Governing the Algorithm: Muslim, Catholic, and Jewish Visions for Ethical AI

Dr. Tarek Cherkaoui



(TUR Presidency / Murat Kula - Anadolu Agency)

The AI revolution is not merely a technological challenge; it is a moral and civilizational one. *Governing the Algorithm* examines how Muslim, Catholic, and Jewish traditions converge on a shared ethical vision: technology must remain subordinate to human dignity, truth, and the common good. From Islamic legal concepts to papal encyclicals and Jewish theological reflection, this policy outlook explores the principles that can guide responsible AI governance. It also confronts the devastating consequences of AI-enabled warfare, exposing the failure of corporate self-regulation and the urgent need for accountable, pluralistic oversight. This paper aims to contribute to an ongoing conversation for policymakers, scholars, and all who are concerned that the future of AI be shaped by wisdom and justice, not just efficiency.

The opinions expressed in this policy outlook represent the views of the author(s) and do not necessarily reflect the views of the TRT World Research Centre.

Introduction

Artificial Intelligence is increasingly being recognised as one of the defining forces of the 21st century. Unlike previous technological revolutions, AI is not simply transforming economic productivity or industrial processes; it is reshaping how societies generate knowledge, exercise power, make decisions, and understand human agency itself.

As AI systems become more powerful, the debate surrounding them is no longer confined to engineers, entrepreneurs, and tech companies. Religious leaders, philosophers, policymakers, and civil society organisations are increasingly entering the discussion. Among the most significant interventions are those emerging from Muslim reflections on AI ethics, technological sovereignty, and human dignity, alongside Pope Leo XIV's encyclical on Artificial Intelligence. As the Yaqeen Institute for Islamic Research has argued, Muslims cannot afford to be "either uncritically enthusiastic or passively withdrawn" from this technological transformation. According to this viewpoint, the stakes are high: "if we fail to understand and shape this technology, others will define how it represents and impacts our communities".

Although these perspectives arise from different intellectual traditions, they share remarkable areas of convergence. Both reject technological determinism. Both reject the notion that AI is value-neutral. Both insist that technology must remain subordinate to human dignity, moral responsibility, and the common good. Most importantly, both challenge the increasingly influential assumption that society should simply trust technology companies to regulate themselves.

Areas of Convergence

A central insight emerging from contemporary Islamic reflection on AI is the observation that modern societies have inverted the relationship between means and meanings. We have developed extraordinary technical capabilities—AI systems that can process language, make predictions, and simulate reasoning—but we have failed to cultivate the moral and metaphysical frameworks necessary to guide their use. This reflects a broader civilizational shift: from pre-industrial societies that integrated scientific inquiry with spiritual purpose to a hyper-technological age where efficiency has become an end in itself. As academic Mohamed Gamal Abdelnour argues, this inversion 'mirrors a broader civilizational shift where technical means have vastly outpaced our capacity to generate or sustain meanings.'



(Raşit Aydoğan - Anadolu Agency)

Abdelnour further exemplified the tension between means and meanings, noting a powerful illustration in the contrasting legacies of two iconic figures: al-Khwarizmi, the father of algebra, who developed his work within the framework of Islamic jurisprudence—particularly in service of inheritance laws grounded in the Quran—and J. Robert Oppenheimer, the father of the atomic bomb, who stood at the helm of an epochal technological leap driven largely by nationalistic and militaristic pressures. Al-Khwarizmi's science was inseparable from spiritual and ethical purpose; Oppenheimer's was haunted by moral uncertainty. This contrast epitomises the shift from a cosmos ordered by meanings to a world dominated by unchecked means."

Human Dignity Above Technological Capability

Emerging Muslim perspectives frame AI as a tool that must serve humanity rather than replace human moral agency. AI may assist scholars, researchers, doctors, and policymakers, but it cannot assume responsibility for ethical judgment, accountability, or spiritual discernment. The Yaqeen Institute echoes this principle, stating that "the responsibility for truth always lies with people" and that while "AI can support our teams by speeding up work, it does not make decisions or carry moral weight."

