From Principles and Standards to Certification

John C. Havens, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Ali Hessami, Vega Systems

How worldwide collaboration among the thought leaders in autonomous and intelligent systems design and application are affecting this important emerging technology.

FROM THE EDITOR

In May 2016, this column discussed the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, which was created in March of that year. Much progress has occurred since that column was published. From the beginning, people with varied backgrounds and skills have engaged to create a body of work on the principles of designing autonomous and intelligent systems based on core ethical principles. Two significant outputs of this effort include a set of standards and a design document titled *Ethically Aligned Design*. Earlier drafts of *Ethically Aligned Design* were distributed in December of 2016 and 2017. *Ethically Aligned Design* is expected to be released in the spring of 2019. Several of the standards projects were initiated early in 2016 but have since grown to include 14 projects. –*F.D. Wright*

D esigning autonomous and intelligent systems is not purely a technical activity. Instead, engineers, scientists, and other designers must understand and consider the ethical considerations involved in their work and make appropriate choices to deal with situations and concerns often involving life and safety. The IEEE P7000 series of standards starts to address those considerations.

ETHICALLY ALIGNED DESIGN

Ethically Aligned Design, created by members of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (https://standards.ieee.org/industry-connections/ec/autonomous-systems.html), has, to date, resulted in 14 approved standardization projects and a major certification initiative. The standards projects, in the IEEE P7000 series, are in response to a host of needs and requirements arising from the extensive exploration of the ethically aligned design paradigm and cover a broad range of emergent features of autonomous and intelligent systems.

IEEE P7000 APPROVED STANDARDIZATION PROJECTS

The IEEE P7000 series of standards projects under development represents a unique addition to the collection of more than 1,900 global IEEE standards and projects. Whereas more traditional standards focus on technology interoperability, functionality, safety, and trade facilitation, the IEEE P7000 series addresses specific issues at the intersection of technological and ethical considerations. Like its technical standards counterparts, the IEEE P7000 series empowers innovation across borders and yields societal benefits.

For more information or to join any working group, please see the following list. The committees that authored *Ethically Aligned Design* as well as other committees within the IEEE that created specific working groups are listed below each project.

› IEEE Standard P7000, *Model Process for Addressing Ethical Concerns During System Design*, inspired by the Methodologies to Guide Ethical Research and Design Committee and sponsored by the IEEE Computer Society; standards.ieee.org /project/7000.html
› IEEE Standard P7001, *Transparency of Autonomous Systems,* inspired by the General Principles Committee and sponsored by the IEEE Vehicular Technology Society; standards.ieee.org /project/7001.html
› IEEE Standard P7002, *Data Privacy Process*, inspired by The Personal Data and Individual Agency Control Committee and sponsored by the IEEE Computer Society; standards.ieee.org /project/7002.html
› IEEE Standard P7003, *Algorithmic Bias Considerations*, sponsored by the IEEE Computer Society; standards.ieee.org/ project/7003.html
› IEEE Standard P7004, *Standard for Child and Student Data Governance*, inspired by the Personal Data and Individual Agency Control Committee and sponsored by the IEEE Computer Society; standards.ieee.org/project/7004.html
› IEEE Standard P7005, *Standard for Transparent Employer Data Governance*, inspired by the Personal Data and Individual Agency Control Committee and sponsored by the IEEE Computer Society; standards.ieee.org/project/7005.html
› IEEE Standard P7006, *Standard for Personal Data Artificial Intelligence (AI) Agent*, inspired by the Personal Data and Individual Agency Control Committee and sponsored by the IEEE Computer Society; standards.ieee.org/project/7006.html
› IEEE Standard P7007, *Ontological Standard for Ethically Driven Robotics and Automation Systems*, sponsored by the IEEE Robotics and Automation Society; standards.ieee.org/project/7007.html
› IEEE Standard P7008, *Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems*, inspired by the Affective Computing Committee and sponsored by the IEEE Robotics and Automation Society; standards.ieee.org/project/7008.html
› IEEE Standard P7009, *Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems*, sponsored by the IEEE Reliability Society; standards.ieee.org /project/7009.html
› IEEE Standard P7010, Well-Being *Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems*, inspired by the Well-being Committee and sponsored by the IEEE Systems, Man and Cybernetics Society; standards. ieee.org/project/7010 .html
› IEEE Standard P7011, *Standard for the Process of Identifying and Rating the Trustworthiness of News Sources*, sponsored by the IEEE Society on Social Implications of Technology; standards. ieee.org/project/7011.html
› IEEE Standard P7012, *Standard for Machine Readable Personal Privacy Terms*, sponsored by the IEEE Society on Social Implications of Technology; standards. ieee.org/project/7012.html
› IEEE Standard P7013, *Inclusion and Application Standards for Automated Facial Analysis Technology*, sponsored by the IEEE Society on Social Implications of Technology; standards.ieee.org /project/7013.html.

Although individual standards projects have been initiated due to the necessities and requirements arising from the IEEE Global Initiative, there is a need for cross-pollination and a common connecting thread between and among all of the P7000 projects. A new initiative aimed at coordinating and ensuring a degree of alignment and consistency between the work of separate working groups has launched. It is anticipated that this coordination work will result in the sharing of best practices while aligning the individual standards as members of a coherent and cohesive family regarding the ethical aspects of autonomous and intelligent systems. This implies having a common reference model for the realization of ethical properties for autonomous and intelligent systems according to the specific focus and perspective for each member of the P7000 family. The basis for this common reference point is the generic ethical design reference model devised as a backbone to the IEEE P7000 standard for addressing ethical concerns in system design.

