

The Pope, the Algorithm, and the Stewards of Capital

Reflections on *Magnifica Humanitas* and the Epworth Ethics Framework

by Rev Dr Andrew Harper

On 25 May 2026, Pope Leo XIV presented his first encyclical, *Magnifica Humanitas*, in person, breaking with Vatican tradition. Standing alongside him was Christopher Olah, co-founder of the AI laboratory Anthropic and one of the pioneers of mechanistic interpretability, the field that asks whether human beings can understand what is actually happening inside the AI systems they are building. The choreography was unusual, and it was meant to be. Leo signed the document on 15 May, 135 years to the day after his predecessor Leo XIII signed *Rerum Novarum*, the encyclical that became the foundation stone of modern Catholic Social Teaching. When a pope speaks at his own encyclical's launch, and chooses to stand beside a leader in the AI space to do it, he is telling the world that what is in the document is urgent.

For Epworth's clients, *Magnifica Humanitas* is more than a curiosity of ecclesiastical news. Our Ethics Framework already engages explicitly with Catholic Social Teaching through *Mensuram Bonam*, the 2022 Vatican document on faith-based measures for Catholic investors, and we work closely with Catholic organisations to apply these principles in practice. We read the new encyclical, then, from within a shared tradition rather than from outside it. This note attempts to do what the moment seems to demand: to tease out what Leo is really saying, to read it alongside our seven pillars, and to consider what is at stake for the capital we are entrusted to steward. There is a great deal at stake here. The encyclical's diagnosis is that humanity is at a turning point in its understanding of itself, and that the choices made by states, by companies, and by investors in this decade will shape what kind of creature we become.

The deeper claim: an anthropological emergency

The single most important sentence in the encyclical is the one Leo XIV repeats in different forms throughout the text: 'The challenge is not technological, but

anthropological.’ This is not a pastoral flourish. It is the document’s load-bearing claim, and unless it is taken seriously the rest of the encyclical reads as a series of disconnected warnings. Taken seriously, it reorganises everything.

What Leo is arguing is that AI is not, fundamentally, a problem about safety benchmarks, alignment techniques, or even market concentration. Those are real, and the encyclical addresses them. But the underlying question is what kind of being a human being is, and what happens when machines begin to perform, simulate, or substitute for the activities through which human beings have historically known themselves: cognition, creativity, judgment, relationship, communication. The Church has a name for this examination of what humans are: anthropology, in its theological sense. Catholic teaching holds that human beings are made in the *imago Dei*, the image of God, and that this image is not reducible to rationality or productive capacity. It includes creativity, vulnerability, relationship, and the irreducible uniqueness of each person. AI systems that simulate these capacities are therefore not merely technically interesting; they are touching something the encyclical considers sacred.

Several of the encyclical’s most striking arguments rotate around this axis. Leo XIV writes that we are ‘experiencing an eclipse of the sense of what it means to be human’ and that ‘by simulating human voices and faces, wisdom and knowledge, consciousness and responsibility, empathy and friendship, the systems known as artificial intelligence not only interfere with information ecosystems, but also encroach upon the deepest level of communication, that of human relationships’ (§14). This is a categorical claim, not a harm-reduction argument. It is not that synthetic media enables fraud (though it does); it is that the medium of human encounter is itself being colonised. The deepfake problem in this framing is not really about misinformation. It is about whether faces and voices, which Leo elsewhere calls ‘unique, distinctive features of every person that reveal a person’s own unrepeatable identity,’ can continue to function as the signs through which human beings recognise one another at all.

In Chapter Three the encyclical confronts transhumanism directly, naming it as the ideological horizon toward which an unchecked AI trajectory tends. The

transhumanist promise, in its purest form, is the elimination of human limitation: disease, aging, cognitive constraint, eventually death itself. Leo XIV refuses this promise. He warns against ‘the prospect of a technology that promises to free us from all weakness’ (§12) and argues that human fulfilment comes not through eliminating weakness but through what the tradition calls integral development: growth through difficulty, relationship, and the exercise of distinctively human capacities. The claim, in plain terms, is that vulnerability is constitutive of personhood. A being that cannot suffer cannot love in the way persons love. An AI that simulates care without vulnerability is offering a simulation of care, and the difference between the two is the kind of difference on which human social life depends. This is the encyclical's deepest philosophical move, and it is the one most likely to be missed by readers who treat the document as a technology policy paper.

