IEEE P70xx, Establishing Standards for Ethical Technology

Ansgar Koene
University of Nottingham
United Kingdom
ansgar.koene@nottingham.ac.uk

Adam Leon Smith
Piccadilly Labs
United Kingdom
adamsmith@piccadillygroup.com

Takashi Egawa
NEC corporation
Japan
t-egawa@ct.jp.nec.com

Sukanya Mandalh
IEEE
India
sukanya@ieee.org

Yohko Hatada
EMLS RI
United Kingdom
y.hatada@emlsri.org

ABSTRACT

In recognition of the increasingly pervasive role of algorithmic decision making systems in corporate and government service, and growing public concerns regarding the black-box nature of many of these systems, the IEEE Standards Association (IEEE-SA) launched the IEEE Global Initiative on Ethics for Autonomous and Intelligence Systems in April 2016. The 'Global Initiative' aims to provide "an incubation space for new standards and solutions, certifications and codes of conduct, and consensus building for ethical implementation of intelligent technologies". As of early 2018 the two main pillars of the Global Initiative are:

- a public discussion document "Ethically Aligned Design: A vision for Prioritizing human Well-being with Autonomous and Intelligent Systems" [2], on establishing ethical and social implementations for intelligent and autonomous systems and technology aligned with values and ethical principles that prioritize human well-being in a given cultural context;
- a set of thirteen working groups to create the IEEE P70xx series ethics standards, and associated certification programs, for Intelligent and Autonomous systems.

  This project showcase will give the KDD community insights into the aims and working practices of the IEEE P70xx "Ethical Technology Standards" working groups, with an open invitation to join our efforts for establishing guiding frameworks to achieve Technology For Humanity.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Software and its engineering** → **Software creation and management**; **Requirements analysis**; **Risk management**; **Empirical software validation**; **Documentation**; • **Social and professional topics** → *Computing / technology policy*;

KDD'18, August, 2018, ExCeL London UK

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

KEYWORDS

Standards, Bias, Ethics, Algorithms

ACM Reference format:
Ansgar Koene, Adam Leon Smith, Takashi Egawa, Sukanya Mandalh, and Yohko Hatada. 2018. IEEE P70xx, Establishing Standards for Ethical Technology. In *Proceedings of KDD, ExCeL London UK, August, 2018 (KDD'18),* 2 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

In recognition of the increasingly pervasive role of algorithmic decision making systems in corporate and government service, and growing public concerns regarding the black-box nature of many of these systems, the IEEE Standards Association (IEEE-SA) launched the IEEE Global Initiative on Ethics for Autonomous and Intelligence Systems [1] in April 2016. The 'Global Initiative' aims to provide "an incubation space for new standards and solutions, certifications and codes of conduct, and consensus building for ethical implementation of intelligent technologies". As of early 2018 the two main pillars of the Global Initiative are:

- a public discussion document "Ethically Aligned Design: A vision for Prioritizing human Well-being with Autonomous and Intelligent Systems" [2], on establishing ethical and social implementations for intelligent and autonomous systems and technology aligned with values and ethical principles that prioritize human well-being in a given cultural context;
- a set of thirteen working groups to create the IEEE P70xx series ethics standards, and associated certification programs, for Intelligent and Autonomous systems.

The IEEE P70xx series of ethical technology standards aims to translate the principles that are discussed in the Ethically Aligned Design document into actionable guidelines or frameworks that can be used as practical industry standards. The thirteen IEEE P70xx standards that are currently under development are:

- 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

A brief paper outlining the aims of IEEE P7003 and its relationship to the other IEEE P70xx series standards working groups was published in [3] and a tech-industry oriented summary of the IEEE P70xx series standards appeared on the technology-industry blog TechEmergence [4]. IEEE P7003 is aimed to be used by people/organizations who are developing and/or deploying automated decision (support) systems (which may or may not involve AI/machine learning) that are part of products/services that affect people. Typical examples would include anything related to personalization or individual assessment, including any system that performs a filtering function by selecting to prioritize the ease with which people will find some items over others (e.g. search engines or recommendation systems). Any system that will produce different results for some people than for others is open to challenges of being biased. Examples could include:

- Image processing systems and their ability to accurately recognise facial characteristics, including facial recognition and tracking, and applications that, for example, detect theft or suspicious behaviour, or flag to law enforcement that an individual should be detained.
- Marketing automation applications that calibrate offers, prices, or content to an individual's preferences and behaviour. Examples have occured where well remunerated executive job adverts are shown to men, but not women with similar experience [5].
- Online purchasing systems can use location data (e.g. zip code) to make different products and services available at different prices, to different groups of customers, who may be predominantly from a specific ethnic group, age group or socio-economic class.

As a practical example, an online retailer developing a new product recommendation system might use the IEEE P7003 standard as follows:

Early in the development cycle, after outlining the intended functions of the new system IEEE P7003 guides the developer through a process of considering the likely customer groups, in order to identify if there are subgroups that will need special consideration (e.g. people with visual impairments). In the next phase of the development, the developer is establishing a testing dataset to validate if the system is performing as desired. Referencing P7003 the developer is reminded of certain methods for checking if all customer groups are sufficiently represented in the testing data to avoid reduced quality of service for certain customer groups.

Throughout the development process IEEE P7003 challenges the developer to think explicitly about the criteria that are being used for the recommendation process and the rationale, i.e. justification, for why these criteria are relevant and why they are appropriate (legally and socially). Documenting these will help the business respond to possible future challenges from customers, competitors or regulators regarding the recommendations produced by this system. At the same time, this process of analysis will help the business to be aware of the context for which this recommendation system can confidently be used, and which uses would require additional testing (e.g. age ranges of customers, types of products).

REFERENCES

[1] IEEE. The ieee global initiative on ethics of autonomous and intelligent systems. URL https://ethicsinaction.ieee.org/.
[2] IEEE. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, version 2, 2017. URL http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.
[3] L. Dowthwaite A. Koene and S. Seth. Ieee p7003 standard for algorithmic bias considerations. URL https://doi.org/10.1145/3194770.3194773.
[4] Daniel Fagella. The ethics of artificial intelligence for business leaders - should anyone care? URL https://www.techemergence.com/ethics-artificial-intelligence-business-leaders/.
[5] Anupam Datta Amit Datta, Michael Carl Tschantz. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. URL https://arxiv.org/abs/1408.6491.