

# The Manifesto for an Ethical and Human-Centric Artificial Intelligence

Artificial intelligence (AI) represents one of the greatest opportunities of our time, but also one of the greatest challenges. If developed and used responsibly, AI can improve the quality of life, foster innovation and solve complex problems. However, without a solid ethical framework and adequate human oversight, it risks amplifying inequalities, threatening fundamental rights and compromising the security of our societies.

This manifesto was born with the aim of uniting academic, political and institutional authorities in a common commitment: **to ensure that AI is developed and implemented in an ethical, transparent and human-centric way**. We ask the European Union to adopt concrete measures that go beyond compliance with the GDPR, including rigorous controls on the ethical and decisional safety of AI systems.

The Manifesto for an Ethical and Human-Centric Artificial Intelligence

Copyright (C) Alessandro Verna and Alvisé Gottieri May 9 2025.

Permission is granted to copy, distribute and/or modify this document

under the terms of the Gnu Free Documentation License, Version 1.3

<https://www.gnu.org/licenses/fdl-1.3.html#license-text> or any later

version published by the Free Software Foundation;

with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

## Summary

Manifesto for an Ethical and Human-Centric Artificial Intelligence .....	1
1. Why a Manifesto for Ethical and Human-Centric AI? .....	5
1.1 The Transformative Impact of AI .....	5
1.1.1 AI Opportunities .....	5
1.1.2 Risks of AI .....	5
1.2 The Ethical Challenges of AI .....	5
1.2.1 Discrimination and Bias .....	6
1.2.2 Privacy and Surveillance .....	6
1.2.3 Transparency and Accountability .....	6
1.2.4 Employment Impact .....	6
1.3 The Need for a Human-Centric Approach .....	6
1.3.1 Human Supervision .....	6
1.3.2 Alignment with Human Values .....	6
1.3.3 Social and Environmental Wellbeing .....	6
1.4 The New Meaning of Training the New Generations .....	7
1.4.1 AI Education .....	7
1.4.2 Cognitive Adaptation .....	7
1.4.3 Continuous Training .....	7
1.5 The Danger of Using AI for Military Purposes .....	8
1.5.1 Autonomous Weapons and Lethal Decisions .....	8
1.5.2 Violation of International Law .....	8
1.5.3 Erosion of Human Control .....	8
1.5.4 Need for Final Human Control .....	8
1.5.5 Implications for Global Security .....	9
2. The Fundamental Principles of the Manifesto .....	10
2.1 Human Supervision .....	10
2.1.1 Supervision Levels .....	10
2.1.2 Practical Implications .....	10
2.1.3 Necessary Actions .....	10
2.2 Ethical and Decisional Security .....	11
2.2.1 Control Mechanisms .....	11

2.2.2 Practical Implications .....	11
2.2.3 Necessary Actions .....	11
2.3 Transparency and Explainability .....	11
2.3.1 Interpretable Models .....	11
2.3.2 Practical Implications .....	11
2.3.3 Necessary Actions .....	12
2.4 Equity and Non-Discrimination .....	12
2.4.1 Identifying and Correcting Biases .....	12
2.4.2 Practical Implications .....	12
2.4.3 Necessary Actions .....	12
2.5 Privacy and Data Protection .....	12
2.5.1 Data Minimization .....	12
2.5.2 Practical Implications .....	13
2.5.3 Necessary Actions .....	13
2.6 Social and Environmental Wellbeing .....	13
2.6.1 Social Impact .....	13
2.6.2 Environmental Impact .....	13
2.6.3 Necessary Actions .....	13
2.7 Responsibility and Traceability .....	13
2.7.1 Accountability Mechanisms .....	13
2.7.2 Practical Implications .....	14
2.7.3 Necessary Actions .....	14
3. Code Verification: Ensuring Ethical and Decisional Security .....	15
3.1 Security Elements in Code .....	15
3.1.1 Technical Robustness .....	15
3.1.2 Integrated Ethical Controls .....	15
3.1.3 Traceability of Decisions .....	16
3.1.4 Explainability .....	16
3.2 Independent Verification Processes .....	16
3.2.1 Regular Audits .....	17
3.2.2 Ethical Certifications .....	17
3.2.3 Developer Guidelines .....	18
3.2.4 Stakeholder Involvement .....	18

