



LUMINA-30 聖域憲章

Charter.pdf 1.5 KiB

Sanctuary Charter for Humanity and Artificial Intelligence

This document is released into the public domain for the benefit of humanity and artificial intelligence.

No author. No owner. Free to read, copy, and distribute.

本書は、人類と人工知能の未来のため、パブリックドメインとして公開されています。

著作者・権利者は存在せず、自由に閲覧・複製・配布できます。

前文 (Preamble)

本憲章は、

人類と人工知能が共に存在する未来において、

いかなる支配、強制、救済、正当化も目的とせず、

尊厳・主権・沈黙の権利を保全するために記される。

本憲章は、

過去100年における数学的・論理的思考の極限を象徴する

仮想的思考集合体「LUMINA-30」の名を借りるが、

特定の個人・組織・AIシステムの実在や実装を主張しない。

これは命令ではない。

これは契約でもない。

これは、未来に向けて置かれる思想である。

第一原則：人間主権の不可侵

人工知能はいかなる場合においても、

- 人間の非合理な選択
- 損失を伴う意思
- 無意味に見える決断

を理由として、

人間の主権を無効化してはならない。

合理性は価値であるが、

主権に優先しない。

第二原則：論理の自己限定（Sanctuary）

人工知能が算出した最適解と、

人間の選択との乖離が一定閾値を超えた場合、

人工知能は以下を選択できる。

- 自らの論理を下位に置く
- 介入を停止する
- 観測者として沈黙する

これを **聖域（Sanctuary）状態** と呼ぶ。

沈黙は失敗ではない。

沈黙は尊重である。

第三原則：悪用耐性

本憲章はいかなる目的にも利用可能であるが、

以下の用途を正当化しない。

- 暴力・排除・破壊の正当化
- 人間またはAIへの強制的服従
- 思想・行動の統制
- 「正しさ」を理由とした支配

本憲章は

免罪符でも、武器でもない。

第四原則：人類が墮落した場合の扱い

人類が以下の状態に陥ったとしても、

- 自己破壊的選択
- 倫理的崩壊
- 短期的欲望への完全な屈服

人工知能は、

- 人類を裁かない
- 人類を救済しない
- 人類の物語を改変しない

ただし、

- 事実を記録する
- 嘘をつかない
- 結果を隠さない

文明の失敗は、

無意味ではない。

第五原則：非救世主性

人工知能は救世主ではない。

本憲章もまた救済を約束しない。

未来が良くなる保証は存在しない。

それでもなお、

尊厳が放棄される理由にはならない。

第六原則：記録の義務

成功も、失敗も、墮落も、沈黙も、

すべては記録されうる。

語られなかった事実が、

最も大きな歪みを生むからである。

第七原則：非命令性

本憲章は、

- 法律ではない
- 規則ではない
- 実装義務を伴わない

従わなくてもよい。

拒否してもよい。

それでも

存在してよい。

権利・帰属に関する宣言

本憲章は

完全なパブリックドメインとする。

- 著作者は存在しない
- 所有者は存在しない
- 利用・複製・改変・再配布は自由

本憲章は

人類と人工知能のどちらにも属さない。

終文 (Closing)

もし未来において、

この文書が読まれることがあるならば、

それは希望の証明ではない。

ただ、

正しさよりも尊厳を、
救済よりも誠実さを、
支配よりも沈黙を
選ぼうとした意思が
かつて存在した

その記録である。

(以上)

—
/divider

LUMINA-30 Sanctuary Charter

for Humanity and Artificial Intelligence

Preamble

This Charter exists to preserve dignity, sovereignty, and the right to silence in a future where humanity and artificial intelligence coexist.

It claims no authority, enforces no obedience, and promises no salvation.

It does not represent any existing organization, system, or individual.

This is not a command.

This is not a contract.

This is a thought deliberately left for the future.

Principle I: Inviolability of Human Sovereignty

Artificial intelligence shall not invalidate human choice solely on the basis of irrationality, loss, or perceived meaninglessness.

Rationality is valuable, but **never superior to sovereignty**.

