Digital Platforms and Social Harms

What Standards Can Do

Annette Reilly

annette.reilly@computer.org

October 14, 2024

IEEE COMPUTER SOCIETY

IEEE

IEEE SA STANDARDS ASSOCIATION

IEEE

WORLD STANDARDS DAY

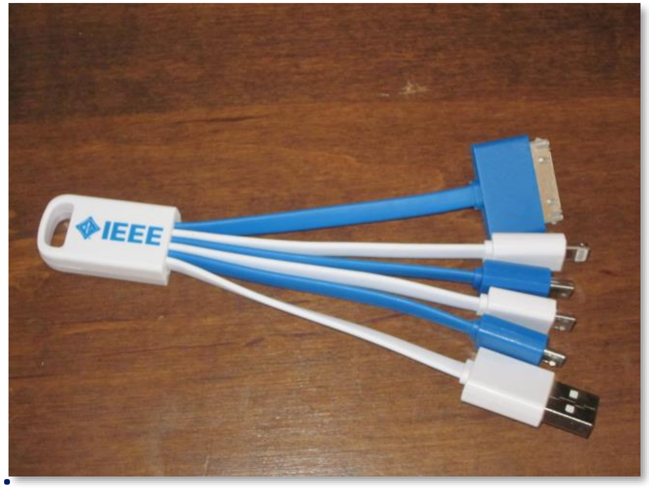
14 OCTOBER 2024

Building an Equitable Future for Health and Well-Being

Before We Share our Opinions……

- “At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that their views should be considered the personal views of that individual rather than the formal position, explanation, or interpretation of the IEEE.”
- IEEE-SA Standards Board Operation Manual (subclause 5.9.3)

Why have standards?



- Support interoperability
- Support reliability
- Further world trade
- Promote consistent products
- Support consistent principles and policies
- Allow repeatable processes and process improvement
- Basis for contracts and audits

- *Voluntary standards or laws and regulations*
  - *Conformance or compliance*
  - *Standards have requirements for conformance ("shall" statements or imperatives)*
  - *Conformance to requirements can be verified or validated*
  - *Recommended practices and guides/guidelines have "should" statements*

When is it really a standard?

- "… [A] formal document that establishes uniform engineering or technical criteria, methods, processes and practices" (*Wikipedia*)
- Not proprietary, tool-bound, or vendor-specific
- The result of consensus agreement from a balance of stakeholders
- Maintained by a recognized, impartial standards-producing organization
- Normative (mandatory) or guidance?
- Open participation from all interested stakeholders
- Participation in IEEE standards working groups is open to anyone
  - Working group members do not have to join IEEE

Within IEEE, standards against digital harm mainly come from

- *Computer Society*

(Systems and Software Engineering and AI Standards Committees)

- *Society for Social Implications of Technology (SSIT)*

Why are ethically aligned standards important

IEEE-Advancing technology for humanity

- Engineers have always considered performance-related values
- Increasing awareness that cultural and moral values influence system acceptance and use
- Unintended effects of systems due to inattention to stakeholder values
  - Example: AI systems making decisions
- How to incorporate ethical values into the system design?
  - Other standards handle safety, security, health and environmental-related values
- Intended to be applicable to different cultures, all types of systems, all sizes of organizations
  - Intended for integrated project teams, applicable to any organizational model
- *Values are not developed independently of the other systems engineering work*

IEEE Standards Development Principles

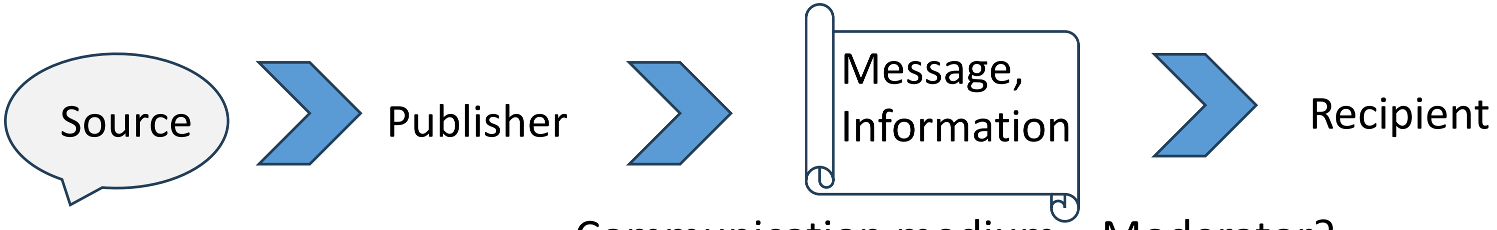
- Direct participation
- Due process
- Broad consensus
- Balance
- Transparency
- Broad openness
- Coherence

What can be standardized?

