



Self-Improving

$$\mathcal{A}_{t+1} = \text{IMPROVE}(\mathcal{A}_t; \mathcal{S}_t), \quad \mathcal{A}_t = (\theta_t, \Sigma_t), \quad \Sigma_t := (p_t, m_t, \mathcal{T}_t, g_t)$$

(1) FM Improvement

$$\theta_{t+1} = \text{IMPROVE}_{\theta}(\theta_t; \mathcal{S}_t), \quad \Sigma_{t+1} = \Sigma_t$$

Key Improvement: $\theta_t \rightarrow \theta_{t+1}$

Intrinsic Demo

- Generation strategies
- Data Formats
- Application Domains

$$\mathcal{S}_t \approx \mathcal{D}_t$$



SFT & RL

Intrinsic Feedback

- Rubric feedback
- Consistency feedback
- Corrective feedback

$$\mathcal{S}_t \approx r_t$$



SFT & RL

Extrinsic Experience

- Grounded Interaction
- Simulated Interaction

$$\mathcal{S}_t \approx \mathcal{T}_t$$



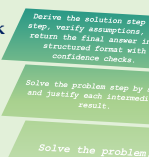
(2) Scaffolding Improvement

$$\Sigma_{t+1} = \text{IMPROVE}_{\Sigma}(\Sigma_t; \mathcal{S}_t), \quad \theta_{t+1} = \theta_t$$

Key Improvement: $\Sigma_t \rightarrow \Sigma_{t+1}$

Prompt

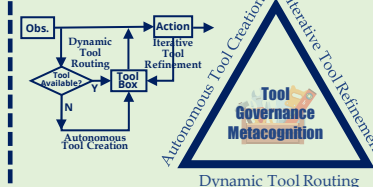
- Scalar-Feedback
- Qualitative-Feedback
- Population-Based
- Gradient-Based



Prompt SI Loop
1 Initialize prompt p_0
2 for $t = 0, \dots, T-1$:
3 $\tau_t = \text{Eval}(\pi_t, p_t)$
4 $S_t = \text{Eval}(\tau_t)$
5 $A_t = \text{Upd}(S_t)$
6 $p_{t+1} = \text{App}(p_t, A_t)$
7 return p_T

$$p_{t+1} = \text{IMPROVE}_p(p_t; \mathcal{S}_t)$$

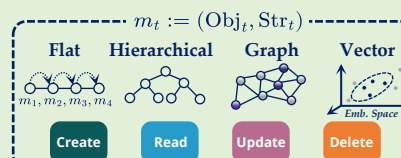
Tool



$$\mathcal{T}_{t+1} = \text{IMPROVE}_{\mathcal{T}}(\mathcal{T}_t; \mathcal{S}_t)$$

Memory

- Object
- Structure
- Processing

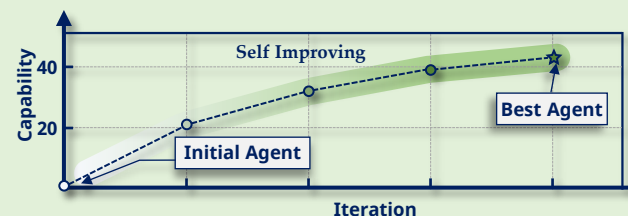


$$m_{t+1} = \text{IMPROVE}_m(m_t; \mathcal{S}_t)$$

Full Scaffolding

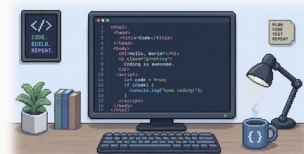
- Self-Referential Code Update
- Generate-Test-Patch Loop
- Open-Ended Search over Agent Designs

$$\Sigma_{t+1} = \text{IMPROVE}_{\Sigma}(\Sigma_t; \mathcal{S}_t)$$



Self-Improving AI Agent

A Survey



Software Engineering



Web Navigation and Automation



Gaming and Strategic Reason



Scientific Discovery



Embodied AI and Robotics



General Computer Control