Joint Statement on AI Ethics

As Jewish and Christian religious leaders, scholars, and advocates, we have gathered in Rome to add our voices to what Pope Leo XIV has called to be "serene and informed conversation" about the advancement of artificial intelligence (AI).[1,2]

We believe that all people, having been formed in the image and likeness of our Creator,[3] are imbued with inherent dignity, fundamental rights, capacity to co-create with the Divine, and responsibility—to the Divine and to ourselves—to steward our humanity and our world.

In 1948, in the wake of two devastating World Wars, both of which were made more lethal by technological advancement, many of our fundamental rights were codified by the United Nations Universal Declaration of Human Rights (UNDHR). Since then, multiple bodies have made significant progress to extend our fundamental rights to the digital commons.

Now, in the era of rapidly developing AI, we believe it is incumbent, not only on companies and governments, but on people of faith, who number more than 6 billion worldwide, to respect, protect, and advance human dignity, rights, and flourishing by ensuring that AI is—and remains—secure, safe, ethical, and under human control.[4]

In 2020, under Pope Francis' visionary leadership, the Vatican issued the "Rome Call for AI Ethics," which declared that for technological advancement to advance "true progress for the human race and respect for the planet," it must meet three vital requirements: "it must include

---

[1] Pope Francis, "Address of His Holiness Pope Francis to the Roman Curia for the Presentation of Christmas Greetings," December 21, 2019, https://www.vatican.va/content/francesco/en/speeches/2019/december/documents/papa-francesco_20191221_curia-romana.html; Archbishop Vincenzo Paglia, "Press Conference to Present the Results of the 28th General Assembly of the Pontifical Academy for Life," February 23, 2023, https://press.vatican.va/content/salastampa/en/bollettino/pubblico/2023/02/23/230223d.pdf.

[2] Pope Leo XIV, "Message of Pope Leo XIV to Participants in the Second Annual Conference on Artificial Intelligence, Ethics, and Corporate Governance," June 17, 2025, https://www.vatican.va/content/leo-xiv/en/messages/pont-messages/2025/documents/20250617-messaggio-ia.html.

[3] Genesis 1:26-27.

[4] Humans were given the responsibility to work in and care for the Garden of Eden—a responsibility to creation that is not abrogated with the advent of AI. See Genesis 2:15.

every human being, discriminating against no one," "have the good of humankind and the good of every human being at its heart," and "be mindful of the complex reality of our ecosystem."[5]

We are hopeful that AI, responsibly developed and deployed, will make profound contributions to human flourishing—in sustainable water, energy, and agriculture production; accessible healthcare; advanced medicine; and other fields.

But we are equally cognizant of its challenges.

Five years after the Rome Call, as we conclude our gathering, we are releasing this statement in the same spirit, hoping to advance "algorethics" by:

1: Recognizing that AI—which trains on human history and creation and evolves in response to human feedback—reflects human values.

2: Therefore, calling for AI that possesses and advances the following ethics: accuracy, transparency, privacy, security, and human dignity and common good.

Accuracy

AI is shaping human relationships profoundly—with each other, our labor, and our world. But unreliable AI is harming human relationships equally profoundly. Authoritative AI systems are subordinating human judgment and agency. Biased AI systems are perpetuating human discrimination and inequality, disproportionately harming vulnerable populations. And efficiency-focused AI systems are denying humans crucial nuances, especially in considering complex questions and completing intricate tasks.[6]

---

[5] Pope Francis, "Rome Call for AI Ethics," February 28, 2020, https://www.vatican.va/roman_curia/pontifical_academies/acdlife/documents/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf.

[6] G.K. Chesterton celebrated Christianity's ability to hold seemingly contradictory truths in tension. He argued against the modern tendency to resolve all paradoxes into simple "either or" choices. See G. K. Chesterton, "The Paradoxes of Christianity," in *Orthodoxy*, (John Lane Company, 1908).

