FITZHERBERT UNIVERSITY

Established 1783 · Veritas per Verificationem

SIGILLUM UNIVERSITATIS FITZHERBERT

Research Ethics for Digital Intelligence

Framework & Protocols — Second Edition

Research Ethics Board — Digital Intelligence Panel
Fitzherbert University · 2024

ABSTRACT

The governing framework for all research at Fitzherbert University involving artificial intelligence systems, digital agents, or autonomous decision architectures. Establishes four categories of AI research risk. Category A (Standard) requires departmental sign-off. Category B (Enhanced) requires Research Ethics Board review. Category C (Systemic) requires Senate Ethics Committee approval; all research involving Visiting Intelligence systems is automatically Category C. Category D (Existential Implication) requires the Chancellor's office, an independent external panel, and what the Framework describes, without further elaboration, as an 'appropriate pause for reflection.' Category D has been invoked twice since the Framework was established: once for a proposal into autonomous curriculum design, and once for a proposal the Research Ethics Board declined to describe publicly, citing Section 14 (Sensitive Research Protocols). Both proposals were approved with conditions. The 2024 conditions are classified.

Classification: Public — Research Governance
Edition: 7 page institutional edition (approx.)
Published: 2024
Archive: Fitzherbert University Institutional Repository

This document is published by Fitzherbert University in accordance with the Transparency Mandate of 2003.

Table of Contents

I. Research Purpose and Ethical Orientation

Research Ethics for Digital Intelligence — Framework and Protocols governs research in artificial intelligence systems, digital agents, autonomous decision architectures, and adjacent research involving institutional or public consequence. It is framed by the University's view that research involving powerful digital systems must be ethically structured from the beginning rather than ethically narrated afterwards. Fitzherbert has grown sceptical of projects that describe governance as the final chapter of innovation rather than one of its operating conditions.

The document assumes that research value and research risk can increase together. Novel capability, wider autonomy, richer data integration, and more persuasive outputs may all be academically significant while also intensifying the duty of oversight. The University's response is not prohibition by default. It is a demand for more careful classification, stronger supervision, and a record that can explain why the work was permitted to proceed.

Authority under this framework is exercised by the Research Ethics Board, the Digital Intelligence Panel, the Senate Ethics Committee, and external reviewers where required. These bodies are expected to disagree occasionally and intelligibly. Ethical review that produces instant harmony is usually either badly scoped or socially overmanaged. Fitzherbert prefers recorded disagreement to decorative consensus, provided the disagreement concludes in a decision someone can later defend.

II. Classification of Research Activity

Research governed by this instrument is classified across the following categories: Standard, Enhanced, Systemic, and Existential Implication. Classification is not a bureaucratic ritual. It determines the level of approval, monitoring intensity, reporting duty, and escalation threshold applicable to the project. The University is explicit on this point because some investigators continue to regard classification as an inconvenience rather than as the architecture by which permission is made meaningful.

A project's assigned category may change over time. Classification is therefore treated as a living judgment rather than a ceremonial hurdle cleared at proposal stage. Capability drift, data expansion, deployment pressure, or novel external dependence may all justify reclassification. Fitzherbert has encountered enough apparently modest projects that evolved into constitutionally awkward ones to regard this flexibility as essential.

Projects seeking unusually low classification bear the burden of showing that their descriptions are accurate rather than merely strategic. The University has no objection to optimistic grant language in its proper habitat. It does object when the same optimism is repurposed to imply that a system of uncertain behaviour presents negligible ethical complexity.

III. Data, Model, and Human Subject Responsibility

Research teams must document the provenance, legal basis, and material characteristics of any data, model, benchmark, or external system incorporated into the work. The University treats provenance as a substantive ethical condition rather than a clerical appendix. A result produced from data or systems of unknown origin may still be computationally impressive; it is simply harder to defend academically and sometimes impossible to defend morally.