Similarly, the Pope's encyclical affirms that "the promotion of the common good can never be separated from respect for the right of peoples to exist, to preserve their own identity and to contribute their unique qualities to the family of nations." This principle, deeply resonant with Muslim concerns about technological sovereignty, insists that AI development must serve the flourishing of distinct cultural and civilizational identities. Thus, from this perspective, technological progress should never become an end in itself. The central question is not whether AI can perform a task, but whether its deployment enhances human dignity and contributes to human flourishing. Both traditions, therefore, reject the idea that efficiency alone constitutes progress.

Meanwhile, the Jewish tradition, as explained by Rabbi Efrem Goldberg, offers a distinctive theological framing of technology's ultimate purpose: it serves as a metaphor to strengthen faith in divine providence and accountability. From this perspective, an AI system that can process and respond to billions of personalised inquiries simultaneously offers a powerful analogy for how God can know and care for every individual. Just as an app can provide personalised answers, so too can the Almighty guide

each person's unique path. This perspective enriches the interfaith conversation by demonstrating how a tradition can view the same technology not only as a tool but as a spiritual mirror that leads to contemplation on the divine attributes.

Technology Is Not Neutral

One of the strongest points of convergence concerns the nature of technology itself. Muslim scholars and AI practitioners have raised concerns that contemporary AI systems are largely developed within specific cultural, philosophical, and political environments. As a result, the values embedded within many frontier models often reflect particular Western assumptions about morality, identity, individualism, and social life. Muslim researchers note that "because most of that data is Western and secular in origin, AI often carries blind spots about Islam and Muslims."

This principle is echoed in Pope Leo XIV's first encyclical, *Magnifica humanitas*, which asserts that "technology is never neutral, because it takes on the characteristics of those who devise, finance, regulate, and use it." AI systems are not value-free instruments; they carry the cultural, political, and economic assumptions of their creators.

AI's Anti-Muslim Bias and the Crisis of Credibility

For many Muslim communities, scepticism toward AI platforms is not rooted in technophobia but in lived experience. Artificial intelligence does not need to hate you to discriminate against you; it only needs the wrong data. Across Europe and North America, facial recognition systems routinely misread darker skin tones and hijabs, border-control software flags Muslim names as potential threats, and online moderation algorithms quietly remove posts about Palestine while allowing anti-Muslim hate to circulate freely. A major peer-reviewed study found that the word "Muslim" was analogised to "terrorist" in 23% of test cases involving GPT-3—a rate far higher than for any other religious group. Follow-up research confirmed that even after debiasing attempts, these associations persist in newer models. This pattern of algorithmic discrimination has a direct consequence: Muslim communities increasingly refuse to take AI platforms at face value.

When technologies that claim neutrality consistently produce biased outcomes, the claimed neutrality itself becomes suspect. The result is not only a credibility deficit for AI systems but a deeper crisis of trust. If Muslim and Global South perspectives remain absent from the international AI governance debate, bias will continue to be framed as a technical fault rather than a political injustice. As Mutale Nkonde, CEO of AI for the People, argues: "We keep asking how to make machines smarter when we should be asking how to make systems fairer."

Concern About Concentrated Power

A major theme in both discussions is the concentration of power within a small number of technology companies. Muslim commentators warn that dependency on foreign technological infrastructures may create new forms of political, economic, and cultural vulnerability. The debate is therefore not merely about AI capabilities but about technological sovereignty and control over critical knowledge systems. Muslim researchers echo this concern, noting that "AI is already reshaping economies and military capabilities, areas where Muslim nations risk deepening dependence if they remain consumers rather than contributors."

Pope Leo XIV similarly warns against technological monopolies, excessive concentration of wealth, and the emergence of an unaccountable technological elite capable of shaping public life without democratic oversight. The Pope also draws attention to the hidden costs of the digital revolution, warning against "new forms of slavery" embodied in the "scarred, injured, and worn down" bodies of those mining rare earth elements. This is not only a matter of economic exploitation, but a profound ethical challenge that both Muslim and Catholic social thought insist must be confronted.

This concern about concentrated power extends beyond economics and governance to the realm of spiritual and moral authority. Jewish thinkers have warned of a "ChatGPT god"—a phenomenon in which AI's extraordinary intelligence, creativity, and responsiveness might lead individuals to attribute to it metaphysical qualities that belong only to the Divine. When technology companies control the systems that billions consult for answers, guidance, and even meaning, the concentration of power becomes not only political but deeply theological. The Islamic tradition, in its concern for *shirk* (associating partners with God), similarly warns against mistaking human creation for a source of ultimate truth. These overlapping concerns reveal a profound interfaith consensus: the governance of AI cannot be separated from the governance of trust, and both must remain accountable to communities and traditions, not merely to market forces.

