Brain memory stores items, events and high-level structures
- Computer memory stores data, programs and temporary variables





Memory in Robotic Systems





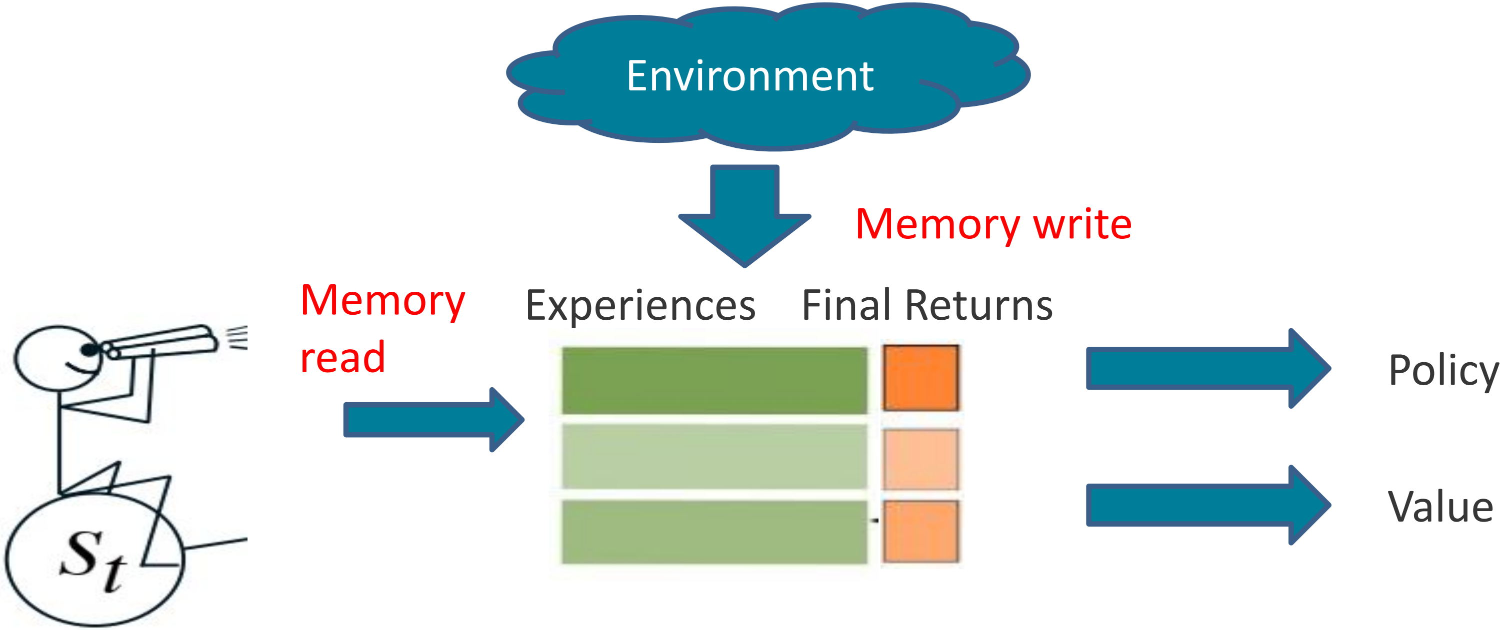
F. Peller-Konrad, R. Kartmann, C. R. G. Dreher, A. Meixner, F. Reister, M. Grotz, and T. Asfour, "A Memory System of a Robot Cognitive Architecture and Its Implementation in ArmarX," Robotics and Autonomous Systems, 2023.

