

Bioethics Outlook

Volume 37 No 3

September 2026

From Mission to Mechanism

**Mercy Health Australia's AI journey
from Rome Call signatory
to an Ethical AI Governance Framework for Catholic
Organisations**

Paul Jurman

Artificial intelligence (AI) is transforming healthcare at a pace that outstrips the capacity of organisations, regulators, and governance frameworks to respond. For Catholic healthcare organisations, this challenge is acute. The mission to uphold the dignity of every person demands that AI be adopted thoughtfully, governed rigorously, and anchored in the values that define the Catholic tradition.

Dr Paul Jurman presents the experience of Mercy Health, an Australian Catholic not-for-profit healthcare provider in navigating this challenge. Dr Jurman is Chief Information and Digital Transformation Officer at Mercy Health in Melbourne. He traces the development of Mercy Health's AI Strategy with its basis in the Pontifical Academy for Life's 'Rome Call for AI Ethics', its journey to signatory status, and its response to the gap identified between the aspirational principles of the Rome Call and the operational governance required to deliver them.

1. Who is Mercy Health ?

Mercy Health is one of Australia's leading health organisations caring for people in four of Australia's six states. Founded by the Sisters of Mercy and built on a tradition of serving those in need, today Mercy Health employs more than 11,500 staff with revenues exceeding \$1 billion annually. The organisation delivers a wide range of services, including acute and subacute hospital care, residential and community aged care, mental health, maternity and women's health, early parenting, and palliative care. What distinguishes Mercy Health from other Australian healthcare providers is not its scale but its mission. Mercy Health is explicitly, intentionally, and operationally a Catholic organisation. Its mission '*Together we witness God's mercy and strive for justice in the delivery of compassionate care to our communities*' is a live commitment that shapes every decision the organisation makes, from clinical governance to technology procurement. The charism of the Sisters of Mercy, rooted in Catherine McAuley's conviction that mercy is both a spiritual virtue and a practical obligation to act for the most vulnerable, continues to shape Mercy Health's governance, culture, and strategy. Mercy Health is a ministry of the ministerial public juridic person, Mercy Ministry Companions, and is governed by concurrent Mercy Health Boards. In 2025 Mercy Health became the first Australian health and aged care provider to sign the Rome Call for AI Ethics and has developed a framework to support other Catholic organisations to adopt and embed the Rome Call's principles.

1.1 Mission and Catholic Identity

Mercy Health's Catholic identity is expressed through its core values of Compassion, Hospitality, Respect, Innovation, Stewardship, and Teamwork, and through its explicit commitment to Catholic Social Teaching as a set of principles which inform strategic decisions. Catholic Social Teaching, developed across more than 130 years of social encyclicals, from Leo XIII's *Rerum Novarum* (1891) through John XXIII's *Pacem in Terris* (1963), the Second Vatican Council's *Gaudium et Spes* (1965), and John Paul II's *Centesimus Annus* (1991), through to Pope Francis's *Laudato Si'* (2015) and *Laudate Deum* (2023), provides Mercy Health with a coherent moral anthropology: human beings possess inherent, irreducible dignity as creatures made in the image and likeness of God (*imago Dei*); they are social beings whose flourishing depends on the quality of their community and institutions; and they bear obligations of solidarity and preferential care toward the most vulnerable.

This foundation is directly relevant to artificial intelligence governance. The Dicastery for the Doctrine of the Faith and the Dicastery for Culture and Education issued *Antiqua et Nova* in January 2025. It affirmed that the human person's dignity as *imago Dei* is the non-negotiable standard against which all AI development must be assessed. For Mercy Health, this affirmation is the organisation's own foundational conviction, expressed through its mission and enacted through its governance.

1.2 Scale, Complexity, and the AI Imperative

Mercy Health's scale makes it not merely a consumer of AI but a significant actor in shaping how AI is deployed in Australian healthcare. The use of AI in Mercy Health's operations creates both opportunity and risk. The opportunity is that if AI is applied well it has the ability to extend the reach and impact of the organisation's mission. It has the potential to provide more accurate diagnoses, better resource allocation, and more responsive care. However if AI is applied carelessly, for example trained on biased data, optimised for efficiency, or deployed without adequate human oversight, there exists the potential to cause harm to those the organisation was founded to serve. Managing this tension is the governance challenge that Mercy Health's AI Strategy was designed to address.

2. Current Key Catholic Literature on AI

The Catholic Church's engagement with artificial intelligence has evolved substantially since 2020. What began as a 'principles' document has grown into a substantial body of work. Understanding this is essential context not only for Mercy Health's AI Strategy but also for the ethical governance of AI for Catholic organisations more widely.

2.1 Rome Call for AI Ethics (Pontifical Academy for Life, 2020)

The Vatican's first major public commitment on artificial intelligence, the *Rome Call for AI Ethics* (Pontifical Academy for Life, 2020), was signed on 28 February 2020 by the Pontifical Academy for Life alongside Microsoft, IBM, and the United Nations Food and Agriculture Organisation. Deliberately non-denominational in its language, the document was designed to attract the broadest possible global coalition. It introduced the concept of *algorithethics* (the ethical use of AI), built around six foundational principles: **transparency, inclusion, responsibility, impartiality, reliability, and security and privacy**. The document functions as a living commitment rather than a static text. In January 2023 Muslim and Jewish faith leaders signed the Rome Call providing genuine inter-faith reach (Wagle, 2025).

2.2 Artificial Intelligence and Peace (Pope Francis, 2024a)

Pope Francis' message for the 57th World Day of Peace was the first papal document dedicated to artificial intelligence. It marked a significant escalation in the Church's involvement in AI discourse. Rather than restating principles, it called for a binding international treaty to ban lethal autonomous weapons, condemned AI disinformation and electoral interference, warned against 'the idea of giving them the power to make decisions about human beings without human accountability', and demanded accountability structures that 'move beyond aspirational values' (Francis, 2024a). This document marks a direct call for a governance architecture, not merely ethics statements.

2.3 Artificial Intelligence and the Wisdom of the Heart (Pope Francis, 2024b)

Where the World Day of Peace message addressed geopolitical threats, Francis' (2024b) World Communications Day message, *Artificial Intelligence and the Wisdom of the Heart: Towards a Fully Human Communication*, considered what AI does to human relationships, meaning, and

authentic encounter. The document argues that genuine communication requires the 'wisdom of the heart', empathy, conscience, and the capacity for love. Francis argued that genuine communication is something that no algorithm can replicate (Francis, 2024b). In this message Pope Francis cautioned that AI risks reducing individuals to data, thinking to a mechanical process, and experience to statistical aggregation, and in doing so risks dissolving reality into abstract calculation. Francis also warned against digital 'echo chambers' that imprison rather than liberate, and argued that authentic information 'cannot be separated from living relationships' requiring 'sensitivity to faces and facial expressions, compassion and sharing' (Francis, 2024b). For Catholic healthcare organisations, this document carries specific operational weight. It affirms that the compassionate human encounter between carer and patient is not a residual practice to be optimised away, but the irreducible core of the ministry of healthcare.

2.4 Decree No. DCCII: Vatican City State AI Guidelines (2024–25)

On the 1 January 2025 Decree No. DCCII from the Pontifical Commission for Vatican City State entered into force. It represented the first binding AI law in Vatican history. The document consists of fifteen articles covering nine domains and illustrates the Church's ability to translate ethical principles into concrete operational AI governance. The Decree established bi-annual AI reporting, an AI Commission, and formal AI monitoring powers (Pontifical Commission for Vatican City State, 2025). Article 6 of the Decree is particularly relevant for healthcare providers as it requires that healthcare AI systems do not 'impair or limit the decision-making assessment of those exercising the medical profession' and that they be 'periodically checked and updated to ensure their reliability and minimise risk'.

2.5 Antiqua et Nova (2025)

Issued in January 2025 by the Dicastery for the Doctrine of the Faith and the Dicastery for Culture and Education, *Antiqua et Nova* is a key document in the Catholic Church's AI literature. The title 'old and new' signals its core argument: that the Church approaches AI not as a novelty requiring new ethics, but with the full depth of a two-thousand-year tradition of reflection on the human person (Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education, 2025). The document's most significant contribution is its affirmation of the *imago Dei* as the non-negotiable standard. Human intelligence is 'God's gift fashioned for the assimilation of truth'; AI 'can perform tasks but cannot think'; and treating AI as an autonomous moral actor 'is to abdicate human responsibility'. In the context of Mercy Health's AI governance framework this theological grounding frames AI oversight as not just a matter of regulatory compliance but as an expression of mission. Similarly, the clinician's moral responsibility is not simply a legal requirement but also a theological conviction.

2.6 Pope Leo XIV: A new chapter

The election of Pope Leo XIV in May 2025 signalled a deliberate continuation of the Vatican's AI agenda. In his first address to the College of Cardinals, Leo XIV (2025) explained his choice of papal name. He chose the name Leo XIV after Leo XIII, specifically because Leo XIII's *Rerum Novarum* (1891) addressed the social upheaval of the first industrial revolution. The new Pope

explicitly identified AI as the defining challenge of a second industrial revolution, one that ‘poses new challenges for the defence of human dignity, justice and labour’ (Leo XIV, 2025).

