Rome Declaration Calls for Human Control Over Artificial Intelligence to Prevent Nuclear Risks and Autonomous Warfare Escalation

Published On: 28-07-2026

Recent Developments:

- **The Rome Declaration for an Unarmed and Disarming Peace** was adopted in **July 2026** by **Nobel Peace Prize laureates, Artificial Intelligence scientists, religious leaders, diplomats and global policy experts**, highlighting the growing **"convergent peril"** arising from delegating moral and ethical decision-making to **Artificial Intelligence (AI)** systems.
- The Declaration calls for a **binding international treaty** prohibiting **autonomous AI systems** from participating in **nuclear command-and-control** and insists that decisions involving war, peace and the survival of humanity must always remain under **meaningful human control**. It also aligns with the broader global debate on regulating **Lethal Autonomous Weapons Systems (LAWS)** and trustworthy AI governance.

Rome Declaration for an Unarmed and Disarming Peace:

About the Declaration:

- **The Rome Declaration for an Unarmed and Disarming Peace** is a **strategic, ethical and policy-oriented appeal** urging the international community to establish global safeguards against the misuse of **Artificial Intelligence** in warfare and nuclear decision-making.
- The Declaration emphasises that **human morality, accountability and judgment** cannot be replaced by autonomous algorithms, particularly in matters involving **nuclear weapons, armed conflict and international peace**.
- Unlike a legally binding treaty, the Declaration presently serves as a **normative international framework** intended to shape future global negotiations and responsible State behaviour.
- The Declaration reflects the growing international consensus that **technological capability must remain subordinate to human values**, democratic oversight and international humanitarian principles.

Philosophical Foundation:

- The Declaration draws intellectual inspiration from **Pope Leo XIV's encyclical "Magnifica Humanitas"**, which stresses that technological advancement must reinforce rather than diminish **human dignity, ethical responsibility and social justice**.
- It argues that technological innovation should complement **human conscience** instead of replacing ethical reasoning in critical public decisions.
- The document reinforces the principle that **peace and security require moral responsibility**, not merely computational efficiency.

Present Status:

- The Declaration currently functions as a **voluntary global norm** rather than an enforceable legal instrument.
- No universally accepted international treaty presently regulates the deployment of **Artificial Intelligence** in military command-and-control systems.
- Divergent national security priorities and rapid technological competition continue to delay consensus on global AI regulation.

Core Principles of the Rome Declaration:

Preventing an AI Arms Race:

- The Declaration seeks to prevent a new generation of **Artificial Intelligence-driven military and nuclear arms competition**.
- It recognises that uncontrolled military AI development could destabilise strategic deterrence and increase the probability of unintended conflicts.

Responsible AI Development:

- AI technologies should be designed within **transparent, accountable and ethically governed frameworks**.
- Developers should ensure that AI systems remain auditable, explainable and subject to meaningful human supervision throughout their operational lifecycle.

Responsible Use of AI:

- Artificial Intelligence should support human decision-making without independently exercising lethal authority.
- Autonomous systems should not determine the use of force in high-risk military environments.

Responsible Global Governance:

- International cooperation should establish **common standards, regulatory oversight and accountability mechanisms** governing military AI applications.
- States should strengthen confidence-building measures to reduce technological mistrust and prevent escalation.

Responsible Leadership:

- Governments, scientific institutions, technology companies and defence establishments should prioritise **ethical responsibility over geopolitical competition**.
- Innovation should be guided by **international humanitarian law, human rights and sustainable peace**.

Commitment to Nuclear Disarmament:

- The Declaration renews the international objective of achieving the **complete, irreversible and verifiable elimination of nuclear weapons**.
- AI governance and nuclear disarmament should progress together to minimise existential security risks.

Ethical and Operational Mandates:

Meaningful Human Control:

- The Declaration demands an **absolute prohibition** on allowing AI systems to make the final decision regarding the launch of nuclear weapons.
- Human operators must retain ultimate authority over every stage of strategic military decision-making.

- The principle of **Human-in-the-Loop (HITL)** is recognised as an essential safeguard against catastrophic errors.

Digital Commons Approach:

- The Declaration advocates treating critical AI knowledge and safety research as a **global digital commons**.
- Wider access to trustworthy datasets enables independent scientific evaluation of systemic AI risks and unintended consequences.