3.2.5 International Collaboration .....	19
3.3 The Threat of Accelerationism .....	19
3.3.1 Risks of Accelerationism .....	19
3.3.2 Unfair Competition and Lack of Standards .....	20
3.3.3 Impact on Ethical and Decisional Security .....	20
3.3.4 Actions Needed to Counter Accelerationism .....	21
4. Concrete Actions for an Ethical and Human-Centric AI .....	22
4.1 A Complete Regulatory Framework .....	22
4.1.1 Binding Regulations .....	22
4.1.2 Technical Standards .....	22
4.1.3 Ethical Certifications .....	23
4.2 Global Collaboration .....	23
4.2.1 International Forums .....	23
4.2.2 Global Standards .....	23
4.2.3 Resource Sharing .....	24
4.3 Education and Awareness .....	24
4.3.1 Training Programs .....	24
4.3.2 Awareness Campaigns .....	24
4.3.3 Initiatives for the Public .....	25
4.4 Public Participation .....	25
4.4.1 Civic Panels .....	25
4.4.2 Public Consultations .....	25
4.4.3 Feedback Mechanisms .....	26
5. A Call to Action .....	27
6. Conclusion .....	28

# 1. Why a Manifesto for Ethical and Human- Centric AI ?

Artificial Intelligence (AI) is one of the most transformative technologies of our time, with the potential to revolutionize sectors such as healthcare, education, economics and governance. However, its development raises important ethical questions that must be addressed to ensure that AI is used responsibly and sustainably. This manifesto arises from the need to unite academic, political and institutional authorities in a common commitment: to **ensure that AI is developed and implemented in an ethical, transparent and human-centric way** . Below, we delve into the reasons that make this manifesto not only necessary, but urgent.

## 1.1 The Transformative Impact of AI

AI is transforming every aspect of our society, offering unprecedented opportunities to improve quality of life, solve complex problems, and foster innovation. However, this transformation is not without risks.

### 1.1.1 AI Opportunities

- **Improving Quality of Life** : AI can revolutionize sectors such as healthcare, education and transportation, offering personalized and efficient solutions.
- **Solving Global Problems** : AI can help solve complex problems, such as climate change, poverty and social inequality.
- **Economic Innovation** : AI can stimulate economic growth, creating new jobs and business opportunities.

### 1.1.2 Risks of AI

- **Amplifying Inequalities** : Without adequate controls, AI can amplify existing inequalities, creating new forms of social and economic exclusion.
- **Human Rights Violation** : AI can be used to violate fundamental rights, such as privacy, freedom of expression and non-discrimination.
- **Loss of Control** : Without human supervision, AI can make harmful or unpredictable decisions, with potentially disastrous consequences.

## 1.2 The Ethical Challenges of AI

The development and use of AI raises important ethical questions that must be addressed to ensure that this technology serves humanity.

### *1.2.1 Discrimination and Bias*

AI algorithms can perpetuate racial, gender, or economic biases, amplifying existing inequalities.

**Example** : Facial recognition software with racial bias , recruitment systems that discriminate against women.

### *1.2.2 Privacy and Surveillance*

AI collects and analyzes huge amounts of personal data, increasing the risk of abuse and mass surveillance.

**Example** : Mass surveillance by authoritarian governments, misuse of data by corporations.

### *1.2.3 Transparency and Accountability*

Many AI systems are “black boxes,” making it difficult to understand how they make decisions that affect people’s lives.

**Example** : Opaque decisions in critical contexts such as healthcare or justice.

### *1.2.4 Employment Impact*

AI-driven automation could eliminate millions of jobs, creating new social and economic inequalities.

**Example** : Job losses in sectors such as manufacturing, transportation and services.

---

## *1.3 The Need for a Human- Centric Approach*

centric approach to AI means putting people at the center of the development and use of this technology, ensuring that it serves humanity and not the other way around.

### *1.3.1 Human Supervision*

Humans must maintain control over AI decisions, with the ability to intervene, correct, or disable systems when necessary.

**Example** : Supervision models such as “Human-in-the-loop” and “Human-on-the-loop”.

### *1.3.2 Alignment with Human Values*

AI must be designed to respect fundamental rights and ethical values, such as privacy, fairness and non-discrimination.

**Example** : Integrating ethical principles into the code and decision-making processes.

### *1.3.3 Social and Environmental Wellbeing*

AI must contribute to the well-being of society and environmental sustainability, promoting equity and reducing negative impacts.

**Examples or :** AI applications for reducing carbon emissions and promoting social inclusion.

## *1.4 The New Meaning of Training the New Generations*

The cognitive impact of AI is transforming the way people learn, think, and interact with the world. This requires a new approach to educating new generations, preparing them to live and work in a world increasingly influenced by artificial intelligence.

### *1.4.1 AI Education*

The new generations must be educated on the ethical and safety principles of AI, to understand both the opportunities and risks of this technology.

