Introduction

As the use and impact of autonomous and intelligent systems (A/IS) become pervasive, we need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity's values and ethical principles. These systems must be developed and should operate in a way that is beneficial to people and the environment, beyond simply reaching functional goals and addressing technical problems. This approach will foster the heightened level of trust between people and technology that is needed for its fruitful use in our daily lives.

To be able to contribute in a positive, non-dogmatic way, we, the techno-scientific communities, need to enhance our self-reflection. We need to have an open and honest debate around our explicit or implicit values, including our imaginary[1] around so-called "Artificial Intelligence" and the institutions, symbols, and representations it generates.

Ultimately, our goal should be *eudaimonia*, a practice elucidated by Aristotle that defines human well-being, both at the individual and collective level, as the highest virtue for a society. Translated roughly as "flourishing", the benefits of eudaimonia begin with conscious contemplation, where ethical considerations help us define how we wish to live.

Whether our ethical practices are Western (e.g., Aristotelian, Kantian), Eastern (e.g., Shinto, 墨家/School of Mo, Confucian), African (e.g., Ubuntu), or from another tradition, honoring holistic definitions of societal prosperity is essential versus pursuing one-dimensional goals of increased productivity or gross domestic product (GDP). Autonomous and intelligent systems should prioritize and have as their goal the explicit honoring of our inalienable fundamental rights and dignity as well as the increase of human flourishing and environmental sustainability.

The goal of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems ("The IEEE Global Initiative") is that *Ethically Aligned Design* will provide pragmatic and directional insights and recommendations, serving as a key reference for the work of technologists, educators and policymakers in the coming years.

Ethically Aligned Design sets forth scientific analysis and resources, high-level principles, and actionable recommendations. It offers specific guidance for standards, certification, regulation or legislation for design, manufacture, and use of A/IS that provably aligns with and improves holistic societal well-being.

[1]The symbols, values, institutions, and norms of a societal group through which people imagine their lives and constitute their societies.

Executive Summary

I. Purpose of *Ethically Aligned Design, First Edition* (EAD1e)

Autonomous and intelligent technical systems are specifically designed to reduce the necessity for human intervention in our day-to-day lives. In so doing, these new systems are also raising concerns about their impact on individuals and societies. Current discussions include advocacy for a positive impact, such as optimization of processes and resource usage, more informed planning and decisions, and recognition of useful patterns in big data. Discussions also include warnings about potential harm to privacy, discrimination, loss of skills, adverse economic impacts, risks to security of critical infrastructure, and possible negative long-term effects on societal well-being.

Because of their nature, the full benefit of these technologies will be attained only if they are aligned with society's defined values and ethical principles. Through this work we intend, therefore, to establish frameworks to guide and inform dialogue and debate around the non-technical implications of these technologies, in particular related to ethical aspects. We understand "ethical" to go beyond moral constructs and include social fairness, environmental sustainability, and our desire for self-determination.

Our analyses and recommendations in *Ethically Aligned Design* address values and intentions as well as implementations, both legal and technical. They are both aspirational, what we hope or wish should happen, and practical, what we—the techno-scientific community and every group involved with and/or affected by these technologies—could do for society to advance in positive directions. The analyses and recommendations in EAD1e are offered as guidance for consideration by governments, businesses, and the public at large in the advancement of technology for the benefit of humanity.

Chapters in *Ethically Aligned Design*, First Edition

1. From Principles to Practice
2. General Principles
3. Classical Ethics in A/IS
4. Well-being
5. Affective Computing
6. Personal Data and Individual Agency
7. Methods to Guide Ethical Research and Design
8. A/IS for Sustainable Development
9. Embedding Values into Autonomous and Intelligent Systems
10. Policy
11. Law

Introduction

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.

III. Ethical Foundations

Classical Ethics

By drawing from over two thousand five hundred years of classical ethics traditions, the authors of *Ethically Aligned Design* explored established ethics systems, addressing both scientific and religious approaches, including secular philosophical traditions, to address human morality in the digital age. Through reviewing the philosophical foundations that define autonomy and ontology, this work addresses the alleged potential for autonomous capacity of intelligent technical systems, morality in amoral systems, and asks whether decisions made by amoral systems can have moral consequences.

