Social implications and ethics challenges of artificial intelligence

Maria Isabel Aldinhas Ferreira
Centre of Philosophy of the University of Lisbon
and
Institute for Robots and Intelligent Systems/IST
University of Lisbon
Lisbon. Portugal

isabelferreira@letras.ulisboa.pt


INTELLIGENCE

- The capacity of any natural or artificial system to evolve in the environment it is embedded in:

- by identifying patterns to which it assigns a meaning
- by (re)acting accordingly
- by learning and adapting to changes according to the present and the previous experiences.

Humans are intelligent

Animals are intelligent

Organisms are intelligent

Artificial systems can be intelligent

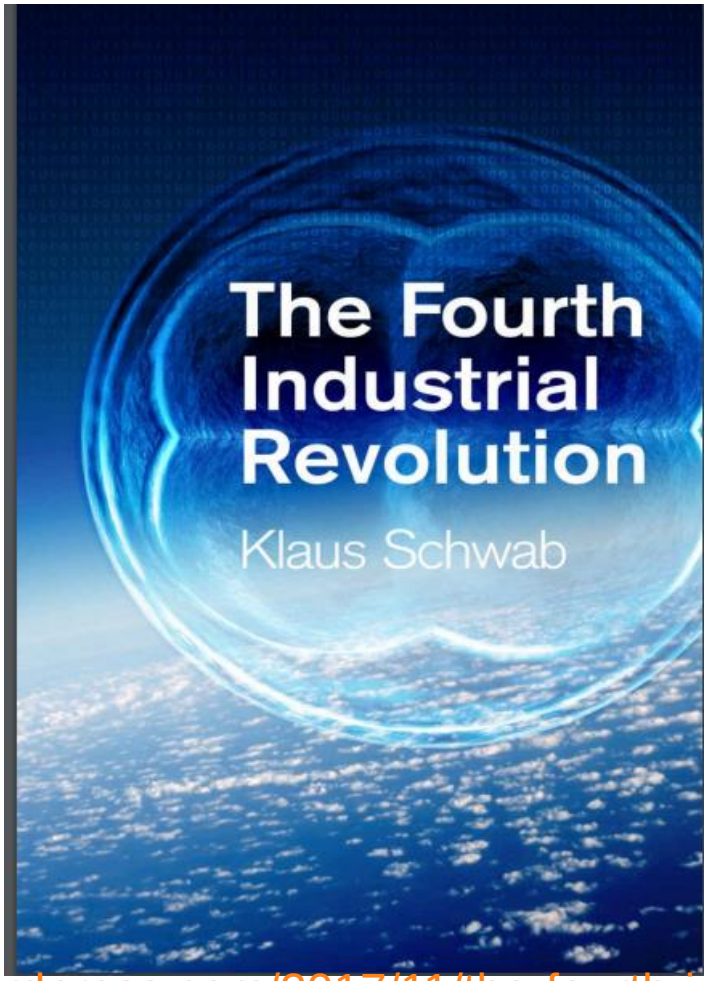
- Natural Intelligence is embodied

- Artificial Intelligence can be either embodied or non-embodied

- Either embodied AI (e.g. robots) or non embodied systems will be shaped according to a human vision of the world and will be embedded in a physical or in a virtual semiosphere – a defined setting, a social economic and cultural context that corresponds to a specific civilizational framework where particular values are actualised.

Artificial intelligence (AI) is no longer sci-fi. From driverless cars to the use of machine learning to improve healthcare services, the financial industry, business……, AI is shaping our daily practices as well as a fast-growing number of fundamental aspects of our societies.

We are at the beginning of a revolution that is fundamentally changing the way we live, work, and relate to one another.(2016:7)



https://luminariaz.files.wordpress.com/2017/11/the-fourth-industrial-revolution-2016-21.pdf

- **Velocity**: Contrary to the previous industrial revolutions, this one is evolving at an exponential rather than linear pace.

- **Breadth and depth**: It builds on the digital revolution and combines multiple technologies that are leading to unprecedented paradigm shifts in the economy, business, society, and individually.

- **Systems Impact**: It involves the transformation of entire systems, across (and within) countries, companies, industries and society as a whole

- We do not yet know exactly how the transformations driven by this industrial revolution will unfold

But

- their complexity and interconnectedness across sectors imply that all stakeholders of global society – governments, business, academia, and civil society – have a responsibility to work together to better understand the emerging trends and antecipate and prevent risks

AI is not merely another utility that needs to be regulated. It is at the core of a profound transformation of our habitat- which is progressively becoming a hybrid world where the digital permeates the physical where natural and artificial intelligence coexist.

The deep ecological nature of the present technological revolution requires its future to be supported by a clear socio-political design, grounded on a regulative ideal that places human beings dignity and well-being at its centre.