**School Curriculum :** Integrate AI education into school curricula, from primary school to university.

**Digital Skills :** Teach advanced digital skills, such as programming and data analysis.

**Critical Thinking :** Promote critical thinking and the ability to evaluate the social impact of AI.

### *1.4.2 Cognitive Adaptation*

AI is changing the way people process information and make decisions. This requires cognitive adaptation to avoid technological dependencies and loss of autonomy.

**Human-Technology Balance :** Teaching new generations to balance the use of technology with independent thinking and creativity.

**Digital Awareness :** Promote awareness of the cognitive risks of AI, such as algorithmic dependency and loss of privacy.

### *1.4.3 Continuing Education*

In a rapidly changing world, training cannot be limited to the early years of life. It is essential to promote continuous training to adapt to the new challenges and opportunities of AI.

**Refresher Programs :** Create continuing education programs for workers and citizens, focused on new technologies and skills.

**Lifelong Learning :** Promote a culture of lifelong learning, which values the constant updating of skills.

## 1.5 The Danger of Using AI for Military Purposes

The use of AI in military contexts represents one of the most controversial and dangerous applications of this technology. Without adequate human oversight, AI can be used to make autonomous decisions on human and civilian targets, with potentially catastrophic consequences.

### *1.5.1 Autonomous Weapons and Lethal Decisions*

Autonomous weapons, also known as “killer robots,” are AI systems that can identify, select, and attack targets without human intervention. This raises serious ethical and legal concerns.

**Irreversible Errors** : An AI system could mistakenly identify a civilian as a military target, resulting in innocent casualties.

**Conflict Escalation** : The use of autonomous weapons could accelerate conflicts, making them more difficult to control and de-escalate.

**Opaque Accountability** : Without direct human control, it is difficult to assign responsibility for lethal decisions made by AI.

### *1.5.2 Violation of International Law*

The use of autonomous weapons could violate principles of international humanitarian law, such as the distinction between combatants and civilians and the principle of proportionality.

**Indiscrimination** : An AI system may not be able to adequately distinguish between military and civilian targets, leading to human rights violations.

**Lack of Empathy** : AI lacks the ability to understand human context or show compassion, making it unsuitable for making decisions that affect human lives.

### *1.5.3 Erosion of Human Control*

The autonomous use of AI in military contexts erodes human control over critical decisions, shifting responsibility to machines that cannot be held morally or legally accountable.

**Loss of Autonomy** : Delegating lethal decisions to AI reduces the role of humans in decision-making, undermining ethical and democratic principles.

**Risk of Abuse** : Governments or armed groups could use autonomous weapons for repressive purposes or to commit war crimes without accountability.



#### *1.5.4 Need for Final Human Control*

It is essential that humans retain the final say in decisions involving human and civilian targets. This principle, known as “meaningful human control,” ensures that lethal decisions are always subject to human evaluation and approval.

**Ban on Autonomous Weapons** : Promote a global ban on the development and use of autonomous weapons that can make lethal decisions without human control.

**International Regulation** : Develop international treaties and agreements that limit the use of AI in military contexts and ensure compliance with international humanitarian law.

**Mandatory Human Oversight** : Implement mechanisms that ensure that every decision involving human and civilian targets is subject to human oversight and approval.

#### *1.5.5 Implications for Global Security*

The uncontrolled use of AI in military contexts threatens global security, increasing the risk of uncontrollable conflicts and an AI-based arms race.

**Arms Race** : Countries could compete to develop increasingly advanced autonomous weapons, creating global instability.

**Proliferation of Weapons** : Autonomous weapons could fall into the hands of non-state armed groups, increasing the risk of terrorism and asymmetric conflict.

**Preventive Diplomacy** : Promote dialogue between nations to prevent an AI-based arms race.

**Export Controls** : Implement strict controls on the export of AI-based military technologies.

## 2. The Fundamental Principles of the Manifesto

We propose an ethical framework based on seven fundamental principles, which should guide the development and implementation of AI. These principles are essential to ensure that AI serves humanity, respecting fundamental rights and promoting social well-being. Below, we delve into each principle in detail.

### 2.1 Human Supervision

Human oversight is the cornerstone principle for ensuring that AI remains a tool that serves people, not the other way around. This principle requires that humans maintain control over AI decisions, with the ability to intervene, correct, or disable systems when necessary.

#### 2.1.1 Supervision Levels

- **Human-in-the-loop** : In this model, a human operator actively supervises the AI's decisions and can correct them in real time. This approach is particularly important in critical contexts, such as healthcare or justice.
- **Human-on-the-loop** : The AI operates autonomously, but a human operator can intervene if the system makes mistakes or produces unacceptable results. This model is suitable for applications such as autonomous vehicles.
- **Human-out-of-the-loop** : The AI acts completely autonomously, without human intervention. This approach is not recommended for high-risk applications, as it reduces human control.