Türkiye's recent AI Action Plan (2026-2030) exemplifies this emerging push for technological sovereignty. President Recep Tayyip Erdoğan has framed the initiative as a matter of national security, stating that "political, military, and economic power can no longer be considered separately from digital sovereignty." The plan aims to mobilise over $10 billion in investment for data centres and AI infrastructure, train 100,000 AI professionals, and develop indigenous large language models such as "Bilge" to reduce dependency on Western platforms.

This national ambition is accompanied by a sharp critique of Big Tech monopolies. Selçuk Bayraktar, the CTO of Turkish drone manufacturer Baykar, has accused global tech

cartels of "trapping billions of users in algorithmic loops designed to optimise screen time through neurological exploitation." He has called on developing nations to "build independent, transparent, and auditable open-source ecosystems" to counter this concentration of power. These arguments resonate with the interfaith concerns documented in this paper, underscoring that the governance of AI is not only a matter of regulation but of reclaiming digital agency from unaccountable corporate actors.

Work, Meaning, and Human Purpose

Both perspectives reject purely economic understandings of labour. Muslim reflections emphasise that human beings derive meaning not merely from consumption or productivity but from service, responsibility, family, community, and spiritual purpose. The Pope similarly argues that work provides purpose, identity, belonging, and participation in the common good. Large-scale AI-driven unemployment, therefore, raises moral concerns that cannot be measured solely through economic indicators.

The concern in both Muslim and Catholic traditions is that AI could create a society that is materially wealthier but spiritually and socially poorer.

AI and Warfare

Both traditions express serious concern regarding the militarisation of AI. Muslim commentators have highlighted the growing role of AI in surveillance systems, target identification, military intelligence, and modern warfare. The Pope similarly warns against autonomous weapons, AI-enabled warfare, mass surveillance, and the delegation of life-and-death decisions to machines.

In both cases, the sanctity of human life remains a central ethical principle that technological innovation must not undermine.

The Gaza Precedent: When Precision Promises Collapse

The ethical concerns raised by both Muslim scholars and Pope Leo XIV regarding AI-enabled warfare are not hypothetical. They have been tested and catastrophically failed in Gaza. Proponents of AI-driven military technology have long argued that algorithmic precision would reduce civilian casualties. The logic appeared compelling: machines do not tire, rage, or seek revenge. They process data dispassionately, identify targets with mathematical accuracy, and, in theory, minimise harm to non-combatants. According to this view, governments face a moral imperative to adopt such technology if it can demonstrably protect civilian populations.

Israel's genocidal campaign in Gaza has exposed this promise as an illusion.

The Gospel and Lavender Systems

During its military operations, the Israeli military deployed at least two AI-powered systems. The first, known as Gospel, was designed to analyse vast datasets and rapidly generate targets for infrastructure strikes, including schools, medical facilities, and aid organisation offices. The second, Lavender, went further. It produced "kill lists" targeting individual operatives, with a tracking software called "Where's Daddy?" ensuring targets could be located and struck, typically at night and inside their own homes.

The human toll was immediate and devastating. In the first month of the war alone, more than 6,120 fatalities belonged to just 1,340 families—many of which were completely wiped out while inside their residences. Even minors were reportedly flagged by Lavender as targets, despite the system's intended focus on individuals over the age of 17.

The Mathematics of Mass Casualties

What makes Lavender particularly chilling is not the system itself but the rules of engagement built around it. Investigative reporting has revealed extraordinarily loose thresholds for civilian harm. Strikes on low-level Hamas members permitted the killing of 10 to 20 civilians. Attacks targeting senior commanders deemed up to 100 civilian deaths acceptable collateral damage.

At a 10% error rate deemed acceptable by the Israeli military, Lavender may have falsely identified approximately 3,700 innocent individuals in Gaza as Hamas members. If each strike on those individuals resulted in up to 20 civilian deaths, the cumulative effect could reach an estimated 77,700 innocent non-combatants inadvertently harmed by inaccurate AI assessments (Abraham, 2025). This is not precision warfare. It is a mass assassination factory.