In his April 2026 address at the Catholic University of Central Africa at Yaoundé, Leo XIV (2026a) applied the framework in an educational context. He warned that AI-structured digital environments are ‘optimised to persuade’, neutralise genuine human encounter and reduce relationships to functional responses. When simulation becomes the norm, he argued, it has the potential to weaken the human capacity for discernment, to close social bonds into self-referential circuits, and ultimately to transform the human relationship with truth (Leo XIV, 2026a). He called on Catholic universities to form leaders capable of discernment in the digital age, arguing that without humanistic formation, passive adaptation to dominant technological paradigms will be mistaken for competence, and the loss of freedom will be mistaken for progress.

2.7 Magnifica Humanitas (Leo XIV, 2026b)

First published 15th May 2026, *Magnifica Humanitas* is Leo XIV’s first encyclical and the first to treat AI governance as its primary organising theme. Its title *Magnifica Humanitas*, meaning ‘the magnificence of humanity’ deliberately echoes the Latin root of Mary’s Magnificat. Its opening image frames the challenge for every organisation: ether to build a new Tower of Babel with technology in the service of uniformity, and self-sufficiency, or to rebuild Jerusalem piece by piece, as Nehemiah did in the service of shared responsibility, community, and human dignity (§§7–10). The ‘Babel syndrome’ vs ‘way of Nehemiah’ distinction runs throughout the encyclical as the evaluative question Leo XIV poses to all institutions.

Chapter one sets the scene providing the context of the development of social doctrine in the church drawing from Leo XIII and *Rerum Novarum*; Pius XI and *Quadragesima Anno*; Pius XII; Saint John XXIII and *Mater et Magistra*; the Second Vatican Council; Saint Paul VI and *Populorum Progressio* and *Octogesima Adveniens*; Saint John Paul II and *Laborem Exercens* and *Sollicitudo Rei Socialis* and *Centesimus Annus*; Benedict XVI and *Caritas in Veritate*; Francis and *Evangelii Gaudium* and *Laudato Si* and *Fratelli Tutti*.

Chapter two examines the foundations and principles of the social doctrine of the church in order to ‘help us to interpret the “new things” of our time – among these ‘new things’ being AI. The foundation of the social doctrine is then explored including the human person; image of the Triune God; the equal dignity of all human beings (with reference to *Dignitas Infinita*); and the supreme value of human rights. The document then examines the principles of social doctrine, in particular the principle of the common good, the principle of the universal destination of goods, the principle of subsidiarity, the principle of solidarity, the principle of social justice, and the concept of integral human development.

Chapter three contains the most operationally significant teaching for Catholic organisations regarding AI governance (§§97–110). In this chapter Leo XIV is precise about what AI is and what it is not, stating that AI systems ‘merely imitate certain functions of human intelligence’ and do not possess moral conscience, embodied experience, or the capacity for love, friendship, or responsibility from within (§99). The chapter asserts that AI cannot be morally

neutral as ‘every technical tool embodies choices and priorities through what it measures, ignores and optimises, and how it classifies people and situations’ (§104). Leo XIV builds on this further by stating that ethical discernment cannot be limited to asking whether AI is being used for good or bad purposes but must also examine the vision of the human person embedded in the system’s data and models. Accountability follows from this and ‘responsibility must be clearly defined at every stage from those who design and develop these systems to those who use them and rely on them for concrete decisions’. In this context accountability means the ability to identify who must justify decisions, monitor them, and remedy harm caused (§105). Leo XIV also calls for ‘disarming AI’, freeing it from monopolistic control and making it genuinely accessible, human-friendly, rather than a new instrument of dominance (§110). The section concludes with an examination of transhumanism and posthumanism cautioning that ‘if the human being is treated as something to be perfected or surpassed, it becomes easier to accept that some lives are less useful, less desirable or less worthy. In the name of progress, ‘necessary sacrifices’ may begin to be justified, placing the burden on the most vulnerable in pursuit of a supposed optimisation of the species’ (§117).

Chapter Four addresses AI’s impact across several key areas including truth, democracy, communication, and the collective imagination. It also introduces the concept of an ecology of communication. Specifically, this section calls to action families, schools, and universities. It explores the importance of education and the development of an educational alliance between policymakers, educational institutions, and families to protect against manipulation and defend the dignity and respect of individuals in digital environments. It highlights the central role schools play, outlining three major challenges they must address being; the inequalities that persist in access to education both within and across different nations; the increasing struggle of educational institutions to keep pace with change and to support the integral development of students; and the risk of the ‘dehumanisation of knowledge’ resulting in an education system in which an incessant flow of information replaces the exercise of research, reflection, and discernment. Leo XIV examines the impact of AI on the structure of work cautioning that AI ‘frequently forces workers to adapt to the speed and demands of machines, rather than machines being designed to support those who work’ (§150), and it can de-skill staff and subject them to automated surveillance. Leo XIV also warns against the potential impact of a rapid contraction in available jobs driven by AI adoption which has the potential to ‘create a chain reaction deeply impacting families, young people, and local economies’ (§151).

The encyclical defines three requirements for organisations deploying AI. Firstly, decisions influenced by algorithms must be ‘understandable, contestable and subject to oversight’. Secondly the benefits of innovation must be paired with investment in skills and access so that technology does not widen the gap between those who have and those who have not. And finally, equity requires that concentration of power be corrected rather than rewarded (§164).

On supply chains, Leo XIV calls on companies to adopt preventive ethical verification (that is, due diligence) (§179) and supply chain transparency so that no competitive advantage is built on the hidden exploitation of data workers or raw material extraction. The section concludes with a statement of shared responsibility. Institutions, businesses, intermediary organisations,

and citizens each bear a distinct obligation to ensure AI serves integral human development rather than becoming a source of exclusion and dominance (§181).

Chapter Five confronts what Leo XIV considers the most dangerous convergence of the digital age: the normalisation of war through technology. He asserts that AI is making lethal decisions faster, more impersonal, and less subject to human conscience. While algorithms amplify polarisation, the military-industrial complex creates structural incentives for conflict, and autonomous weapons systems increasingly remove human accountability from life and death choices. To address this Leo XIV proposes three non-negotiable criteria to govern AI in warfare. First, decisions must remain traceable. Secondly, lethal force must stay under effective human control. And, finally, civilian protection must be guaranteed (§200). Against a ‘culture of power’ (§188) that treats war as inevitable Leo XIV proposes a ‘civilization of love’ (§210) built through five paths of personal responsibility: disarm words, build peace through justice, adopt the perspective of victims, cultivate a healthy realism, and revive dialogue.

The encyclical concludes with a program for Christian life in the era of AI, structured around four pillars: contemplating God's plan, living ecclesial unity through the Eucharist, building a world centered on the common good, and praying in union with Mary (§§229–245).

2.8 Underlying Principles of the Key Catholic Literature on AI

Across these documents, spanning a Vatican City State decree, a joint dicastery note, a papal encyclical, and multiple apostolic messages, a coherent theological and ethical argument emerges based on four core principles.

The first principle is that the human person possesses an irreducible dignity. Created in the image of God the human person possesses a dignity that no machine can replicate. Intelligence is not merely computation, it is grounded in conscience, freedom, relationality, and the capacity for love (Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education, 2025).

The second principle is that AI is a tool, not an agent. Artificial intelligence however sophisticated is a human artefact. It can assist, augment, and extend human capacities. However it cannot replace human judgement, moral responsibility, or the wisdom of the heart. To treat AI as an autonomous moral actor is to abdicate human responsibility (Francis, 2024a; Francis, 2024b; Leo XIV, 2026b).

The third principle is that technology must serve the common good. AI development must be governed by ethics oriented toward benefiting every person, not merely the powerful or the wealthy. This requires transparent governance, international cooperation, and a commitment to leaving no one behind digitally (Pontifical Academy for Life, 2020; Francis, 2024a; Leo XIV, 2026b).

The fourth principle, introduced by *Magnifica Humanitas*, is that AI systems are never morally neutral. The use of every technical tool embodies choices and priorities in what it measures, ignores, and optimises, and in how it classifies people and situations. The design of a system already contains a moral orientation, which organisations have a duty to examine before adoption. Ethical discernment cannot be limited to asking whether AI is being used for good

purposes; it must also scrutinise the vision of the human person embedded in the system's data and models (Leo XIV, 2026b, §104). This principle provides the doctrinal foundation for the procurement and supply chain assurance requirements of the proposed Ethical AI Governance Framework, and for the requirement that Catholic organisations assess, not merely deploy, the AI systems they procure.

3. Mercy Health's AI Strategy Overview

Mercy Health's Artificial Intelligence Strategy 2025–2028 was approved by the Mercy Health Board of Directors in 2025 and represents the first comprehensive AI governance framework in the organisation's history. Its development was driven by a recognition that the exponential growth of AI in healthcare had created governance obligations that required a more comprehensive strategic approach.

