IEEE SA
STANDARDS ASSOCIATION

IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems

STANDARDS

IEEE Society for Social Implications of Technology

Developed by the
Standards Committee

IEEE Std 7014™-2024

IEEE

IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems

Developed by the

Standards Committee
of the
IEEE Society for Social Implications of Technology

Approved 20 May 2024

IEEE SA Standards Board

Abstract: Guidance and actions for the ethical development, deployment, or decommission of autonomous and intelligent systems that attempt to emulate aspects of human empathy are provided by this standard.

Keywords: affect, affective computing, AI, AI ethics, artificial intelligence, autonomous systems, emotion, emotion recognition, empathy, ethics, IEEE 7014™, intelligent systems, machine learning, software engineering

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

PDF: ISBN 979-8-8557-0849-3 STD27025
Print: ISBN 979-8-8557-0850-9 STDPD27025

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE Standards documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page (https://standards.ieee.org/ipr/disclaimers.html), appear in all IEEE standards and may be found under the heading "Important Notices and Disclaimers Concerning IEEE Standards Documents."

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents are developed within IEEE Societies and subcommittees of IEEE Standards Association (IEEE SA) Board of Governors. IEEE develops its standards through an accredited consensus development process, which brings together volunteers representing varied viewpoints and interests to achieve the final product. IEEE standards are documents developed by volunteers with scientific, academic, and industry-based expertise in technical working groups. Volunteers involved in technical working groups are not necessarily members of IEEE or IEEE SA and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE makes no warranties or representations concerning its standards, and expressly disclaims all warranties, express or implied, concerning all standards, including but not limited to the warranties of merchantability, fitness for a particular purpose and non-infringement IEEE Standards documents do not guarantee safety, security, health, or environmental protection, or compliance with law, or guarantee against interference with or from other devices or networks. In addition, IEEE does not warrant or represent that the use of the material contained in its standards is free from patent infringement. IEEE Standards documents are supplied "AS IS" and "WITH ALL FAULTS."

Use of an IEEE standard is wholly voluntary. The existence of an IEEE standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity, nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document should rely upon their own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus balloting process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English language version published by IEEE is the approved IEEE standard.

Use by artificial intelligence systems

In no event shall material in any IEEE Standards documents be used for the purpose of creating, training, enhancing, developing, maintaining, or contributing to any artificial intelligence systems without the express, written consent of IEEE SA in advance. "Artificial intelligence" refers to any software, application, or other system that uses artificial intelligence, machine learning, or similar technologies, to analyze, train, process, or generate content. Requests for consent can be submitted using the Contact Us form.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE SA Standards Board Operations Manual is not, and shall not be considered or inferred to be, the official position of IEEE or any of its committees and shall not be considered to be, or be relied upon as, a formal position of IEEE or IEEE SA. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that the presenter's views should be considered the personal views of that individual rather than the formal position of IEEE, IEEE SA, the Standards Committee, or the Working Group. Statements made by volunteers may not represent the formal position of their employer(s) or affiliation(s). News releases about IEEE standards issued by entities other than IEEE SA should be considered the view of the entity issuing the release rather than the formal position of IEEE or IEEE SA.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE or IEEE SA. However, **IEEE does not provide interpretations, consulting information, or advice pertaining to IEEE Standards documents**.

Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its Societies and subcommittees of the IEEE SA Board of Governors are not able to provide an instant response to comments or questions, except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in evaluating comments or revisions to an IEEE standard is welcome to join the relevant IEEE SA working group. You can indicate interest in a working group using the Interests tab in the Manage Profile and Interests area of the IEEE SA myProject system.[1] An IEEE Account is needed to access the application.

Comments on standards should be submitted using the Contact Us form.[2]

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not constitute compliance to any applicable regulatory

[1]Available at: https://development.standards.ieee.org/myproject-web/public/view.html#landing.
[2]Available at: https://standards.ieee.org/about/contact/.

requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

Data privacy

Users of IEEE Standards documents should evaluate the standards for considerations of data privacy and data ownership in the context of assessing and using the standards in compliance with applicable laws and regulations.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under US and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, neither IEEE nor its licensors waive any rights in copyright to the documents.