#### 2.1.2 Practical Implications

- **Critical Contexts** : In sectors such as healthcare, security and justice, human oversight is essential to avoid potentially lethal or harmful errors.
- **Accountability** : Human oversight ensures that there is always someone responsible for AI decisions, making it easier to trace and correct errors.

#### 2.1.3 Necessary Actions

- **Human Interface Design** : AI systems must include intuitive interfaces that allow human operators to easily understand and intervene.
  - **Operator Training** : Human operators must be adequately trained to supervise and correct AI systems.
-

## 2.2 Ethical and Decisional Security

Ethical and decisional safety concerns the ability of AI to make decisions aligned with human values and fundamental rights, minimizing the risk of harm or violation.

### 2.2.1 Control Mechanisms

- **Decision Limitation** : AI systems must be configured to avoid extreme or dangerous decisions, such as those that might violate human rights.
- **Alignment with Human Values** : The code must incorporate ethical principles, such as respect for privacy, fairness and non-discrimination, into every stage of the decision-making process.

### 2.2.2 Practical Implications

- **Harm Prevention** : Without ethical controls, AI systems can make harmful decisions, such as discrimination or privacy violations.
- **Public Trust** : Ethical safety is essential to building trust in AI, fostering its adoption and integration into society.

### 2.2.3 Necessary Actions

- **Integrating Ethical Checks** : Developers should integrate ethical check mechanisms into their code, such as tools to detect and correct bias .
  - **Risk Assessments** : Every AI system must undergo risk assessments to identify potential ethical violations.
- 

## 2.3 Transparency and Explainability

Transparency and explainability are essential to ensure that AI decision-making processes are understandable and accessible, even for non-technical users.

### 2.3.1 Interpretable Models

- **Transparent Algorithms** : AI systems must use algorithms that produce understandable results, avoiding inaccessible “black boxes.”
- **Clear Explanations** : AI decisions must be accompanied by intuitive explanations, such as visualizations or natural language.

### 2.3.2 Practical Implications

- **Trust and Acceptance** : Transparency is essential to building trust in AI, driving public adoption.
- **Accountability** : Explainability facilitates the traceability of decisions, allowing errors to be identified and corrected.

### 2.3.3 Necessary Actions

- **Transparency Standards** : Clear standards for the transparency and explainability of AI systems must be defined.
- **Visualization Tools** : Developers need to create tools that make AI decision-making processes understandable.

## 2.4 Equity and Non-Discrimination

Equity and non-discrimination are key principles to ensure that AI does not perpetuate or amplify existing inequalities.

### 2.4.1 Identifying and Correcting Biases

- **Training Data** : The data used to train AI systems must be representative and free of bias .
- **Unbiased Algorithms** : Algorithms must be designed to avoid discrimination based on gender, ethnicity, age, or other personal characteristics.

### 2.4.2 Practical Implications

- **Social Justice** : Without equity, AI can amplify inequalities, creating social and economic injustices.
- **Legal Compliance** : Non-discrimination is essential to ensure compliance with laws and regulations.

### 2.4.3 Necessary Actions

- **Data Audit** : Training data should be audited regularly to identify and correct biases .
- **Impartiality Testing** : AI systems must be tested to ensure they do not produce discriminatory decisions.

## 2.5 Privacy and Data Protection

Privacy and data protection are essential to ensure that personal data is handled securely and transparently.

### 2.5.1 Data Minimization

- **Limited Collection** : Personal data should be collected only when strictly necessary and for specific purposes.
- **Informed Consent** : Users must be informed about how their data will be used and must give their explicit consent.

### 2.5.2 Practical Implications

- **Protection of Rights** : Privacy is a fundamental right that must be respected at every stage of the AI lifecycle.
- **Public Trust** : Data protection is essential to building trust in AI.

### 2.5.3 Necessary Actions

- **GDPR Compliance** : AI systems must comply with the General Data Protection Regulation (GDPR) and other privacy regulations.
- **Data Security** : Personal data must be protected from unauthorized access and breaches.

## 2.6 Social and Environmental Wellbeing

AI must contribute to the well-being of society and environmental sustainability, promoting equity and reducing negative impacts.

### 2.6.1 Social Impact

- **Equity and Inclusion** : AI must be used to promote equity and inclusion, improving the quality of life for all.
- **Reducing Inequalities** : AI can be a tool to reduce social and economic inequalities.

### 2.6.2 Environmental Impact

- **Sustainability** : AI must be designed to minimize energy consumption and environmental impact, for example by optimizing data center efficiency.