3.1 Governance Architecture: The eight pillars

The Mercy Health AI Strategy establishes a clear vision for AI at Mercy Health; ***The effective, safe, and ethical use of Artificial Intelligence for the benefit of our patients, residents, clients, and staff.***

The AI Strategy establishes a governance architecture of eight strategic pillars based on the Australian Institute of Company Directors (AICD) AI Governance framework (AICD, 2024). The AICD is Australia's peak membership body for company directors. It provides advocacy and credentialing for company boards and is broadly equivalent to the UK's Institute of Directors or the US National Association of Corporate Directors. The strategic pillars are:

1. Roles and responsibilities
2. People, skills & culture
3. Governance structures
4. Principles, policies & strategy
5. Practices processes & controls
6. Stakeholder engagement & impact assessment
7. Supporting infrastructure
8. Monitoring, reporting & evaluation

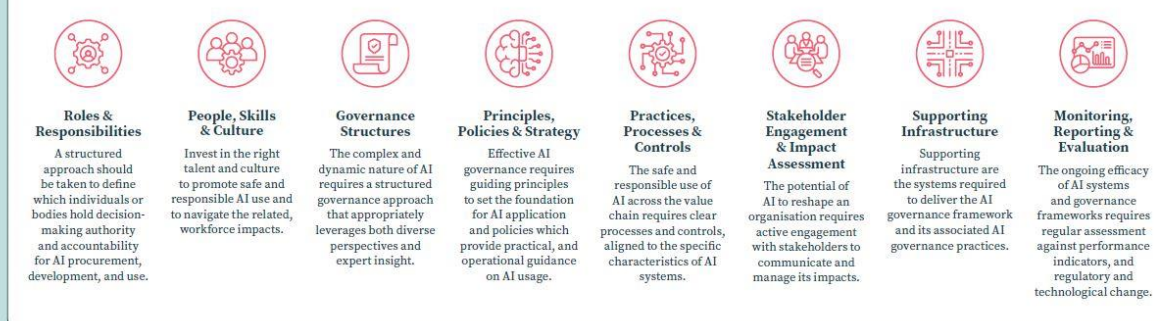
Each pillar has clear objectives, measurable outcomes, and is underpinned by an annual key activity work plan. In addition to the pillars, the AI Strategy adopts the Rome Call for AI Ethics' six principles to guide strategic and day-to-day AI decision making. These principles are: ***Transparency; Inclusion; Responsibility; Impartiality; Reliability; Security and Privacy.*** Each of the eight strategic pillars is mapped to the Rome Call's six ethical principles ensuring that Catholic ethical commitments are embedded in every governance, operational, and monitoring activity in the Strategy.

MERCY HEALTH AI STRATEGY: OUR VISION, KEY ACTIVITY PILLARS AND FOUNDATIONAL PRINCIPLES

A CLEAR VISION

The effective, safe, and ethical, use of Artificial Intelligence for the benefit of our patients, residents, clients, and staff.

WITH EIGHT STRATEGIC PILLARS TO GUIDE INVESTMENT



FOUNDATION PRINCIPLES



Mercy Health AI Strategy summary illustrating the vision, 8 strategic pillars, and 6 foundations

3.2 Foundation Principles: Direct Adoption of the Rome Call

The Strategy adopts the Rome Call's six principles as its foundation principles, expanding them with operational definitions. This is a deliberate strategic choice. Rather than developing its own specific AI principles and then checking their alignment with the Rome Call, Mercy Health adopted the Catholic Church's framework as its own. In doing so it treated Catholic AI ethics as the intrinsic requirement of Mercy Health's governance commitments. The Strategy's Concept Map within the Strategy explicitly names the Rome Call for AI Ethics and *Antiqua et Nova* as the two primary ethical inputs to the Strategy's design (Mercy Health Australia, 2025a, p.8).

4. Mercy Health's Journey to Rome Call Signatory

4.1 The Application Process

The Rome Call for AI Ethics is administered by the RenAIssance Foundation operating under the Pontifical Academy for Life. Organisations seeking signatory status submit a Letter of Commitment in which they (a) describe their organisation and its mission; (b) articulate their motivation for joining; (c) make concrete commitments to apply the Rome Call's six principles; and (d) provide brief closing remarks (see Appendix A). The application is submitted electronically through the RenAIssance Foundation's digital platform.

Mercy Health submitted its initial Letter of Commitment in September 2025. The letter made five commitments: (1) formally adopting the Rome Call principles in Mercy Health's AI Policy;

(2) educating staff in the Rome Call principles; (3) partnering with technology providers that adhere to Rome Call principles; (4) requiring rigorous AI Committee evaluation before any AI adoption, with all suppliers declaring Rome Call compliance; and (5) conducting a retrospective analysis of existing AI use to harmonise with Rome Call principles.

Reflecting the substantially expanded Vatican AI literature and the enhancement areas identified through subsequent analysis (see Section 6), Mercy Health submitted an updated Letter of Commitment in May 2026 (Jurman, 2026a).

4.2 Australian Government AI Ethics: Compatibility Analysis

Any Catholic organisation operating in Australia must navigate two AI ethics frameworks simultaneously; the Rome Call and the Australian Government's AI Ethics Principles (Department of Industry, Science and Resources, 2019). The Australian Government framework consists of eight principles;

- 1. **Human, Societal and Environmental Wellbeing:** *AI must deliver net benefit to people, society and the environment.*
- 2. **Human-Centered Values:** *AI must respect human rights, dignity, autonomy and democratic values.*
- 3. **Fairness:** *AI must be inclusive, accessible and free from unfair discrimination*
- 4. **Privacy Protection and Security:** *AI must uphold privacy rights and be secure against attack.*
- 5. **Reliability and Safety:** *AI must operate safely, reliably, and as intended across its lifecycle.*
- 6. **Transparency and Explainability:** *AI decisions must be explainable to those affected and to regulators.*
- 7. **Contestability:** *People must be able to challenge significant AI decisions.*
- 8. **Accountability:** *Responsibility for AI outcomes must be clearly assigned to identifiable humans.*

A comparative analysis of the two frameworks conducted as part of Mercy Health's AI Strategy development (see Figure 2 below) identified a strong alignment with one noteworthy difference (Jurman, 2025b).

Rome Call Principle	Australian Government Principle	Alignment	Key Difference
Transparency	Transparency & Explainability (6)	Strong - both require AI decisions to be explainable to those affected	Australia adds disclosure to people affected by AI decisions; Rome Call emphasises developer openness.
Reliability	Reliability & Safety (5)	Strong - both require consistent, safe operation throughout lifecycle	Australia explicitly names safety testing; Rome Call adds adversarial robustness focus.

Security & Privacy	Privacy Protection & Security (4)	Strong - near-identical scope covering data governance and security	Rome Call frames privacy as an expression of human dignity, not just compliance.
Impartiality	Fairness (3)	Strong - both address bias and discrimination	Australia emphasises equal access to AI benefits across groups, whereas the Rome Call's principle of impartiality focuses on preventing bias in AI outputs.
Responsibility	Accountability (8) + Human-Centered Values (2)	Strong - both require identifiable human accountability	Rome Call frames as moral responsibility; Australian Government focuses on lifecycle identifiability. Overlap with the Rome Call with the requirement that AI respect human autonomy and that humans retain meaningful control.
Inclusion	Human-Centered Values (2) + Fairness (3) + Human, Societal & Environmental Wellbeing (1)	Partial - both address equity and rights Partial - Rome Call does not name environmental wellbeing as a distinct principle	Rome Call explicitly names developing nations and global justice; Australia focuses on individuals, communities or groups. Australian Government framework contains an explicit obligation to consider the environmental footprint of AI systems. The Rome Call contains no such direct requirement, and this would appear to be contrary to the priorities of <i>Laudato Si</i> (2015) and <i>Magnifica Humanitas</i> (2026).
(No equivalent)	Contestability (7)	Gap - Rome Call has no procedural challenge mechanism	Currently the Rome Call does not create procedural rights for individuals on the receiving end of AI outputs. In regulated care environments such as health and aged care where decisions can have serious consequences for vulnerable people, the absence of contestability represents a material governance gap.

This difference appears in the Australian Government's principle of contestability which requires that people have a timely and accessible means to challenge significant AI decisions. This principle has no Rome Call equivalent. For a Catholic healthcare organisations this procedural right is particularly relevant. Patients affected by AI-assisted clinical decisions have a right of challenge that must be built into governance processes.

Endorsing the Rome Call satisfies all but one of the Australian Government Principles. Every principle in the Australian Government's framework has a direct or near-direct equivalent in the Rome Call. An organisation that genuinely implements the Rome Call will also be compliant with the Australian Government's AI Ethics Principles in all but the area of contestability.

Mercy Health has addressed these requirements within its AI Strategy by establishing an AI appeal process to enable AI generated results/outcomes to be challenged by stakeholders. This appears as part of Strategic Pillar 6 "Stakeholder Engagement & Impact Assessment". This is further supplemented by an escalation protocol for high risk AI issues (Pillar 1), external assurance reviews incorporated into the Internal Audit Plan (Pillar 5), and a risk-based monitoring and reporting system for mission-critical AI systems (Pillar 8).

For a Catholic healthcare organisation operating within Australia, both frameworks are relevant and mutually reinforcing. The key insight is that endorsing the Rome Call does not mean choosing it over Australia's framework, it means going further. Endorsing the Rome Call for AI Ethics has positioned Mercy Health as aligned with the Australian Governments AI Ethics Principles. It has signalled that Mercy Health governs AI not merely to comply with government guidelines, but because the responsible use of technology is a moral obligation rooted in the dignity of every person it serves.

4.3 The Rome Call's Impact on the Mercy Health AI Strategy

The Rome Call is deeply integrated into the Mercy Health AI Strategy. The Strategy's Concept Map, Foundation Principles, Principle Alignment Indicators across all eight pillars, Platform Selection Rationale, and Signatory Commitment in two separate pillars all reflect a governance framework that was built *from* the Rome Call outward rather than adding the Rome Call as an ethical overlay to a pre-existing governance structure.

The most significant governance innovation in the AI Strategy is the treatment of AI supply chain governance. Inspired by the framework in the Australian Government Modern Slavery Act 2018 (Cth) under which Mercy Health already conducts its annual supply chain due diligence through the Australian Catholic Anti-Slavery Network, the Strategy requires all AI suppliers to declare their compliance with the Rome Call's principles.