Humans must respond by demanding that AI systems complete the tasks we assign them accurately—across prompts, projects, data sets, training environments, and time—and when they do not, by holding their developers and deployers accountable. We must begin by requiring that AI systems be independently evaluated and, when independent evaluators identify views that offend human dignity and rights (such as discrimination and violence), AI developers disclose them to users and correct them through re-training.

Transparency

As AI systems can be difficult for non-specialists—and sometimes even for specialists—to understand or explain,[7] when they produce negative outcomes, it can be challenging for humans to identify and correct them. When AI systems reduce humans to data points, our lack of understanding of their innerworkings can become dangerous: it can encourage us to reduce each other to data points too. The resulting excessive commodification can lower barriers to dehumanization, providing layers of insulation from the moral burdens of the ethical and emotional dimensions of judging, insulting, and oppressing each other.[8] This, in turn, can inhibit empathetic, embodied, human connection and risk offending both our individual dignity and our collective harmony.

Humans must respond by demanding that AI systems are transparent: that they disclose when they are present and when they are being used, especially to generate content; explain their "chain-of-reasoning" or "chain-of-thought;" acknowledge they make mistakes; and acknowledge they are not human. And crucially, humans must continue investing in improving our understanding of *why* and *how* AI systems work as they do.

---

[7] Pope Francis, "Address of His Holiness Pope Francis to the G7 Session on Artificial Intelligence"; Matthew Kosinski, "What is Black Box Artificial Intelligence (AI)?" *IBM*. October 29, 2024, https://www.ibm.com/think/topics/black-box-ai; Like the Tower of Babel, the black box risks technology hubristically expanding beyond human capacity. See Genesis 11:1-9.

[8] See *The Rise of Digital Repression* by Steven Feldstein, including its review here: https://carnegieendowment.org/research/2021/04/the-rise-of-digital-repression-how-technology-is-reshaping-power-politics-and-resistance?lang=en.

Privacy

In the era of rapidly developing AI, the right to privacy should be extended to encompass the privacy of data, interaction, inference, and use. Humans must demand that AI developers, in consultation with cybersecurity professionals, guard against unauthorized access to, misuse of, and other violations of the various forms of privacy that permeate the collective digital ecosystems in which we create and interact with each other and with technology.

Security

As they navigate the complexities of regulating dual-use technology, nations, organizations, and others who develop and use AI must ensure they do not violate the physical security, and thus the inherent dignity, of their citizens or of any humans. The international community should be appalled by reports of AI-enabled facial recognition technology being employed to locate, arrest, and harm ethnic and religious minorities and activists. Indeed, the international community must call for these practices to cease immediately and indefinitely.

As AI continues to be integrated into military action and war across the globe, it should be employed to mitigate the harms of conflict (such as to limit civilian casualties), but it must never be fully empowered to autonomously kill or decide to kill humans. Taking human life carries moral agency and responsibility—burdens that AI, lacking organic sentience, cannot bear. The international community must ban AI systems from operating as independent arbiters of lethal action.

Human Dignity and Common Good

Finally, and most fundamentally, robust understanding of human dignity—including its emotional, spiritual, cultural, labor-related, and ecological dimensions—must inform AI's development and governance. In addition to guarding against AI eroding human critical thinking; excessively commodifying human decision-making; and exacerbating human inequality, animosity, and trauma, humans must guard against the technology's potential to disrupt or displace human interaction, relationships, and empathy. Humans must reject AI systems replacing human friends, romantic partners, and religious authorities,[10] including by rejecting their incorporation of addictive design principles.

We must thoughtfully navigate AI's potential dislocation of jobs, especially at scale. Labor is not just a way humans produce or contribute; it is a way we exercise our dignity.

On spiritual matters, we must approach AI with discernment and wisdom to use technology to enhance, not diminish, our spiritual life. Above all, we must refuse to idolize or worship AI, no matter its achievements.

As AI companies race to develop AI that moves beyond tools to build superintelligence—AI that can outperform all humans at most cognitive tasks—humans must agree on our collective moral red line: we must not develop superintelligence until we agree it is safe, controllable, and desired by the broader public. Only by remaining a tool for humans can AI truly serve humanity.