Photocopies

Subject to payment of the appropriate licensing fees, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400; https://www.copyright.com/. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every 10 years. When a document is more than 10 years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit IEEE Xplore or contact IEEE.[3] For more information about the IEEE SA or IEEE's standards development process, visit the IEEE SA Website.

Errata

Errata, if any, for all IEEE standards can be accessed on the IEEE SA Website.[4] Search for standard number and year of approval to access the web page of the published standard. Errata links are located under the Additional Resources Details section. Errata are also available in IEEE Xplore. Users are encouraged to periodically check for errata.

[3]Available at: https://ieeexplore.ieee.org/browse/standards/collection/ieee.
[4]Available at: https://standards.ieee.org/standard/index.html.

Patents

IEEE standards are developed in compliance with the IEEE SA Patent Policy.[5]

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE SA Website at https://standards.ieee.org/about/sasb/patcom/patents.html. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

IMPORTANT NOTICE

Technologies, application of technologies, and recommended procedures in various industries evolve over time. The IEEE standards development process allows participants to review developments in industries, technologies, and practices, and to determine what, if any, updates should be made to the IEEE standard. During this evolution, the technologies and recommendations in IEEE standards may be implemented in ways not foreseen during the standard's development. IEEE standards development activities consider research and information presented to the standards development group in developing any safety recommendations. Other information about safety practices, changes in technology or technology implementation, or impact by peripheral systems also may be pertinent to safety considerations during implementation of the standard. Implementers and users of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, data privacy, and interference protection practices and all applicable laws and regulations.

[5]Available at: https://standards.ieee.org/about/sasb/patcom/materials.html.

Participants

At the time this standard was completed, the IEEE P7014 Working Group had the following membership:

Ben Bland, *Chair*
Sumiko Shimo, *Vice Chair*
Karen Bennet, *Secretary*
Gregg Gunsch, *Former Vice Chair*

Elizabeth Adams
Aladdin Ayesh
Ken Bell
Scott Bennet
Andrew Bolster
Felix Burkhardt
Kevin B. Clark
Gokce Cobansoy
Brandt Dainow
Lubna Dajani
Pamela Dixon
Jayfus Doswell
Hassan El Shazly
Angelo Ferraro
Victoria (Vicky) Hailey
Khan Iftekharuddin
Faiz Ikramulla
Kondaine Mark Kailiwo
Carolyn Matheus
Quintin McGrath
Gary McKeown
Andrew McStay
Gawain Morrison
Mathana
Jade Nester
Temitayo Olugbade
Ellas Papadopoulou
Annette Reilly
Pablo Rivas
Daniel Schwarz
Ido Shamun
Randy Soper
Robert Stratton
Abd-Elhamid Taha

The following members of the individual Standards Association balloting group voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

Boon Chong Ang
Lee Barford
Karen Bennet
Curtis Blais
Ben Bland
Kerry Blinco
Pieter Botman
Michelle Calabro
Diego Chiozzi
Jackie Csonka-Peeren
Lubna Dajani
Howard Deiner
Hassan El Shazly
Angelo Ferraro
Deborah Hagar
Jon Hagar
Werner Hoelzl
Tyler Jaynes
Piotr Karocki
T. Leopold
Ruth Lewis
Jun Li
Quintin McGrath
Joanna Olszewska
Bansi Patel
Howard Penrose
George Percivall
Cam Posani
Edson Prestes
R.K. Rannow
Peter Reid
Annette Reilly
Gopalakrishnan Renganathan
Pablo Rivas
Peter Saunderson
Robert Schaaf
Jhony Sembiring
Sumiko Shimo
Wayne Stec
Eugene Stoudenmire
Robert Stratton
Walter Struppler
Abd-Elhamid Taha
Stephen Webb
Yu Yuan

When the IEEE SA Standards Board approved this standard on 20 May 2024, it had the following membership:

David J. Law, *Chair*
Jon Walter Rosdahl, *Vice Chair*
Gary Hoffman, *Past Chair*
Alpesh Shah, *Secretary*

Sara R. Biyabani
Ted Burse
Stephen Dukes
Doug Edwards
J. Travis Griffith
Guido R. Hiertz
Ronald W Hotchkiss
Hao Hu
Yousef Kimiagar
Joseph L. Koepfinger*
Howard Li
Xiaohui Liu
John Haiying Lu
Kevin W. Lu
Hiroshi Mano
Paul Nikolich
Robby Robson
Lei Wang
F. Keith Waters
Sha Wei
Philip B. Winston
Don Wright

*Member Emeritus

Introduction

This introduction is not part of IEEE Std 7014-2024, IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems.