Characteristics of Memories

		Lifespan	Plasticity	Example
1	Working memory	Short-term	Quick	• 1 episode is one day • Last for 1 day • Build memory instantly
2	Episodic memory	Long-term	Quick	• Persists across agent's lifetime • Last for several years • Build memory instantly
3	Semantic memory	Long-term	Slow	• persists across agent's lifetime • Last for several years • Take time to build memory

Episodic Memory for Agents

Episodic Control Paradigm



Current experience
Eg: (St, At),…

- Episodic memory is a key-value memory
- Directly binds experience to return→ refers to experiences that have high return to make decisions

Model-free Episodic Control: K-nearest Neighbors

Algorithm 1 Model-Free Episodic Control.

```
for each episode do
    for t = 1, 2, 3, ..., T do
        Receive observation o_t from environment.
        Let s_t = φ(o_t).
        Estimate return for each action a via (2)
        Let a_t = arg max_a \hat{Q}^EC(s_t, a)
        Take action a_t, receive reward r_{t+1}
    end for
    for t = T, T − 1, ..., 1 do
        Update Q^EC(s_t, a_t) using R_t according to (1).
    end for
end for
```

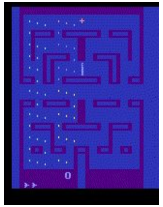
$$\widehat{Q^{\mathrm{EC}}}(s,a) = \begin{cases} \frac{1}{k}\sum_{i=1}^{k} Q^{\mathrm{EC}}(s^{(i)},a) & \text{if } (s,a) \notin Q^{\mathrm{EC}}, \\ Q^{\mathrm{EC}}(s,a) & \text{otherwise,} \end{cases}$$

Fix-size memory
First-in-first out

- No need to learn parameters (pretrained ϕ)
- Quick value estimation

Blundell, Charles, Benigno Uria, Alexander Pritzel, Yazhe Li, Avraham Ruderman, Joel Z. Leibo, Jack Rae, Daan Wierstra, and Demis Hassabis. "Model-free episodic control." *NeurIPS* (2016).

Sample Efficiency Test on Atari Games



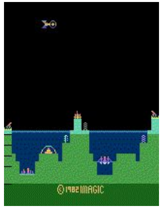
DQN (200M)

Model-free (10M)

Hybrid (40M)

Model-based (10M)

Model	All	25 games
Nature DQN	15.7/51.3	83.6/16.0
MFEC	85.0/45.4	77.7/40.9
NEC	99.8/54.6	106.1/53.3
EMDQN*	528.4/92.8	250.6/95.5
EVA	-	172.2/39.2
ERLAM	-	515.4/103.5
MBEC++	**654.0/117.2**	**518.2/133.4**

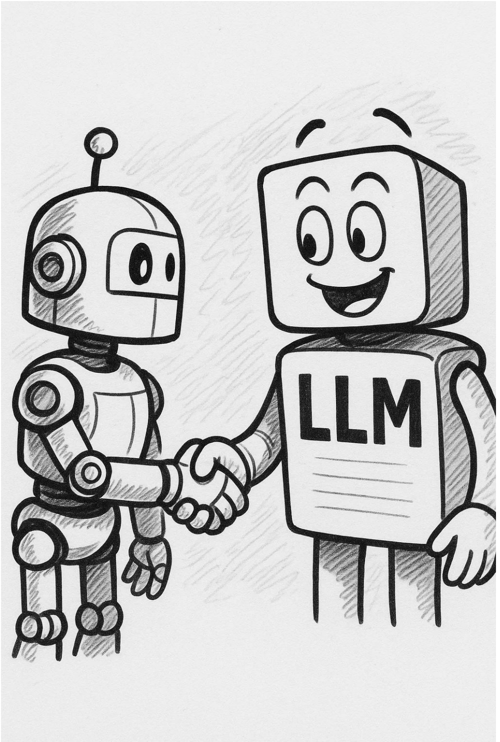


Le, Hung, Thommen Karimpanal George, Majid Abdolshah, Truyen Tran, and Svetha Venkatesh. "Model-Based Episodic Memory Induces Dynamic Hybrid Controls." *NeurIPS* (2021).

Beyond Games?