Because AI poses fundamental questions concerning its ethical, social, and economic impact several actions from academia, civil society and policy makers have been taking place:

In 2016

• The White House Office of Science and Technology Policy (OSTP) released the US report on AI, entitled '**Preparing for the Future of Artificial Intelligence.**

•On 31st May, 2016, the European Parliament's Committee on Legal Affairs (JURI) published the draft report on **Civil Law Rules on Robotics.**

- Artificial Intelligence has the potential to help address some of the biggest challenges that society faces. Smart vehicles may save hundreds of thousands of lives every year worldwide, and increase mobility for the elderly and those with disabilities. Smart buildings may save energy and reduce carbon emissions. Precision medicine may extend life and increase quality of life. Smarter government may serve citizens more quickly and precisely, better protect those at risk, and save money. AI-enhanced education may help teachers give every child an education that opens doors to a secure and fulfilling life. **These are just a few of the potential benefits if the technology is developed with an eye to its benefits and with careful consideration of its risks and challenges".**
- In PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE – The Obama Report (Oct 2016)

The report focuses on the impacts of AI-driven automation on the US job market and economy*. It hypothesises that low wage and middle income workers are most likely to be negatively impacted by AI, and that the government should develop public policy to ensure that AI does not increase economic inequality. It presents three specific policy responses to the perceived impact of AI on the US economy:

(1) Invest in and develop AI for its many benefits;

(2) Educate and train Americans for the jobs of the future;

(3) Aid workers in the transition and empower workers to ensure broadly shared growth.

*The impact on the job market is interpreted to be so significant that, in December 2016, a new report entitled ‘‘Artificial Intelligence, Automation and the Economy’’ was published by a team from the White House Executive Office of the President

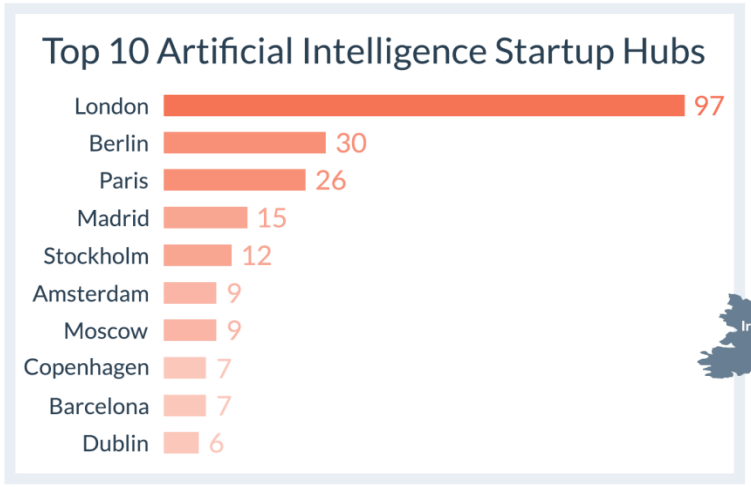
However according to the report regulation of AI should happen in a light-handed fashion, and, where applicable, the government should aim to fit AI into existing regulatory schemes, for example in the automotive and aviation industries. However, the report also calls upon the relevant agencies to ensure that—in evolving regulation on the basis of existing schemes—they ‘‘remain mindful of the fundamental purposes and goals of regulation to safeguard the public good, while creating space for innovation and growth in AI’’.

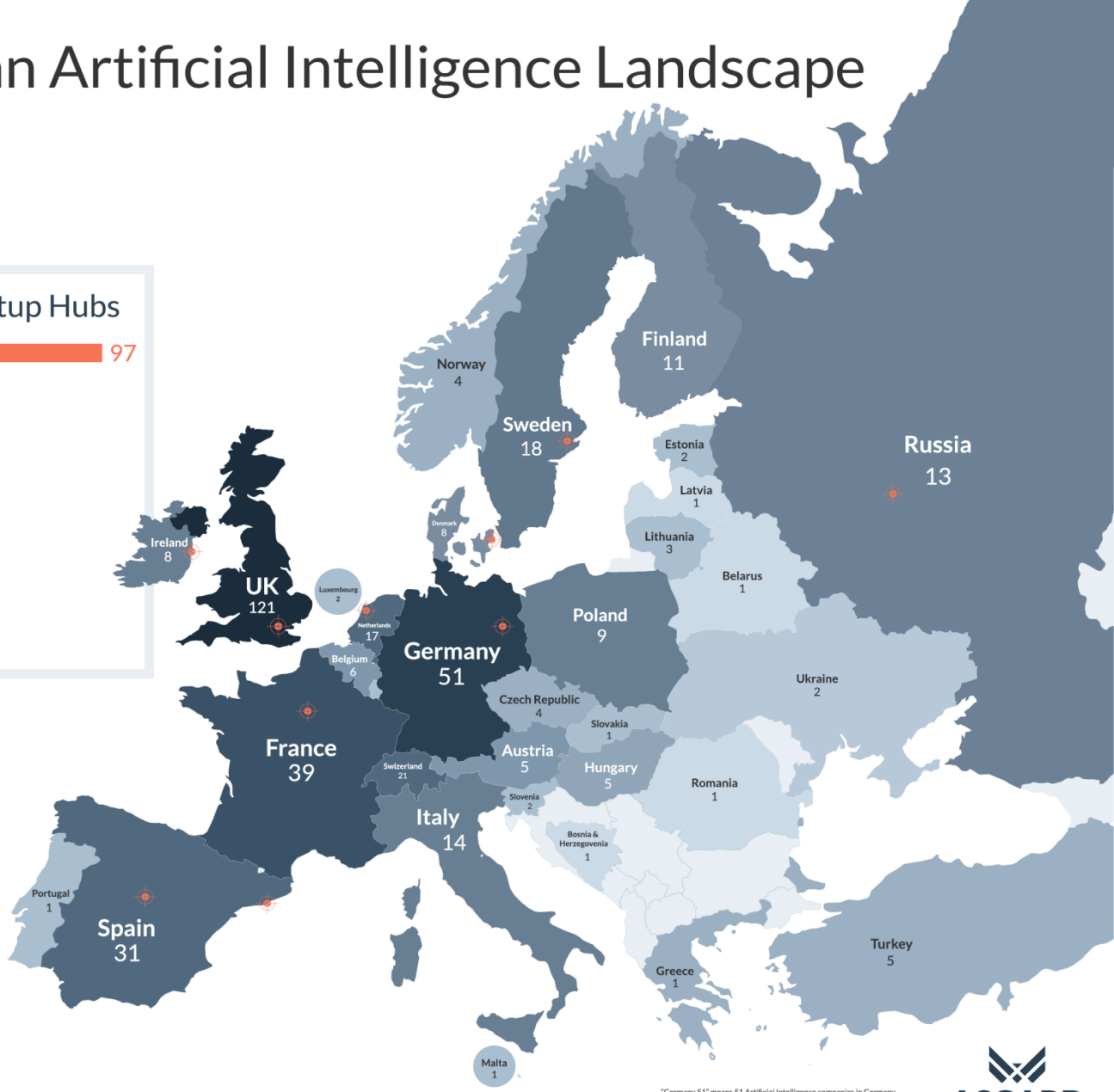
The US government’s vision of its own role, as a regulator, is limited and focused on ensuring that it does not hinder the development of AI technologies, ‘‘allowing a thousand flowers to bloom”