IV. Areas of Impact

A/IS for Sustainable Development

Through affordable and universal access to communications networks and the Internet, autonomous and intelligent systems can be made available to and benefit populations anywhere. They can significantly alter institutions and institutional relationships toward more human-centric structures, and they can address humanitarian and sustainable development issues resulting in increased individual societal and environmental well-being. Such efforts could be facilitated through the recognition of and adherence to established indicators of societal flourishing such as the United Nations Sustainable Development Goals so that human well-being is utilized as a primary success criteria for A/IS development.

Introduction

Personal Data Rights and Agency Over Digital Identity

People have the right to access, share, and benefit from their data and the insights it provides. Individuals require mechanisms to help create and curate the terms and conditions regarding access to their identity and personal data, and to control its safe, specific, and finite exchange. Individuals also require policies and practices that make them explicitly aware of consequences resulting from the aggregation or resale of their personal information.

Legal Frameworks for Accountability

The convergence of autonomous and intelligent systems and robotics technologies has led to the development of systems with attributes that simulate those of human beings in terms of partial autonomy, ability to perform specific intellectual tasks, and even a human physical appearance. The issue of the legal status of complex autonomous and intelligent systems thus intertwines with broader legal questions regarding how to ensure accountability and allocate liability when such systems cause harm. It is clear that:

- Autonomous and intelligent technical systems should be subject to the applicable regimes of property law.
- Government and industry stakeholders should identify the types of decisions and operations that should never be delegated to such systems. These stakeholders should adopt rules and standards that ensure effective human control over those decisions and how to allocate legal responsibility for harm caused by them.
- The manifestations generated by autonomous and intelligent technical systems should, in general, be protected under national and international laws.
- Standards of transparency, competence, accountability, and evidence of effectiveness should govern the development of autonomous and intelligent systems.

Policies for Education and Awareness

Effective policy addresses the protection and promotion of human rights, safety, privacy, and cybersecurity, as well as the public understanding of the potential impact of autonomous and intelligent technical systems on society. To ensure that they best serve the public interest, policies should:

- Support, promote, and enable internationally recognized legal norms.
- Develop government expertise in related technologies.
- Ensure governance and ethics are core components in research, development, acquisition, and use.
- Regulate to ensure public safety and responsible system design.
- Educate the public on societal impacts of related technologies.

Introduction

V. Implementation

Well-being Metrics

For autonomous and intelligent systems to provably advance a specific benefit for humanity, there need to be clear indicators of that benefit. Common metrics of success include profit, gross domestic product, consumption levels, and occupational safety. While important, these metrics fail to encompass the full spectrum of well-being for individuals, the environment, and society. Psychological, social, economic fairness, and environmental factors matter. Well-being metrics include such factors, allowing the benefits arising from technological progress to be more comprehensively evaluated, providing opportunities to test for unintended negative consequences that could diminish human well-being. A/IS can improve capturing of and analyzing the pertinent data, which in turn could help identify where these systems would increase human well-being, providing new routes to societal and technological innovation.

Embedding Values into Autonomous and Intelligent Systems

If machines engage in human communities as quasi-autonomous agents, then those agents must be expected to follow the community's social and moral norms. Embedding norms in such quasi-autonomous systems requires a clear delineation of the community in which they are to be deployed. Further, even within a particular community, different types of technical embodiments will demand different sets of norms. The first step is to identify the norms of the specific community in which the systems are to be deployed and, in particular, norms relevant to the kinds of tasks that they are designed to perform.

Methods to Guide Ethical Research and Design

To create autonomous and intelligent technical systems that enhance and extend human well-being and freedom, values-based design methods must put human advancement at the core of development of technical systems. This must be done in concert with the recognition that machines should serve humans and not the other way around. Systems developers should employ values-based design methods in order to create sustainable systems that can be evaluated in terms of not only providing increased economic value for organizations but also of broader social costs and benefits.

Affective Computing

Affect is a core aspect of intelligence. Drives and emotions such as anger, fear, and joy are often the foundations of actions throughout our lives. To ensure that intelligent technical systems will be used to help humanity to the greatest extent possible in all contexts, autonomous and intelligent systems that participate in or facilitate human society should not cause harm by either amplifying or dampening human emotional experience.