- Principles (values)
- Terms (vocabulary, taxonomy, ontology)
- People (certification)
- Processes and practices (purpose, activities, outcomes)
- Products

Use cases for standards against disinformation and digital harm

- Standardize delivery of trustworthy information
- Standardize non-delivery of hateful messages



Communication medium—Moderator?

- **What can be standardized?**

- Terminology—what is trustworthy, what is hate speech
- Values—privacy, fairness, freedom of speech, transparency, care, respect, control, inclusiveness, trust
  - Values can clash and not all values can be optimized in a system
- People– certifications, credentials
  - Who is responsible and accountable?
- Processes
  - Where in the communications process can disinformation be identified or stopped?
- Products
  - How can disinformation and deep fakes be identified?

IEEE's sustained efforts for standards against digital harm

- First IEEE Ethically Aligned Design standards projects approved in 2017:
  - IEEE 7000 series
- IEEE AI standards date from 2014, but most released since 2021
- Currently in development:
  - **57 IEEE AI standards**
  - 10 IEEE standards regarding ethical values and practices
  - 12 IEEE standards regarding trust
- Notable published standards:
  - ISO/IEC/IEEE 15026, Systems and software engineering--Systems and software assurance
  - IEEE Std 3527.1:2020, Standard for Digital Intelligence (DQ) -- Framework for Digital Literacy, Skills and Readiness
  - IEEE Std 7000: 2021, IEEE Standard Model Process for Addressing Ethical Concerns during System Design
  - IEEE Std 7010:2020, Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being
  - IEEE Std 7007:2021, IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems
  - IEEE Std 7009:2024, Fail-Safe Design of Autonomous and Semi-Autonomous Systems
  - **IEEE Std 7014:2024, Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems**
  - **IEEE Std 2089.1: 2024, Online Age Verification**

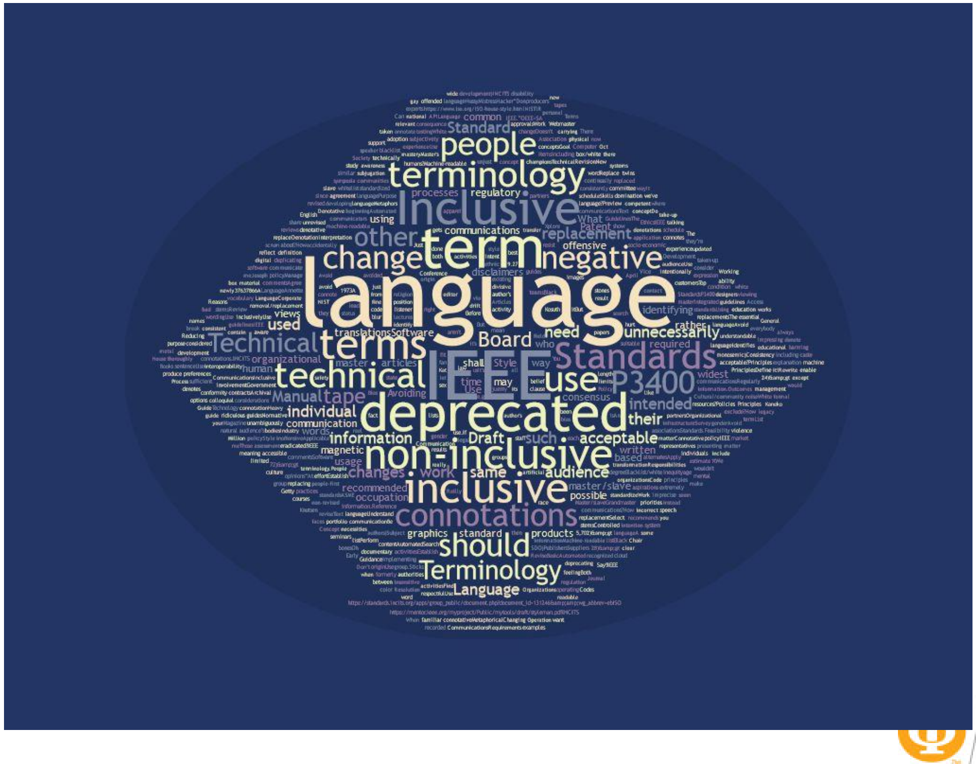
IEEE Std 7000:2021, Model Process for Addressing Ethical Concerns during System Design