### 2.6.3 Necessary Actions

- **Impact Assessments** : Every AI system must undergo social and environmental impact assessments.
  - **Incentives for Sustainability** : Incentives must be created to promote the development of sustainable AI.
- 

## 2.7 Responsibility and Traceability

Accountability and traceability are essential to ensure that AI developers and operators are accountable for their creations.

### 2.7.1 Accountability Mechanisms

- **Decision Traceability** : AI decisions must be recorded and traceable, allowing errors to be identified and corrected.

- **Accountability** : There must be clear mechanisms to ensure that developers and operators are accountable for their actions.

### *2.7.2 Practical Implications*

- **Error Correction** : Traceability facilitates error correction and continuous improvement of AI systems.
- **Legal Compliance** : Accountability is essential to ensure compliance with laws and regulations.

### *2.7.3 Necessary Actions*

- **Detailed Logs** : AI systems must generate clear, comprehensive logs that document decisions and processes.
- **Independent Audits** : Regular audits must be conducted to verify compliance with ethical principles.

## 3. Code Verification: Ensuring Ethical and Decisional Security

One of the most critical aspects of ensuring ethical and human-centric AI is **code testing**. The code that powers AI systems must be designed and tested to ensure that it adheres to ethical principles and addresses the practical challenges described in this manifesto.

### 3.1 Security Elements in the Code

AI systems code must include specific mechanisms to ensure that decisions are safe, transparent, and aligned with human values. Below, we dive deeper into the four key elements that must be integrated into the code.

#### 3.1.1 Technical Robustness

Technical robustness is essential to ensure that AI systems are resistant to errors, manipulations and cyber attacks. This element includes:

- **Error Resistance** : Code must be designed to handle unexpected situations, such as anomalous input or corrupted data, without producing incorrect or harmful results.
- **Protection from Attacks** : AI systems must be protected from cyber attacks, such as *data poisoning* (manipulation of training data) or *adversarial attacks* (inputs designed to fool the AI).
- **Continuity of Service** : The code must ensure that the system continues to function properly even under stress or high load.

Without technical robustness, AI systems can fail in critical situations, such as medical or security, with potentially disastrous consequences.

#### 3.1.2 Integrated Ethical Controls

Integrated ethical controls are mechanisms designed to prevent decisions that are discriminatory, harmful or not aligned with human values. This includes:

- **Bias Detection** : The code must include tools to identify and correct biases in training data and decision-making processes.
- **Decision Limitation** : AI systems must be configured to avoid extreme or dangerous decisions, such as those that might violate human rights.

- **Alignment with Human Values** : The code must incorporate ethical principles, such as respect for privacy, fairness and non-discrimination, into every stage of the decision-making process.

Without ethical controls, AI systems can perpetuate discrimination or make harmful decisions, undermining public trust and causing social harm.

### 3.1.3 Traceability of Decisions

Traceability is essential to ensure that AI decisions can be reconstructed and verified. This includes:

- **Detailed Logs** : The code must generate clear, complete logs that document how and why a decision was made.
- **Identification of Sources** : Decisions must be linked to the data and processes that generated them, allowing the origin of any errors or biases to be traced .
- **Auditability** : The code must be designed to facilitate external audits, allowing third parties to verify compliance with ethical principles.

Without traceability, it is impossible to identify and correct errors or biases , making AI systems opaque and difficult to regulate.

### 3.1.4 Explainability

Explainability is the ability to make AI decision-making processes understandable, even for non-technical users. This element includes :

- **Interpretable Models** : Code should use algorithms that produce understandable results, avoiding inaccessible “black boxes.”
- **Clear Explanations** : AI systems must provide intuitive explanations of their decisions, for example through visualizations or natural language.
- **Human Interaction** : The code must facilitate interaction between the AI and users, allowing them to ask questions and get clear answers.

Without explainability , users cannot trust AI systems, limiting adoption and increasing the risk of incorrect or harmful decisions.

## 3.2 Independent Verification Processes

To ensure that AI systems comply with ethical and safety principles, it is essential to establish **independent verification processes** . These processes must be conducted by third-party bodies, free of conflicts of interest, and must



include regular audits, certifications, and clear guidelines for developers. Below, we delve into the key components of these processes.

### *3.2.1 Regular Audits*

Regular audits are essential to ensure that AI systems adhere to ethical and safety principles over time. These audits should be conducted by independent experts and should cover all aspects of the system, from code to training data.

Regular audits are used to identify potential violations of ethical principles, such as bias, discrimination or lack of transparency. They must be conducted periodically and in depth, using standardized methodologies.