The European AI Road Map

- Countries within the EU have been emphasizing the importance of joining forces and showing a unified “European AI Alliance”. This strategy is seen as a means to increase competitiveness with countries like the United States and China, and also as a means to ensure the respect of “European values”.

The European Artificial Intelligence Landscape





2017 | Berlin | Asgard - Human Venture Capital for Artificial Intelligence | www.asgard.vc

https://medium.com/cityai/the-europea[illegible]ial-intelligence-landscape-more-than-400-ai-companies-build-in-europe-bd17a3d499b

The European AI Road Map

The 2016 report of the European Commission calls for the creation of a “European Agency for Robotics and AI” consisting of regulators and external technical and ethical experts, who can monitor AI and robotics-based trends, identify standards for best practice, recommend regulatory measures, define new principles, and address potential consumer protection issues

According to the report, this Agency will also manage an EU-wide registration system for all smart robots

The European AI Road Map

Additionally, the report calls for the creation of “a guiding ethical framework for the design, production and use” of AI and robotics, “based on the principles of beneficence, non-maleficence and autonomy, as well as on principles enshrined in the EU Charter of Fundamental Rights, such as human dignity and human rights, equality, justice and equity, non-discrimination and non-stigmatisation, autonomy and individual responsibility, informed consent, privacy and social responsibility.”

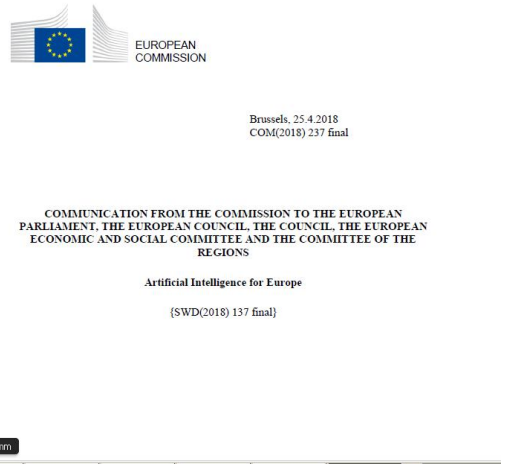
This approach clearly envisions a role for governments and policy-makers in setting a long-term strategy for the ‘good AI society’, instead of leaving it to industry and the research sector.

The European AI Road Map

- The European Council of October 2017 sets out a European initiative on AI, which aims to:
- ▪ Boost the EU's technological and industrial capacity and AI uptake across the economy, both by the private and public sectors.
- ▪ Prepare for socio-economic changes brought about by AI by encouraging the modernisation of education and training systems, anticipating changes in the labour market, supporting labour market transitions and adaptation of social protection systems.
- ▪ Ensure an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the EU and the development of AI ethics guidelines

The European AI Road Map

In April 2018, the EU Commission adopted the Communication on Artificial Intelligence for Europe: a 20-page document that lays out the EU's approach to AI.


EUROPEAN COMMISSION

Brussels, 25.4.2018
COM(2018) 237 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

Artificial Intelligence for Europe

{SWD(2018) 137 final}

I
In Embracing Change (Introduction)

Like the steam engine or electricity in the past, AI is transforming our world, our society and our industry

The way we approach AI will define the world we live in. A solid European framework is needed.

The European AI Road Map

- The EU Commission again sets as goals to:
- (1) increase the EU's technological and industrial capacity and AI uptake by the public and private sectors;
- (2) prepare Europeans for the socioeconomic changes brought about by AI;
- (3) ensure that an appropriate ethical and legal framework is in place.

