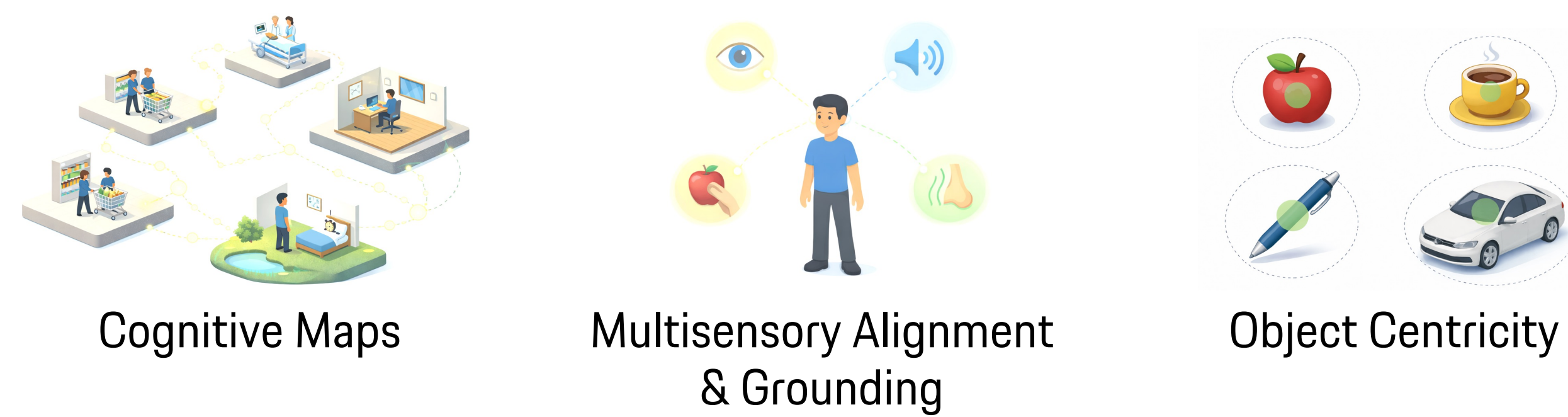
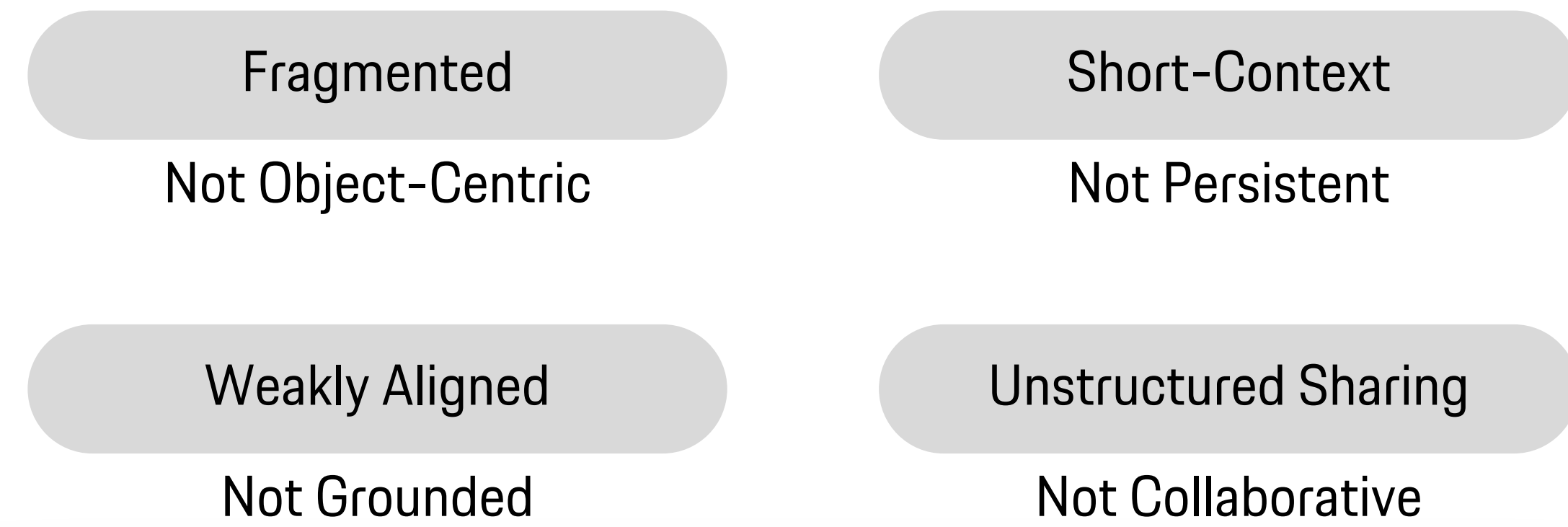