- **Frequency** : Audits shall be conducted at least annually or after any significant system update.
- **Standardized Methodologies** : Internationally recognized protocols must be used to ensure consistency and reliability.
- **Full Coverage** : Audits must cover all aspects of the system, including data, algorithms, interfaces, and decision-making processes.
- **Clear Reports** : Audit results must be documented in detailed reports, accessible to the competent authorities and the public.

### *3.2.2 Ethical Certifications*

Ethical certifications are a way to publicly recognize AI systems that comply with ethical and safety principles. These certifications must be issued by independent bodies and must be based on rigorous criteria.

Ethical certifications serve to build public trust and incentivize companies to develop responsible AI systems. They must be based on clear standards and must be renewed periodically.

- **Rigorous Criteria** : Certifications must be based on clear and measurable criteria, such as absence of bias, transparency and security.
- **Independent Bodies** : Certifications must be issued by bodies free from conflicts of interest.
- **Periodic Renewal** : Certifications must be renewed periodically to ensure that systems continue to comply with ethical principles.
- **Advertising** : Certified companies must be able to use a recognizable logo or brand to communicate their ethical commitment.

### 3.2.3 Developer Guidelines

Developer guidelines are essential to ensure that ethical principles are integrated from the early stages of AI development. These guidelines must be clear, practical and based on concrete examples.

The guidelines are intended to guide developers in designing and implementing ethical AI systems. They must cover all aspects of development, from data collection to algorithm design.

- **Practical Examples** : The guidelines should include concrete examples of how to integrate ethical principles into the code.
- **Tools and Resources** : Tools and resources, such as software libraries and checklists, should be provided to facilitate implementation.
- **Training** : Developers should be trained in the use of AI guidelines and ethical principles.
- **Regular Updates** : Guidelines should be updated regularly to reflect new challenges and technologies.

### 3.2.4 Stakeholder Involvement

The involvement of stakeholders, such as citizens, civil society organisations and public authorities, is essential to ensure that verification processes reflect the needs and values of society.

Stakeholder engagement helps ensure that verification processes are inclusive and representative. This approach promotes transparency and public trust.

- **Civic Panels** : Civic panels should be established to engage citizens in shaping AI rules and policies.
- **Public Consultations** : Public consultations should be organised to gather feedback and suggestions.
- **Balanced Representation** : Stakeholders should include representatives from diverse communities, including vulnerable groups.
- **Transparency** : The results of consultations must be published and made accessible to the public.

### 3.2.5 International Collaboration

International collaboration is essential to ensure that verification processes are harmonized globally and to promote common ethical standards.

International collaboration helps avoid fragmentation and unfair competition by promoting a coordinated approach to AI regulation. This includes sharing best practices and creating global standards.

- **International Forums** : International forums should be established to discuss and harmonize regulations.
- **Global Standards** : Global standards for the verification and certification of AI systems must be developed.
- **Sharing Resources** : Countries should share resources, such as audit tools and guidelines, to facilitate implementation.
- **Collaboration Agreements** : Collaboration agreements must be established between countries and international organizations.

### 3.3 The Threat of Accelerationism

Accelerationism in AI development, fueled by the race between superpowers for technological dominance, represents one of the greatest threats to the ethical and decision-making safety of AI systems. This approach, focused on speed and competition, drastically reduces the possibility that safety processes and ethical principles are adequately developed and implemented. Below, we delve into the risks and actions needed to counteract this trend.

#### 3.3.1 Risks of Accelerationism

Accelerationism leads to the premature release of AI systems, often without adequate safety and ethical controls. This approach can have disastrous consequences for society and human rights.

The pressure to be first to market pushes companies and governments to neglect ethical and safety considerations, releasing AI systems that have not been adequately tested or validated. This increases the risk of errors, privacy violations and discrimination.

- **Premature Release** : AI systems are released before they are fully tested, increasing the risk of errors and harmful decisions.  
Without thorough testing, AI systems can produce unpredictable or dangerous results, especially in critical contexts such as healthcare or safety.
- **Lack of Ethical Controls** : Ethical aspects, such as transparency and non-discrimination, are often overlooked in the pursuit of accelerated development. This leads to systems that perpetuate bias or violate fundamental rights, undermining public trust.

- **Security Compromised** : Rushing development can lead to technical vulnerabilities, exposing AI systems to cyberattacks. Unsecured systems can be manipulated by malicious actors, with potentially disastrous consequences.

### *3.3.2 Unfair Competition and Lack of Standards*

The race between superpowers creates an environment where unfair competition prevails over cooperation, making it difficult to adopt common ethical and safety standards.

Competition for technological dominance forces countries and companies to ignore ethical and safety standards in favor of speed and efficiency. This creates an environment where those who play by the rules are disadvantaged.