The European AI Road Map

- Key initiatives include a commitment to increase the EU's investment in AI from €500 million in 2017 to €1.5 billion by the end of 2020, the creation of the European AI Alliance (which people can now join), and a new set of AI ethics guidelines to address issues such as fairness, safety, and transparency. A High-Level Group on Artificial Intelligence will act as the steering group for the European AI Alliance.

The European AI Road Map

- In March 2018, the European Commission established a High-Level Expert Group to gather expert input and develop guidelines for AI ethics

The European AI Road Map

- On 10 April 2018, 24 Member States and Norway committed to working together on AI in order to ensure that:
- **Europe is competitive in the AI landscape**
- **No one is left behind in the digital transformation** -a **New Skills Agenda for Europe** http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0381
- **New technologies are based on values** and fundamental rights as well as ethical principles such as accountability and transparency. The **General Data Protection Regulation (May 2018)** ensures a high standard of personal data protection, including the principles of data protection by design and by default.
- Within this legislative framework, assumes particular importance the principle of transparency as stated in the Article 22: "In particular, the controller must allow for a human intervention and the right for individuals to express their point of view, *to obtain further information about the decision that has been reached on the basis of this automated processing*, and the right to contest this decision."

The European AI Road Map

Building Trust in Human-Centric AI

- The High-Level Expert Group published ethical guidelines for trustworthy AI in April 2019

The European AI Road Map

Trust worthy AI has three components, which should be met throughout the system's entire life cycle:

- (1) it should be lawful, complying with all applicable laws and regulations
- (2) it should be ethical, ensuring adherence to ethical principles and values
- (3) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm. Each component in itself is necessary but not sufficient for the achievement of Trustworthy AI. Ideally, all three components work in harmony and overlap in their operation. If, in practice, tensions arise between these components, society should endeavour to align them.
- These Guidelines set out a framework for achieving Trustworthy

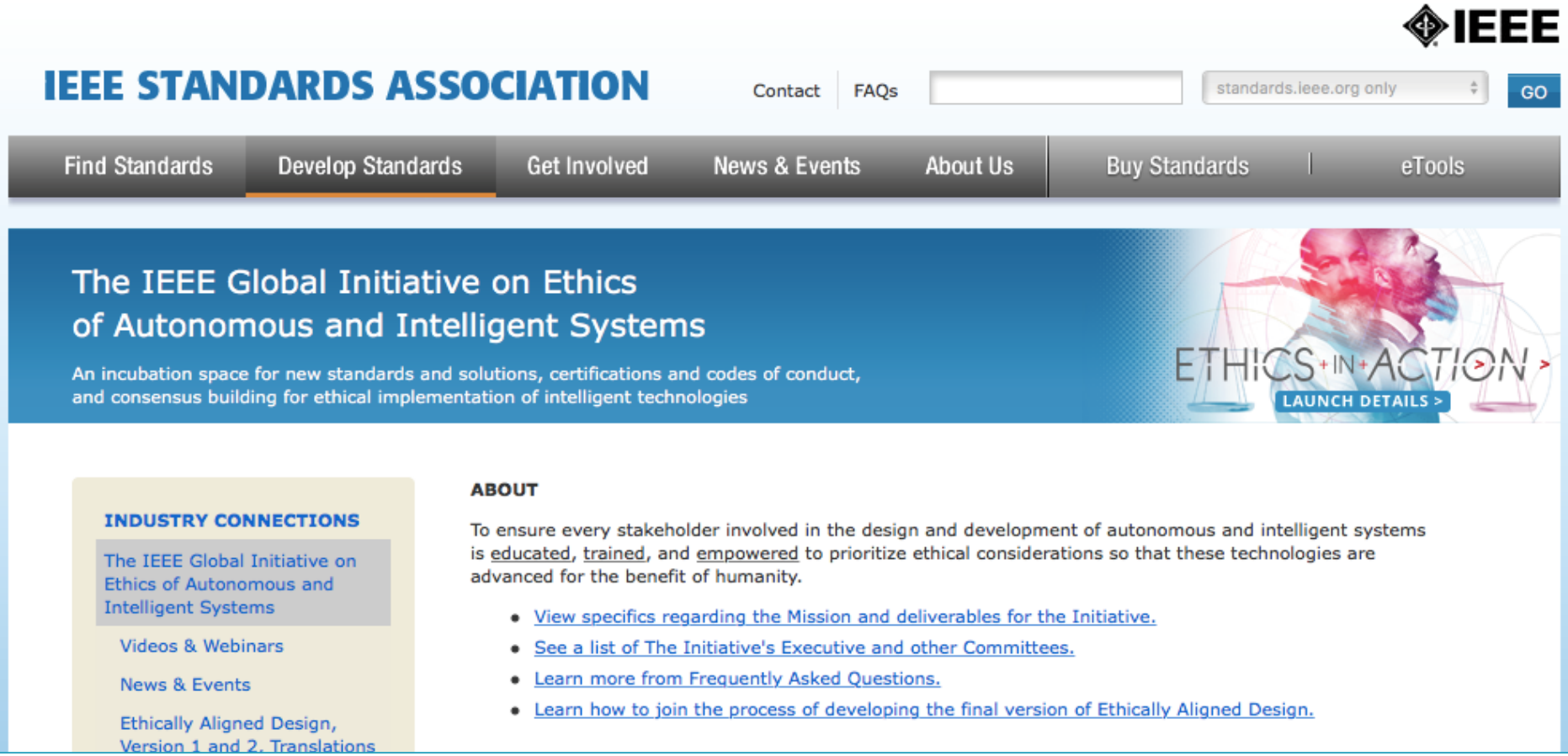
- Shaping the fourth industrial revolution to ensure that it is empowering and human-centred, rather than divisive and dehumanizing, is not a task for any single stakeholder or sector or for any one region, industry or culture.
- The fundamental and global nature of this revolution means it will affect and be influenced by all countries, economies, sectors and people.