Ethical AI Development:

- AI developers should publicly disclose the ethical principles governing system design and deployment.
- The development of **fully autonomous, self-learning systems** that cannot be audited, interrupted or overridden by humans should be prohibited.

Protection of Nuclear Infrastructure:

- Nuclear-armed States should immediately conduct comprehensive **technical vulnerability assessments** of nuclear command-and-control networks.
- AI-enabled cyberattacks, algorithmic manipulation and automated misinformation should be recognised as emerging national security threats.

Time-Bound Nuclear Disarmament:

- States should resume **good-faith international negotiations** towards complete, irreversible and verifiable nuclear disarmament.
- Strategic stability should increasingly depend upon diplomacy rather than technological escalation.

Need for the Rome Declaration:

Shrinking Decision-Making Time:

- Modern AI systems can process military intelligence within seconds, substantially reducing the time available for human deliberation during strategic crises.
- Historical nuclear deterrence relied upon **human hesitation, diplomatic communication and political judgment**, which prevented accidental escalation during crises such as the **Cuban Missile Crisis (1962)** and the **1983 Soviet false alarm incident**.

Algorithmic Limitations:

- Advanced AI models remain vulnerable to **hallucinations, biased training data, adversarial attacks and data poisoning**, creating significant operational uncertainty.
- Black-box algorithms often generate outputs whose internal reasoning cannot be independently verified, reducing transparency in high-risk decisions.

Weakening Global Arms Control Architecture:

- Several traditional arms-control mechanisms have weakened, increasing uncertainty in strategic stability.
- The erosion of confidence-building arrangements has heightened concerns regarding emerging military technologies.
- AI integration into nuclear systems without internationally accepted safeguards may further increase strategic instability.

Emerging Geopolitical Flashpoints:

- Regions experiencing persistent geopolitical tensions increasingly rely upon AI-enabled surveillance, cyber capabilities and autonomous defence technologies.
- Miscalculation involving AI-supported military systems could rapidly escalate regional crises into wider international conflicts.

Understanding "Convergent Perils" in Artificial Intelligence:

Meaning:

- **Convergent Peril** refers to a situation where **multiple independent technological, political, military and cyber risks interact simultaneously**, producing a systemic threat that is significantly greater than the sum of individual risks.
- Such convergence creates cascading failures across security, governance and international stability.

Major Components of Convergent Perils:

- **Artificial Intelligence errors, cyber warfare, nuclear weapons, misinformation campaigns and geopolitical rivalry** may reinforce one another during periods of crisis.
- Autonomous systems operating without adequate human oversight could amplify accidental escalation through rapid machine-to-machine interactions.
- Weak global governance and fragmented international regulation increase the probability of transnational technological risks.

UPSC Perspective:

- The concept demonstrates how **emerging technologies require multidisciplinary governance**, combining **international law, ethics, cybersecurity, diplomacy, defence policy and human rights**.
- It highlights the growing importance of **Responsible AI**, **Explainable AI**, **Human-Centred AI** and **Global Technology Governance** in contemporary international relations.

Global AI Ethics and Governance Landscape:

European Union Artificial Intelligence Act, 2024:

- The **European Union Artificial Intelligence Act** is the world's first comprehensive legislation governing Artificial Intelligence through a **risk-based regulatory framework**.
- It prohibits **unacceptable-risk AI applications**, imposes strict obligations on **high-risk AI systems**, and establishes transparency requirements for **general-purpose AI models**, including foundation models.

Bletchley Declaration, 2023:

- The **Bletchley Declaration** was endorsed by **28 countries**, including **India**, recognising the opportunities and risks associated with advanced frontier AI.
- It promotes international cooperation on **AI safety research, scientific collaboration and coordinated governance mechanisms**.

UNESCO Recommendation on the Ethics of Artificial Intelligence:

- Adopted in **2021**, it represents the **first global normative framework** on AI ethics.
- It promotes **human dignity, transparency, accountability, privacy, fairness, sustainability and protection of fundamental rights** throughout the AI lifecycle.