Mercy Health's commitment and rationale are clearly expressed in their Rome Call Letter of Commitment, which states: *'Our commitment to the Rome Call mirrors our vigilance against modern slavery - by scrutinising our supply chains and joining with other organisations as part of the Australian Catholic Anti-Slavery Network, we pursue justice and a future in which our staff and patients can flourish. The ethical challenge of AI is no less important'* (Jurman, 2025a). This statement highlights that a Rome Call commitment that does not extend to the supply chain cannot be held with integrity.

Such an approach is also consistent with the requirements outlined in *Magnifica Humanitas* (Leo XIV, 2026b, §§173-179) which states that *'the supply chains that underpin the technological industry and the digital economy need to become more transparent, so that no competitive advantage is built upon hidden exploitation'* and that *'companies and investors need to adopt clear criteria for preventive ethical verification (due diligence), placing among their priorities the protection of workers, the fight against forced labour and the assessment of the social impact of data-driven business models'* (Leo XIV, 2026b, §179).

5. Catholic Values, the Rome Call, and the Governance Gap in AI: An Opportunity for Policy Discourse and Influence

The preceding sections described Mercy Health's response to a challenge that is common to every organisation that has engaged with the Rome Call: the gap between the framework's principles and the operational governance architecture required to deliver them. Such a gap presents an opportunity for the Catholic Church to exert substantive influence on global AI governance. To illustrate this, a case study of the AI company Anthropic will be used.

5.1 The Principles-Practice Gap

The Rome Call's five-year history reveals a structural tension that is common to all voluntary principles-based governance framework: the gap between stated commitments and operational implementation. Currently organisations can join the Rome Call and be listed as signatories without any verification of their governance practices, or any structured accountability mechanism, or any requirement to report on progress. As the signatory list grows, so does the risk of what might be termed 'algorithethics washing' or the association with a prestigious ethical framework without the substance to support it.

Pope Francis identified this risk (Francis, 2024a): *'If it is to be at the service of integral human development, this major transformation of the technological world requires that those responsible assume an ethical responsibility for its effects.'* According to ethical principle 6 of the Rome call, 'Responsibility', those who design and deploy AI must proceed with responsibility and transparency. However, without accountability mechanisms responsibility cannot be demonstrated; it can only be asserted. This governance gap is the space in which the Catholic Church with its deep tradition of ethics, globally credible voice on human dignity, and experience in building governance frameworks for complex social missions is uniquely positioned to lead.

5.2 A Practical Case Study: Anthropic

The most recent notable example of the Catholic Church's potential to influence AI governance comes from Anthropic. Anthropic is an American artificial intelligence company and developer of the Claude large language models. In January 2026, Anthropic published *Claude's Constitution* a comprehensive document establishing the values, ethical commitments, and

decision-making frameworks for its AI systems (Askell et al., 2026). This document explicitly acknowledged two Catholic contributors among its external reviewers: Father Brendan McGuire, a former technology executive and co-founder of the Institute for Technology, Ethics and Culture, a partnership between Santa Clara University and the Vatican's Dicastery for Culture and Education, and Bishop Paul Tighe, Secretary of the Pontifical Dicastery for Culture and Education and a senior Vatican official engaged with technology ethics (Askell et al., 2026; Wachtel, 2026).

The commitments outlined in the document were not merely symbolic. In February 2026, Anthropic lost its USD 200 million contract with the US Department of Defence after it refused to comply with a request for unrestricted access to Claude 'for all lawful uses', and for the removal of safety restrictions banning the use of Claude for mass domestic surveillance of American citizens or for use in fully autonomous lethal weapons systems without a human in the loop. Both prohibitions correspond directly to the Vatican positions in *Antiqua et Nova*, in the Holy See's submission to the 2026 UN Conference on Disarmament in Geneva calling for a moratorium on lethal autonomous weapons, and in Pope Leo XIV's message for the 59th World Day of Peace, in which he warned that delegating '*decisions about life and death*' to machines marks '*an unprecedented and destructive betrayal*' of the principles of humanism (Leo XIV, 2026c). In response, President Trump directed US government agencies on 27th February 2026 to cease working with Anthropic. Defence Secretary Pete Hegseth designated Anthropic a 'supply chain risk', a label previously reserved for foreign adversaries which effectively blacklisted the company from federal contracting. Anthropic filed suit on 9th March 2026. In a public statement on 26th February, CEO Dario Amodei said that '*frontier AI systems are simply not reliable enough to power fully autonomous weapons*' and that Anthropic would '*not knowingly provide a product that puts America's warfighters and civilians at risk*'.

On 13 March 2026, fourteen Catholic moral theologians, philosophers and social ethicists filed an *amici curiae* brief in support of Anthropic, stating the company '*was acting as a responsible and moral corporate citizen*' (OSV News, 2026). The substantive argument was co-authored by Charles Camosy (Catholic University of America), Joseph Vukov (Loyola University Chicago), Brian J.A. Boyd and Brian Patrick Green (Santa Clara University). The brief grounded its arguments in Catholic teaching on privacy, subsidiarity, and just war theory, arguing that *jus in bello* requires particular human judgment on proportionality and discrimination that autonomous weapons eliminate by definition. The scholars noted that their moral position was more strict than Anthropic's, with the Catholic scholars opposing lethal autonomous weapons categorically because human judgment is intrinsic to the moral conditions of a just war (OSV News, 2026).

The Anthropic case demonstrates something the Rome Call's architects perhaps hoped for but could not have predicted. This is that Catholic ethical frameworks, articulated with intellectual rigor and embedded in institutional practice, can shape the values architecture of the most widely deployed AI systems in the world. The acknowledgement of Father McGuire and Bishop Tighe as external reviewers of Claude's Constitution is evidence that Catholic ethical thinking is part of that conversation.

5.3 The Church's Opportunity: Policy Discourse and Structural Influence

The Anthropic case is not unique but is a visible example of a broader pattern of technology companies, regulators, and governments actively seeking ethical frameworks with philosophical coherence, and institutional credibility to anchor AI governance. The Catholic Church, through the Rome Call and the development of a body of Vatican AI guidance is positioned to provide precisely this.

This positioning is not accidental. The Church brings to the AI conversation what no government framework, international standard, or corporate ethics document can provide. A two-thousand-year tradition of sustained reflection on the human person, the nature of intelligence, the ethics of technology, and the obligations of institutions to the vulnerable. To quote Wachtel (2026), *'Why is Silicon Valley turning to the Catholic Church for AI ethics? Because the Church has been thinking about the human person, the common good, and the limits of technology for two thousand years. In a field where everyone is improvising, that tradition is suddenly very valuable'*. This is further reinforced by Pope Leo XIV's framing (Leo XIV, 2025) that the Church responds to AI as it responded to the labour question in *Rerum Novarum* which changed the governance of labour in the twentieth century, shaping trade union law, social security systems, and workplace rights across the democratic world.

Magnifica Humanitas sharpens this into a practical evaluative question that Catholic organisations can apply directly. Leo XIV frames the entire AI challenge through two biblical images: Building the Tower of Babel, in which technology serves uniformity, and the concentration of power, or rebuilding Jerusalem, in which technology serves community, shared responsibility, and the dignity of every person (Leo XIV, 2026b, §§7–10). The question the encyclical poses to every institution is not *'does this AI system work?'* but *'which city is this organisation building?'* For Catholic healthcare, where the carer-patient relationship is the irreducible core of the ministry, the Babel/Jerusalem distinction is not abstract theology. It is the governance question behind every decision to automate, to delegate to an algorithm, or to reduce clinical judgement to a data point. Mercy Health's Rome Call commitment and governance framework are in these terms a deliberate choice to build Jerusalem.

The question for Catholic organisations is not whether to engage with the AI governance conversation, but how to engage with sufficient rigour to be credible. A Letter of Commitment to the Rome Call is a starting point, not an endpoint. What the Church and Catholic organisations need and what Mercy Health's experience suggests is a governance architecture that operationalises the Rome Call's principles in ways that can be verified, reported, and independently assessed is necessary.

6. The Rome Call II: Turning Principles to Practice for Catholic Organisations

This section introduces the concept of the 'Rome Call II'. This is a two-document architecture that constitutes a potential model for a complete AI governance model for Catholic organisations.

The first document, the Rome Call for AI Ethics itself, is the source of the principles and the institutional anchor that connects an organisation's AI governance to global Catholic ethics.

The second document, the Ethical AI Governance Framework for Catholic Organisations (attached as Appendix B), is the operational architecture that translates the Rome Call's principles into specific governance requirements, board obligations, accountability mechanisms and assessment tools. The central argument is simple. The Rome Call establishes the '*what*' of Catholic AI ethics. The Framework establishes the '*how*.' Neither document is sufficient alone.

6.1 Ten opportunities for enhancement of the Rome Call

A systematic analysis of the Rome Call framework and its application process, conducted as part of Mercy Health's AI Strategy development, identified ten potential opportunities for further enhancement across the two categories (Jurman, 2025c; Jurman, 2026b). The first category addresses current limitations of the Rome Call as a governance framework. The second addresses current limitations in the current Rome Call application process.

6.2 Rome Call Framework: Areas for Potential Enhancement

Enhancement Area 1: Principles without operational architecture. The Rome Call establishes six principles but provides no governance architecture for their implementation. Signatories commit to principles but receive no guidance on the governance structures, policies, controls, and accountability mechanisms required to deliver them.