THE IEEE INITIATIVE

- The Institute of Electrical and Electronic Engineers' (IEEE) Global Initiative on Ethics of Autonomous and Intelligent Systems (A/IS) was launched in April 2016 to incorporate ethical aspects of human well-being that may not automatically be considered in the current design and manufacture of A/IS technologies, and to reframe the notion of success so that human progress can include the intentional prioritization of individual, community and societal ethical values.

THE IEEE INITIATIVE

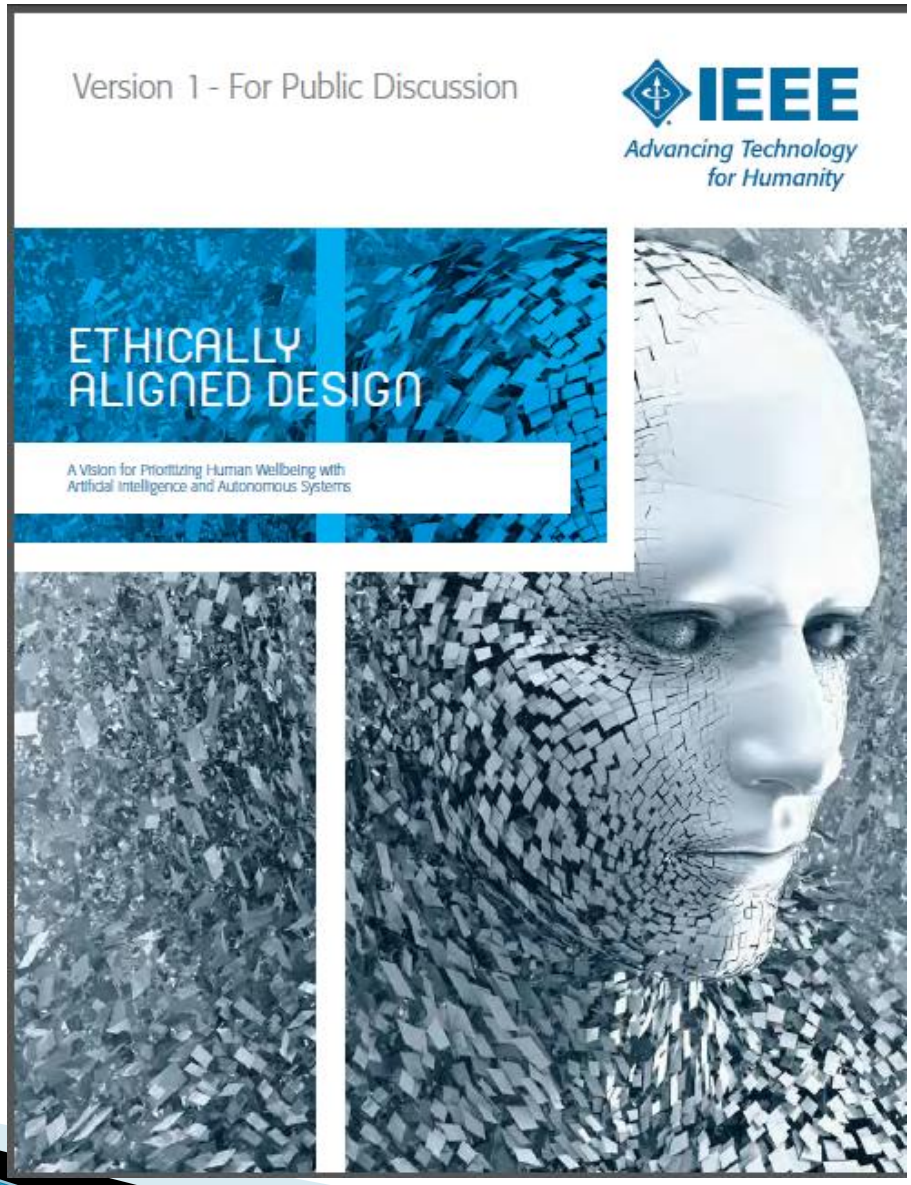
- The initiative seeks to ensure that every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained and allowed to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity



The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (April 2016)

Mission: To ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity

***Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*,.**



Ethically Aligned Design:

- Goal:

  to create ethical principles for AI/AS that:

1. Embody the highest ideals of human rights. (Charter of Human Rights…..)

2. Prioritize the maximum benefit to humanity and the natural environment. (Prioritize Well-Being)

3. Mitigate risks and negative impacts as AI/AS evolve as socio-technical systems.

Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems,.