How Humans memorize the Environment



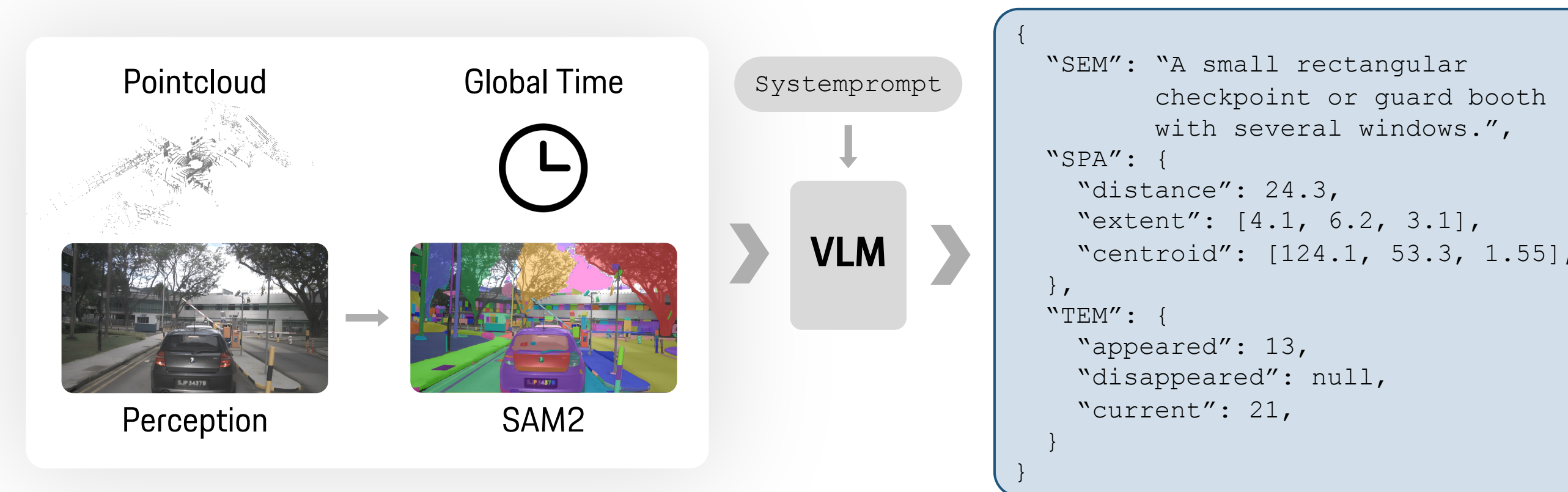
How current Methods memorize the Environment



Key Contributions

- **R⁴ Framework:** Training-free retrieval-augmented system grounding semantics in a persistent 4D spatiotemporal world model.
- **Continuous 4D Memory:** Incrementally refined database tracking object features & evolution for collaborative embodied reasoning.
- **Structured 4D Retrieval:** First to decompose queries into semantic/spatial/temporal keys, iteratively integrating evidence for human-like reasoning.