This standard addresses the ethics of those autonomous and intelligent systems that attempt to emulate aspects of human empathy, such as estimating a person's mood or simulating an emotional state. Emotions and cognitive states are closely related to decision-making, health, and general wellbeing. Thus, as these "empathic" autonomous and intelligent systems (EA/IS) become increasingly prevalent, advanced, and affordable, they present significant potential for improved products and services with both social and economic benefits, and also a risk of bias, exploitation, invasion of privacy and other potential harms, for both individuals and groups.

A standard for the ethics of EA/IS was envisaged to provide guidance on ethically aligned development, deployment or decommission of such systems, which aim to facilitate positive outcomes for users and other people, and contribute to human flourishing, while mitigating potential negative outcomes. However, it is important to recognize that ethical decisions and actions are not always simple and may not produce positive outcomes for some or all stakeholders.

This standard was developed following the creation of the IEEE 7000 series of standards, each addressing a different area of the ethics of autonomous and intelligent systems. Commencing in July 2019, this standard was drafted by a voluntary, global working group of independent individuals with backgrounds in academia, industry, policy, and other areas.

The standard is divided into three major sections, as shown in Table 1.

Table 1—Major sections of this standard

Section	Content
Context	Narrative information to aid the developer (e.g., Overview, Scope, Conformance, Intended audience).
System life cycle processes	The normative statements (i.e., rules and guidelines) related to each process, which are required for conforming to the standard.
Annexes	Further informative materials.

Contents

IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems

1. Overview

1.1 Scope

This standard defines a model for ethical considerations and practices in the design, creation, and use of empathic technology, incorporating systems that have the capacity to identify, quantify, respond to, or simulate affective states, such as emotions and cognitive states. This includes coverage of "affective computing," "emotion artificial intelligence," and related fields.

1.2 Purpose

The purpose of this standard is to provide clear and practical guidance for the design and implementation of empathic systems that are prioritized to maximize human flourishing and protect users from bias, abuse, or exploitation.

1.3 Word usage

The word *shall* indicates mandatory requirements strictly to be followed in order to conform to the standard and from which no deviation is permitted (*shall* equals *is required to*).[6,7]

The word *should* indicates that among several possibilities one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required (*should* equals *is recommended that*).

The word *may* is used to indicate a course of action permissible within the limits of the standard (*may* equals *is permitted to*).

The word *can* is used for statements of possibility and capability, whether material, physical, or causal (*can* equals *is able to*).

[6]The use of the word *must* is deprecated and cannot be used when stating mandatory requirements; *must* is used only to describe unavoidable situations.
[7]The use of *will* is deprecated and cannot be used when stating mandatory requirements; *will* is only used in statements of fact.

1.4 Examples of potential risks and harms

Empathic autonomous and intelligent systems (EA/IS) have the potential to revolutionize the way people interact with machines, but there can be potential risks and harms associated with their use. The following are a few examples of what can happen with EA/IS:

a) *Bias:* Unfair or discriminatory outcomes for certain individuals or groups may occur due to inherent biases in the EA/IS, such as inadequate data set(s) used in system training (e.g., images of faces from a narrow sample of ethnicities and genders) or a lack of proper data set screening.

b) *Manipulation:* Upon establishing an emotional connection with an empathic system, users may be exploited to perform actions for which they might not otherwise provide consent (e.g., empathic chatbots could be used to convince users to purchase a product or service they may not actually need or normally afford or to provide sensitive personal information that leads to identity theft).

c) *Invasiveness:* EA/IS may be able to read and interpret signals of human emotions and intentions without the user's prior informed consent or understanding. This can be seen as an invasion of privacy and could lead to further data exploitation or manipulation.

d) *Overuse:* EA/IS can, by-design or unintentionally, engender affective responses in the user such that it generates an otherwise unnatural attachment to the system. This can lead to misuse, overuse, or overreliance on the technology or desensitize the user to real-life emotional experiences.

e) *Misinterpretation:* A system may misinterpret or misdiagnose emotions, leading to incorrect conclusions or recommendations. This can be particularly harmful in the context of mental health diagnoses or treatment.

f) *Dependence:* EA/IS can, by-design or unintentionally, isolate a user, resulting in the system becoming an otherwise unnaturally prominent source of emotional, intellectual, moral, romantic, or spiritual support for the user, with further potential for a resulting impairment to their social skills, or their ability to recognize and respond appropriately to real-life emotional experiences.