The argument about children carries the same anthropological weight in a more practical register. Leo XIV has said that he is concerned for ‘the freedom and inner life of our children and young people’ and for AI's potential effect on their ‘intellectual and neurological development.’ To a stadium of teenagers, he said: use AI ‘in such a way that if it disappeared tomorrow, you would still know how to think.’ This is not Luddism. It is a claim that certain cognitive capacities, namely sustained attention, original synthesis, the willingness to sit with difficulty rather than delegate it, are not fixed endowments but muscles that atrophy when unused. A generation that never has to think through a problem because a machine will always do it for them may grow up without capacities that are constitutive of what we have meant by being human.

Finally, on warfare, the encyclical's language is the language of moral emergency. Leo XIV speaks of a ‘spiral of annihilation’ in which AI in military applications is not an optimisation to be managed but an escalation dynamic to be condemned. The expected call for a categorical ban on lethal autonomous weapons systems that operate without meaningful human oversight is grounded in the same anthropology: decisions to end human life cannot be delegated to systems that cannot understand the moral weight of what they are deciding. An autonomous

weapon processes targets. It cannot know that the person in its sights is unrepeatable. The encyclical argues that this incapacity is not a bug to be engineered away; it is a reason to prohibit the class of action entirely.

Where the framework already speaks Leo's language

Read at this depth, *Magnifica Humanitas* recognises in advance many of the convictions Epworth has been articulating across its seven pillars.

Epworth's Labour pillar holds that 'all people have the right to employment at a living wage and in conditions that do not violate their human dignity or deny their opportunity to flourish.' Leo extends this directly into the AI economy, observing that AI systems are 'putting much of the human creative industry at risk of being dismantled and replaced,' turning people into 'passive consumers of unthought thoughts and anonymous products without ownership or love.' The displacement question is no longer hypothetical; it is the central labour question of the next decade, and our framework's emphasis on human dignity in work positions us to engage with it on theological as well as economic ground.

Our Society pillar already names 'AI and automation' and the 'surveillance economy' as areas of concern, and the encyclical gives this concern a deep doctrinal foundation. Leo's extension of the universal destination of goods to 'patents, algorithms, digital platforms, technological infrastructure and data' (§67) is a remarkable theological move: the building blocks of the AI economy are being treated as analogous to common natural resources. When these are concentrated in a few hands, he argues, a new imbalance is created that contradicts the principle that the earth's goods are given for all. Our existing concerns about market concentration in technology now have an explicit doctrinal warrant.

Our Fairness, Responsibility and Transparency pillar identifies algorithmic decision-making and AI ethics as governance challenges. The encyclical's argument here is sharper still. The presence of Christopher Olah at the launch was not incidental. Olah's work in mechanistic interpretability is the technical face of an argument the encyclical is making in moral terms: governance of AI is fictional if humanity cannot understand the systems it is governing. The Vatican's framing was explicit:

‘understanding machines is not merely an engineering problem.’ It is a question of epistemic sovereignty, of whether human beings retain the capacity to direct the tools they build. Our framework’s emphasis on transparent governance and independent oversight maps directly onto this, and we should be willing to escalate engagement with companies whose AI systems are operationally opaque, even where the technology itself appears beneficial.

Our Conflict pillar holds that ‘all people and companies should work actively toward peace’ and already excludes companies involved in controversial weapons. The encyclical’s emphasis on lethal autonomous weapons strengthens the case for extending this scrutiny into dual-use AI systems, particularly those marketed for defence applications without clear human-in-the-loop safeguards. Our existing framework anticipates this; the encyclical sharpens its application.

What the encyclical leaves for the investor to resolve

For all its weight, the encyclical leaves several questions either unanswered or answered in ways that do not quite map onto how active stewards of capital should operate.