4D Retrieval-Augmented Reasoning



The VLM-Agent populates...

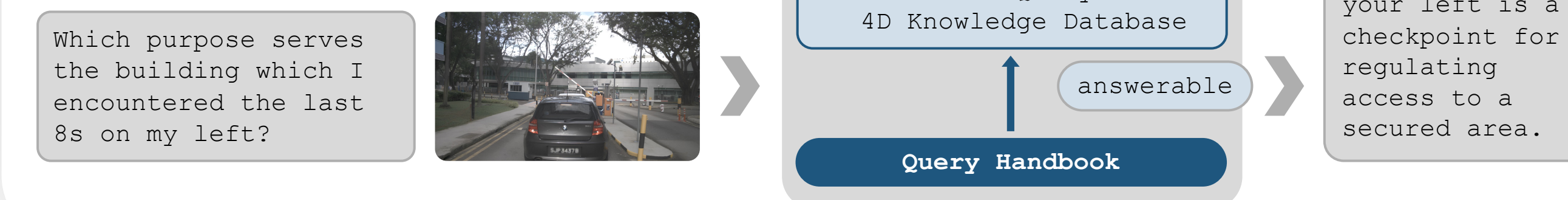
- semantic object descriptions
- precise 3D spatial attributes
- temporal intervals of visibility and action dynamics

continuously into a 4D Knowledge Database.

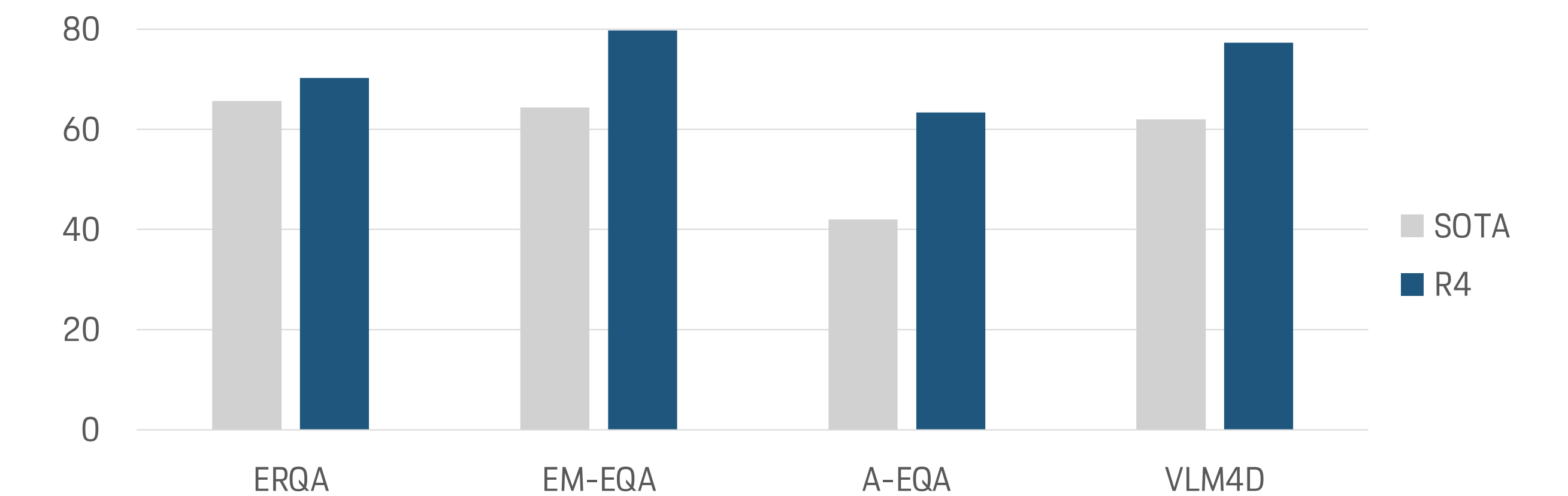
The 4D Knowledge Database is collaboratively accessible for VLMs via...