These issues are potentially increased for vulnerable parties, such as the young, the elderly, people with cognitive disabilities, and marginalized groups. Additional care is required for such vulnerable parties. For informative examples and use cases of EA/IS, see Annex A.

1.5 Further explanation of scope

This standard applies to all systems that acquire visual, sonic, biometric, textual, or personal data of an individual or population, use that acquired data to compute inferences about the supposed affect of an individual or population, and then. either promptly or with delay, attempt to influence the affect of an individual or population. Influence may involve either simulated emotion by the system or stimulation of the subject's emotion. See 3.1 for definition of "subject."

Regardless of the developer's intention for the system, the system is EA/IS if it can reasonably be expected that the subject may interpret the following about the system:

a) It gives a reasonably informed subject the illusion or impression of understanding or relationship at an affective level.

b) It processes data semantically with respect to affect.

c) It gives the subject the sense that they are being heard in an affective sense.

This standard addresses requirements, processes, and practices for any size of system or subsystem or any level of the developer's organization, from high-level policy to detailed operational procedures.

Some developers may already have ethical frameworks in place while others may not. Those with existing frameworks in place may benefit from a gap assessment to identify what they already have in place and what is needed to expand their frameworks to the context of empathic artificial intelligence (AI) to bring their frameworks into conformance with this standard. For informative examples and use cases of EA/IS, see Annex A.

1.6 Digital/virtual domains, extended reality, and the "metaverse"

The concept of empathic AI—and, therefore, the scope of this standard—extends beyond uses on physical human subjects and applies to the digital property attributed to or owned by individuals (whether personal data, virtual avatars or digital twins, and Internet profiles). If a system performs affective inferences, or attempts to influence through affective computing, in a physical or digital domain, it shall be considered an EA/IS and, therefore, within the scope of this standard.

1.7 Out-of-scope artifacts

A device or system that was designed using AI principles to enhance stimulation or the simulation of emotion, but does not acquire data, can be considered an artifact, and is outside the scope of this document. Examples of such out-of-scope artifacts may include:

a) An emotive toy that has no sensor inputs for affective data or does not adapt its behavior based on affective training

b) Compelling media (e.g., video, music, or signage) that is designed to elicit mood in the subject but does not adapt to the subject's estimated mood

c) Environments with a persistent build that are designed to elicit a given "mood" in the subject

The intention or act of acquiring contemporaneous data about an interaction is one aspect that distinguishes an EA/IS from an artifact. Therefore, systems that acquire or process data after simulated emotion or stimulation would be in the scope of this document. Even if the data was acquired long after the original stimulus or simulation, the intention to acquire the data would have been contemporaneous with the creation of the system. Examples may include the affective steering of individuals or groups to specific political, financial, or societal ends.

1.8 Further explanation of purpose

As part of the IEEE Society for Social Implications of Technology (SSIT),[8] the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (Global Initiative) and the IEEE 7000™ series of standards[9], this standard is ethical in nature. Its orienting suite of fundamental ethics broadly come from the United Nations Universal Declaration of Human Rights [B46], the United Nations Convention on the Rights of the Child [B9], and values established in the wider IEEE Global Initiative, particularly the Global Initiative's document, Ethically Aligned Design, First Edition [B14].[10]

The goal of this standard is to provide applied ethics that are important in the design, creation, and usage of empathic technologies. Overall, this standard promotes the development of empathic autonomous and intelligent systems (EA/IS) that are designed with an ethically aligned approach, and with the intention to be responsible, promote a high degree of agency for subjects of the system, and are focused on improving the user experience.

[8]More about the IEEE Society for Social Implications of Technology at: https://sagroups.ieee.org/ssit/
[9]More about the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (Global Initiative) and the IEEE 7000™ series of standards at: https://standards.ieee.org/industry-connections/ec/autonomous-systems/
[10]The numbers in brackets correspond to those of the bibliography in Annex D.

