2026 START Program
CFP Brief

THEME: **02.** **Agentic, Artificial Intelligence**

SUB-THEME: **2.2.** **Next-Generation Self-Evolving Artificial Agents**

Context/ Overview

As artificial intelligence transitions from static, pre-trained models to dynamic, interactive systems, the frontier of AI research lies in **Self-Evolving Agents.** We must pioneer architectures, algorithms, and frameworks that enable AI agents to autonomously learn, adapt, and improve their reasoning capabilities over time. The goal of this initiative is to advance research that bridges the gap between static foundational models and fully autonomous, lifelong-learning agents capable of self-improvement in open-ended environments.

Problem Statement

Current agent systems face key technical bottlenecks:

1. Memory limitations: interaction histories are poorly structured, leading to retrieval errors and inconsistent long-term behavior
2. Reasoning instability: multi-step planning degrades under changing context or accumulated uncertainty.
3. Weak self-evaluation: agents lack reliable internal scoring of task success, preventing autonomous improvement.
4. Unsafe learning loops: continuous policy updates risk behavioral drift, reward hacking, or regression

These issues prevent deployed agents from achieving true self-evolution

Specific Topics & focus areas*

1. **Autonomous Self-Improvement & Recursive Optimization**:
   Mechanisms allowing agents to iteratively refine their own code, reasoning pathways, or neural weights without human intervention.

2. **Long-Term Memory & Lifelong Learning (Continual Learning)**:
   This research will focus on retaining and synthesizing knowledge over extended time horizons while preventing catastrophic forgetting. It involves designing episodic and semantic memory systems for agents, ensuring memory consistency, and implementing effective updating and forgetting policies. Additionally, explore self-evolving parametric memory with architectural plasticity, focusing on continual learning capabilities.

3. **Reasoning Under Dynamic Context:**
   The focus is on developing adaptive planning strategies for extended multi-step workflows, incorporating self-reflection and reasoning verification methods. Agents will interact with dynamic, complex environments to gather feedback and learn from novel situations. These agents will go beyond using existing APIs by writing, testing, and deploying new tools to address emergent problems effectively.

4. **Multi-Agent, Decentralized and Emergent Coordination – Collective Intelligence:**
   Coordination among multiple adaptive agents operating in shared digital or physical environments. Mechanisms for shared memory, knowledge exchange, and distributes reasoning across agent populations. This includes emergent communication, adversarial Self-Play and collaborative problem solving

5. **Safety and Controlled Evolution:**
   Principled constraints for safe autonomous policy updates and behavioral adaptation. Mechanisms for detecting and preventing unsafe behavioral drift during continuous learning

※ The topics are not limited to the above examples and the participants are encouraged to propose other original ideas.

References

1. Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language Agents with Verbal Reinforcement Learning. Advances in Neural Information Processing Systems (NeurIPS)

2. Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST). 3. Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., ... & Wang, C. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework. arXiv preprint arXiv:2308.08155.

4. Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). Constitutional AI: Harmlessness from AI Feedback. arXiv preprint arXiv:2212.08073.

END OF DOCUMENT