- Principle 1 – Human Benefit
- 1. AI/AS should be designed and operated in
- a way that respects human rights, freedoms,
- human dignity, and cultural diversity.

- 2. AI/AS must be verifiably safe and secure
- throughout their operational lifetime.

- 3. If an AI/AS causes harm it must always
- be possible to discover the root cause
- (*traceability) for said harm (see also Principle*
- *3 – Transparency*

Ethically Aligned Design

- Principle 2 – Responsibility
- To best address issues of responsibility:
- 1. Legislatures/courts should clarify issues of responsibility for autonomous and intelligent systems where possible during development and deployment (to free manufacturers and users to understand what their rights and obligations should be).

- 2. Designers and developers of autonomous and intelligent systems should remain aware of, and take into account when relevant the diversity of existing cultural norms among the groups of users of these AI/AS.
- ….

Ethically Aligned Design

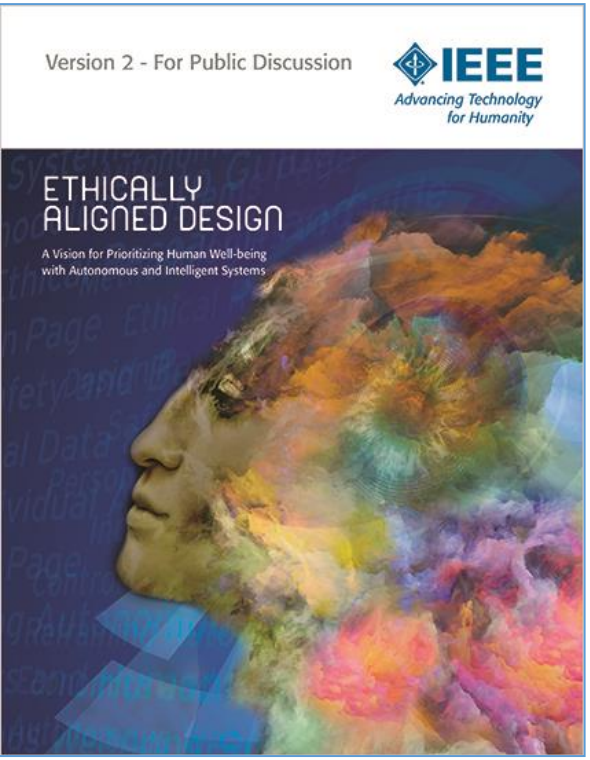
- Principle 3 – Transparency

- a *transparent AI/AS is one in which it is possible* to discover how and why the system made a particular decision, or in the case of a robot, acted the way it did.

Ethically Aligned Design

- Principle 4 – Education and Awareness
- 1. Providing ethics education and security awareness that sensitizes society to the potential risks of misuse of AI/AS.
- 2. Delivering this education in new ways, beginning with those having the greatest impact that also minimize generalized (e.g.,
- non-productive) fear about AI/AS (e.g., via accessible science communication on social media such as Facebook or YouTube).
- 3. Educating law enforcement surrounding these issues so citizens work collaboratively with them to avoid fear or confusion (e.g., in the same way police officers have given public safety lectures in schools for years, in the near future they could provide workshops on safe AI/AS).