1.9 Need for this standard

While the technologies with which EA/IS are built and operated may not be unique, they are used to profile and interact with human subjectivity in novel ways. It is this use of data about human subjectivity for human-system interaction, in particular, that raises the need for a globally applicable set of ethical standards for the development, deployment, or decommission of these systems.

Human affect is closely related to decision-making, health, and general well-being. By interacting with humans at an emotional or cognitive level, ethically aligned EA/IS can contribute to healthier, happier, more satisfying lives for individuals, groups, and wider society.

However, EA/IS also carry the potential to manipulate people, whether individually, in groups and communities, or collectively, with particular and heightened risks for vulnerable parties such as the young, people with cognitive disabilities, and marginalized groups. As empathic technology becomes increasingly prevalent, advanced, and affordable, a widespread base of subjects are encountering systems that have the potential to monitor, measure, and interact with these subjects in highly intimate and personal ways. This creates the potential for exploitation of subjects through classification or manipulation of their emotions and cognitive states. Misuse of such systems can produce both predictable and unexpected harm for subjects as well as wider social consequences.

Creating a system with artificial affect (e.g., changing the emotional tone of a voice assistant's voice, or a chatbot using emojis) requires intentional consideration and care from the developer, as it can lead to deception, manipulation and potential psychological harm for the subject. EA/IS can create an interaction loop between system and subject that affects the subject (e.g., at an emotional or social level). Thus, the development, deployment, or decommissioning of such systems inherently entails ethical choices on the part of the developer.

1.10 Intended audience

The primary audience for this standard is EA/IS developers (see definition of developer in 3.1). However, secondary stakeholders (such as policymakers, assessors, end users, data subjects, and so forth) will also find this standard useful in their activities or engagement with EA/IS developers or in their interaction with EA/IS.

1.11 Structure

This standard enables the developer (i.e., the person or organization using this standard) to apply an ethical approach to the entire system life cycle. There are several frameworks for defining system life cycles, but this standard makes particular reference to the system life cycle standard ISO/IEC 5338, Information technology—Artificial intelligence—AI system life cycle processes [B27], which is in turn based on ISO/IEC/IEEE 12207-2017, Systems and software engineering—Software life cycle processes [B30] and ISO/IEC/IEEE 15288, First Edition 2015–05–15, Systems and software engineering—System life cycle processes [B31].

NOTE—ISO/IEC 5338 [B27] was still in development and unpublished when it was reviewed for the development of this standard. Thus, ISO/IEC 5338 has only been used for informative purposes, it is not a normative reference.

Based on the processes of the above-mentioned system life cycle standards, this standard is broadly structured in a series of system life cycle processes that include "agency," "transparency, interpretability, and explainability," and others, as mentioned in Clause 4. These processes are further grouped into "high-level processes" in accordance with the above system life cycle standards, such as "Organizational project-enabling processes" and "Technical management processes."

Each process is divided into a structure that closely matches ISO/IEC 5338 [B27], namely the following:

a) Purpose

b) Activities and tasks

c) EA/IS-specific particularities.

NOTE—Whereas ISO/IEC 5338 [B27] includes "AI-specific particularities," for this standard, this has been changed to "EA/IS-specific particularities" to be specific EA/IS.

For convenience, the entire life cycle is sometimes abbreviated into the phases of "development, deployment, and decommission," which are broadly chronological in order.

Some of the processes in this standard (e.g., risk, issue, and impact management) apply to multiple, or all, stages of the system life cycle. Any processes can be applied to any appropriate stages, in any order or repeated where appropriate. This standard does not prescribe a specific order in which each system life cycle process should be conducted. However, the processes below are presented in an order that is likely to be approximately chronological for the development, deployment, and decommission of the system.

At times, it may be appropriate for the developer to repeat a process due to a change in the system's context (e.g., a change to the system design, a change in the policy environment, or discovery of new stakeholder insights).

1.12 Conformance

This standard contains statements that take the form of "shall," "should" or "may" statements (see 1.3).

The normative statements (those with "shall") and recommendations (those with "should" or "may") in this standard are listed within the "System life cycle processes" (Clause 4), under subsections entitled "Activities and tasks." Any text outside of those subsections that is presented as a recommendation, rule, guideline, etc. is not considered part of the normative requirements for conformance to this standard.