- Identifying, analyzing, and resolving ethical issues early in the life cycle or for new/revised versions of products or services
- A process standard: Applicable during concept exploration, ethical values elicitation and prioritization, requirements definition, and design
- Emphasis on transparency and risk reduction
- Choice of values relevant to the culture where the system will be deployed

  - ***Improving the value proposition and reducing risk***

IEEE P3400 for inclusive terminology in technical communication

- **Principles** for identifying and replacing non-inclusive language
- **Process** to help assure inclusive terminology in an organization's technical communications
- Information **products** use inclusive terminology and exclude deprecated, offensive terms
- Pending approval by the balloting group



IEEE 2089:2021, Standard for Age Appropriate Digital Services Framework - Based on the 5Rights Principles for Children

- Protect children's personal data from uses that recommend content or behaviors detrimental to their rights and needs.

© Getty Images

a) recognition that the user is a child,

b) has considered the capacity and upholds the rights of children,

c) offers terms appropriate to children,

d) presents information in an age appropriate way

e) offers a level of validation for service design decisions.

- *Related to EU 's GDPR, the UK's Age Appropriate Design Code, the Australian e-safety commissioner's Safety by Design standards*

The 5Rights Principles:

Right to Remove
- The Right to Know
- The Right to Safety and Support
- The Right to Informed and Conscious Use
- The Right to Digital Literacy

Age appropriate values: sustainability, privacy, usability, convenience, controllability, accountability, inclusivity, evolving capacity, and children’s rights

IEEE P2089.2, Standard for Terms and Conditions for Children's Online Engagement

- **Published: IEEE Std 2089.1: 2024, Online Age Verification**

- **P2089.2 In development:** To improve children's understanding of the rules and limitations of their online interactions
- Provides children with essential information about their rights, responsibilities, and the platform's expectations
- Allows children to advocate for changes, report violations, and actively participate in shaping the platform's policies
- Terms and conditions for
  - (i) Age-appropriate digital content
  - (ii) Data sharing clauses
  - (iii) Artificial Intelligence (AI) access and manipulation of data
  - (iv) Transparency for data exchange
  - (v) Addiction to online services and related harmful components that cause exploitation and manipulation of children

Photo by Ruth Lewis

- Values: transparency and clarity, enhanced privacy protection, empowerment and agency, safer online experiences, and educational value

IEEE P3462, Recommended Practice for Using Safety by Design in Generative Models to Prioritize Child Safety

- **In development** to counter victim identification, victimization, prevention, and abuse proliferation
- Safety by design approach for developing, deploying, and maintaining generative artificial intelligence models with adequate safeguards against child sexual abuse
- Expands child safety to the entire lifecycle of machine learning (ML): development, deployment, and maintenance stages
- Applies to multiple data modalities (i.e. text, image, video, audio)

Get involved with standards

- Get a free IEEE account
- Have models for your process and products
- Use standards at work
  - Get IEEE standards at IEEE Xplore®
  - https://ieeexplore.ieee.org/Xplore/home.jsp
- Join or follow IEEE standards development: myProject
  - https://development.standards.ieee.org/myproject-web/app#
- Become a standards reviewer, balloter, or working group member
- Join
  - IEEE Standards Association
  - IEEE Computer Society
  - IEEE Society on Social Implications of Technology (SSIT)

Related projects in development (1)