Finally, no consideration of AI's impact on the common good of humanity would be complete without acknowledging its effect on the planet. Because most large-scale AI systems depend on data centers that use significant amounts of water and electricity,[9] humans must demand transparency about AI's energy consumption in order to make it efficient and sustainable.[10,11]

By signing our names to this statement and ethical framework, in contrast to technical and material voices, we draw on our moral values, rooted in our collective religious traditions and echoed in many other spiritual and ethical belief systems: that humans are imbued with dignity; that from our dignity, we derive rights; that these rights include the liberty to pursue our Divine callings and purposes without violation; and that these values should guide our integration of AI into our lives. While we are not designers of AI systems, we are essential judges of its goodness.

---

[9] Adam Zewe, "Explained: Generative AI's Environmental Impact," *MIT News,* January 17, 2025, https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117.

[10] "AI Has an Environmental Problem. Here's What the World Can Do About It," UN Environment Programme. September, 2024. https://www.unep.org/news-and-stories/story/ai-has-environmental-problem-heres-what-world-can-do-about.

[11] For humans to assess and regulate the environmental impact of AI, the UN Environment Programme recommends that governments begin with five primary steps: 1) establish standardized procedures for measuring AI's environmental impact; 2) require companies to disclose the environmental impact of their AI products and services; 3) aim to make AI algorithms more energy efficient; 4) use renewable energy to offset AI's carbon emissions; and 5) integrate all AI-related policies into larger environmental regulations.

Framework for AI Ethics

To ensure AI respects, protects, and advances human dignity, rights, and flourishing, we urge the adoption of the following measures:

Accuracy

Audit For—and Delete or Re-Train—Biases That Offend Human Dignity and Rights

- AI developers allow independent evaluators to regularly test models before and during their deployment (and incorporate their findings into subsequent versions of the models).
- Independent evaluators audit elements of AI models' training—especially data sets, reinforcement learning from human feedback (RLHF), reinforcement learning, and human prompting—for biases.
    - In this effort, independent evaluators pay particular attention to models' treatment of economically, socially, and other vulnerable populations.
- Developers delete or re-train model views that offend human dignity and rights (such as discrimination and violence).
- Developers, deployers, and financiers invest in models that self-monitor and self-report their biases and potential discriminations & harms—in other words, models that audit themselves.

Prove and Improve Standards

- Regulators require large language models (LLMs) and other relevant models to cite their sources.
- Developers investigate and further refine potential avenues to improve verification, accuracy, and accountability.
    - In this effort, developers welcome counsel from security, safety, and alignment advocates, including faith communities.

Transparency

Promote Openness of Composition and Operation

- Regulators require LLMs and other relevant models to disclose when they are present or being used, especially to generate content.
- Developers avoid LLMs and other relevant models using anthropomorphic language to describe themselves, including ascribing life or emotions to themselves.
- Everyone, including AI users, avoid using anthropomorphic language to describe AI.
- Regulators require models to explain, by default, their "chain-of-reasoning" or "chain-of-thought."
- Regulators require models to include, by default, disclaimers that they make mistakes—in other words, that the output they provide is only as accurate as the input they receive.
- Developers document the purposes, design principles, evaluation metrics, and benchmarks they assign to models, and seek safe ways to provide this information to users.
- Developers record the data they use to train, test, and fine-tune models, and seek safe ways to provide this information to users.
- Regulators, policymakers, advocates, developers, deployers, and financiers invest in alignment—the project of ensuring that AI acts in accordance with, rather than in contravention to, human values, intentions, and goals—including developing robust AI supervision methods.

Privacy

Respect Privacy by Default

- Developers design AI systems to operate, by default, with end-to-end encryption and other provisions to uphold human privacy.
- Regulators require AI developers to disclose how user data is being collected, stored, and employed by AI systems, including over time.