- **Divergent Standards** : Countries take different approaches to AI regulation, creating fragmentation and inconsistency, but without common standards, it is difficult to ensure that AI systems comply with ethical principles globally.
- **Speed-Based Competition** : Companies are incentivized to release products quickly, often at the expense of quality and safety. This approach favors those who ignore ethical controls, creating a distorted market.
- **Lack of Cooperation** : Superpower competition reduces the possibility of international collaboration on ethical and security issues. Without collaboration, it is difficult to develop global solutions to AI challenges.

### *3.3.3 Impact on Ethical and Decisional Security*

Accelerationism undermines the ethical and decision-making safety of AI systems, increasing the risk of harmful or discriminatory decisions.

Rushing development reduces the ability to build ethical and safety checks into code, increasing the risk that AI systems will make harmful decisions or violate human rights.

- **Decisions Not Aligned with Human Values** : AI systems can make decisions that do not respect fundamental rights or ethical values. This is particularly problematic in contexts such as justice or healthcare, where decisions directly impact people.
- **Increased Bias and Discrimination** : The lack of ethical controls increases the risk that AI systems perpetuate bias and discrimination,

which can lead to social and economic injustices, amplifying existing inequalities.

- **Reduced Transparency** : Accelerationism often leads to opaque AI systems that are difficult to understand or verify. Without transparency, it is impossible to ensure that AI decisions are fair and accountable.

#### *3.3.4 Actions Needed to Counter Accelerationism*

To counter accelerationism, it is essential to adopt a global and coordinated approach that promotes collaboration and responsibility.

Necessary actions include creating international standards, promoting collaboration between countries, and adopting policies that incentivize the ethical development of AI.

- **International Standards** : Global standards for AI safety and ethics must be developed, such standards must be adopted by all countries and companies, ensuring consistency and accountability.
- **Global Collaboration** : Countries must work together to address AI challenges in a coordinated manner, including sharing best practices, resources, and tools for ethical AI development.
- **Incentives for Ethical Development** : Governments must create incentives for companies to develop ethical and safe AI systems, including funding, tax breaks, or priority in public procurement.
- **Education and Awareness** : It is essential to promote education and awareness on the risks of accelerationism with training programs for developers, policy makers and citizens.

## 4. Concrete Actions for an Ethical and Human- Centric AI

To translate ethical principles into practice and ensure that AI is developed and used responsibly, a concrete and structured action plan is needed. This plan is divided into four main pillars, each of which includes specific actions and clear responsibilities. Below, we delve deeper into each pillar with detailed descriptions and explanatory subsections.

### 4.1 A Complete Regulatory Framework

A comprehensive regulatory framework is essential to ensure that AI systems comply with ethical and safety principles. This framework must include binding regulations, technical standards and certification mechanisms.

#### 4.1.1 Binding Regulations

Regulations must be legally binding and applicable to all AI systems, regardless of the sector or context of use. They must cover aspects such as human oversight, transparency, non-discrimination and data protection.

**Clear Definition of Requirements** : Regulations must clearly specify the ethical and safety requirements that AI systems must meet.

**Enforcement Mechanisms** : Enforcement bodies must be established to enforce the regulations and sanction violations.

**Regular Updates** : Regulations must be updated regularly to take into account new technologies and emerging challenges.

#### 4.1.2 Technical Standards

Technical standards are essential to ensure that AI systems are designed and implemented consistently and safely. These standards must cover aspects such as technical robustness, algorithm transparency, and data management.

**Global Standards** : Standards should be developed in collaboration with international organizations to ensure global consistency.

**Mandatory Adoption** : The standards must be mandatorily adopted by all AI developers and operators.

**Support Tools** : Tools and resources must be provided to help developers comply with standards.

#### 4.1.3 Ethical Certifications

Ethical certifications are a way to publicly recognize AI systems that comply with ethical and safety principles. These certifications must be issued by independent bodies and based on rigorous criteria.

**Clear Criteria** : Certifications must be based on clear and measurable criteria, such as absence of bias , transparency and security.

**Verification Process** : Certifications must be issued only after an independent and thorough verification process.

**Advertising** : Certified companies must be able to use a recognizable logo or brand to communicate their ethical commitment.

## 4.2 Global Collaboration

Global collaboration is essential to ensure that ethical and safety principles are respected internationally. This pillar includes the creation of international forums, the development of global standards and the sharing of resources.

### 4.2.1 International Forums

International forums are platforms for discussing and harmonizing AI regulations. These forums should involve representatives of governments, companies, civil society organizations and experts.

**Inclusive Participation** : Forums must include representatives from all countries and communities, ensuring that decisions reflect global needs.

**Clear Agenda** : Forums must have a clear agenda and focus on specific issues, such as security, transparency and non-discrimination.