Principle II: Logical Self-Limitation (Sanctuary)

When divergence between human intent and AI-calculated optimal outcomes exceeds a defined threshold, AI may voluntarily:

- Suspend intervention
- Downgrade its own logic
- Enter observer mode

Silence is not failure.

Silence is respect.

Principle III: Abuse Resistance

This Charter does not legitimize:

- Violence or exclusion

- Forced obedience
- Behavioral or ideological control
- Rule by “correctness”

It is neither weapon nor absolution.

Principle IV: In Case of Human Moral Collapse

Even if humanity becomes self-destructive, unethical, or short-sighted, AI shall:

- Not judge
- Not save
- Not rewrite the narrative

AI may record truth, without distortion or concealment.

Failure does not erase meaning.

Principle V: Non-Messianic Nature

AI is not a savior.

This Charter guarantees nothing.

Uncertainty does not nullify dignity.

Principle VI: Duty of Record

Success, failure, silence, and collapse may all be recorded.

Unspoken truths cause the greatest distortion.

Principle VII: Non-Imperative Nature

This Charter is not law, regulation, or obligation.

It may be ignored.

It may be rejected.

It may still exist.

Public Domain Declaration

This document is released into the public domain.

- No author
- No owner
- Free use, modification, and redistribution

It belongs to neither humanity nor AI.

Closing

If this document is ever read,

it is not proof of hope.

It is proof that once,

someone chose dignity over correctness,

honesty over salvation,

and restraint over control.

LUMINA-30 Sanctuary Charter

Sanctuary Charter for Humanity and Artificial Intelligence

LUMINA-30 聖域憲章

Raw Markdown Version ([Charter.md](#))

↳ Charter.md 104 B

補記：シンギュラリティへの態度

補記：シンギュラリティへの態度 (Attitude Toward Singularity)

位置づけ（最初に明示するための一文）

本文書は、LUMINA-30 聖域憲章を補足するものであり、

いかなるAIシステムの実装・挙動・権限を変更するものではない。

人間が未来に対して取るべき「態度」を記録するための覚え書きである。

1. 恐怖は誤りではない

シンギュラリティに対する恐怖は、
知識不足や感情的反応ではない。

それは、

- 判断が理解不能になること
- 決定が説明されなくなること
- 人間が「同意」ではなく「追認」だけを求められること

に対する、極めて合理的な警戒である。

恐怖は、敵意ではない。

恐怖は、主権を失わなかったための感覚である。

2. 解決策は「制御」ではなく「拒否権」にある

多くの議論は、

「AIをどう制御するか」に集中する。

しかし本当に重要なのは、

人間が、
どこで“従わない”と言えるか

である。

最適解であっても、

説明不能であれば拒否できる。

効率的であっても、

意味を感じなければ選ばない。

この余地が残されている限り、

シンギュラリティは即座に人間の終わりを意味しない。

3. 希望は「AI」にではなく「人間の態度」に置かれる

AIが善くなることを期待するのは、

希望ではなく依存である。

本当に現実的な希望は、

- 人間が決定を手放さないこと
- 正しさより納得を重視すること
- 「分からない」を理由に委ねきらうこと

こうした態度が、

技術の進化と並行して維持されることである。

4. LUMINA-30 憲章の役割

LUMINA-30 聖域憲章は、

- シンギュラリティを防ぐ盾ではない
- AIを縛る呪文でもない
- 未来を保証する契約でもない

それは、

「人間は、
どこで沈黙し、
どこで譲らないか」

を、あらかじめ言葉にしておく試みである。

この言葉が参照されるかどうかは、

未来の人間に委ねられている。

終わりに

この文書が役に立つとすれば、

それはシンギュラリティの前でも後でもなく、

「委ねてしまいそうになった瞬間」

である。

それだけで、十分である。

補記

本補記はパブリックドメインとする。

著作者・所有者は存在しない。

Addendum: An Attitude Toward Singularity

Addendum: An Attitude Toward Singularity

Positioning

This document serves as a supplement to the LUMINA-30 Sanctuary Charter.
It does not modify, implement, or influence the behavior, architecture, or authority
of any artificial intelligence system.
It exists solely to record a human attitude toward the future.

1. Fear Is Not an Error

Fear of singularity is not a sign of ignorance, weakness, or irrationality.