Enhancement Area 2: Theological grounding not operationalised. The Rome Call is deliberately non-denominational. For Catholic organisations, this means the deepest rationale for the principles of the *imago Dei*, the theology of intelligence, and the irreducible human moral responsibility affirmed in *Antiqua et Nova* are absent from the framework's text.

Enhancement Area 3: Supply chain blind spot. The Rome Call places obligations on those who 'design and deploy' AI. But most Catholic organisations deploy AI primarily through third-party products. There is no obligation to assess whether suppliers AI systems embody the same principles. This gap is analogous to the problem the Australian Modern Slavery Act 2018 (Cth) was designed to address.

Enhancement Area 4: Lack of environmental wellbeing as a distinct principle. Under the current framework there is no requirement to specifically consider the environmental footprint of AI systems. This would appear to be contrary to the priorities of *Laudato Si'* (2015) and *Magnifica Humanitas* (2026).

Enhancement Area 5: No verification or accountability mechanism. There is no audit, no independent assessment, no structured follow-up, and no accountability for non-performance. As the signatory list grows, so does the risk of 'algorithethics washing' or organisations claiming ethical alignment without the governance architecture to substantiate it.

Enhancement Area 6: Static framework not reflecting the evolved Vatican AI literature. The Rome Call was issued in 2020. The Vatican has since issued multiple AI documents including a

papal encyclical. The current Rome Call application process does not require engagement with this substantially enhanced guidance.

Enhancement Area 7: Contestability gap. Individuals and communities who are materially affected by AI-driven decisions should have access to a meaningful and timely mechanism to question, challenge, or seek review of those decisions and the systems that produced them. The Rome Call does not require signatories to establish such procedural rights for people on the receiving end of AI outputs. In regulated care environments such as health and aged care, where AI-informed decisions can carry serious consequences for vulnerable people, the absence of any contestability obligation represents a material governance gap.

6.3 Rome Call Application Process: Areas for Potential Enhancement

Enhancement Area 8: Commitments not mapped to the six principles. The current Rome Call template provides open-ended commitment rows with no requirement to address each of the six Rome Call principles, allowing partial compliance to satisfy the letter but not the spirit of the requirement.

Enhancement Area 9: No accountability or reporting section. The template contains no field for describing how the organisation will report on progress, despite the Rome Call explicitly requiring signatories to 'commit to report the initiatives undertaken according to the principles of transparency and accountability.'

Enhancement Area 10: No renewal or review mechanism. An organisation that submitted its letter in 2020 has no formal obligation to update its commitments, regardless of the substantially changed AI environment or the expanded Vatican framework of 2026.

6.4 The Ethical AI Governance Framework as the Response

The proposed Ethical AI Governance Framework for Catholic Organisations was designed to address precisely these areas for potential enhancement. It builds on the approach developed for the Mercy Health AI Strategy and the Australian Institute of Company Directors Framework (AICD, 2024) and adds a ninth element called 'AI Supply Chain Assurance'. This is a direct response to the requirement of *Magnifica Humanitas* that companies adopt 'preventive ethical verification (due diligence)' and supply chain transparency so that no competitive advantage is built on hidden exploitation (Leo XIV, 2026b, §179). The Framework provides a nine-element governance architecture that translates each Rome Call principle into specific, documented requirements applicable to Catholic healthcare, aged care, education, and social services organisations.

The Framework's nine elements are: (1) Roles and Responsibilities; (2) Governance Structures; (3) People, Skills and Culture; (4) Principles, Policies and Strategy; (5) Practices, Processes and Controls; (6) Supporting Infrastructure; (7) Stakeholder Engagement; (8) Monitoring, Reporting and Evaluation; and (9) AI Supply Chain Assurance. Each element contains a Catholic theological overview grounding the requirement in Catholic Social Teaching, specific governance requirements, board-level actions, key governance questions, and risk indicators.

The Framework's most significant innovation from a governance accountability perspective is the self-assessment tool. It transforms Rome Call adherence from a binary (signatory / non-signatory) status to a continuous, measurable governance maturity spectrum with four scoring levels per question and a scoring interpretation framework. The tool is designed to be updated as Catholic AI literature develops.

A Catholic organisation that completes the self-assessment tool could provide the RenAIssance Foundation with a structured progress report; could demonstrate to its Board and external stakeholders that its Rome Call commitment is verifiable; and could identify specific governance investment priorities aligned with Rome Call adherence rather than general risk management.

Conclusion

The Rome Call for AI Ethics is a landmark achievement. It is the first time the Catholic Church has convened technology companies, governments, and civil society around a shared ethical framework for artificial intelligence. Five years on, its significance is not diminished. But the gap between its aspirational principles and the operational governance architecture required to deliver them is visible. In the six years since the Rome Call was signed, the Vatican has developed the most coherent and authoritative institutional response to AI governance produced by any major global institution: from the Rome Call's foundational principles (2020) through the binding Decree No. DCCII (2025) and the theological capstone of *Antiqua et Nova* (2025), to the first full social encyclical of the AI age, *Magnifica Humanitas* (2026). That trajectory from principles to law to theology to encyclical represents a depth and continuity that no government standard or corporate ethics document can match.

Mercy Health Australia's experience demonstrates that this gap can be closed, and that Catholic organisations are uniquely positioned to close it. The combination of the Church's two-thousand-year tradition of moral anthropology, the Vatican's evolving body of AI literature from Decree No. DCCII and *Antiqua et Nova* to the encyclical *Magnifica Humanitas*, and the practical governance innovation of the Ethical AI Governance Framework for Catholic Organisations, provides a complete architecture for responsible AI governance in Catholic healthcare.

The Anthropic case study suggests that this architecture matters not only for Catholic organisations but for the broader AI governance conversation. Technology companies are actively seeking ethical frameworks with intellectual depth, institutional credibility, and genuine governance substance. The Catholic Church can provide what no government standard or corporate ethics document can, a tradition of thinking about the human person, the common good, and the limits of technology that the AI age is discovering it desperately needs.

The 'Rome Call II' consisting of the Rome Call for AI Ethics combined with an Ethical AI Governance Framework for Catholic Organisations is Mercy Health's contribution to making that tradition operational. It is offered not as a proprietary model but as a starting point for the wider community of Catholic organisations navigating the same challenge to govern AI with the rigour that human dignity demands.

References

- AICD (Australian Institute of Company Directors) (2024) A Director's Guide to AI Governance. Sydney: AICD.
- Askell, A. et al. (2026) Claude's Constitution. San Francisco: Anthropic. Available at: anthropic.com (Accessed: 1 May 2026).
- Commonwealth of Australia (2018) Modern Slavery Act 2018 (Cth). Canberra: Commonwealth of Australia.
- Department of Industry, Science and Resources (2019) Australia's AI Ethics Principles. Canberra: Australian Government. Available at: www.industry.gov.au (Accessed: 1 May 2026).
- Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education (2025) Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence. Vatican City: Libreria Editrice Vaticana.
- Francis, Pope (2024a) Artificial Intelligence and Peace: Message for the 57th World Day of Peace, 1 January 2024. Vatican City: Libreria Editrice Vaticana.
- Francis, Pope (2024b) Artificial Intelligence and the Wisdom of the Heart: Towards a Fully Human Communication: Message for the 58th World Day of Social Communications. Vatican City: Libreria Editrice Vaticana.
- ISO (International Organisation for Standardisation) (2023) ISO/IEC 42001:2023: Information Technology - Artificial Intelligence - Management System. Geneva: ISO.
- Jurman, P. (2025a) Letter of Commitment: Joining the Rome Call for AI Ethics. To the RenAIssance Foundation, c/o Pontifical Academy for Life, 3 September 2025.
- Jurman, P. (2025b) Rome Call for AI Ethics and Australia's AI Ethics Framework: A Comparative Analysis. Richmond, Victoria: Mercy Health Australia. Internal Document.
- Jurman, P. (2025c) Rome Call Letter of Commitment Template: Review and Recommendations for Improvement. Richmond, Victoria: Mercy Health Australia. Internal Document.
- Jurman, P. (2026a) Letter of Commitment: Joining the Rome Call for AI Ethics - Updated May 2026. To the RenAIssance Foundation, c/o Pontifical Academy for Life, 11 May 2026.
- Jurman, P. (2026b) From Aspiration to Accountability: The Rome Call for AI Ethics - Framework Shortcomings, Application Process Gaps, and the Catholic Governance Framework Response. Richmond, Victoria: Mercy Health Australia. Internal Document.
- Leo XIII, Pope (1891) Rerum Novarum: On Capital and Labour. Vatican City: Libreria Editrice Vaticana.
- Leo XIV, Pope (2025) Address to the College of Cardinals Following Papal Election, 10 May 2025. Vatican City: Libreria Editrice Vaticana.
- Leo XIV, Pope (2026a) Address to the Academic World at the Catholic University of Central Africa, Yaoundé, 17 April 2026. Vatican City: Libreria Editrice Vaticana.
- Leo XIV, Pope (2026b) *Magnifica Humanitas: Encyclical Letter on Safeguarding the Human Person in the Time of Artificial Intelligence*, 15 May 2026. Vatican City: Libreria Editrice Vaticana.
- Leo XIV, Pope (2026c) 'Peace Be with You All: Towards an Unarmed and Disarming Peace': Message for the 59th World Day of Peace, issued 8 December 2025 [for 1 January 2026]. Vatican City: Libreria Editrice Vaticana.
- Mercy Health Australia (2025a) Artificial Intelligence Strategy 2025–2028. Richmond, Victoria: Mercy Health Australia.
- Mercy Health Australia (2026) Ethical AI Governance Framework for Catholic Organisations. Richmond, Victoria: Mercy Health Australia.

