



Token-level Memory

Features:

- Symbolic, addressable, transparent
- Swift adding/deleting/updating

Suitable Applications:



Multi-turn chatbot



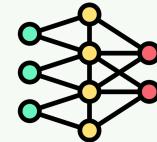
Personalized agents



Recommender system



High-stake domains
(law, finance, medical)



Parametric Memory

Features:

- Implicit, abstract, and generalizable
- Slower memory update
- (Typically) better performance gain
- More severe catastrophic forgetting

Suitable Applications:



Role-playing



Reasoning-intensive task



Tasks that require
fundamentally new
capabilities



Latent Memory

Features:

- Implicit, human-unreadable
- Trade-off between efficiency/flexibility
- Convenient modality fusion
- Machine-native and token-efficient

Suitable Applications:



Multimodal memory



On-device or edge deploy



Low-resource or small-data
setting