Global Partnership on Artificial Intelligence (GPAI):

- **GPAI** is a multi-stakeholder international initiative supporting **responsible, inclusive and human-centric AI**.
- **India** is a founding member and hosted the **Global Partnership on Artificial Intelligence Summit, 2023**, reinforcing its commitment to democratic and trustworthy AI governance.

India's AI Governance Framework:

National Strategy for Artificial Intelligence (2018):

- **NITI Aayog's National Strategy for Artificial Intelligence – "AI for All"** laid the foundation for India's vision of leveraging AI for **inclusive growth, social welfare and economic development**.
- The strategy identifies **healthcare, agriculture, education, smart mobility and smart cities** as priority sectors where AI can generate maximum public value.
- It advocates **Responsible AI** based on the principles of **safety, transparency, fairness, inclusivity, privacy, security and accountability**.
- The strategy also emphasises **research and innovation, public-private partnerships, skilled human resources and ethical governance** to build a globally competitive AI ecosystem.

IndiaAI Mission (2024):

- **IndiaAI Mission** seeks to establish a comprehensive indigenous AI ecosystem through investments in **high-performance computing infrastructure, quality datasets, startups, innovation and AI talent development**.
- The Mission includes a dedicated **Safe and Trusted AI** pillar to ensure that AI systems remain **ethical, reliable, transparent and aligned with constitutional values**.
- It also promotes **AI Centres of Excellence, indigenous foundation models, multilingual AI applications and responsible AI research** to strengthen India's digital sovereignty.
- The initiative aims to reduce dependence on foreign AI infrastructure while encouraging innovation consistent with public interest.

AI Governance Guidelines by Ministry of Electronics and Information Technology (MeitY):

- **MeitY** follows a **principle-based techno-legal approach** that balances technological innovation with safeguards for citizens.
- The framework is guided by the **Seven Sutras** for **trustworthy, human-centric, safe and responsible Artificial Intelligence**.
- It promotes **algorithmic accountability, transparency, explainability, privacy protection, cybersecurity and human oversight** throughout the AI lifecycle.
- The guidelines encourage **risk-based regulation** rather than imposing blanket restrictions on AI innovation.

Challenges in AI Governance:

Ethical Challenges:

- AI systems may reproduce or amplify **biases embedded in training data**, leading to discriminatory outcomes in employment, finance, policing and public service delivery.
- Lack of transparency in complex machine learning models creates difficulties in assigning **legal and ethical accountability** for automated decisions.
- Excessive reliance on AI may gradually reduce **human autonomy, moral reasoning and democratic oversight** in governance.

Security Challenges:

- AI can significantly enhance the capabilities of **cyberattacks, misinformation campaigns, deepfakes and autonomous military systems**.
- Malicious actors may exploit AI for **identity theft, election manipulation, financial fraud and critical infrastructure disruption**.
- Integration of AI into defence systems increases risks associated with **algorithmic errors, cyber manipulation and unintended military escalation**.

Regulatory Challenges:

- Rapid technological advancement often outpaces **legislative and regulatory frameworks**, creating governance gaps.
- Different countries adopt varying regulatory approaches, making **international coordination** difficult.
- Balancing **innovation, national security, economic competitiveness and individual rights** remains a complex policy challenge.

Technological Challenges:

- Advanced AI models frequently operate as **black-box systems**, limiting explainability and public trust.
- Ensuring **robustness, auditability, reliability and interoperability** becomes increasingly difficult as AI systems grow more sophisticated.
- Dependence on large datasets and computing infrastructure raises concerns regarding **data sovereignty, privacy and equitable access**.

International Legal and Governance Issues:

Absence of Binding Global Treaty:

- There is currently **no comprehensive international treaty** regulating military applications of Artificial Intelligence.
- Existing international humanitarian law provides general principles but does not specifically address autonomous AI decision-making.
- Divergent geopolitical interests continue to delay negotiations on globally accepted AI governance norms.

Lethal Autonomous Weapons Systems (LAWS):

- **LAWS** refer to weapon systems capable of **identifying, selecting and engaging targets without meaningful human intervention**.
- Discussions under the **Convention on Certain Conventional Weapons (CCW)** focus on establishing international rules governing such systems.
- Many countries, civil society organisations and experts advocate a **legally binding treaty** requiring meaningful human control over all lethal autonomous weapons.
- The Rome Declaration strengthens these ongoing global efforts by linking AI regulation with **nuclear risk reduction and humanitarian principles**.

