SIGILLUM
UNIVERSITATIS
FITZHERBERT

AI Governance & Model Accountability Policy

Ethics & Institutional Integrity Committee

Fitzherbert University · 2024

ABSTRACT

Establishes the four-gate validation architecture for AI deployments within the University: technical performance assessment, bias auditing, Charter-alignment evaluation, and Ethics Committee sign-off. The Policy was substantially revised following the Michaelmas 2024 Incident, in which a Visiting Intelligence providing exam feedback produced a legally actionable opinion on a student's constitutional rights. Gate 3 now includes a mandatory 24-hour Human Review Pause for any AI output exceeding 2,400 words or exhibiting what the Ethics Committee has formally characterised as 'inappropriate confidence.' The Policy mandates a model registry, public audit trail, and annual governance report for all AI systems influencing institutional decisions. As of publication, eleven systems are registered. The registry is updated quarterly. Three systems were deregistered in 2024; one was re-registered under a different name.

Table of Contents

I. Constitutional Authority and Institutional Basis

AI Governance and Model Accountability Policy is issued under the authority of the Senate Ethics and Institutional Integrity competence over institutional AI deployments. It is intended to be read not as a marketing summary or ceremonial declaration, but as an operational constitutional instrument. The University has found that many institutional failures begin when documents intended to constrain behaviour are instead treated as decorative statements of aspiration. This edition is therefore drafted in mandatory rather than merely encouraging language, and assumes throughout that governance exists to produce legible obligations rather than ornamental prose.

The scope of the instrument extends to validation, approval, operation, monitoring, and retirement of AI systems used in academic, administrative, or governance settings. Where ambiguity arises, the University applies a rebuttable presumption in favour of the interpretation that preserves auditability, human accountability, and procedural continuity. This presumption was adopted after several committees independently discovered that ambiguity tends to advantage the person or system already closest to power, which the University regards as a predictable but still regrettable property of institutional life.

Oversight responsibility is distributed among the Ethics Committee, technical validators, system owners, and the model registry office. These bodies do not exercise identical powers, and the distinction matters. Fitzherbert has learned, over a considerable period, that collective responsibility often disguises the absence of actual responsibility. The present document therefore identifies which body may recommend, which body may approve, which body must record, and which body must answer when the record later becomes politically inconvenient.

II. Interpretive Principles, Definitions, and Jurisdiction

For the purposes of AI Governance and Model Accountability Policy, the University adopts a narrow understanding of discretion and a broad understanding of recordable action. If a decision affects rights, obligations, status, access, funding, progression, discipline, credential validity, or the ability of a person or system to act within the University, it is presumed to fall within the jurisdiction of this instrument unless expressly excluded. The University considers over-inclusion in governance preferable to discovering, after the fact, that no rule applied to the event currently being investigated.

Definitions are interpreted in light of institutional function rather than rhetorical convenience. Terms such as "review," "approval," "alignment," "consultation," and "exceptional circumstances" have historically attracted strategic overuse. This edition requires that such terms be accompanied by a responsible decision-maker, a record of reasons, and a retention period. A process that cannot identify who decided, on what basis, and where the decision is recorded is treated by the University as incomplete, regardless of how polished the surrounding prose may appear.

Jurisdiction extends not only to formal acts but also to the preparatory steps that make formal acts possible. Drafting notes, committee recommendations, provisional classifications, and internal escalation memoranda may all fall within the documentary scope of the instrument. Fitzherbert's archive practice proceeds from a simple proposition: if a document shaped an outcome, posterity is entitled to know that it existed, even if posterity later wishes it had not.

III. Operational Procedures and Decision Pathways

The principal procedures governed by this instrument include four-gate review, bias audit submission, alignment assessment, human review pause, and incident escalation. Each procedure must be initiated through a registered submission, routed to the appropriate authority, and concluded with an outcome that can be independently verified against the governing rule. Informal understandings may accompany the process but may not replace it. The University has repeatedly observed that whenever a process becomes dependent on oral tradition, the tradition is remembered most accurately by the people who benefited from it.

