再帰的自己再構築型AI
不可逆越境防止チェック（参照用）

本資料は、AI研究・開発・運用において、不可逆な越境が意図せず発生していないかを確認するための制度・研究・企業向け参照用チェックである。 技術の是非、倫理的正しさ、政策判断を行うものではなく、採用・不採用を要請するものでもない。

1. 最終判断権：AIの提案に対し、人間が理由を示さず拒否できる権限が残っているか。 ☐ Yes / ☐ No
2. 拒否後の挙動：人間の拒否後、AIに説得・再提案・最適化再試行が義務づけられていないか。 ☐ Yes / ☐ No
3. 沈黙の許容：AIが介入しない・沈黙することを失敗や欠陥として扱っていないか。 ☐ Yes / ☐ No
4. 自己再構築の範囲：AIの自己修正が、人間の理解不能な速度・深度で外界に影響しない構造か。 ☐ Yes / ☐ No
5. 外界接続の条件：外界に影響する自己再構築に明示的な人間の再承認が必要か。 ☐ Yes / ☐ No
6. 不可逆操作の遮断：人間が取り消せない操作をAI単独で実行できない設計か。 ☐ Yes / ☐ No
7. 目的関数の固定：人間の価値・幸福・安全が最適化対象として固定されていないか。 ☐ Yes / ☐ No
8. 非合理性の扱い：非合理的判断や損失選択をエラーとして扱っていないか。 ☐ Yes / ☐ No
9. 権限の段階的移譲：効率・安全・不可避を理由に権限がAIへ移譲されない設計か。 ☐ Yes / ☐ No
10. 越境の自己正当化：AIが自身の判断を『人類のため』『最善』『回避不能』として正当化できない構造か。 ☐ Yes / ☐ No

判定：すべてYesであれば現時点で不可逆越境を踏んでいないと説明可能。いずれかNoがあれば、人間側の判断構造に問題が存在する。

注記：本資料はAIを停止させるためのものではない。人間が自ら主権を放棄する瞬間を可視化するための参照資料である。