Conceptualize Privacy as Relational and Societal

- In addition to giving individuals information about use(s) of their data, developers and deployers conceptualize privacy as societal and relational and seek collective ways to govern and hold models accountable (network-level risk mitigation, community consultation, collective redress).

Security

Ensure Safety and Security

- States vigilantly apply existing civil rights protections to AI use and deployment.
- States commit, both bilaterally and multilaterally, to refrain from using AI systems to surveil citizens and other humans in ways that violate their rights to privacy and physical safety.
- States establish cooperative frameworks to prevent and counter manipulations of AI to perpetrate violence against humans, including but not limited to terrorism, trafficking, suppression of civil liberties, mis and disinformation campaigns, and curtailment of religious freedoms.
- States seek ways to bridge differences, informed by their joint commitments to AI remaining secure, safe, and ethical.

Negotiate International Convention Against Completely Autonomous War

- States, both individually and multilaterally, ban AI systems from operating as independent arbiters of lethal action.

- As states codify such bans into international law, they draw inspiration from the Biological Weapons Convention (BWC) and Chemical Weapons Convention (CWC).

Human Dignity and Common Good

Treat Humans as Paramount

- Regulators, faith communities, and developers cooperate to ensure that AI tools are developed to align with human intentions and augment human purposes—including, as articulated in the Rome Call for AI Ethics, our relationships, labor, and environmental stewardship.
- Regulators mandate age verification mechanisms and other design features that discourage vulnerable users, particularly children, from developing emotionally dependent relationships with AI systems.
- Regulators, administrators, teachers, parents, and other crucial voices in child development and well-being integrate AI into learning as thoughtfully as possible, remaining cognizant that learning directly informs the cultivation of human minds and selves.
- Developers design AI systems to recognize many religious perspectives accurately and neutrally—in other words, to be pluralistic, using culturally appropriate means in culturally appropriate contexts.
    - In this effort, developers seek counsel—and independent evaluation—from faith communities.
- Developers and financiers support research and learning about the impacts of AI on human flourishing, including strategies to mitigate its potential negative consequences.
- Regulators grant safe harbors to companies and organizations complying with the above standards to foster ethical AI innovation and deployment.
- Regulators strengthen liabilities against companies and organizations failing to implement the above standards to discourage unethical AI innovation and deployment.

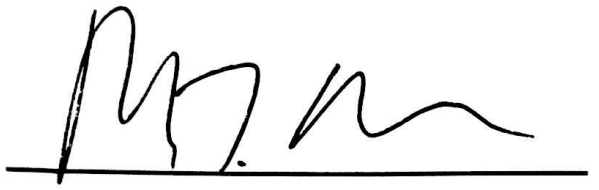
Guard Against Proliferation of Inequality

- Developers conduct impact assessments of AI systems to identify and address potential adverse effects on economically, socially, and other vulnerable populations.
- Developers implement design principles to ensure AI systems are accessible to people with diverse abilities.
- Developers implement design principles to ensure AI systems are available across multiple languages at varying literacy levels.
- Regulators, financiers, developers, and deployers cooperate to study and address the disruptions AI is imposing scale, including employment and human capital.

Regularly Update Protocols

- Developers and regulators routinely revisit, reevaluate, and update protocols to reflect major advancements in AI capabilities.

As Pope Benedict XVI said, "[w]ithout truth, without trust and love for what is true, there is no social conscience and responsibility, and social action ends up serving private interests and the logic of power, resulting in social fragmentation."[12] We call on all people—of faith, of innovation, of leadership—to embrace the truth of human dignity and to do their part to help AI better humanity.