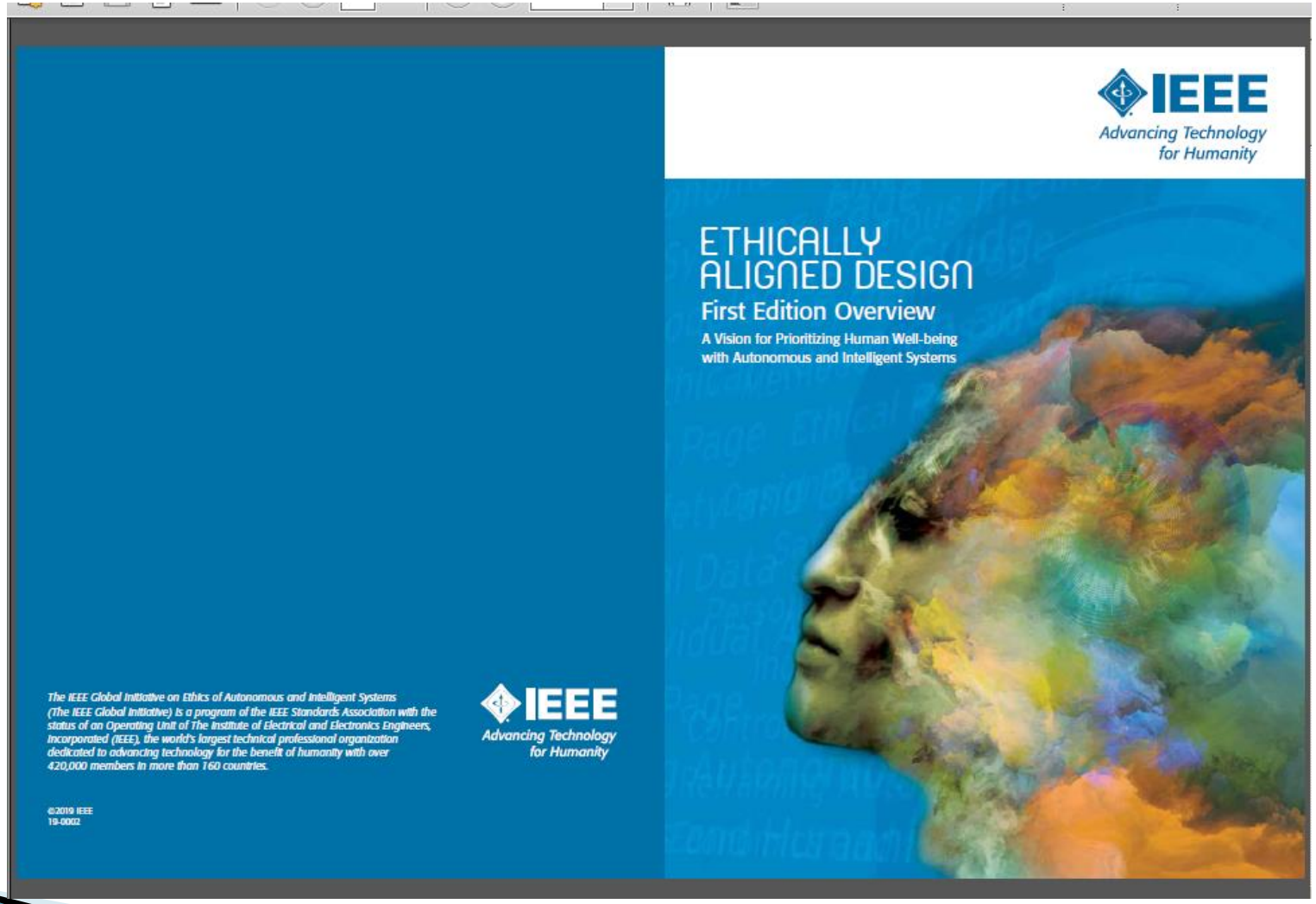
Ethically Aligned Design–version 2



Version 2

- Launched December 2017 as a Request for Input
- Created by over 250 Global A/IS & Ethics professionals, in a bottom up, transparent, open and increasingly globally inclusive process.
- Incorporates over 200 pages of feedback from public RFI and new Working Groups from China, Japan, Korea and more.
  ~ 300 pages very pertinent input received for EAD V2.
- Thirteen Committees / Sections
- Contains **over one hundred twenty** key Issues and Candidate Recommendations