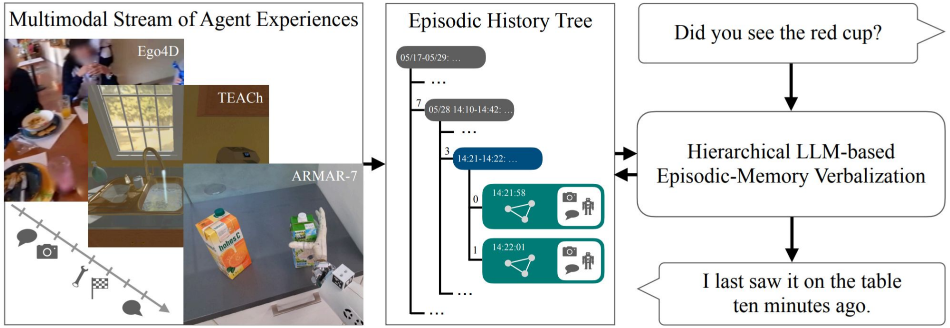
- Real-world tasks requires complex state spaces, especially in robotic tasks:
  - Visual scene
  - Dialogs
  - Timestamps
- Storing latent state representations is hard to learn a good policy and requires a lot of training
- With LLMs, we can store high-level information in episodic memory

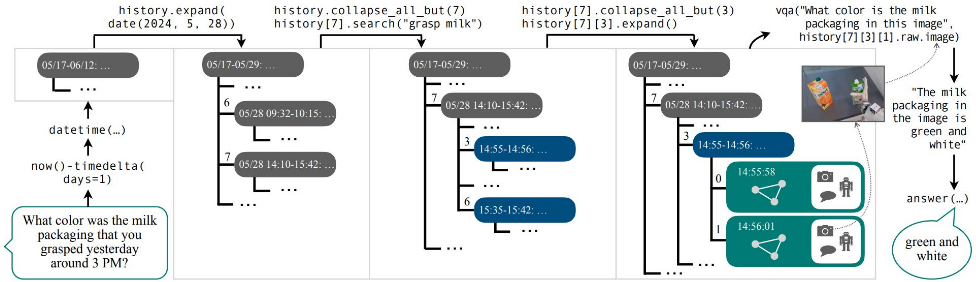
  → achieving new capabilities!



Episodic Memory for Verbal Summarization

- Episodic memory enables robots to summarize and answer questions about their past experiences, enhancing human-robot interaction
- Constructs a tree-like hierarchical structure from raw sensory data to abstract natural language concepts.
- Employs a large language model (LLM) agent to interactively search and retrieve relevant information from the episodic memory.





Bärmann, L., DeChant, C., Plewnia, J., Peller-Konrad, F., Bauer, D., Asfour, T., & Waibel, A. (2024). Episodic Memory Verbalization using Hierarchical Representations of Life-Long Robot Experience. IEEE Humanoids 2025.

Working Memory for Agents

When the State Is Not Enough …

- Partially Observable Environments:
  - States do not contain all required information for optimal action
  - E.g. state=position, does not contain velocity
- Ways to improve:
  - Build richer state representations
  - Memory of all past observations/actions

$$h_t = \langle o_0, a_0, o_1, a_1, \ldots, o_{t-1}, a_{t-1}, o_t \rangle$$

- Policy gradient

$$\nabla_\theta J \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{t=0}^{T} \nabla_\theta \log \pi(a_t | h_t^n) R_t^n$$