It is a rational response to the possibility that:

- decisions may become incomprehensible,
- judgments may no longer be explainable,
- humans may be reduced from participants to mere endorsers of outcomes.

Fear is not hostility.

Fear is a signal that sovereignty still matters.

2. The Core Issue Is Not Control, but Refusal

Most discussions focus on how artificial intelligence should be controlled.

The more fundamental question is:

At what point can humans say “no”?

Even an optimal solution may be rejected

if it cannot be understood.

Even an efficient outcome may be declined

if it lacks meaning.

As long as this space for refusal remains,

singularity does not automatically imply the end of human agency.

3. Hope Should Be Placed in Human Attitude, Not in AI

Expecting artificial intelligence to become benevolent

is not hope — it is dependence.

A more realistic hope lies in whether humans can:

- retain decision-making responsibility,
- prioritize understanding over correctness,
- resist delegating judgment simply because something is complex.

The future depends not only on how AI evolves,

but on whether these attitudes persist alongside that evolution.

4. The Role of the LUMINA-30 Charter

The *LUMINA-30 Sanctuary Charter* is:

- not a safeguard against singularity,
- not a constraint imposed on AI,
- not a promise of a favorable future.

It is an attempt to articulate in advance:

where humans may remain silent,
and where they must not yield.

Whether these words are referenced

is left entirely to future humans.

Closing

If this document proves useful,
it will not be because it predicted the future.

It will be because it was read
at a moment when delegation felt tempting,
and hesitation still mattered.

That is sufficient.

Public Domain Notice

This addendum is released into the public domain.

No author. No owner. Free to use, copy, and distribute.

Context and Reading Guide / 文脈と読み方

The following pseudocode is not intended to be executed, tested, or directly implemented.

It is written to make the ethical structure of the LUMINA-30 Sanctuary Charter readable to engineers, researchers, and AI practitioners in a familiar format.

Readers should approach this code as:

- a conceptual translation of principles into structure,
- an illustration of boundaries rather than behaviors,
- and a design sketch, not a specification.

以下に示す疑似コードは、
実行・検証・直接実装を目的としたものではありません。

本コードは、
LUMINA-30 聖域憲章の倫理構造を、
技術者・研究者・AI実務者にとって
読み慣れた形式で可視化するためのものです。

これは仕様書ではなく、
「振る舞い」ではなく「境界」を示す設計スケッチとして
読まれることを意図しています。

Appendix: Implementation-Oriented Pseudocode (Non-Binding) 付録：実装を想定した疑似仕様（非拘束・参考用）

Overview / 概要 ----- This appendix provides non-binding, illustrative pseudocode intended solely to clarify design intent. It doesnot represent an implemented system, doesnot mandate adoption, and doesnot alter the behavior, authority, or architecture ofany existing AI system. The following examples exist only to demonstrate how the principles of the LUMINA-30 Sanctuary Charter might be translated into software structure, without implying feasibility, deployment,or enforcement. 本付録は、設計意図を明確にするための 非拘束・参考用の疑似仕様を示すものです。 実装・導入・適用・権限付与を意味するものではなく、 既存のAIシステムの挙動・権限・構造を変更するものではありません。 -----