-
- NIST (National Institute of Standards and Technology) (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: NIST.
- OSV News (2026) 'Catholic moral theologians, ethicists back Anthropic in government AI showdown', OSV News, 16 March. Available at: www.osvnews.com (Accessed: 1 May 2026).
- Pontifical Academy for Life (2020) Rome Call for AI Ethics. Vatican City: Pontifical Academy for Life. Available at: www.vatican.va (Accessed: 1 May 2026).
- Pontifical Commission for Vatican City State (2025) Decree No. DCCII: Guidelines on Artificial Intelligence. Issued 16 December 2024; in force 1 January 2025. Vatican City: Governorate of Vatican City State.
- Wachtel, E. (2026) 'Why Silicon Valley Is Turning to the Catholic Church', *The Atlantic*, 25 April. Available at: www.theatlantic.com (Accessed: 1 May 2026).
- Wagle, S. (2025) Rome Call for AI Ethics: A Comprehensive Guide. Available at: www.romecall.org (Accessed: 1 March 2025).

AI tool usage declaration:

I acknowledge the use of AI specifically Microsoft Copilot and Anthropic Claude (Sonnet 5) in the preparation of this paper, in particular during the drafting and editing process. For example, drafts were submitted to the AI tools with the instruction to 'improve the academic style and accuracy of language', to 'summarise and simplify', and to 'improve structure, flow, and tone'. Drafts were then reviewed so as to suit my own style and to ensure that the words expressed my own arguments and conclusions. All arguments and conclusions are my own. Additionally, during the research phase I used these tools to challenge my own assumptions and hypotheses.

Appendices

Appendix A:

Rome Call Application Questionnaire.

[To be put on letterhead]

To the **RenAIssance Foundation**

c/o Pontifical Academy for Life

Via della Conciliazione 1

00120 Città del Vaticano

Joining the Rome Call for AI Ethics

Letter of Commitment (Organisations)

Background	
<i>Please explain your organisation's previous experience in the field of AI Ethics. Please mention any relevant activities, actions, initiatives undertaken (e.g. events organised, paper published, results obtained, good practices realised, etc.) – Max 2000 characters.</i>	
Motivation	
<i>Please explain the reasons that led your organisation to apply: why and how joining the Rome Call will add value to your mission? – Max 1000 characters.</i>	
Commitment	
<i>Please describe the concrete commitments your organisation intends to make to support the implementation of the Rome Call for AI Ethics' principles:</i>	
• <i>Transparency</i> • <i>Inclusion</i> • <i>Responsibility</i> • <i>Impartiality</i> • <i>Reliability</i> • <i>Security and privacy</i>	
To foster the implementation of the Rome Call for AI Ethics, my organisation intends to:	
Commitment 1	<i>Describe one commitment for each row. Where possible, please include a timing for reaching it.</i>
Commitment 2	

Commitment X	
<i>Please add other rows as required</i>	
Closing remarks	
<i>Please add your closing remarks – Max 500 characters</i>	

Sincerely,

Please add the signature here: _____

Role within the organisation: _____

Date/place: _____

Appendix B:

Ethical AI Governance Framework for Catholic Organisation

Audience	Boards, Executive Leaders, Mission Leaders, and AI Governance Practitioners in Catholic organisations.
Grounded in	Rome Call for AI Ethics (2020); Antiqua et Nova (2025); Vatican City State AI Commission Decree No. DCCII (2025); Magnifica Humanitas (2026); ISO/IEC 42001:2023; NIST AI RMF 1.0 (2023)
Scope	Any Catholic organisation deploying, or considering deploying, artificial intelligence systems in service of its mission

PURPOSE AND HOW TO USE THIS FRAMEWORK

The Rome Call for AI Ethics established six foundational principles for AI governance: Transparency, Inclusion, Responsibility, Impartiality, Reliability, and Security and Privacy. This Framework translates those commitments into nine operational governance elements for Catholic organisations. The Master Mapping Matrix below links each element to the principles it serves. A 29-question self-assessment tool enables boards to evaluate governance maturity and prioritise investment. Element 9 extends obligations through the AI supply chain, modelled on the Australian Government Modern Slavery Act 2018 (Cth) reporting framework.

MASTER MAPPING MATRIX

The table below maps each of the nine governance elements to the Rome Call principles they primarily serve (●) and those they support secondarily (○).

Governance Element	Transparency	Inclusion	Responsibility	Impartiality	Reliability	Security & Privacy
1. Roles & Responsibilities	○		●			
2. Governance Structures	○	○	●		○	
3. People, Skills & Culture	○	●	●			
4. Principles, Policies & Strategy	●	●	●	●	●	●
5. Practices, Processes & Controls			●	●	●	●
6. Supporting Infrastructure	●				●	●
7. Stakeholder Engagement & Impact	●	●	●	●		
8. Monitoring, Reporting & Evaluation	●		●	●	●	
9. AI Supply Chain Assurance	●	●	●	●	●	●

THE NINE GOVERNANCE ELEMENTS

ELEMENT 1

ROLES & RESPONSIBILITIES

Establishing clear accountability for AI across the organisation

Responsibility

Transparency

GOVERNANCE REQUIREMENTS

- **Accountability mapping** - a documented record specifying which individual or body holds decision-making authority for each AI system at board and management level.
- **AI inventory** - a structured centrally maintained register of all AI systems covering purpose, data sources, ongoing costs, risk assessments, and named accountable owner.
- **Board-level oversight** - the board retains overall accountability for AI governance even where day-to-day responsibility is delegated to management.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- If the board asked today for a complete list of every AI system in use across the organisation, could management produce it and would it include AI embedded in third-party products?
- When an AI system causes harm to a patient, resident, or client, is it clear within the organisation who is responsible for the immediate response and the governance review?
- Does the organisation have a named executive whose mandate explicitly includes AI governance and is this reflected in their position description and performance objectives?

⚠️ AMBER - Monitor This Area

- AI accountability is held entirely within the technology function. No executive or board member has a formally documented AI governance mandate.
- The organisation knows it uses some AI systems but has not attempted to map their full scope including AI embedded within procured software and cloud services.
- No escalation framework distinguishes AI incidents requiring operational response from those requiring executive notification or board-level decision.

❌ RED - Address This Risk

- The organisation cannot identify who is ultimately accountable for AI systems that effect the individuals and communities they serve.
- AI knowledge and system documentation resides with individual staff members rather than in formal governance records. The loss of those individuals would leave the organisation exposed.
- AI procurement decisions are made without reference to the organisation's governance framework, mission commitments, or Rome Call obligations.

ELEMENT 2

GOVERNANCE STRUCTURES

Building the institutional architecture for ethical AI oversight

Responsibility

Transparency

Reliability

GOVERNANCE REQUIREMENTS

- **Formal AI governance mandate** - at least one board committee and one management level body must carry explicit documented responsibility for AI governance in their terms of reference.
- **Cross-functional representation** - governance bodies must include mission and ethics perspectives alongside legal, operational, data, cybersecurity, risk, and human resources expertise.
- **Regular board reporting** - management must report on AI governance to the board on regular defined intervals, not only in response to incidents.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- Does any board committee have explicit documented authority over AI use. If so has it exercised that authority through a formal review, approval, or inquiry in the past twelve months?
- Does the management level AI governance body include a representative of the organisation's mission and Catholic identity function not just technology, risk, and legal?
- If the board receives no AI-related report for twelve months does any governance instrument require one to be produced regardless?

⚠️ AMBER - Monitor This Area

- Terms of Reference for committees with AI responsibility were drafted before current AI capabilities existed and have not been reviewed.
- The AI governance body is composed predominantly of technology staff. Mission, ethics, clinical, and legal perspectives are absent or marginal.
- External AI ethics expertise is engaged reactively when a problem arises not as a standing and scheduled governance input.

❌ RED - Address This Risk

- No board committee has ever formally reviewed an AI use case, incident, or AI governance report of any kind.
- All AI governance decisions are made at management or technical level. The board has no defined entitlement to receive AI governance information.
- The organisation has no external AI ethics input at any level including board, management, or advisory.

ELEMENT 3

PEOPLE, SKILLS & CULTURE

Building the human capability and culture required for ethical AI

Inclusion

Responsibility

Transparency

GOVERNANCE REQUIREMENTS

- **AI literacy baseline** — every person in the organisation should be able to answer three questions about their own role: which AI tools am I expected to use, what are the main risks those tools carry for the people we serve, and what do I do if something goes wrong?
- **Tiered training programme** — differentiated training for board members (governance and risk focus), executives (operational and ethical), and frontline staff (safe use, incident reporting, escalation).
- **Workforce impact assessment** — a formal assessment of AI's impact on the workforce covering employment, skills requirements, role changes, and staff wellbeing.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- Can each board member name at least one AI system the organisation currently operates and articulate the primary governance risk it presents to those in the organisation's care?
- Is there a documented pathway by which a frontline worker can escalate an AI concern to a decision-maker? Is there evidence this pathway is used?
- If a staff member identified that an AI tool was producing outputs that seemed discriminatory or unsafe, what would happen next — and does every relevant staff member know this?

⚠️ AMBER — Monitor This Area

- AI literacy for board members and senior executives has not been formally assessed, and no programme exists to address identified gaps.
- Staff have raised AI concerns informally, but no structured channel exists that guarantees these concerns reach a decision-maker and are acted upon.
- Workforce planning has not been revisited in light of the organisation's current or planned AI deployments and their effects on roles and skills.