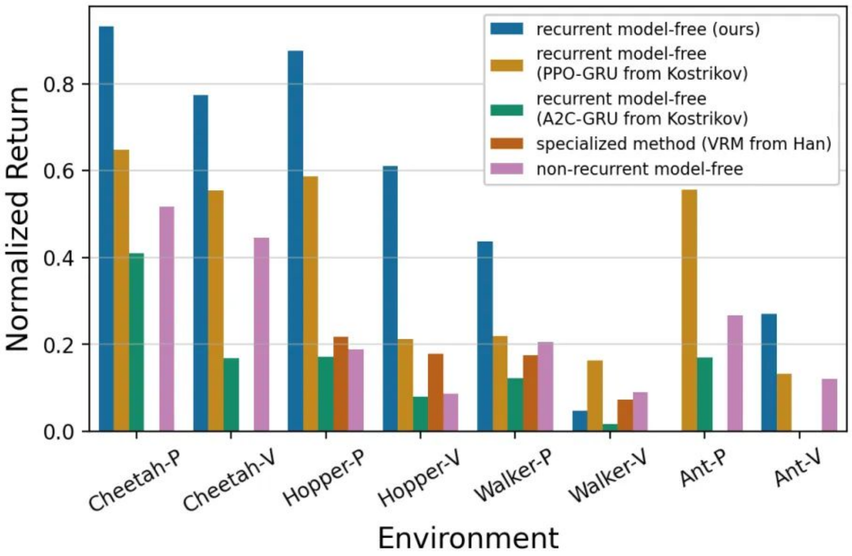
Full map

Observed state

RNN hidden state

RNN as policy model

Working Memory: Recurrence or Attention?



Ni, Tianwei, Benjamin Eysenbach, and Ruslan Salakhutdinov. "Recurrent Model-Free RL Can Be a Strong Baseline for Many POMDPs." In International Conference on Machine Learning, pp. 16691-16723. PMLR, 2022

Performance | Speed

RNN vs Transformer

$$M_t = M_{t-1} \odot C_\theta(x_t) + U_\phi(x_t)$$

Calibration | Update

Product Accumulation

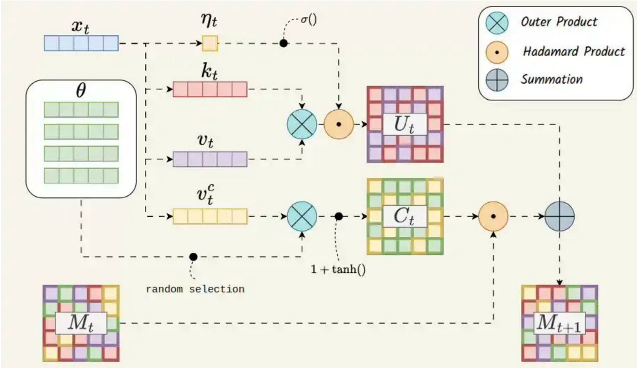
$$M_t = M_0 \prod_{i=1}^{t} C_i + \sum_{i=1}^{t} U_i \odot \prod_{j=i+1}^{t} C_j$$

Stable Working Memory: Fast and Powerful

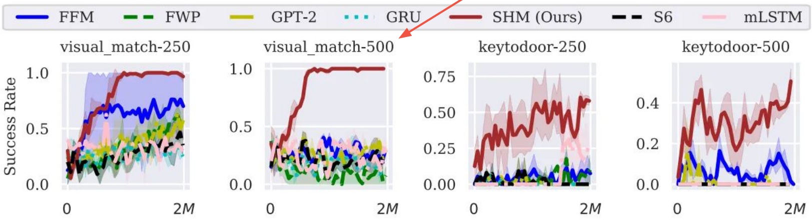
- Stable Hadamard Memory (SHM) introduces a special calibration matrix Ct defined as:

$$C_\theta(x_t) = 1 + \tanh(\theta_t \otimes v_c(x_t))$$

- θt: Trainable parameters that are randomly selected for each timestep.
- vc(xt): A mapping function (e.g., a linear transformation)
- This dynamic design keeps updates stable by ensuring that the cumulative product of calibration matrices is bounded: $\mathbb{E}\left[\prod_{t=1}^{T} C_t\right] \approx 1$



