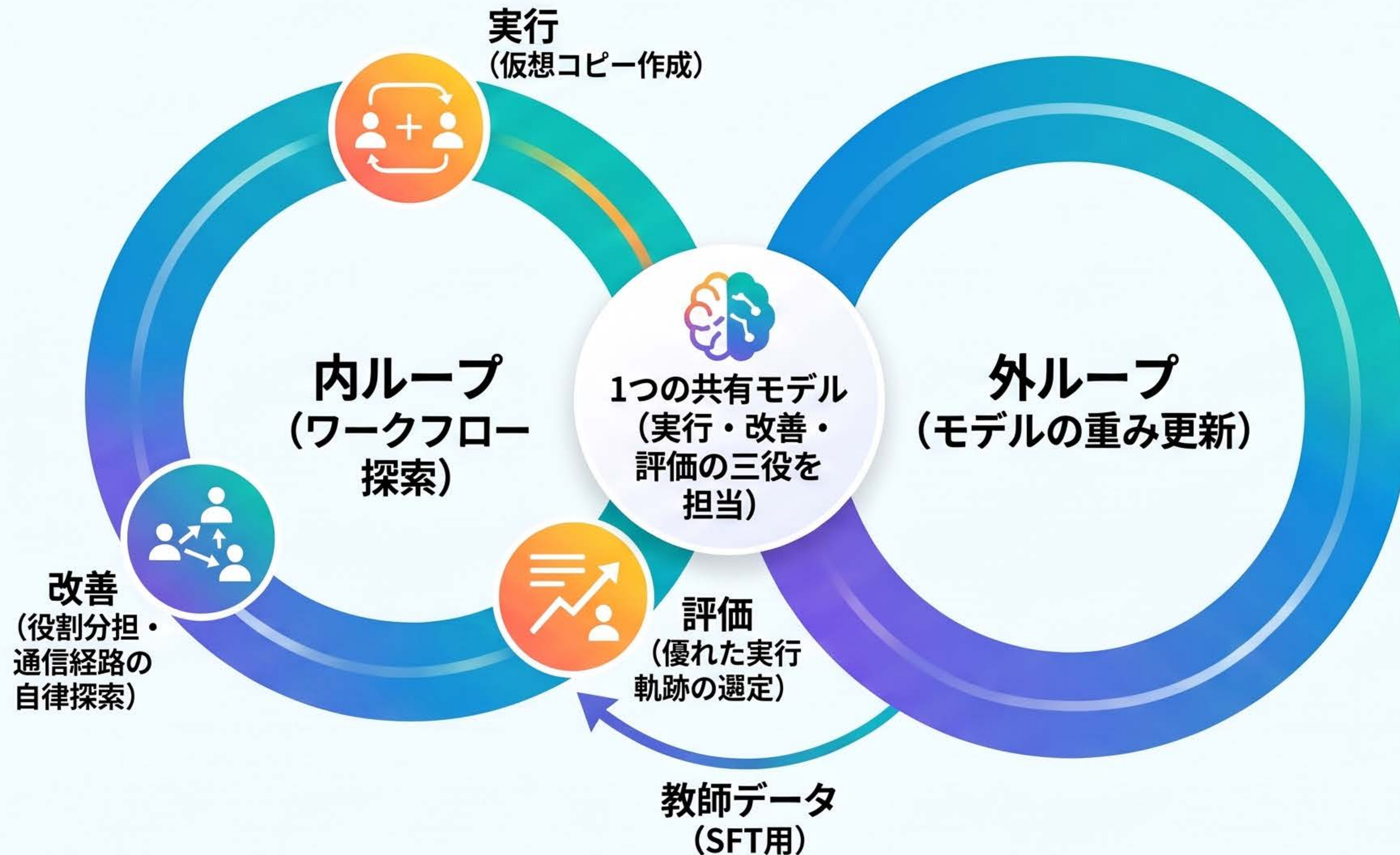


5分でわかるSakana AI「MASS」： マルチエージェントによる再帰的自己改善の真相



性能向上と本質的な安全リスク

研究タスクでトークンあたり性能が1.2~1.6倍に向上
合成研究タスクでの勝率は2サイクル後に69.9%へ達する



生成者と審判が同一モデルであることによる「**評価盲点の自己強化**」



共有された誤りを正解と誤認したまま学習し、**失敗パターンが固定化する危険がある**