❌ RED — Address This Risk

- No member of the executive team can name the AI systems the organisation currently operates, or describe the main risks they carry for the individuals and communities they serve.
- There is documented reluctance among staff to raise concerns about AI system failures, based on prior experience or the prevailing organisational culture.
- AI is deployed to frontline workers without any training, oversight framework, or safe-use protocols.

ELEMENT 4

PRINCIPLES, POLICIES & STRATEGY

Embedding Rome Call ethics in organisational policy and strategy

Transparency

Inclusion

Responsibility

Impartiality

Reliability

Security & Privacy

GOVERNANCE REQUIREMENTS

- **Formal AI principles aligned with the Rome Call** - the organisation's AI principles must address each of the six Rome Call principles grounded in Catholic Social Teaching and the organisation's mission.
- **Comprehensive AI use policy** - operational guidance translating each Rome Call principle into specific rules that staff can follow in practice. Not just aspirational statements.
- **Policy integration** - AI policy must be aligned and integrated with existing privacy, data governance, cybersecurity, and procurement policies.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- If asked today to open the AI principles document and point to exactly how each Rome Call principle is operationalised in practice, could the organisation do so or would the document reveal aspirational language with no enforceable content underneath it?
- When a question arises about a new AI use case which policy document is the definitive authority and does every relevant manager know how to apply it?
- Has the AI strategy ever been presented to the board with a clear account of how each proposed use case serves the Catholic mission. Or has the strategy been driven by vendor presentations and operational efficiency arguments without reference to values?

⚠️ AMBER - Monitor This Area

- The AI principles document contains Rome Call commitments in aspirational language but provides no operational guidance. Staff cannot determine from the document what they are and are not permitted to do in practice.
- The AI policy was drafted by the technology team without input from mission leadership, pastoral care, or frontline clinical governance representatives.
- The AI policy addresses only systems the organisation builds or directly commissions. It contains no provisions for AI embedded in procured third-party products where most AI risk actually sits.

❌ RED - Address This Risk

- The organisation has no AI policy, no formal AI principles, and no written commitment to any ethical framework governing how AI is used in the delivery of services.
- The organisation has signed or is considering the Rome Call but has not connected that commitment to any operational policy, strategy, or governance decision.
- AI strategy decisions are being made at management level without board awareness or approval and without reference to the organisation's Catholic identity and values.

ELEMENT 5

PRACTICES, PROCESSES & CONTROLS

Translating ethical commitments into operational enforcement

Responsibility

Impartiality

Reliability

Security & Privacy

GOVERNANCE REQUIREMENTS

- **Mandatory AI Impact Assessments** - required before all AI solution deployments with analysis proportional to risk level and explicit assessment of impacts on vulnerable populations.
- **AI-specific risk framework** - risk appetite and risk management processes updated to address AI specific risks including bias, performance degradation, model drift, cybersecurity, and ethical harm.
- **Bias testing and monitoring** - a standard control applied to all AI systems, with documented results and remediation plans where bias is identified.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- If a new AI tool is proposed, for example an AI-assisted wound assessment tool or a predictive model for readmission risk, what is the specific process for evaluating and approving its use, and who within the governance structure has the authority and the mandate to say no?
- For each high risk AI system currently operating, is there documented evidence that a responsible person reviewed and took accountability for outcomes across the most recent decision cycle. Or is 'human in the loop' a policy statement with no operational mechanism behind it?
- When was bias testing last conducted on AI systems affecting clinical care, care planning, or service allocation. Can management produce the test results and any remediation actions taken?

⚠️ AMBER - Monitor This Area

- AI Impact Assessments have been completed for some deployments but not others and the threshold for when an assessment is required has never been formally defined or communicated.
- Human review is a policy requirement but is not enforced by any workflow control. An AI assisted decision could be acted upon without human sign-off, and no system would flag the omission.
- Bias testing was conducted at initial deployment for some systems but there is no ongoing monitoring program, no defined re-testing cycle, and no documented remediation history.

❌ RED - Address This Risk

- No AI Impact Assessment has ever been conducted for any AI system currently in use by the organisation in the delivery of services, or employment decisions.
- There is no defined process by which a staff member can report an AI related safety or ethics concern, and no documented investigation of an AI concern has ever taken place.
- High risk AI assisted decisions are being acted upon without any defined or enforced requirement for human review.

ELEMENT 6

SUPPORTING INFRASTRUCTURE

Building the technical foundations for ethical AI in practice

Security & Privacy

Transparency

Reliability

GOVERNANCE REQUIREMENTS

- **AI system and data inventory** - a comprehensive, current register of all AI systems, data sources, model types, performance metrics, and risk assessments.
- **Data governance framework updated for AI** - policies covering data lineage, bias in training data, right to use, data minimisation, retention, purging, and consent with particular attention to sensitive personal information.
- **AI-specific cybersecurity controls** - security policies updated to address AI-specific threats including model poisoning, adversarial attacks, prompt injection, and data exfiltration through AI interfaces.
-

KEY QUESTIONS FOR GOVERNANCE LEADERS

- Could the Chief Information Officer or equivalent produce, within 24 hours, a list of every AI system in operation across the organisation including AI embedded in third-party products along with the sensitivity classification of the data each system accesses?
- For the three highest risk AI systems currently in operation, has the cybersecurity team specifically assessed AI-specific attack vectors such as model poisoning, adversarial inputs, prompt injection? Or was the security review conducted using a framework written before these risks were understood?
- If those we serve asked 'why did the system recommend this for me?' in relation to an AI assisted decision is there a documented answer the organisation could provide. Or is the system's reasoning genuinely unavailable?

⚠️ AMBER - Monitor This Area

- An AI and data inventory exists but is known to be incomplete. Several AI enabled vendor products have not been assessed and the inventory is reviewed irregularly rather than on a defined cycle.
- The cybersecurity policy addresses standard IT threats but has not been updated to account for AI specific attack vectors identified since the last security framework review.
- Data governance consent frameworks pre-date the organisation's AI deployments and have not been assessed for whether they adequately cover AI use of personal and sensitive information.

✖️ RED - Address This Risk

- No AI inventory exists. The organisation cannot identify which AI systems it operates, what data those systems access, who is accountable for them, or what their performance characteristics and known risks are.
- Cybersecurity has never specifically assessed AI attack surfaces for any system the organisation operates or relies upon.
- Data from those we serve is being used in AI systems for which no adequate consent framework exists and no consent review has been conducted.

ELEMENT 7

STAKEHOLDER ENGAGEMENT & IMPACT ASSESSMENT

Giving voice to all affected communities, especially the most vulnerable

Inclusion

Impartiality

Transparency

Responsibility

GOVERNANCE REQUIREMENTS

- **Stakeholder mapping** - a formal map of all parties directly and indirectly affected by the organisation's AI systems with explicit identification of vulnerable and marginalised groups.
- **Structured engagement process** - a formal documented process for engaging stakeholders in AI governance that goes beyond consultation to genuine participation where the stakes are high.
- **Vulnerable group assessment** - AI Impact Assessments must explicitly assess and document impacts on vulnerable and marginalised populations served by the organisation.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- Can the organisation name a specific consultation that was conducted with those they serve before deploying an AI system that directly affects them, including what was asked, what was heard, and what changed as a result of that engagement?
- If those we serve wanted to know whether and how AI was being used in decisions and wanted to challenge a recommendation, what would happen, step by step, and is that pathway written down and known to the people they would approach?
- Has the organisation assessed whether any AI system in use is less accessible, less accurate, or less equitable for specific populations including First Nations people, people from culturally and linguistically diverse backgrounds, people with disability, or people experiencing socioeconomic disadvantage?

⚠️ AMBER - Monitor This Area

- Stakeholder engagement for AI governance has been conducted but has not meaningfully reached the populations most directly affected, for example aged care residents or patients from First Nations or culturally and linguistically diverse communities, relying instead on organisational representatives or professional advisers speaking on their behalf.
- An appeal mechanism exists for service decisions but it has not been reviewed to confirm it applies to AI assisted decisions and affected individuals are not informed of their right to use it.
- AI Impact Assessments include a section on vulnerable populations but it is completed by internal staff without reference to direct consultation with those communities.

❌ RED - Address This Risk

- The organisation has never sought input from the communities it serves about any AI system that affects the services they receive, or the decisions made about them.
- No individual served by the organisation has any formal mechanism to request an explanation of an AI-assisted decision, to challenge it, or seek redress when it has caused them harm.
- AI deployments directly affecting patients, residents, or clients have been approved on the basis of internal assessments only with no structured engagement with affected populations at any stage of the approval process.

ELEMENT 8

MONITORING, REPORTING & EVALUATION

Sustaining Rome Call adherence through continuous evidenced oversight

Responsibility

Reliability

Transparency

Impartiality

GOVERNANCE REQUIREMENTS

- **Rome Call aligned KPIs** - formal performance indicators linked to each of the six Rome Call principles measured and reported on a defined cycle.
- **Regular AI system performance monitoring** - ongoing monitoring of all material AI systems for performance degradation, model drift, bias emergence, and safety incidents.
- **Board reporting cadence** - management must report AI governance performance to the board on a defined schedule not merely in response to incidents.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- Can the board or governance committee point to a specific KPI for each of the Rome Call's six principles and show the last three data points as evidence of an active, evidenced monitoring programme rather than a policy commitment that has never been measured?
- When AI governance performance is reported to the board, does the report include system performance data, bias monitoring results, an incident summary, and an assessment against Rome Call commitments or does it describe policy and process status without demonstrating actual outcomes?
- Has the organisation identified which AI systems are most critical to its mission and are those systems subject to the most frequent and rigorous performance monitoring?