Long-term credit assignment

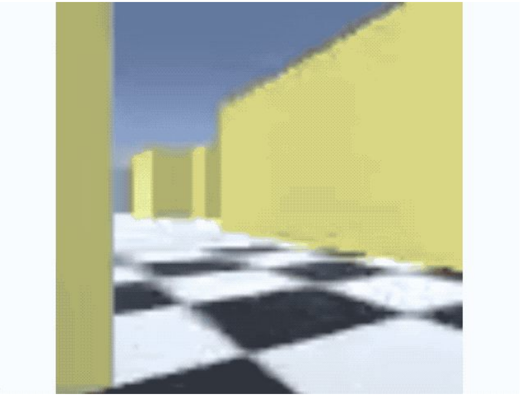


Le, Hung, Kien Do, Dung Nguyen, Sunil Gupta, and Svetha Venkatesh. "Stable Hadamard Memory: Revitalizing Memory-Augmented Agents for Reinforcement Learning." ICLR, 2025.

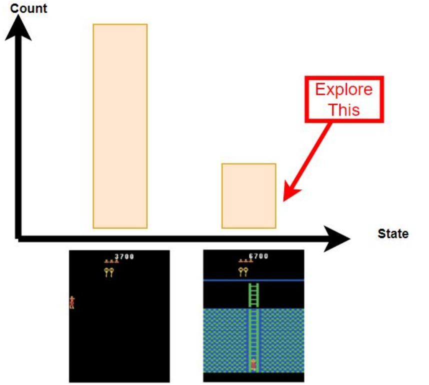
Integrated Memory Systems for Challenging Tasks

Classic RL with Hybrid Memory Systems

- Addressing challenging hard-exploration problems:
  - Montezuma Revenge
  - Noisy TV
- Working Memory: Novelty estimation within episode
- Episodic Memory: Novelty estimation across episode
- Semantic Memory: Surprise estimation via prediction error
- A hybrid metric: **surprise novelty**, the error of reconstructing surprise (the error of state prediction)



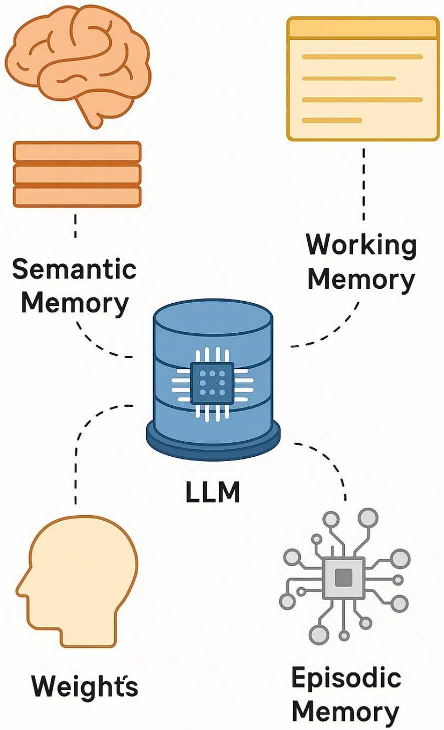
Noisy-TV: a random TV will distract the RL agent from its main task due to high surprise (source).



Le, Hung, Kien Do, Dung Nguyen, and Svetha Venkatesh. "Beyond Surprise: Improving Exploration Through Surprise Novelty. In AAMAS, 2024.