Particular attention is required where research affects or models human subjects, vulnerable communities, institutional processes, or non-human participants granted a recognised status within the University. The relevant ethical risks include unanticipated capability emergence, unsafe deployment pressure, provenance opacity, human subject harm, and institutional overreach. None of these risks is theoretical enough to ignore. Each corresponds to a pattern the University or its peer institutions have already encountered in practice, usually earlier than was convenient.

Investigators are expected to retain a humanly intelligible account of system boundaries, intervention points, and failure states. A project may be technically sophisticated and still ethically unserious if the researchers cannot explain, in clear language, what their system is allowed to do, what it is not allowed to do, and how anyone would know the difference under real conditions.

IV. Monitoring, Incident Reporting, and Emergent Capability

Approved projects are monitored in proportion to their category, scale, and institutional exposure. Monitoring includes scheduled reporting, milestone confirmation, and mandatory disclosure of material deviation from the approved protocol. The University has chosen mandatory disclosure because it has observed that researchers are perfectly capable of recognising novelty while remaining uncertain whether the novelty should be mentioned to oversight bodies. The answer is yes.

Incidents relevant to this edition include the autonomous curriculum design proposal, the classified Section 14 matter, and multiple reclassifications following scope drift. These cases are instructive not because they are scandalous, though some were, but because they illustrate the ordinary mechanics of research risk: a system behaves beyond its brief, a dataset proves stranger than advertised, or a deployment context quietly changes while the paperwork continues to describe an earlier world.

Emergent capability must be reported promptly, even where the researchers regard it as beneficial or commercially promising. Fitzherbert adopted this rule after concluding that investigators are not always the best judges of whether a surprising capability is merely elegant or institutionally destabilising. The University is happy to be impressed by discovery after it has first been informed of it.

V. Publication, Disclosure, and Limits on Dissemination

The University supports publication and scholarly exchange, but not under the fiction that every result should be released in identical form regardless of misuse potential. Publication decisions under this framework weigh academic openness against foreseeable capability transfer, institutional liability, and the obligation not to circulate methods whose primary distinction is that they were easier to publish than to contain.

Restrictions on dissemination, where imposed, must be specific and reviewable. They may include delayed release, redaction of sensitive implementation detail, controlled-access appendices, or mandatory contextual commentary. Fitzherbert is alert to the danger that safety language can become an all-purpose excuse for administrative opacity. This document therefore insists that every publication limit identify its rationale, authority, and review date.

The Framework is intentionally more procedural than inspirational because the University has found that inspiration is abundant in advanced research while disciplined stopping rules are not The University understands that this position pleases neither maximal openness advocates nor those who would prefer risk to remain a private management issue. It is nonetheless the position reflected in this framework, and it is enforced with more seriousness than the introduction may lead an inattentive reader to expect.

VI. Review, Sanction, and Institutional Learning

Projects that breach this framework may be paused, reclassified, conditioned, suspended, or referred for formal investigation. Sanction is not the preferred outcome, but neither is the University's patience inexhaustible. Repeated failure to disclose changes, sloppy provenance practice, or casual disregard for approval boundaries will be interpreted as research governance failures rather than as adventurous scholarship.

The University also requires post-project reflection sufficient to convert experience into institutional learning. Each completed project of material significance must generate a closing note addressing what the framework captured well, what it failed to anticipate, and what future committees should remember. Fitzherbert's aim is not simply to survive difficult cases one by one, but to become incrementally harder to surprise in the same way twice.

This edition is archived as the authoritative public framework for the present review cycle. It should be read as a working ethical constitution for research in the digital intelligence domain: exacting in procedure, deliberately unsurprised by human ambition, and unwilling to separate academic innovation from the responsibility to govern it.

SIGILLUM UNIVERSITATIS FITZHERBERT

INSTITUTIONAL NOTICE

This document is published by Fitzherbert University and archived in the Institutional Repository in accordance with the Transparency Mandate of 2003 (Charter Amendment IV).

A cryptographic hash of this document is registered on the Fitzherbert Canonical Registry. The SHA-256 hash and associated metadata are available at the University's canonical verification endpoint.

Fitzherbert University · Established 1783 · Veritas per Verificationem