To conform to this standard, the developer is required to achieve all mandatory normative statements (those statements that use the word "shall") and, where explicitly stated, publish evidence of their conformance (see 4.2.3.2 for details of publishing requirements, and Annex C for a complete list of required materials to be published). The developer is also required to publish a signed statement of conformance (see 4.2.3).

The developer may optionally act beyond these mandatory requirements ("shall" statements) by displaying conformance to recommended ("should," "may," or "can") via published evidence.

In select contexts, normative statements only apply to "high-risk" EA/IS (see 4.2.2.1). If the developer, in their risk, issue, and impact management activities (see 4.2.2) assesses any aspect of the system to fall within a "high risk" (or an equivalent) category, they are thus subject to these selective normative statements.

Where this standard states that an activity (such as a risk assessment) shall be conducted, it does not prescribe specific models or processes to perform those activities. A suitable, recognized model or process is deemed sufficient, insofar as the developer publishes an explanation for their choice of model or process.

Conformance is required by any person or group responsible for the impacts of the system, who are collectively referred to as developers in this standard. This includes resellers, licensees, etc. See 3.1 for a definition of developer.

Where this standard requires the developer to "publish" (e.g., publish a risk assessment), there are requirements for publishing which are outlined in 4.2.3.

1.13 EA/IS key differentiators

The following aspects of EA/IS are key factors that differentiate its design and use from those of other forms of A/IS or traditional systems:

— *Unique or heightened ethical considerations:* EA/IS raise important ethical considerations, such as in relation to privacy, data collection, and the potential for emotional manipulation.

— *Affective interaction:* EA/IS can provoke affective states, or can produce models that attempt to infer the cognition or mood of subjects. These potentially intrusive processes can introduce intrinsic design-based risk related to psychosocial harm (e.g., embarrassment, manipulation, coercion).

— *Heightened capacity for influence:* EA/IS are more complex, ubiquitous, intelligent, and intrusive than traditional computer systems given that EA/IS can influence subjects more rapidly, with greater specificity and subtlety, and through a greater frequency of repetition.

— *Rapid system evolution:* The ability of EA/IS to use affective data and cause both good and potential harm continues to increase rapidly as these systems evolve.

— *Accuracy and ground truth are elusive:* With contention and subjectivity underlying affective theories, affective inferences made by the system are not objective fact(s). Thus, "accuracy" and "ground truth" can be more contentious and elusive than in traditional computer systems or other forms of autonomous and intelligent systems (A/IS). The "accuracy" of the system may be called into question and its inferences contended respective to the subject's interpretation of their feelings. See 1.16, 4.2.6, and B.4.2 for more on the nature of accuracy and ground truth in EA/IS.

— *Affect detection:* EA/IS are designed to detect human emotions, moods, cognitive states, or other modes of affect via the monitoring and analysis of facial expressions, tone of voice, and other nonverbal cues.

— *Intimacy of data and interactions:* Because common social conventions lead many societies to enforce privacy rights (particularly relating to individual thoughts and emotions), attempting to measure or simulate human affect can intrude on these rights. Hence, there is an intrinsic design-based risk of psychosocial harm (e.g., embarrassment, manipulation, coercion) and private or public rejection of affective systems via a perception that intimate data is not treated with an adequate level of care and delicacy.

— *Personal interaction:* In general, EA/IS tend to rely on personal data and make personal inferences about people. This entails a heightened level of care, which is both technical (requiring advanced privacy and security processes) and ethical (requiring benefit and risk management).

— *Adaptive:* EA/IS are often designed to learn about individual users over time to personalize interaction, or to adapt interaction based on the user's evolving emotive state (e.g., dynamically adjusting the "tone" of the system during an interaction). This adaptive behavior entails a requirement for heightened care from the developer, to maintain continuous monitoring, and continuous consideration for harm mitigation and human-in-the-loop interventions.

— *Heightened connection:* Like other AI systems, EA/IS can collect feedback from users and use it to improve their responses and interactions over time. But the emotive nature of EA/IS entails a heightened connection with people and the human experience.