Acknowledgements

Our progress and the ongoing positive influence of this work are due to the volunteer experts serving on all our Committees and IEEE P7000™ Standards Working Groups, along with the IEEE professional staff who support our efforts. Thank you for your dedication toward defining, designing, and inspiring the ethical principles and standards that will ensure that autonomous and intelligent systems and the technologies associated with them will positively benefit humanity.

We wish to thank the Executive Committee and Committees of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems:

Executive Committee Officers

Raja Chatila, *Chair*
Kay Firth-Butterfield, *Vice Chair*
John C. Havens, *Executive Director*

Executive Committee Members

Dr. Greg Adamson, Karen Bartleson, Virginia Dignum, Danit Gal, Malavika Jayaram, Sven Koenig, Eileen M. Lach, Raj Madhavan, Richard Mallah, AJung Moon, Monique Morrow, Francesca Rossi, Alan Winfield, and Hagit Messer Yaron

Committee Chairs

- **General Principles:** Mark Halverson, and Peet van Biljon
- **Embedding Values into Autonomous Intelligent Systems:** Francesca Rossi and Bertram F. Malle
- **Methodologies to Guide Ethical Research and Design:** Raja Chatila and Corinne Cath
- **Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI):** Malo Bourgon and Richard Mallah
- **Personal Data and Individual Agency:** Katryna Dow and John C. Havens
- **Reframing Autonomous Weapons Systems:** Peter Asaro
- **Sustainable Development:** Elizabeth Gibbons
- **Law:** Nicolas Economou and John Casey
- **Affective Computing:** John Sullins and Joanna J. Bryson
- **Classical Ethics in A/IS:** Jared Bielby
- **Policy:** Peter Brooks and Mina Hannah
- **Extended Reality:** Monique Morrow and Jay Iorio
- **Well-being:** Laura Musikanski and John C. Havens
- **Editing:** Karen Bartleson and Eileen M. Lach
- **Outreach:** Maya Zuckerman and Ali Muzaffar
- **Communications:** Leanne Seeto and Mark Halverson
- **High School:** Tess Posner
- **Global Coordination:** Victoria Wang, Arisa Ema, Pavel Gotovtsev

Introduction

Programs and Projects Inspired by The IEEE Global Initiative:

- ***Ethically Aligned Design* University Consortium:** Hagit Messer, *Chair*
- ***Ethically Aligned Design* Community:** Lisa Morgan, Program Director, Content and Community
- **Ethics Certification Program for Autonomous and Intelligent Systems:** Meeri Haataja, Chair; Ali Hessami, *Vice-Chair*
- **Glossary:** Sara M. Jordan, *Chair*

People

We would like to warmly recognize the leadership and constant support of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems by Dr. Ing. Konstantinos Karachalios, Managing Director of the IEEE Standards Association.

We would also like to thank Stephen Welby, Executive Director and Chief Operating Officer of IEEE for his generous and insightful support of the *Ethically Aligned Design*, First Edition process and The IEEE Global Initiative overall.

We would especially like to thank Eileen M. Lach, the former IEEE General Counsel and Chief Compliance Officer, whose heartfelt conviction that there is a pressing need to focus the global community on highlighting ethical considerations in the development of autonomous and intelligent systems served as a strong catalyst for the development of the Initiative within IEEE.

Finally, we would like to also acknowledge the ongoing work of three Committees of The IEEE Global Initiative regarding their chapters of *Ethically Aligned Design* that, for timing reasons, we were not able to include in *Ethically Aligned Design*, First Edition. These Committees include: Reframing Autonomous Weapons Systems, Extended Reality (formerly Mixed Reality) and Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI). We would like to thank Peter Asaro, Monique Morrow and Jay Iorio, Malo Bourgon and Richard Mallah for their leadership in these groups along with all their Committee Members. Once these chapters have completed their review and been accepted by IEEE they could either be included in *Ethically Aligned Design*, published by The IEEE Global Initiative, or in other publications of IEEE.

For information on disclaimers associated with EAD1e, see *How the Document Was Prepared*.