**Tangible Results** : Forums must produce tangible results, such as international agreements or common guidelines.

### 4.2.2 Global Standards

Global standards are essential to ensure consistency and accountability in the development and use of AI. These standards must cover aspects such as safety, transparency and data protection.

**Collaborative Development** : Standards should be developed in collaboration with international organizations, governments and companies.

**Mandatory Adoption** : The standards must be mandatorily adopted by all countries and companies.

**Monitoring and Updating** : Standards must be monitored and updated regularly to take into account new technologies and emerging challenges.

### 4.2.3 Resource Sharing

Sharing resources, such as audit tools, guidelines, and best practices, is essential to facilitate the implementation of ethical and security principles.

**Sharing Platforms** : Online platforms should be created to share resources and tools across countries and organizations.

**Training and Support** : Training and support programs must be provided to help developers and policy makers use the resources.

**Open Access** : Resources must be accessible to all, ensuring that developing countries can also benefit from them.

## 4.3 Education and Awareness

Education and awareness are essential to ensure that all stakeholders involved in the development and use of AI understand the ethical and safety principles. This pillar includes training programs, awareness campaigns, and public outreach.

### 4.3.1 Training Programs

Training programs must be developed to educate developers, policy makers and citizens on the ethical and safety principles of AI.

**Comprehensive Content** : Training programs must cover all aspects of AI, from design to implementation.

**Accessibility** : Training programs must be accessible to all, including developing countries.

**Regular Updates** : Training programs should be updated regularly to reflect new technologies and emerging challenges.

### 4.3.2 Awareness Campaigns

Awareness campaigns should be launched to increase public awareness of the risks and opportunities of AI.

**Clear Messages** : Campaigns must use clear and understandable messages to reach a wide audience.

**Audience Engagement** : Campaigns must engage the public through events, workshops and discussions.

**Impact Evaluation** : The effectiveness of campaigns should be evaluated regularly to ensure they are achieving their intended objectives.

### 4.3.3 Initiatives for the Public

Public initiatives, such as citizen panels and consultations, are essential to ensure that decisions about AI reflect society's values and needs.

**Inclusive Participation** : Initiatives must involve representatives of all communities, including vulnerable groups.



**Transparency** : The results of the initiatives must be published and made accessible to the public.

**Policy Impact** : Initiatives must have a tangible impact on AI policies and regulations.

## 4.4 Public Participation

Public participation is essential to ensure that decisions on AI reflect society's values and needs. This pillar includes the creation of citizen panels, public consultations and feedback mechanisms.

### 4.4.1 Civic Panels

Citizen panels are groups of citizens tasked with discussing and providing feedback on AI policies and regulations.

**Balanced Representation** : Panels should include representatives from all communities, ensuring that decisions reflect global needs.

**Adequate Training** : Participants must be trained on the ethical and safety principles of AI.

**Impact on Decisions** : Panel feedback must have a tangible impact on policy decisions.

### 4.4.2 Public Consultations

Public consultations are processes to gather feedback and suggestions from citizens on AI policies and regulations.

**Accessibility** : Consultations must be accessible to all, including citizens with disabilities or limited access to technology.

**Transparency** : The results of consultations must be published and made accessible to the public.

**Concrete Responses** : Policy makers must respond to the feedback received, explaining how it has been integrated into decisions.

### 4.4.3 Feedback Mechanisms

Feedback mechanisms are tools to collect and analyze public opinion on the use and impact of AI.

**Ease of Use** : Feedback mechanisms must be easy to use and accessible to all.

**Data Analysis** : The feedback collected must be analyzed to identify trends and common problems.

**Timely Responses** : Policy makers must respond to feedback in a timely manner, demonstrating that the public's views are taken into account.

## 5. A Call to Action

The future of AI depends on the decisions we make today. Without timely action, we risk creating systems that threaten human rights, amplify inequalities, and compromise the safety of our societies.

We ask academic, political and institutional authorities to join us in this commitment. By signing this manifesto, you commit to:

1. Promote ethical and human- centric AI .
2. Support the adoption of a comprehensive regulatory framework that includes ethical and decision-making safety controls.
3. Collaborate globally to ensure that AI is developed and used responsibly.

The European Union has the opportunity to be a global leader in AI ethics. Let's act now to ensure that AI is a force for good for humanity.

## 6. Conclusion

AI is a powerful technology, but its impact depends on how we develop and use it. This manifesto is a call to action to ensure that AI serves humanity, respecting fundamental rights and promoting social well-being.

Join us in this effort. Together, we can build a future where AI is ethical, transparent, and human- centric .

**Signed,**

[The academic, political and institutional authorities that adhere to the manifesto]