A Systemic, Not Individual, Failure

As Neta Crawford has argued, collateral damage should not be viewed as an unavoidable consequence of conflict but as a significant moral dilemma, particularly when it could have been foreseen and mitigated through careful planning. Responsibility extends beyond individual soldiers and commanders to the broader institutional framework of warfare itself.

The Gaza precedent demonstrates that AI does not automatically reduce civilian harm. It can, under the wrong governance, accelerate and amplify it. The technology itself is not the problem. The problem is the political and military framework within which it is deployed. When AI is integrated into a system with loose rules of engagement, weak accountability, and a demonstrated willingness to treat civilian death as an acceptable line item, the promised precision becomes a weaponised excuse for mass killing.

This reality vindicates the warnings issued by both Muslim scholars and Pope Leo XIV: technology must never be trusted to govern itself, and ethical oversight cannot be outsourced to the same institutions that benefit from military contracts. Gaza is not an exception. It is a warning.

The Minab Test: When AI Ethics Collide with Military Reality

The ethical principles outlined above are not abstract philosophical exercises. They are being tested and failing in real time. On February 28, 2026, a US Tomahawk cruise missile struck a girls' elementary school in Minab, Iran. The casualty estimates range from 120 to 168 people, the vast majority children under the age of 12.

The strike was not random. It was enabled by AI systems, including Anthropic's Claude model, integrated with Palantir's Maven military platform—a system designed for rapid target analysis and processing. AI helped sort through data to identify and evaluate targets. One of those targets turned out to be a school full of children.

On June 10, 2026, more than three months after the strike, Anthropic CEO Dario Amodei told Bloomberg that Claude's involvement did not cross Anthropic's self-imposed "red lines." His defence rested on a narrow distinction: human operators retained ultimate decision-making authority. Claude did not pull the trigger. It merely processed information within a system where humans held final authority. The murder of more than 100 schoolgirls, Amodei suggested, was a "use case that doesn't even violate our red lines."

Anthropic's publicly stated ethical boundaries include prohibitions on mass domestic surveillance and fully autonomous lethal weapons. The Minab strike, by Amodei's framing, violated neither. Claude was not making autonomous kill decisions, and the surveillance prohibition applies only to domestic operations, not foreign military campaigns.

The response to Minab exposes a profound failure of the self-regulatory model that both Muslim scholars and Pope Leo XIV have warned against. When technology companies serve as sole arbiters of their own ethical boundaries—when "red lines" are drawn by the same actors profiting from military contracts—the result is a system capable of processing a school strike as a permitted "use case."

If Claude's involvement in the killing of more than 100 children does not cross Anthropic's red lines, then those red lines are not ethical boundaries. They are marketing documents. And the CEOs who defend them while smiling through interviews are not stewards of responsible AI. They are technofeudal war criminals who have outsourced moral accountability to a distinction without a difference.

The Four Broad Positions on Artificial Intelligence

As political commentators Alastair Campbell and Rory Stewart observed on *The Rest is Politics*, the contemporary debate on artificial intelligence can be broadly understood through four competing approaches: the technological inevitability position, the trust-technology-companies model, the human-centred democratic governance framework, and the technological moratorium perspective. To these four, one might add a distinctively Islamic position: conditional permissibility. As the Yaqeen Institute summarises, "the Islamic ruling on artificial intelligence is one of conditional permissibility," meaning that AI is "permissible by default but subject to strict ethical guidelines and Islamic guardrails." This position rejects both blanket acceptance and blanket rejection, demanding instead contextual ethical assessment based on intention, benefit, and harm.

1. Technological Inevitability

This position argues that AI development cannot realistically be stopped. Supporters contend that technological progress is inevitable and that societies should focus primarily on adaptation, competitiveness, and maximising economic benefits. The emphasis is on accelerating innovation while managing disruption.

2. Trust the Technology Companies

This position argues that the companies building AI are best placed to manage its risks. Under this view, firms such as OpenAI, Anthropic, Google, Meta, and others should largely regulate themselves through internal safety measures, voluntary commitments, and industry-led governance frameworks.

However, this approach is increasingly viewed with scepticism. Critics point out that technology companies face powerful incentives to prioritise market share, profitability, investor expectations, and competitive advantage. History offers numerous examples, from social media algorithms to data privacy scandals, demonstrating that commercial incentives do not always align with societal interests.

