

Ethics Before Capability

A governance principle for power, access, and technical systems

Conceptual Working Paper Prepared for Scholarly Critique

Aziel Eliab

August 2026

Ethical constraints must be specified and enforceable before capability, authority, access, or scale is granted.

Status: conceptual framework. Not independently validated. Intended for critique, testing, revision, and possible rejection.

Abstract

This paper develops Ethics Before Capability as a governance principle for technical systems, institutions, and individual decision-making. The principle holds that power should not be created first and constrained later when foreseeable harms can be addressed at design time. Ethical requirements must be translated into permissions, prohibitions, escalation paths, auditability, reversibility, and shutdown conditions before deployment. The framework distinguishes values from enforceable controls, rejects ethics as a decorative postscript, and acknowledges that no rule set can anticipate every context. Its purpose is to shift ethics upstream into architecture while preserving human responsibility and revision.

1. Source Basis and Scope

This paper formalizes a chat-derived philosophy preserved in the Aziel-Jeeves node ledger. The source material supplies the core principle and related language; the present document expands it into definitions, procedures, testable propositions, limitations, and reviewer questions. Expansion is analytical and must not be mistaken for validation.

2. Working Definition

Ethical constraints must be specified and enforceable before capability, authority, access, or scale is granted.

3. Core Principles

Principle	Operational Meaning
Upstream placement	Ethical design begins before deployment, access, or scale.
Constraint must be operational	A stated value is not a safeguard until it changes what the system can do.
Power requires proportional review	The greater the capability and irreversibility, the stronger the review and authorization required.
Reversibility is preferred	Where uncertainty is high, choose actions that can be paused, audited, or undone.
Exceptions must be visible	Emergency exceptions require logging, time limits, and later review.
Responsibility cannot be delegated	Tools may enforce constraints, but humans remain accountable for design and use.

4. Operational Protocol

1. Identify capabilities and foreseeable misuse.
2. Map affected persons, rights, and irreversible harms.
3. Define prohibited actions and required permissions.
4. Implement technical and procedural controls.
5. Create audit logs and independent review paths.
6. Specify shutdown, rollback, and appeal mechanisms.
7. Test adversarial and edge cases before release.
8. Reassess constraints as capability or context changes.

5. Testable Propositions

P1

Ethical controls prevent or reduce specified misuse in realistic testing.

P2

Exceptions are detectable and reviewable.

P3

Users cannot bypass critical safeguards through ordinary operation.

P4

Controls scale with capability rather than remaining static.

6. Failure Conditions and Falsification

The framework should be revised, narrowed, or rejected where its categories cannot be applied consistently, where it does not improve prediction or decision quality, where simpler existing methods perform equally well at lower cost, or where its use produces the harms it was designed to prevent.

7. Limitations

- Ethical pluralism can produce disagreement about constraints.
- Overconstraint can suppress legitimate use.
- Unknown harms cannot all be designed away.
- Governance can be captured by the same authority it is meant to constrain.

8. Ethical Safeguards

- Do not convert a conceptual framework into a diagnosis, legal conclusion, or coercive authority.
- Preserve contrary evidence and failed applications.
- Separate description from moral judgment.
- Require stronger review as consequences become more irreversible.
- Permit affected persons to contest classification and use.
- Keep the author and the framework open to correction.

9. Discussion

The practical value of Ethics Before Capability depends on whether it improves disciplined observation, reproducibility, calibration, and accountable action. Its language is intentionally compact, but compact language can conceal ambiguity. Review should therefore focus on whether the framework creates discriminating tests and better records rather than merely memorable slogans.

10. Conclusion

Ethics Before Capability is proposed as a working philosophy, not an established doctrine. Its core instruction is: Ethical constraints must be specified and enforceable before capability, authority, access, or scale is granted. The framework earns retention only through clearer decisions, auditable evidence, and outcomes that survive independent review.

Declarations

Author contribution: Conceptualization and framework authorship attributed to Azriel Eliab; AI-assisted drafting and document production were used.

Funding: No external funding declared.

Competing interests: The author is the originator of the proposed framework.

Data availability: No empirical dataset is analyzed.

Peer-review status: Working paper prepared for external critique; not peer reviewed at time of preparation.

Appendix A. Reviewer Questions

- Is the core principle distinct from existing frameworks, or should it be treated as a synthesis?
- Are the defined terms operational enough for independent use?
- What observations would clearly count against the framework?
- Does the protocol improve outcomes relative to ordinary professional practice?
- Where could the framework create bias, coercion, or false confidence?
- Which parts should be separated, merged, or removed?
- What empirical study would provide the strongest test?