— *Increased potential drift:* Due to the subjective, subtle, and varying nature of human affect, EA/IS often deal with particularly large or rapid deviations of the production data (data drift), or deviations to or from the desired system behavior based on some agreed measure of objective fact (concept drift), or change in the AI's processing of the same data as a result of learning or technical enhancements (algorithmic drift). Such drift can lead to unexpected negative outcomes. See 3.1 for definition of drift.

— *Empathic interaction is frequently a dialog:* Whether or not an EA/IS is intended to elicit a particular response from a subject, they can experience effects from the system, which can be positive or negative, and predictable or unforeseeable, leading to unexpected, unplanned, or unpredictable interactions between system and subject in either direction.

— *Empathic interaction can entail unexpected feedback or disruption:* The individual subject, and related stakeholders, can experience positive or negative effects from an EA/IS regardless of the intended target of the system (e.g., individual or group). Furthermore, the interaction may be interrupted upon the explicit notification of monitoring by the system. Either scenario can influence the thoughts, feelings, and behavior of the intended target of the system or related stakeholders. These interruptions in interaction, or shifts in attitude or situation for the subject, can influence the data that the system uses for training or empathic interaction, which may blind or bias the EA/IS in manners that pose intrinsic design-specific risks.

— *Context awareness is important:* Due to the subjective and personal nature of human affect, as well as the nuanced manners in which affect is socially expressed, contextual considerations are necessary to ensure that EA/IS performs as designed or intended. Understanding the context of an empathic interaction or affecive experience can help to improve the quality and appropriateness of the EA/IS's behavior in that context. This context can include situational factors (e.g., activity, interaction setting, professional versus personal interactions, time of day, presence of language barriers), the background of the individual (e.g., sex or gender identity, cultural identity or race, age, socioeconomic status), or other contextual factors. However, regardless of contextual understanding, the relative performance of the system is intrinsically dynamic and vulnerable to critiques based upon subjective opinion. Furthermore, any attempt to assess the context (e.g., by measuring a contextual factor via a sensor) can entail ethical considerations such as violating the subject's privacy through inappropriate surveillance.

— *Individual versus group dynamics:* EA/IS raise heightened and unique considerations respective to the degree of influence they have on an indvidual or group level. While this standard encourages the developer to give primacy to the individual subject, rather than making Utilitarian calculations to benefit the group on aggregate over the individual, it nevertheless also encourages care to refrain from generalizing inferences to groups or lowering quality in the group context (See 1.14.2 and 1.14.3). The potential negative impacts of EA/IS, such as psychological manipulation, can be just as pertinent in groups and social contexts as with individual people.

1.14 Outcomes

All forms of A/IS can produce significant outcomes, with varying degrees of positive and negative impacts for different stakeholders. Empathic A/IS can produce several unique or heightened outcomes. Some of these outcomes are discussed in this standard.

This standard requires developers to proceed with caution, do their best to educate themselves on these outcomes, and account for potential issues.

Outcomes of EA/IS that are broadly positive may include the following:

a) Greater levels of measurement, understanding and development of human emotions, for individuals and groups

b) Easier, more naturalistic, and more dynamic interaction between humans and machines, leading to improved products and services with social and economic benefits

c) Advancements in mental and physiological health, well-being, and provision of care

d) The development of more "compassionate" machine systems that are programmed to judge when it is appropriate to use emulated empathy

e) Supportive data and technologies for social development, peace work, and public security

f) Increased efficiency and cost savings due to automation of manual processes, which account for affect

g) Improved decision-making due to better data analysis and insights

EA/IS entail unique and heightened risks compared to other AI systems or other technologies. Outcomes of EA/IS that are broadly negative may include:

h) Psychological or social manipulation, coercion, or control, of individuals or groups

i) Violations of the human right to freedom of thought and feeling and, thus, human autonomy and free agency

j) Bias, prejudice, or unfair influence toward people, due to the predictive and assumptive basis of EA/IS (e.g., assigning discrete labels to affect that are, in reality, felt to be more complex and subjective)

k) Invasion of privacy (e.g., by attempting to predict or expose how a person feels)

l) Automation bias, such as inappropriately high levels of trust in the system, by subjects who are inadequately informed or resourced (e.g., being unaware of when empathic AI is in use, or believing that the system's outputs are more accurate or correct than they are)

m) Unfair changes in, or expectations of, the attitudes and behaviors of subjects, such as where people alter their emotional expressions to avoid being tracked by the system

n) Bias, inaccuracy, and over-reliance resulting from the use of EA/IS in support of high-risk decisions (e.g., policing, judging, career development, education)

o) Over-reliance on inadequate training data and models, as there can be no global training data sets or models for affect or behavior that account for all personal and cultural differences

p) Perceived certainty, uncertainty and trust in the system could negatively affect the subject's agency and well-being