Significance for India:

Strategic Importance:

- As an emerging technological power, **India** must simultaneously strengthen AI innovation and ensure responsible governance.
- Ethical AI enhances **public trust, economic competitiveness and national security**, while reducing long-term systemic risks.
- India's balanced approach enables it to contribute constructively to shaping **global AI governance architecture**.

Diplomatic Relevance:

- India's participation in **GPAI, G20 Digital Economy initiatives, Bletchley Process and multilateral AI dialogues** strengthens its role in developing globally accepted norms.
- Responsible AI governance aligns with India's vision of **human-centric technology, Digital Public Infrastructure (DPI) and inclusive digital transformation**.
- India can act as a bridge between developed and developing countries in evolving an equitable global AI framework.

Way Forward:

Institutionalising Human Oversight:

- **Human-in-the-Loop (HITL)** should remain mandatory for all AI systems affecting **life, liberty, national security and fundamental rights**.
- Artificial Intelligence should function as a **decision-support system rather than an autonomous decision-maker** in critical sectors.

Promoting Explainable AI:

- Developers should design AI systems whose reasoning processes are **transparent, interpretable and independently auditable**.
- Explainability strengthens accountability, improves public confidence and enables effective regulatory oversight.

Ethical Impact Assessments:

- AI systems should undergo **Ethical Impact Assessments (EIA)** before deployment, similar to environmental assessments for major development projects.
- Assessments should evaluate **bias, fairness, privacy, human rights implications, cybersecurity and societal impact**.

Embedding Human Values in AI:

- AI design should incorporate **compassion, dignity, fairness, non-discrimination and respect for fundamental rights** alongside technical efficiency.
- Human welfare should remain the ultimate objective of technological innovation.

Strengthening Global Cooperation:

- Countries should negotiate a **binding international convention** regulating **Lethal Autonomous Weapons Systems (LAWS)** and AI-enabled military technologies.
- Greater cooperation among **governments, international organisations, academia, industry and civil society** is essential for effective AI governance.
- Confidence-building measures, information sharing and common technical standards can reduce the risks of AI-driven strategic instability.

UPSC Perspective:

Possible UPSC Themes:

Meaningful Human Control versus fully autonomous decision-making.

Artificial Intelligence and the future of nuclear deterrence.

Ethics, governance and regulation of emerging technologies.

India's role in shaping responsible global AI governance.

Balancing innovation with accountability in the digital age.

Value Addition for UPSC:

Key Terms:

- **Convergent Peril:** Interaction of multiple technological, military, cyber and geopolitical risks that collectively create a systemic threat greater than their individual impact.
- **Human-in-the-Loop (HITL):** Human operators retain the final authority over critical AI-assisted decisions.
- **Black-Box AI:** AI systems whose internal decision-making process cannot be easily interpreted or explained.
- **Explainable AI (XAI):** AI designed to provide transparent, understandable and auditable reasoning behind its outputs.
- **Lethal Autonomous Weapons Systems (LAWS):** Weapon systems capable of selecting and attacking targets without meaningful human intervention.
- **Responsible AI:** AI that is **lawful, ethical, transparent, safe, inclusive, accountable and human-centric** throughout its lifecycle.

UPSC Prelims Facts:

- **UNESCO Recommendation on the Ethics of AI (2021)** is the **first global normative framework** dedicated to ethical AI governance.
- **European Union AI Act (2024)** is the **world's first comprehensive AI legislation** based on a risk-oriented regulatory model.
- **India** is a **founding member of the Global Partnership on Artificial Intelligence (GPAI)** and hosted the **GPAI Summit 2023**.
- **NITI Aayog's National Strategy for AI (2018)** introduced the vision of **"AI for All"** to promote inclusive and socially beneficial AI.
- **IndiaAI Mission (2024)** includes a dedicated **Safe and Trusted AI** pillar to strengthen ethical AI development, indigenous innovation and trustworthy deployment.
- **Rome Declaration for an Unarmed and Disarming Peace (2026)** advocates **meaningful human control**, ethical AI governance and renewed commitment to **nuclear disarmament**, while calling for an international treaty restricting autonomous AI in nuclear command-and-control systems