- semantic cosine similarity
- spatial nearest neighbor (local/global)
- temporal register search

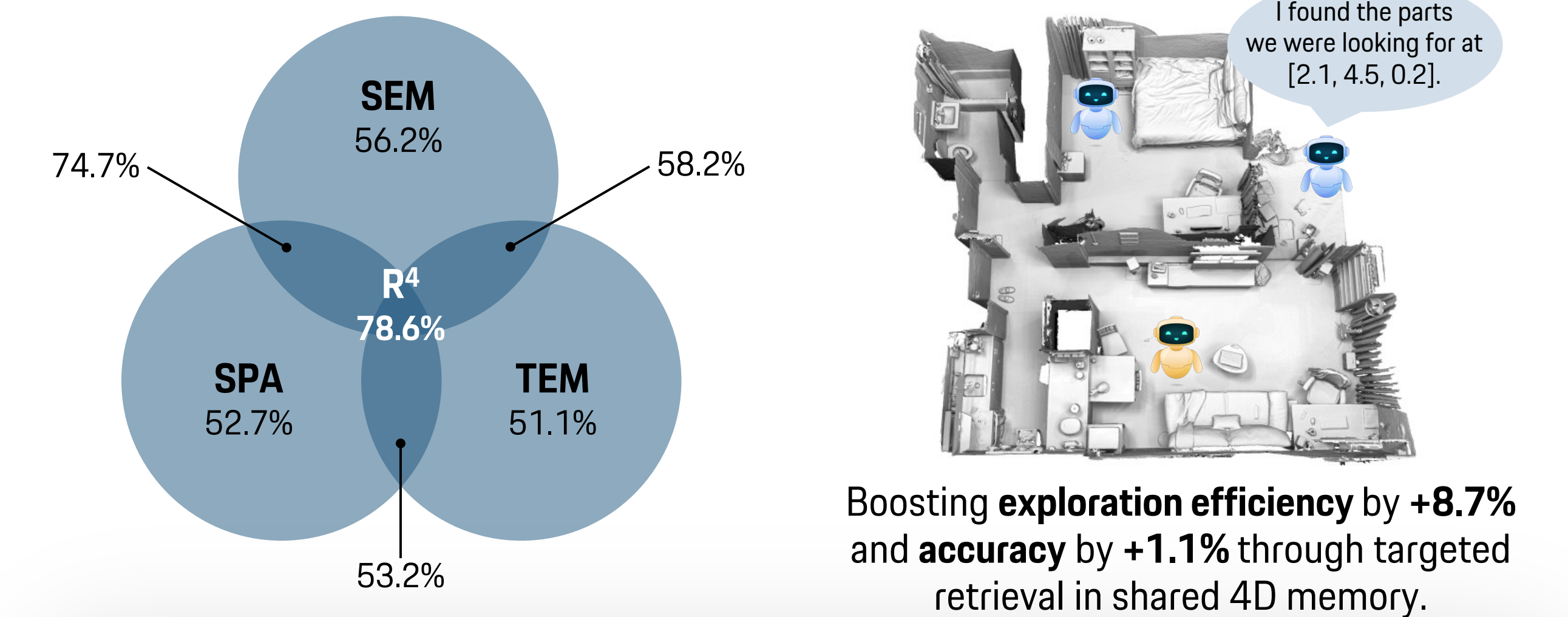
acting as retrieval keys for RAG.



Experiments



Key Insights: Ablation & Collaboration



Conclusion

- **Training-free 4D reasoning** via human-inspired persistent memory
- **Structured 3-axis retrieval** for complex, long-horizon queries
- **SOTA results** on EQA, episodic-memory & navigation benchmarks
- **Multi-agent ready:** Shared 4D memory enables collaboration
- **Ablation-validated:** Semantic, spatial & temporal axes all critical