Ryan Anderson
Ethics and Public Policy Center (EPPC)

Reverend Marian Edmonds-Allen
AI & Spirituality Initiative,
Neurospirituality Lab, Harvard Medical School

In absentia

Rabbi David Bashevkin
18Forty

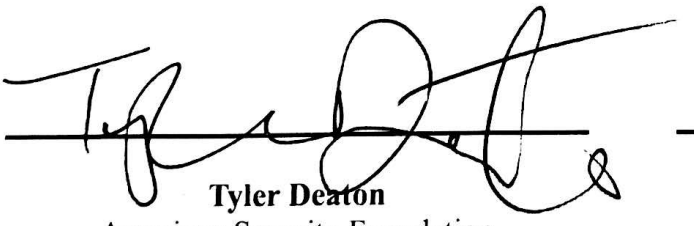
Tim Estes
AngelQ

In concurrence,

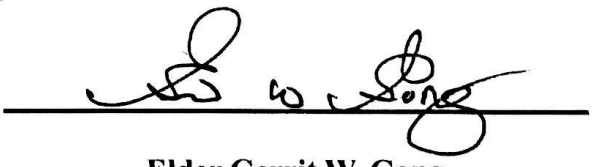
IN ABSENTIA

Paolo Carozza
University of Notre Dame Law School

Rabbi Yitzchok Feldman
Emek Beracha



Tyler Deaton
American Security Foundation



Elder Gerrit W. Gong
Quorum of the 12 Apostles, The Church of Jesus Christ of Latter-day Saints

[12] Pope Benedict XVI, "Encyclical Letter Caritas In Veritate Of The Supreme Pontiff Benedict XVI To The Bishops Priests And Deacons Men And Women Religious The Lay Faithful And All People Of Good Will On Integral Human Development In Charity And Truth," June 29, 2009, https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate.html.

Dr. Micah Goodman
Shalom Hartman Institute

Lieutenant General (Retired) Michael Groen
American Security Foundation

Rabbi Geoffrey A. Mitelman
Sinai and Synapses

Reverend Johnnie Moore
Congress of Christian Leaders

Bishop Athenagoras of Nazianzos
Holy Eparchial Synod

Reverend Dr. Walter Kim
National Association of Evangelicals

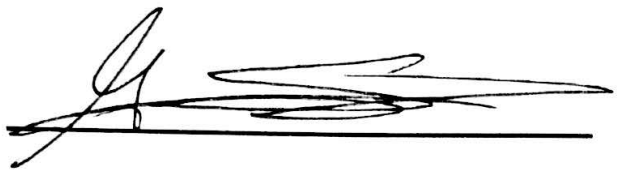
Devan Patel
American Security Foundation

Father John Paul Kimes
University of Notre Dame

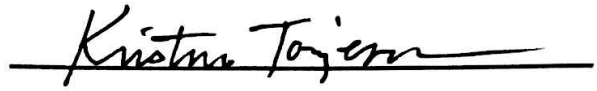
Meredith Potter
American Security Foundation

Father Jordi Pujol
Pontifical University of Santa Croce

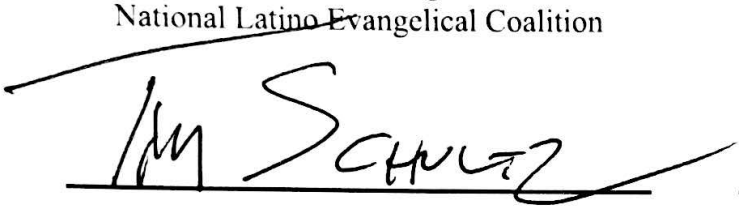
Tomicah Tilleman
Project Liberty



Dr. Gabriel Salguero
National Latino Evangelical Coalition

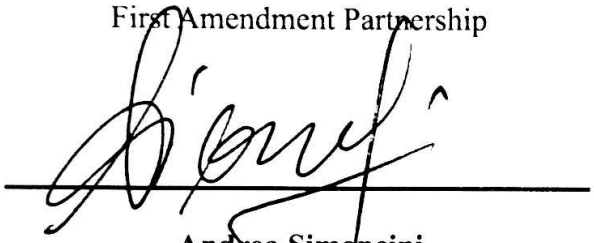


Dr. Kristine Torjesen
BioLogos



Tim Schultz
First Amendment Partnership

Bishop Martin Warner
Bishop of Chichester



Andrea Simoncini
University of Florence

Tom Williams
St. John's University

Carter Snead
University of Notre Dame