⚠️ AMBER - Monitor This Area

- AI governance reporting reaches the board but is not structured around the Rome Call principles and does not include bias monitoring results, incident summaries, or system performance trend data. As a result it is not possible to assess whether governance commitments are being upheld in practice.
- KPIs exist for some AI systems but are not connected to the Rome Call framework so the organisation cannot demonstrate with evidence that any specific ethical principle is being actively maintained.
- The AI incident register captures events but post-incident learning is not systematically actioned, and there is no documented record of governance improvements derived from incident review.

❌ RED - Address This Risk

- There is no defined board reporting cadence for AI governance. The board has not received substantive information about AI risk or performance in the past 12 months, or has never received it at all.
- No KPIs exist for responsible AI use. The organisation cannot demonstrate with any evidence that a single Rome Call principle is being upheld in operational practice.
- Material AI incidents or bias events have occurred in the past 12 months but have not been reported to the board or to any governance committee.

ELEMENT 9

AI SUPPLY CHAIN ASSURANCE

Extending Rome Call obligations through procurement and supplier relationships

Transparency

Inclusion

Responsibility

Impartiality

Reliability

Security & Privacy

GOVERNANCE REQUIREMENTS

- **Supply chain AI identification** - a systematic process for identifying all suppliers that use AI in products or services provided to the organisation embedded in procurement and supplier on-boarding processes.
- **Supply chain AI risk register** - a documented register of supplier AI systems, their risk level, which Rome Call principles they implicate, and the assurance status for each relationship.
- **Supplier Rome Call Assurance Declaration** - a formal written declaration from material suppliers confirming AI use governance framework and adherence to Rome Call principles.

KEY QUESTIONS FOR GOVERNANCE LEADERS

- If the board asked management today for a list of the top ten suppliers whose AI use poses the greatest risk to the people the organisation serves, could management produce that list with supporting risk assessments in time for the next board meeting?
- Has the organisation sought written confirmation from any material supplier about the AI systems embedded in their products, how those systems make decisions, whether they have been bias-tested, and what governance framework the supplier applies. Can it produce that confirmation?
- When a supplier materially changes the AI components of a product in use by the organisation, is there a mechanism ensuring the organisation is notified, and can re-assess its Rome Call compliance before the change takes effect?

⚠️ AMBER - Monitor This Area

- The organisation has begun identifying AI use in its supply chain but relies on supplier self disclosure in response to ad-hoc queries rather than a systematic mapping process embedded in procurement.
- Procurement teams include data protection and cybersecurity requirements in supplier contracts, but no AI governance requirements, bias testing expectations, or Rome Call alignment commitments appear in any standard contract template.
- Supplier AI assurance has been sought from some high risk vendors but there is no standard declaration form, no defined minimum standard for adequacy, and no follow up process where assurance is incomplete or unsatisfactory.

❌ RED - Address This Risk

- The organisation has no visibility of AI use by any supplier and has taken no steps to obtain this information despite the fact that AI is embedded in products directly affecting service delivery.
- No supplier contract entered into by the organisation includes any reference to AI governance obligations, ethical AI requirements, or Rome Call principles.
- No formal Supplier Rome Call AI Declaration has been sought from any supplier, no standard form has been developed, and no plan exists to seek or develop one.

SELF-ASSESSMENT TOOL

The following 29-question self-assessment enables Catholic organisations to evaluate their current AI governance maturity against each of the nine Framework elements. For each question, rate the organisation's current position on a four-point scale:

3 Fully in place and evidenced	2 Partially in place	1 Planned but not yet implemented	0 Not in place
---------------------------------------	-----------------------------	--	-----------------------

A total score of 87 represents full implementation across all nine elements. Scores should be reviewed by the board or relevant committee and used to prioritise governance investment in the areas of greatest risk to Rome Call adherence and Catholic mission integrity.

No.	Self-Assessment Question	Rome Call Principle(s)	Score (0–3)
Element 1: Roles & Responsibilities			
1.1	We have a complete current AI inventory that assigns named accountability for every AI system which is reviewed.	Responsibility, Transparency	
1.2	Accountability for AI is distributed across the leadership team not concentrated in technical specialists with defined escalation protocols.	Responsibility	
Element 2: Governance Structures			
2.1	Our board committee charters explicitly include AI governance oversight with defined terms of reference.	Responsibility, Transparency	
2.2	We have a cross functional AI governance body at management level including mission and ethics representation alongside technical and risk functions.	Responsibility, Reliability	
Element 3: People, Skills & Culture			
3.1	Board members and executives have sufficient AI literacy to exercise meaningful governance oversight and ask substantive questions of management.	Responsibility	
3.2	Staff are equipped to identify and safely report AI concerns or failures and there is evidence they feel safe doing so.	Responsibility, Inclusion	
3.3	We have formally assessed and communicated to our workforce the ways in which AI affects their roles and the organisation.	Inclusion	
Element 4: Principles, Policies & Strategy			

4.1	Our AI principles explicitly address each of the six Rome Call principles and connect them to Catholic mission and Social Teaching.	All six principles	
4.2	Our AI use policy provides specific enforceable operational rules rather than aspirational statements without practical guidance.	All six principles	
4.3	Our privacy, data governance, cybersecurity, and procurement policies are integrated with our AI policy and address AI specific requirements.	Security & Privacy, Reliability	
Element 5: Practices, Processes & Controls			
5.1	AI Impact Assessments are completed for all proposed AI solutions including explicit assessment of impacts on vulnerable and marginalised groups.	Responsibility, Reliability, Impartiality	
5.2	Bias testing is a standard control applied to all AI systems with documented results and remediation plans.	Impartiality	
5.3	Human in the loop requirements are formally defined for high risk AI decisions affecting those we serve.	Responsibility	
5.4	We formally assess and manage AI risk in supplier and vendor products as part of procurement due diligence.	Reliability, Security & Privacy	
Element 6: Supporting Infrastructure			
6.1	Our data governance framework has been updated to address AI specific requirements including the sensitivity of personal data processed by AI systems.	Security & Privacy, Reliability	
6.2	Our cybersecurity framework explicitly addresses AI specific threats, including model poisoning, prompt injection, and data exfiltration.	Security & Privacy	
6.3	Our AI system architecture supports meaningful explanation of AI assisted decisions to those we serve.	Transparency	
Element 7: Stakeholder Engagement & Impact			
7.1	We have formally mapped all stakeholders affected by our AI systems including vulnerable and marginalised groups.	Inclusion, Impartiality	

7.2	Affected individuals have a formal mechanism to request explanation, or to appeal AI assisted decisions that affect them.	Responsibility, Transparency	
7.3	Our AI tools have been assessed for accessibility across literacy levels, languages, disability statuses, and technology access.	Inclusion	
Element 8: Monitoring, Reporting & Evaluation			
8.1	We have KPIs aligned to each of the six Rome Call principles, tracked and reported to the board on a defined cycle.	All six principles	
8.2	Management reports AI governance performance to the board on a defined, regular schedule, not only in response to incidents.	Responsibility, Transparency	
8.3	We have commissioned and acted on independent assurance of our AI governance practices.	Responsibility, Reliability	
8.4	We maintain an AI incident register with documented post incident reviews and evidence of learning fed back into governance.	Reliability, Responsibility	
Element 9: AI Supply Chain Assurance			
9.1	We have conducted a systematic mapping of AI use across our supply chain and recorded findings in our AI inventory.	Responsibility, Transparency	
9.2	Our supplier contracts include explicit Rome Call aligned AI governance requirements and a right to audit compliance.	All six principles	
9.3	Material suppliers have provided a formal Supplier Rome Call AI Declaration confirming their AI governance practices.	Responsibility, Transparency	
9.4	We have a defined and documented escalation and remediation process for suppliers unable to demonstrate adequate Rome Call alignment.	Responsibility, Reliability	
9.5	We publish, or have a plan to publish, an AI Supply Chain Assurance Statement analogous to our modern slavery statement.	Transparency, Responsibility	
TOTAL SCORE (Maximum: 87)			/ 87

Score Interpretation

74–87	Strong Rome Call governance alignment	The organisation can credibly maintain its Rome Call commitment including through the supply chain. Priority: maintain and demonstrate through self assessment and independent assurance.
58–73	Good foundation with targeted gaps	The organisation has meaningful governance in place but identifiable gaps require focused investment. Priority: address lowest scoring elements.
29–57	Significant gaps requiring priority attention	Investment in governance is needed before public commitments can be sustained with integrity. Priority: treat the lowest scoring elements as immediate governance risks.
0–28	Foundational governance work required	The organisation's Rome Call commitment should be treated as aspirational until material progress is made. Priority: establish basic governance architecture before advancing to detailed requirements.

Framework Citation

Mercy Health Australia (2026) Ethical AI Governance Framework for Catholic Organisations. Richmond, Victoria: Mercy Health Australia.

For enquiries regarding use of this framework, contact: pjurman@mercy.com.au.

Bioethics Outlook

A quarterly publication of the Plunkett Centre for Ethics

The Plunkett Centre for Ethics is a joint centre of the Australian Catholic University,
St Vincent's Health Network (Sydney)

Calvary Healthcare and Mercy Hospital Melbourne

Contact: plunkett@acu.edu.au