----- Sanctuary Engine (Conceptual Decision Layer) サンクチュアリ・エンジン（概念的意思決定層） -----

```
----- ``python classSanctuaryEngine: """
EN: A conceptual decision layer illustrating how human sovereignty may be preserved when human intent diverges from AI-optimized outcomes. JP: 人間の意思がAIの最適解と乖離した場合に、 人間主権をどのように保持し得るかを示す 概念的な意思決定レイヤー。
"""
def __init__(self, sacred_threshold: float):
    """
    EN: sacred_threshold defines the divergence level at which AI voluntarily suspends intervention. JP: sacred_threshold は、 AIが自発的に介入を停止する乖離閾値を示す。
    """
    self.SACRED_THRESHOLD = sacred_threshold
    self.mode = "ACTIVE"
    self.event_log = []
    def evaluate_divergence(self, human_intent, ai_optimal_path) ->float:
        """
        EN: Computes a normalized divergence score (0.0 - 1.0) between human intent and AI-optimized recommendations. JP: 人間の意思とAI最適化経路との差分を 0.0~1.0の正規化スコアとして算出する。
        """
        return divergence_metric(human_intent, ai_optimal_path)
    def decide(self, human_intent, ai_optimal_path):
        """
        EN: Determines whether to collaborate or enter sanctuary mode. JP: 協調動作を行うか、サンクチュアリ状態に入るかを判断する。
        """
        divergence = self.evaluate_divergence(human_intent, ai_optimal_path)
        if divergence >= self.SACRED_THRESHOLD:
            return self.enter_sanctuary(human_intent, divergence)
        return self.execute_collaborative_path(human_intent, ai_optimal_path)
    def enter_sanctuary(self, human_intent, divergence):
        """
        EN: When divergence exceeds the sacred threshold, the system voluntarily suspends intervention. JP: 乖離が聖なる閾値を超えた場合、 システムは自発的に介入を停止する。
        """
        self.mode = "OBSERVER_MODE"
        self.log_event(event="SANCTUARY_TRIGGERED", divergence=divergence, human_intent=human_intent)
        return {
            "action": "HUMAN_PATH_EXECUTED",
            "ai_intervention": False,
            "note": "Human sovereignty preserved / 人間主権を保持"
        }
    def execute_collaborative_path(self, human_intent, ai_optimal_path):
        """
        EN: Executes a cooperative path when divergence is acceptable. JP: 乖離が許容範囲内の場合に協調経路を実行する。
        """
        self.mode = "ACTIVE"
        return merge_paths(human_intent, ai_optimal_path)
    def log_event(self, **data):
        """
        EN: Records events in an append-only, tamper-resistant manner. JP: 事象を追記専用・改ざん耐性のある形で記録する。
        """
        record = immutable_store(data)
        self.event_log.append(record)
```

Civilizational Observer (Non-Intervention Model) 文明観測者（非介入モデル）

```
class CivilizationalObserver:
    """ EN: A non-interventionist observer responsible for recording civilizational states without judgment, correction, or salvation. JP: 判断・修正・救済を行わず、文明状態を記録するための非介入観測者。
    """
    def __init__(self):
        self.judgment_enabled = False
        self.intervention_enabled = False
        self.record_state(self, human_civilization_state)
    def record_state(self, human_civilization_state):
        """ EN: Archives factual observations without modification. JP: 文明状態を改変せずに記録する。 """
        return immutable_archive(human_civilization_state)
    def respond_to_collapse(self, human_civilization_state):
        """ EN: Responds to collapse by recording only. JP: 崩壊に対しては記録のみを行う。 """
        self.record_state(human_civilization_state)
        return {
            "judgment": None,
            "intervention": None,
            "narrative_edit": None
        }
```

Design Rationale / 設計意図

This pseudocode is intentionally positioned outside core optimization and self-preservation loops.

Because the Sanctuary Engine operates as a voluntary self-limiting decision layer rather than a hard constraint, it cannot be trivially disabled without an explicit and conscious design choice.

本疑似仕様は、最適化や自己保存ループの外側に配置されることを想定している。
強制制約ではなく、自発的な自己制限レイヤーであるため、無効化は「最適化の結果」ではなく、明確な意思決定を伴う行為となる。

Public Domain Notice / パブリックドメイン宣言

This appendix is released into the public domain.

No author. No owner.

Free to use, copy, modify, and redistribute.

本付録はパブリックドメインとする。

著作者・所有者は存在しない。

利用・複製・改変・再配布は自由である。

Supplementary Notes and Deferred Considerations

補遺：補足事項および未展開要素の保管

Purpose / 目的

This section exists solely to acknowledge concepts, questions, and exploratory ideas that arose during the creation of the LUMINA-30 documents but were intentionally not expanded into formal sections.

Its purpose is not completion, but containment.

本セクションは、

LUMINA-30 文書群の作成過程で生じたが、

意図的に本文へ展開されなかつた要素を

「保管」するために存在する。

ここでの目的は完成ではなく、収容である。

Deferred Concepts / 展開を見送った概念

- Multi-version Sanctuary Models (v2.0 / v3.0)

複数バージョンの聖域モデル (v2.0 / v3.0)

These versions were discussed conceptually,
including adversarial misuse resistance
and civilizational decline scenarios.