CERTIFICATION PROGRAM

The other key initiative related to *Ethically Aligned Design* and the IEEE P7000 series is the Ethics Certification Program for Autonomous and Intelligent

Systems (ECPAIS) (https://standards .ieee.org/industryconnections/ecpais .html). This initiative is driven by the realities and requirements that arise from public and private enterprises procuring and implementing autonomous and intelligent products and systems. The initial remit for the scope of work is to develop audit processes and transparency certification, accountability, and freedom from algorithmic bias. Achieving these objectives requires collaboration with the P7000 series and adoption of the generic processes in the core P7000 project as the reference model for the development of the requisite certification processes. This is intended to accelerate the initiation of the ECPAIS work while also providing a common thread between the three certification processes that are likely to be extended in number or scope.

The ECPAIS Working Group currently comprises public and private members. The work of process and criteria development is being carried out in three expert focus groups. Expert competence criteria have been developed to ensure that these focus groups are resourced by volunteers with the appropriate knowledge, understanding, skills, experience, and qualities pertinent to each group's aims and objectives. The current program for ECPAIS work anticipates the initial draft delivery of the three certification processes by the end of 2019, with trials beginning by early 2020.

FROM PRINCIPLES TO PRACTICE

In addition to creating the IEEE P7000 series of standards and *Ethically Aligned Design*, the IEEE Global Initiative, independently or in cooperation with the IEEE Standards Association, has directly inspired the following works and activities:

› *Artificial Intelligence and Ethics in Design*: These 10 courses are designed for global professionals as well as their managers working in engineering, IT, computer science, big data, artificial intelligence, and related fields across all industries who require up-to-date information on the latest technologies. The courses explicitly mirror content from *Ethically Aligned Design* and feature numerous experts as instructors who helped create *Ethically Aligned Design*. More information can be found at innovationatwork.ieee.org /courses/artificial-intelligence-and-ethics-in-design/.

› *The creation of an autonomous and intelligent system ethics glossary*: The glossary features more than 200 pages of terms that help define the context of autonomous and intelligent system ethics for multiple stakeholder groups, specifically, engineers, policy makers, philosophers, standards developers, and computational disciplines experts. It is currently in its second iteration and input is being provided by the IEEE P7000 standards working groups. The glossary can be downloaded from standards. ieee.org/content/dam/ieee-standards/standards/web /documents/other/ead1e_ glossary.pdf.

› *The launch of the Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS)*: The IEEE Standards Association, inspired by the work of the IEEE Global Initiative, has significantly contributed to the establishment of OCEANIS, a global forum for discussion, debate, and collaboration for organizations interested in the development and use of standards to further the creation of autonomous and intelligent systems. OCEANIS members are working together to enhance the understanding of the role of standards in facilitating innovation while addressing problems that expand beyond technical solutions to approaching ethics and values. More information can be found at ethicsstandards .org/.

› *The launch of the Council on Extended Intelligence (CXI)*: CXI was directly inspired by the work of the IEEE Global Initiative and the MIT Media Lab around extended intelligence and was launched jointly by the IEEE Standards Association and the MIT Media Lab. CXI's mission is to proliferate the ideals of responsible participant design, data agency, and metrics of economic prosperity, prioritizing people and the planet over profit and productivity. Membership includes thought leaders from the European Union Parliament and Commission, the U.K. House of Lords, the Organisation for Economic Cooperation and Development, the United Nations, local and national administrations, and renowned experts in economics, data science, and multiple other disciplines from around the world. Visit global-cxi.org for more information.

› *The launch of the Ethically Aligned Design University Consortium (EADUC)*: EADUC is being established with the aim of reaching engineers at the beginning of their studies to help them prioritize values-driven, applied ethical principles at the core of their

> The work of process and criteria development is being carried out in three expert focus groups.

work. Working in conjunction with philosophers, designers, social scientists, academics, data scientists, and the corporate and policy communities, EADUC also has the goal that *Ethically Aligned Design* will be used in teaching at all levels of education globally as the new vision for design in the algorithmic age.

› *The launch of AI Commons*: The work of the IEEE Global Initiative has delivered key ideas and inspiration that are rapidly evolving toward establishing a global collaborative platform around autonomous and intelligent systems. The mission of AI Commons is to gather a true ecosystem to democratize access to AI capabilities and, thus, allow anyone, anywhere, to benefit from the possibilities that AI can provide. In addition, the group will be working to connect problem owners with the community

of solvers to collectively create solutions with AI. The ultimate goal is to implement a framework for participation and cooperation to make using and benefiting from AI available to all. More information can be found at www .aicommons.com.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems was formed to ensure that 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. The standards projects and certification programs emerging from this global effort have already been recognized around the world. As these projects become standards, it is hoped that future implementers will have the resources

necessary to make the right decisions in designing their autonomous and intelligent systems. **C**

JOHN C. HAVENS is the executive director of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the Council on Extended Intelligence. Contact him at john.havens.us@ieee.org.

ALI HESSAMI is the director of Innovation and Research and Development at Vega Systems UK. Contact him at Hessami@VegaGlobal Systems.com.

ERRATUM

In [1], the biography for Jeffrey Voas was incorrect. It should have read

Jeffrey Voas is with the National Institute of Standards and Technology. He is an IEEE Fellow. Contact him at j.voas@ieee.org.

REFERENCE

1. R. Kuhn, D. Yaga, and J. Voas, "Rethinking distributed ledger technology," *Computer*, vol. 52, no. 2, pp. 68–72, 2019.