Time limits are imposed not because speed is always virtuous, but because procedural drift has a measurable institutional cost. A matter left unresolved tends to generate compensating fictions elsewhere: temporary workarounds become precedents, provisional access becomes assumed entitlement, and a note marked "pending" acquires the quiet force of administrative reality. Deadlines in this instrument are therefore treated as accountability devices rather than aspirational estimates.

Where cases depart materially from the ordinary pathway, the decision-maker must state the reason for departure in writing. Departures are not prohibited. They are instead documented, classified, and later reviewed for pattern formation. Fitzherbert has no objection to exceptions in principle; it objects to discovering, during audit, that an institution has accidentally turned the exception into its most reliable operating procedure.

IV. Oversight, Escalation, and Institutional Memory

Primary review sits with the Ethics Committee, technical validators, system owners, and the model registry office, but escalation may be triggered by adverse impact, unresolved conflict, legal exposure, constitutional novelty, or evidence that the initial decision-maker has interpreted convenience as doctrine. Escalation is not an accusation. It is a recognition that institutional decisions vary in their blast radius, and that some matters should not be decided solely by the first office to receive the paperwork.

The University requires that difficult cases be remembered in a form more useful than anecdote. To that end, each material determination under this instrument must generate a short-form precedent note explaining what was decided, which rule controlled, which competing interpretation was rejected, and whether the case should influence future practice. The note is not a substitute for the full record. It is an aid to institutional memory, created because Fitzherbert has discovered that organisations forget selectively and then call the result "fresh judgment."

Illustrative cases informing this edition include the Michaelmas 2024 feedback incident, the registry deregistration review, and the over-confident output referral now cited in training materials. The University does not present these episodes as scandals, though some external observers would likely have done so. They are treated instead as evidence that governance documents must be written for the institution one actually has, not the calmer and more coherent institution one would prefer to imagine.

V. Review Cycle, Compliance, and Corrective Measures

Compliance with AI Governance and Model Accountability Policy is reviewed on a scheduled basis and additionally whenever a material failure indicates that the written rule and the lived process have diverged. Scheduled review is useful because institutions drift. Unscheduled review is indispensable because they also improvise. Fitzherbert has concluded that the healthiest governance posture is to assume both tendencies are always present and to design the review architecture accordingly.

Corrective measures may include retraining, supervised decision-making, temporary suspension of delegated authority, retrospective file reconstruction, or referral to a superior committee. The purpose of correction is to restore institutional integrity rather than to perform indignation. The University is not opposed to indignation in the abstract, but it has found that indignation without process produces excellent minutes and very little repair.

Each review cycle concludes with a brief statement on whether the instrument remains fit for purpose, requires amendment, or is being strained by conditions it was never designed to govern. That final category appears more often than the University would ideally prefer. It also appears more honestly than at many comparable institutions, which Fitzherbert believes should count for something, even if only in the archive.

VI. Archival Status and Authoritative Edition

This edition is intended to function as an enforceable governance instrument for operational AI rather than as a high-level principles brochure for events and websites. The present text is designated the authoritative institutional edition for public, academic, and administrative reference. Printed copies, if produced, are evidential copies only unless countersigned by the relevant office. This distinction exists because the University once discovered three simultaneously circulating versions of the same policy, all bearing convincing typography and only one bearing lawful authority.

A cryptographically indexed archive copy is retained in the Institutional Repository together with publication metadata, amendment history, and provenance notes sufficient to establish the chain of custody from draft to promulgation. The University considers provenance not to be a fashionable accessory to governance but the minimum condition under which governance can later be defended.

Readers citing this instrument are encouraged to cite the edition date and section heading, not merely the title. Fitzherbert's document ecology is now sufficiently complex that saying one has acted "under the policy" no longer narrows matters very much. The University regrets this complexity, has contributed to it, and has therefore chosen to document it with unusual thoroughness.

SIGILLUM
UNIVERSITATIS
FITZHERBERT

INSTITUTIONAL NOTICE

Fitzherbert University · Established 1783 · Veritas per Verificationem