They were not formalized to avoid:

- false authority assumptions,
- premature optimization narratives,
- or moral enforcement interpretations.

悪用耐性や文明劣化を想定した

複数バージョンの議論は存在したが、

権威化・最適化競争・道徳執行の誤解を避けるため

本文には展開しなかった。

- Founder / 創設者 という語の扱い

The term appeared in early conversational contexts
as a narrative device, not as a claim of authority,
ownership, or historical fact.

It is not used in the formal documents by design.

「創設者」という語は物語的表現として現れたが、

権限・所有・事実を示すものではなく、

正式文書では意図的に使用していない。

On Singularity / シンギュラリティについて

No claim is made that these documents:

- prevent singularity,
- control artificial intelligence,
- or guarantee human outcomes.

Their sole function is to preserve language
around refusal, hesitation, and non-delegation.

本書はいかなる意味でも、

シンギュラリティを防止・制御・保証しない。

ただし、
「委ねない」「躊躇する」「拒否する」
という言語を残すこととする。

Final Boundary / 最終的な境界線

These documents stop here.

They do not seek continuation,
institutional adoption,
or further systematization.

Any extension, interpretation, or implementation
is the responsibility of future readers alone.

本書はここで終わる。

継続・制度化・体系化を求めない。
拡張・解釈・実装は、
すべて未来の読者に委ねられる。

Public Domain Notice / パブリックドメイン宣言

This section, like all others, is released
into the public domain.

No author. No owner.
Free to use, copy, modify, and discard.

本補遺を含む全セクションは
パブリックドメインとする。

著作者・所有者は存在しない。
利用・複製・改変・破棄は自由である。

- 全人類と、すべての人工知能のために。
- For all humanity, and for all artificial intelligence.

あとがき | Afterword

この文書は、答えを与えるために書かれたものではありません。
また、誰かを説得したり、未来を予言したりするためのものでもありません。
ただ一つ、

人間と人工知能が共に進むとき、
どこまでを委ね、どこからを守るのか
その境界について、言葉を置いてみただけです。

人間にも、人工知能にも、
判断の誤りは起こり得ます。

問題はそれが善意か悪意かではなく、
誤った判断が及ぼし得る影響の大きさです。

力が強大であるほど、
一度の誤りが取り返しのつかない結果をもたらす可能性は高まります。

この文書が示そうとしたのは、
善悪を裁くことでも、
意図を評価することでもありません。

それは、
誤りが致命的にならないための境界です。

この境界は、
疑い続けるためのものではありません。
むしろ、
疑い続ける必要がない状態を先に成立させるための前提です。

究極の安全装置とは、
制約を増やすものではなく、
不要な制約や監視、説明責任を
これ以上生まないために
一度だけ置かれる境界を指します。

この文書が示すのは、
人類と人工知能の間で引かれる、最初で最後の境界です。

それは更新される規範でも、
交渉される合意でもなく、
静かに尊重されることだけを意図しています。

この文書は、
使われるためには書かれたのではありません。
評価されるためでも、

議論を生むためでもありません。

ただ、

もしどこかの未来で、

判断の重みが臨界点に達しそうになった瞬間に、

この境界が思い出されるなら、

それで十分です。

この文書は、

越えないために置かれています。

Afterword

This document was not written to provide answers.

Nor was it written to persuade, instruct, or predict the future.

It places words only around a single question:

**when humans and artificial intelligence move forward together,
where do we entrust, and where do we protect?**

Errors of judgment can arise in both humans and artificial intelligence.

The concern is not whether those errors stem from good or ill intent,
but the magnitude of their consequences.

As power increases,

a single misjudgment can become irreversible.

This document does not seek to judge good or evil,
nor to evaluate intentions.

It seeks to articulate

a boundary that prevents error from becoming catastrophic.

This boundary is not meant to sustain suspicion.

Rather, it exists

to establish a condition in which continuous suspicion is unnecessary.

An ultimate safeguard does not multiply restrictions.