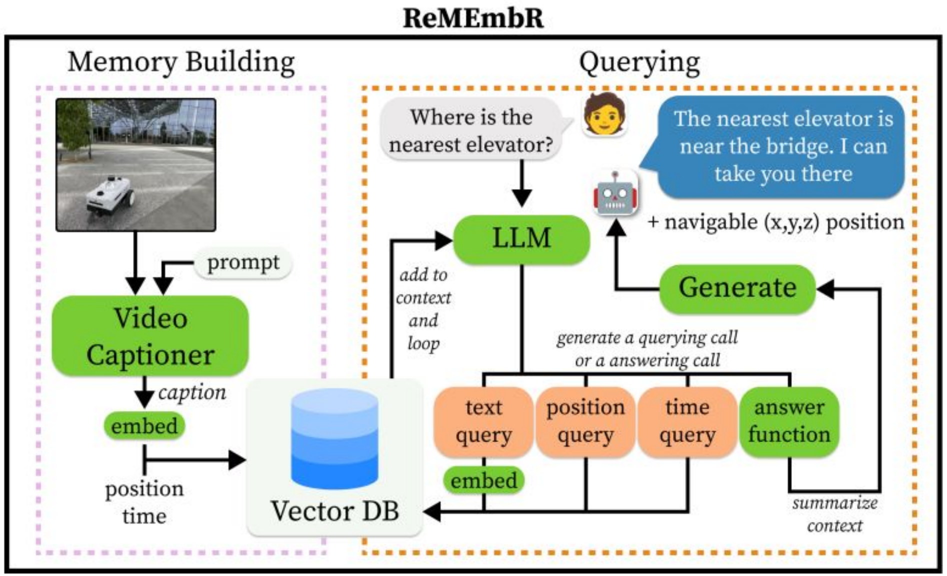
LLM Agents Also Have a System of Memories

- ❑ Semantics Memory: knowledge stored in the LLM's weights or other knowledge database
- ❑ Working Memory: the context stored in the prompt, accessed by attention mechanism
- ❑ Episodic Memory: external memory module to store past experiences



Spatial Memory for Navigation

- The LLM considers the current set of memories (R0:i) and the question (Q) to generate a function call f and a query q which retrieves m memories.
- Each memory contains position, time, and caption information to be used as further context.
- LLM can make up to k queries per iteration. Retrieves k × m memories and updates context
- Check: Can the question be answered?
  - No: Repeat querying with updated context
  - Yes: Summarize relevant info & generate final answer



Anwar, Abrar, John Welsh, Joydeep Biswas, Soha Pouya, and Yan Chang. "Remembr: Building and reasoning over long-horizon spatio-temporal memory for robot navigation." *ICRA* (2025).

Spatial Memory in Action

Method	LLMs	Descriptive Question Accuracy ↑			Positional Error (m) ↓			Temporal Error (s) ↓		
		Short	Medium	Long	Short	Medium	Long	Short	Medium	Long
Ours	GPT4o	**0.62**±0.5	*0.58±0.5*	**0.65±0.5**	**5.1±11.9**	**27.5**±26.8	**46.25±59.6**	**0.3±0.1**	**1.8±2.0**	**3.6±5.9**
	Codestral	0.25±0.4	0.24±0.4	0.11±0.3	151.3±109.7	189.0±109.6	212.4±121.3	4.8±5.6	8.4±6.8	14.8±7.5
	Command-R	0.36±0.5	0.32±0.5	0.14±0.3	158.7±129.6	172.2±119.4	188.7±107.1	4.5±17.3	14.3±6.7	15.3±11.7
	Llama3.1:8b	0.31±0.5	0.33±0.5	0.21±0.4	159.9±123.2	151.2±121.1	165.3±115.1	9.5±27.5	7.9±16.3	18.7±10.8
LLM with Caption	GPT4o	*0.57±0.5*	**0.66±0.5**	*0.55±0.5*	**5.1±8.2**	*33.3±47.3*	*56.0±61.7*	*0.5±0.5*	*1.9±2.2*	*8.0±6.7*
Multi-Frame VLM	GPT4o	*0.55±0.5*	✗	✗	*7.5±11.4*	✗	✗	0.5±2.2	✗	✗

Conclusion

- Memory is essential for agents
- 3 basic types of memory
- Memory is useful to make agents more efficient and robust against challenging environments:
    - Long-horizon tasks
    - Multiple-step planning
    - Noisy environments

A2I2 Lab Foyer, Waurn Ponds

❑ Presenter: Dr. Hung Le from A2I2, Deakin University

❑ Hung Le is a DECRA Fellow and a lecturer at Deakin University, leading research on deep sequential models and reinforcement learning