The first is the question of register. Leo’s prescriptions, including the call to ‘disarm’ AI, the emphasis on regulation, and the warning about ‘a race for ever more powerful algorithms... driven by the desire to secure geopolitical or commercial dominance,’ sit in the register of public policy rather than capital allocation. The encyclical addresses legislators, regulators, and ‘all men and women of goodwill,’ but it has little direct guidance for the investor who must decide whether to hold a particular technology stock. Our framework’s preference for engagement over exclusion (voting, dialogue, and escalation before divestment) is calibrated to the position we actually occupy as a minority shareholder. The encyclical speaks in a prophetic register addressed to states and societies; the investor must also act within a narrower remit, where walking away from a company is sometimes right but almost always cedes influence to a less ethically engaged shareholder.

The second question is one of discrimination. The encyclical’s framing risks collapsing distinct issues into a single warning. Autonomous weapons, surveillance

regimes, generative AI's impact on labour markets, the data privacy of children, and transhumanist ideology are all genuine concerns, but they are also analytically separate problems requiring different stewardship responses. Our framework treats them as such, and a thoughtful application of the encyclical requires us to maintain that discrimination rather than letting a prophetic mode authorise a blanket retreat.

The third question is perhaps the most important. Our framework's standalone exclusion for artificial intelligence is, on closer reading, more discriminating than the encyclical's broad posture might suggest. We do not exclude AI as a technology. We exclude companies 'engaged in the development or deployment of artificial intelligence where there is inadequate governance to safeguard human rights, dignity, and safety.' This is a governance test, not a technology test. There is a profound implication here that bears stating clearly. The encyclical, read at its deepest, is not asking investors to abandon the AI sector. It is asking us to discriminate, with theological seriousness, between companies that are building this technology with genuine regard for the human person and those that are not. That discrimination is the work the framework was built for. It is also work that, if done well, may have more consequence than divestment.

What this means for portfolios

The practical implications follow from the framing. *Magnifica Humanitas* reinforces a direction Epworth was already travelling, and it sharpens several engagement themes that are already active in our voting and stewardship work.

First, governance and interpretability. We will press AI developers harder on transparent governance, independent oversight, and the technical question of whether their systems can be meaningfully audited. Companies whose AI systems are operationally opaque, and whose governance frameworks treat this opacity as acceptable, are increasingly difficult to hold within our framework.

Second, labour and creative work. The displacement of human creative and knowledge work is the *Rerum Novarum* question of our age. We aim to engage portfolio companies on how they are sharing the productivity gains of AI with the

workers whose tasks the technology absorbs, and on whether their deployment of AI respects what Leo calls the dignity of the worker.

Third, weapons and lethal autonomy. Our zero-tolerance position on controversial weapons should be applied with full vigour to the emerging category of lethal autonomous systems. Dual-use AI sold into defence applications without clear human-in-the-loop safeguards warrants escalated engagement and, where appropriate, exclusion.

Fourth, children and the surveillance economy. Companies whose business models depend on capturing the attention of minors, or on extractive data practices in environments where children are present, should expect intensified scrutiny.

The encyclical also offers something less tangible but worth naming. It supplies a language (Babel and Jerusalem, the eclipse of the human, the duty to remain profoundly human) that proves useful in conversations where the technical vocabulary of ESG falls flat. Our framework already invokes John Wesley's call to 'do all the good you can.' Leo gives us complementary imagery for an age in which the most consequential investment decisions are being made about technologies whose moral significance we are still learning to articulate.

Where we should remain alert is to the encyclical's implicit invitation toward simpler answers than the situation warrants. The companies building this technology include some of the most thoughtful and some of the most reckless actors in modern capitalism, and the difference between them matters enormously. The thoughtful ones are precisely the ones engaging with the kinds of questions the encyclical raises, and capital that withdraws from them indiscriminately may do more harm than good. Discerning that difference, neither retreating from the sector nor embracing it uncritically, is the work the framework was built for. Leo XIV has given us a powerful theological articulation of what is at stake. The patient business of voting, engagement, and selective exclusion remains ours.