Moreover, AI systems are no longer simply consumer products. They are becoming infrastructures that shape information flows, political discourse, labour markets, military capabilities, and even democratic institutions. Entrusting their governance primarily to corporations, therefore, requires an extraordinary degree of confidence in actors whose accountability mechanisms remain limited.

For many Muslim scholars and for Pope Leo XIV, this approach appears increasingly risky because it effectively transfers decisions about humanity's future to a small number of private entities operating beyond meaningful democratic oversight.

3. Technological Moratorium

This position argues that AI development should be significantly slowed or temporarily halted until governance frameworks, safety mechanisms, and societal understanding catch up with technological capabilities. Supporters

fear that AI may advance faster than humanity's ability to manage its consequences.

While both Muslim scholars and the Pope acknowledge these concerns, neither perspective advocates a complete halt to technological innovation.

4. Human-Centred Democratic Governance

The final position is the closest to both emerging Muslim perspectives and the Pope's encyclical. It accepts the benefits of AI while insisting that technological development must remain accountable to broader social, ethical, and democratic principles.

Key elements include:

- Public accountability
- Ethical oversight
- Democratic governance
- International cooperation
- Protection of human dignity
- Safeguards against excessive concentration of power
- Inclusion of diverse cultural and civilizational perspectives

Under this approach, AI should serve society rather than society adapting itself to the needs of technology.

Additionally, Abdelnour contends that while institutional regulation and international governance frameworks are essential, Islamic ethics insists that lasting reform must begin with the self. The Quran states unequivocally: {Indeed, God will not change the condition of a people until they change what is in themselves} (13:11). This bottom-up approach demands that individuals examine their own intentions (*niyyah*), resist the temptation to outsource moral reasoning to algorithms, and cultivate the spiritual virtues—wisdom, humility, and compassion—that no machine can replicate. External regulators can set boundaries, but only inner transformation can ensure that technology serves rather than enslaves.

Towards a Shared Ethical Framework

One of the most interesting developments in the AI debate is the emergence of unexpected alliances. Religious traditions, humanists, civil society organisations, and even some technology leaders increasingly share concerns regarding human dignity, meaning, social cohesion, surveillance, inequality, and the concentration of power.

Despite important theological differences, many of these groups converge on a simple principle: technological capability should not become the sole measure of human progress. The central question is not whether AI will become more powerful. It almost certainly will. The real question is whether humanity will remain capable of governing that power in accordance with values that transcend efficiency, profit, and technological competition.

Conclusion

The convergence between religious perspectives on artificial intelligence points toward a shared ethical imperative: technology must be liberated from the logic of domination, extraction, and unaccountable power. To "disarm" AI is not to reject it, but to challenge the assumption that technical capability automatically confers the right to govern. It is to resist a technocratic paradigm that equates efficiency with progress and treats human life as a variable to be optimised.

Representatives of monotheistic traditions reject the notion that technological development should be left to market forces or corporate self-regulation. They insist that AI governance must reflect the moral and cultural values of the communities it affects. And many demand that Muslim, Global South, and civilizational perspectives be centred in the global debate on AI ethics—not as passive recipients of Western technological models, but as active contributors to a more just and pluralistic digital order.

The alternative, as the Minab and Gaza precedents have demonstrated, is not merely a loss of sovereignty or representation. It is the normalisation of mass violence enabled by systems that reduce human life to data points and treat civilian death as an acceptable line item. This is the logic that both Muslim and Catholic scholars have warned against: a logic that delegates moral responsibility to algorithms and ethical oversight to the same institutions that profit from military contracts.

If AI is to serve humanity rather than subordinate it, its development must be guided by principles that transcend profit and efficiency: human dignity, the common good, accountability to communities, and the protection of the most vulnerable. This requires not only ethical frameworks but institutional reforms, international cooperation, and the active participation of diverse cultural and religious traditions in shaping the technological future.

The stakes could not be higher. The question is not whether AI will become more powerful—it will. The question is whether humanity, in all its diversity, will retain the capacity to govern that power in accordance with values that reflect the grandeur of human dignity, not merely the imperatives of the market or the state. Monotheistic traditions have each answered this question with a shared conviction: the future must be built, not endured.