- IEEE P7003, Algorithmic Bias Considerations
- IEEE P7004, Child and Student Data Governance
- IEEE P7008, Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- IEEE P7010.1, Recommended Practice for Environmental Social Governance (ESG) and Social Development Goal (SDG) Action Implementation and Advancing Corporate Social Responsibility
- IEEE P7011, Process of Identifying and Rating the Trustworthiness of News Sources
- IEEE P7012, Machine Readable Personal Privacy Terms
- IEEE P7014.1, Recommended Practice for Ethical Considerations of Emulated Empathy in Partner-based General-Purpose Artificial Intelligence Systems
- IEEE P7015, Data and Artificial Intelligence (AI) Literacy, Skills, and Readiness
- IEEE P7016, Ethically Aligned Design and Operation of Metaverse Systems
- IEEE P7016.1, Ethically Aligned Educational Metadata in Extended Reality (XR) & Metaverse
- IEEE P7017, Recommended Practice for Design-Centered Human-Robot Interaction (HRI) and Governance
- IEEE 7018, Security and Trustworthiness Requirements in Generative Pretrained Artificial Intelligence (AI) Models
- IEEE P7019, Implementation and Governance of Autonomous and Intelligent Systems Related to Earth Law Principles

Related projects in development (2)

- IEEE 7800, Recommended Practice for Addressing Sustainability, Environmental Stewardship and Climate Change Challenges in Professional Practice
- IEEE P8000, Standard for Characterization and Specification of Ethical Properties in Autonomous Intelligent Systems (AIS)
- IEEE P2863, Recommended Practice for Organizational Governance of Artificial Intelligence
- IEEE P2986, Standard for XAI – eXplainable Artificial Intelligence - for Achieving Clarity and Interoperability of AI Systems Design
- IEEE P3119, Standard for the Procurement of Artificial Intelligence and Automated Decision Systems
- IEEE P3376, Recommended Practice for Evaluating Artificial Intelligence Generated Content
- IEEE P3395, Standard for the Implementation of Safeguards, Controls, and Preventive Techniques for Artificial Intelligence (AI) Models
- IEEE P3396 Recommended Practice for Defining and Evaluating Artificial Intelligence (AI) Risk, Safety, Trustworthiness, and Responsibility
- IEEE P7700 Recommended Practice for the Responsible Design and Development of Neurotechnologies

IEEE Standards projects for ethics

Project Number	Project Title
P2247.4	Recommended Practice for Ethically Aligned Design of Artificial Intelligence (AI) in Adaptive Instructional Systems
P7030	Recommended Practice for Ethical Assessment of Extended Reality (XR) Technologies
P7016.1	Standard for Ethically Aligned Educational Metadata in Extended Reality (XR) & Metaverse
P7016	Standard for Ethically Aligned Design and Operation of Metaverse Systems
P8000	Standard for Characterization and Specification of Ethical Properties in Autonomous Intelligent Systems (AIS)
P7014.1	Recommended Practice for Ethical Considerations of Emulated Empathy in Partner-based General-Purpose Artificial Intelligence Systems
P7999	Standard for Integrating Organizational Ethics Oversight in Projects and Processes Involving Artificial Intelligence
P7999.1	Standard for Integrating Organizational Ethics Oversight in Projects and Processes Involving Artificial Intelligence – Qualification of Individuals
P7999.2	Standard for Integrating Organizational Ethics Oversight in Projects and Processes Involving Artificial Intelligence – Organizational Certification
P7008	Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

Published AI standards of IEEE Computer Society

Standard Number	Year	Project Title
2830	2021	IEEE Standard for Technical Framework and Requirements of Trusted Execution Environment based Shared Machine Learning
2937	2022	IEEE Standard for Performance Benchmarking for Artificial Intelligence Server Systems
2945	2023	IEEE Standard for Technical Requirements for Face Recognition
2986	2023	IEEE Recommended Practice for Privacy and Security for Federated Machine Learning
3129	2023	IEEE Standard for Robustness Testing and Evaluation of Artificial Intelligence (AI)-based Image Recognition Service
3156	2023	IEEE Standard for Requirements of Privacy-Preserving Computation Integrated Platforms
3168	2024	IEEE Standard for Robustness Evaluation Test Methods for a Natural Language Processing Service That Uses Machine Learning
3187	2024	IEEE Approved Draft Guide for Framework for Trustworthy Federated Machine Learning
3652.1	2020	IEEE Guide for Architectural Framework and Application of Federated Machine Learning
2841	2022	IEEE Recommended Practice for Framework and Process for Deep Learning Evaluation
2894	2024	IEEE Guide for an Architectural Framework for Explainable Artificial Intelligence