Ethically Aligned Design- First Edition



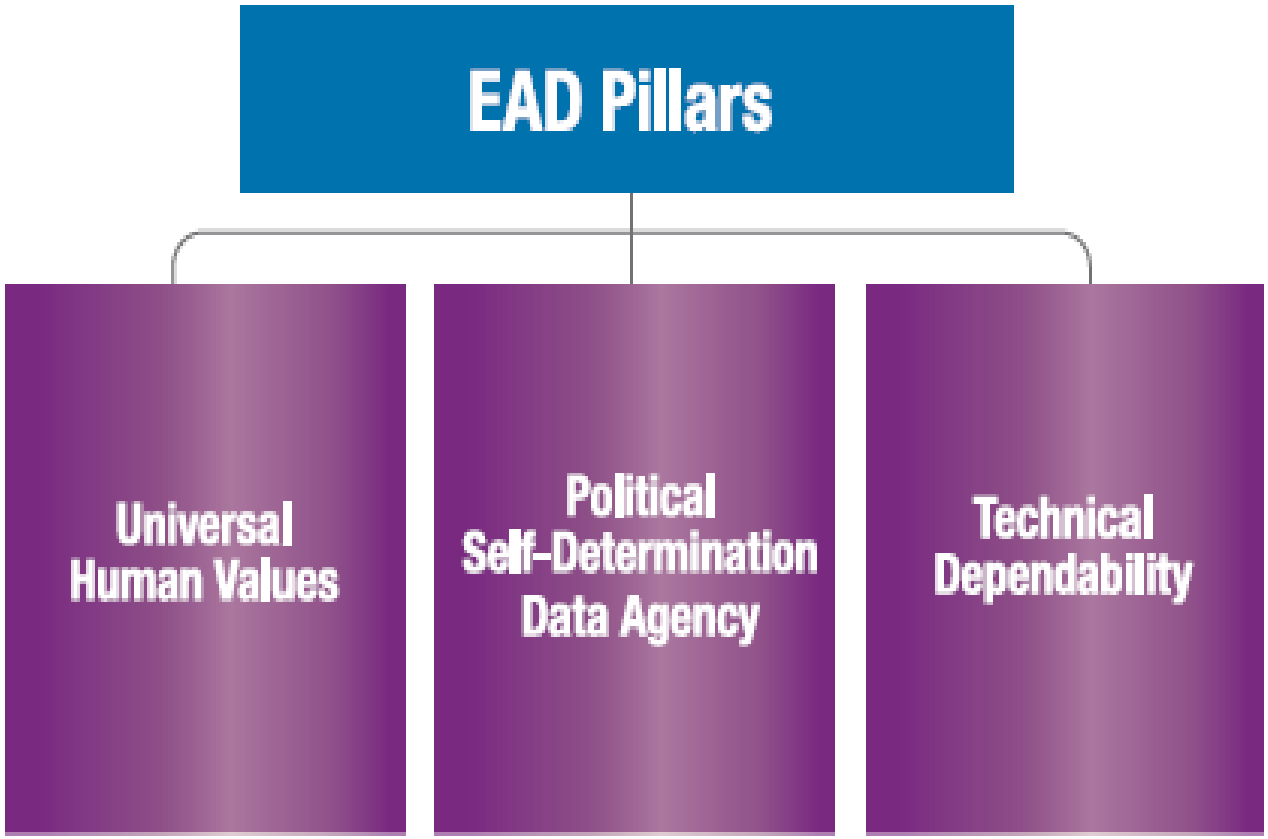
https://ethicsinaction.ieee.org/

The Three Pillars of the *Ethically Aligned Design* Conceptual Framework

The Pillars of the *Ethically Aligned Design* Conceptual Framework fall broadly into three areas, reflecting anthropological, political, and technical aspects:

1. **Universal Human Values:** A/IS can be an enormous force for good in society provided they are designed to respect human rights, align with human values, and holistically increase well-being while empowering as many people as possible. They should also be designed to safeguard our environment and natural resources. These values should guide policy makers as well as engineers, designers, and developers. Advances in A/IS should be in the service of all people, rather than benefiting solely small groups, a single nation, or a corporation.
2. **Political Self-Determination and Data Agency:** A/IS—if designed and implemented properly—have a great potential to nurture political freedom and democracy, in accordance with the cultural precepts of individual societies, when people have access to and control over the data constituting and representing their identity. These systems can improve government effectiveness and accountability, foster trust, and protect our private sphere, but only when people have agency over their digital identity and their data is provably protected.
3. **Technical Dependability:** Ultimately, A/IS should deliver services that can be trusted.[2] This trust means that A/IS will reliably, safely, and actively accomplish the objectives for which they were designed while advancing the human-driven values they were intended to reflect. Technologies should be monitored to ensure that their operation meets predetermined ethical objectives aligning with human values and respecting codified rights. In addition, validation and verification processes, including aspects of explainability, should be developed that could lead to better auditability and to certification[3] of A/IS.

Ethically Aligned Design- First Edition



Ethically Aligned Design- First Edition

- II. General Principles
- The ethical and values-based design, development, and implementation of autonomous and intelligent
- systems should be guided by the following General Principles:
- 1. Human Rights
- A/IS shall be created and operated to respect, promote, and protect internationally recognized
- human rights.
- 2. Well-being
- A/IS creators shall adopt increased human well-being as a primary success criterion for development.
- 3. Data Agency
- A/IS creators shall empower individuals with the ability to access and securely share their data,
- to maintain people's capacity to have control over their identity.
- 4. Effectiveness
- A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
- 5. Transparency
- The basis of a particular A/IS decision should always be discoverable.
- 6. Accountability
- A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
- 7. Awareness of Misuse
- A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
- 8. Competence
- A/IS creators shall specify and operators shall adhere to the knowledge and skill required
- for safe and effective operation.

THE IEEE INITIATIVE

It has two primary outputs:

- the creation and iteration of a body of work known as Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems;
- the identification and recommendation of ideas for standards projects focused on prioritizing ethical considerations in A/IS.

IEEE P7000 Standard Series

Model Process for Addressing Ethical Concerns During System Design

Purpose

- Help to create a shared mission around values and value priorities.
- Help to assure value based system engineering, by building a bridge between the value mission and the actual development of a complex technological system.

- **IEEE P7000™** - Model Process for Addressing Ethical Concerns During System Design
- **IEEE P7001™** - Transparency of Autonomous Systems
- **IEEE P7002™** - Data Privacy Process
- **IEEE P7003™** - Algorithmic Bias Considerations
- **IEEE P7004™** - Standard on Child and Student Data Governance
- **IEEE P7005™** - Standard on Employer Data Governance
- **IEEE P7006™** - Standard on Personal Data AI Agent Working Group
- **IEEE P7007™** - Ontological Standard for Ethically driven Robotics and Automation Systems
- **IEEE P7008™** - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- **IEEE P7009™** - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
- **IEEE P7010™** - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems
- **IEEE P7011™** - Standard for the Process of Identifying and Rating the Trustworthiness of News Sources
- **IEEE P7012™** - Standard for Machine Readable Personal Privacy Terms
- **IEEE P7013™** - Inclusion and Application Standards for Automated Facial Analysis Technology.

To fully benefit from all this potential, we need to go beyond prioritizing exponential growth in developing AI applications and develop them in full respect of human values, simultaneously promoting the education of all the stakeholders involved according to ethical frameworks that stand on unquestionable human rights.

Thank you for your attention!

References:

- https://luminariaz.files.wordpress.com/2017/11/the-fourth-industrial-revolution-2016-21.pdf
- Ethically Aligned Design- First Editionhttps://ethicsinaction.ieee.org/
- **Preparing for the Future of Artificial Intelligence.** https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- **Civil Law Rules on Robotics-** http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html
- https://www.consilium.europa.eu/media/21620/19-euco-final-conclusions-en.pdf
- Communication on Artificial Intelligence: for Europe https://ec.europa.eu/knowledge4policy/publication/communication-artificial-intelligence-europe_en
- **General Data Protection Regulation (May 2018)** https://gdpr-info.eu/
- ethical guidelines for trustworthy AI in April 2019