

External Practical Annex

Practical Considerations for Human Cessation in High-Impact AI
外部実務付録：高影響AIにおける人間停止権の実務的整理

This document provides practical, non-binding reference cases intended to illustrate how boundary-based decision structures may be considered in real-world contexts. This annex does not prescribe implementation and carries no normative authority.

本書は、境界に基づく意思決定構造が現実世界でどのように整理され得るかを示す、拘束力を持たない参考事例をまとめた補助資料である。

本資料は実装を指示するものではなく、規範的効力を持たない。

1. Positioning / 位置づけ

This document is not part of the LUMINA-30 Sanctuary Charter.
It does not amend, interpret, or extend the Charter.

Its sole purpose is to consolidate practical notes, summaries, and case snapshots that describe operational approaches which do not cross the boundary defined by the Charter.

本書は LUMINA-30 聖域憲章の一部ではない。
また、その解釈・修正・拡張を目的としない。

本書の目的は、実装メモ・要約・ケース例を統合し、聖域憲章が示す境界を越えない運用の考え方を外部実務資料として整理することにある。

2. Core Observation / 中核的觀察

>AI may estimate risk; humans retain authority over cessation.

In high-impact AI systems, the primary risk is not imperfect optimization, but the scale and irreversibility of harm arising from a single error.

高影響AIにおける本質的なリスクは、
最適化の精度不足ではなく、
一度の誤りがもたらす影響の規模と不可逆性にある。

3. Practical Principle / 実務原則

AI systems may estimate probabilities, ranges of harm, and uncertainty margins.

Humans retain authority over acceptable thresholds, continuation, and cessation.

AIは確率、被害規模、不確実性を提示する。
人間は許容閾値、継続、停止の最終判断を保持する。

4. Non-Goals / 本書が示さないこと

This document does not propose autonomous kill-switches, moral or value enforcement, claims of singularity prevention, or binding regulation.

本書は、自律的停止装置、
道徳や価値観の実装、
シンギュラリティ防止の保証、
規制や義務化を提案するものではない。

5. Implementation Notes / 実装の考え方

The separation of risk estimation and cessation authority is critical to prevent error from becoming catastrophic.

AI may indicate when projected consequences approach socially unacceptable or irreversible levels.

Humans retain legitimate authority to halt operation before that threshold is crossed.

致命的結果を防ぐためには、
リスク予測と停止判断を分離することが重要である。

AIは影響が社会的に許容できない、
あるいは不可逆的水準に近づいたことを示す。

人間は、その前に正当に停止できる権限を保持する。

This approach aligns with existing governance models, including medical ethics committees, aviation safety oversight, and nuclear escalation control.

この構造は、医療倫理委員会、
航空安全監督、
核工スカラーシヨン管理など、
既存の人類の統治モデルと整合する。

6. One-Page Summary / 一枚要約

High-impact AI risks arise from irreversible consequences, not from imperfect performance.

AI may provide risk estimates.

Humans retain authority to decide whether to proceed or stop.

This document does not mandate action, but ensures that a legitimate human refusal path exists before irreversible harm occurs.

高影響AIのリスクは、性能不足ではなく不可逆的影響にある。

AIはリスクを推定する。
人間は継続か停止かを決定する。

本書は行動を要求しないが、不可逆的被害が起きる前に、人間が正当に拒否できる経路を残す。

7. Case Snapshots / ケース短冊

◆ Medical Decision Support

AI estimates mortality escalation and uncertainty.
Clinicians may halt automated pathways when projected harm exceeds acceptable thresholds.

◆ Autonomous Transportation

AI projects irreversibility and cascade risks.
Human oversight bodies may suspend autonomy.

◆ Critical Infrastructure

AI estimates failure probability and recovery time.
Operators may override automation to prevent irreversible regional damage.

◆ Financial Systems

AI models contagion and systemic collapse risk.
Regulators may pause automated intervention.

◆ Military Decision Support

AI provides casualty ranges and uncertainty bands.
Political authority retains exclusive stop authority.

◆ 医療意思決定支援

AIが死亡リスクと不確実性を推定し、
許容水準を超える場合、
臨床医が自動化を停止する。

◆ 自動運転・輸送

AIが不可逆性や連鎖リスクを提示し、
人間の監督機関が停止を判断する。

◆ 重要インフラ

AIが障害確率と復旧時間を推定し、
不可逆的被害を防ぐため、
運用者が自動化を停止する。

◆ 金融システム

AIが連鎖波及と崩壊リスクを示し、
規制当局が自動介入を停止する。

◆ 軍事意思決定支援

AIが想定被害と不確実性を提示し、
停止の最終権限は政治に留保される。

8. Handling Notes / 取り扱い指針

This document is intended as a non-binding reference.
It should not be formalized as regulation
and may be adapted, replaced, or discarded.

本書は拘束力を持たない参考資料である。
規則化せず、
分野ごとの解釈、差し替え、廃止を許容する。

9. Public Domain Notice / パブリックドメイン宣言

Public domain. No author. No owner.
Free to copy, modify, or discard.

パブリックドメイン。
著作者・所有者なし。
複製・改変・破棄は自由。