1.14.1 A positive approach

This standard requires the developer to take a positive approach to EA/IS, such that the design and use of the system is aimed at contribution to human well-being and flourishing from the outset. To this end, this standard requires the developer to, among other things, conduct a well-being assessment (see 4.2.1).

1.14.2 A precautionary approach

This standard requires the developer to commit to an ethical and socially responsible approach to the development, deployment, and decommission of the system. This standard assumes that all EA/IS carry a heightened potential for harm, and, as such, all EA/IS should be considered as high-risk systems from the outset, unless demonstrated to be otherwise. Thus, this standard requires the developer to use of the Precautionary Principle[11] as follows:

a) Establish and implement a primary objective to first do no harm

b) Go beyond risk management and establish evidence that the system contributes to the subject's well-being and to wider human flourishing (see 4.2.1)

c) Not release the system prior to acquiring evidence of its safety

Such a precautionary approach can be beneficial to innovation rather than detrimental. Significant costs to innovation and product development can be incurred when unexpected impacts occur later.

The Precautionary Principle should guide design and use of the system insofar as harm should be limited, even when concrete proof of that harm is elusive. It is not enough that harm should only be remedied after the fact if it is predictable and avoidable.

At the same time, the developer is encouraged to avoid utilitarian approaches, such as only assessing impact on aggregate over a population and not adequately addressing individuals or small groups.

[11] Various descriptions of the precautionary principle are available online, such as at https://en.wikipedia.org/wiki/Precautionary_principle

1.14.3 Balancing benefits and potential harms

This standard aims to support the developer in committing to an ethical approach, such that the design and use of EA/IS are prioritized to produce positive outcomes for all affected stakeholders. As no system can ever be operated completely free of consequences, the developer is required to consider potential outcomes and consistently act with the intention of producing outcomes in which the benefits greatly outweigh the potential harms.

However, by promoting consideration of the balance of benefits and potential harms, this standard does not promote utilitarian[12] approaches, such as where the system's outcomes are only assessed on aggregate over a population, and do not adequately address individuals or small groups. In general, this standard considers the individual subject as primary, over calculations of wider social benefits.

1.14.4 Ethical explainability

In consideration of the heightened risks, issues and impacts of EA/IS (see 1.4), this standard requires the developer to adopt a strong stance with respect to making the system transparent, interpretable and explainable ("transparent") for all relevant stakeholders. This standard then goes further, by presenting a concept of "ethical explainability," by which the developer is required to publish a rationale for their ethical stance with respect to the purpose of, and approach to, the system's development, deployment, and decommission. See 4.2.3 for more on transparency, interpretability, and explainability.

1.15 Limitations and issues

For anyone wishing to build or operate ethically aligned EA/IS, it is important to recognize that there are significant limitations, uncertainties, biases, and blind spots in the application and foundational knowledge (and technology) of EA/IS.

This standard provides further discussion and guidance on some of these issues, such as the disagreements on the underlying science of emotions, common biases, linguistic challenges in describing emotions, and the conceptual variations in affect and empathy across different human groups. However, no single publication can provide complete background information or guidance on all the potential issues related to this field.

This standard requires developers to proceed with caution, do their best to educate themselves on these limitations, consider the issues that could arise from them, and account for potential problems.

1.16 Caution regarding the underlying concepts of EA/IS

Much debate exists in the scientific community concerning emotions and empathy (see Annex B for background on this debate). In particular, this debate concerns the nature of the internal felt-states at the core of these phenomena, and how these felt-states are outwardly expressed as measurable phenomena.

The relatively new science of Affective Computing has given rise to a suite of technologies that measure visual, acoustic, physiological, and biomechanical human behavior, and subsequently pertain to infer interior states. This is becoming both inexpensive and pervasive. The various strands and theories of emotion and empathy research each recommend differing interpretations of how these measurements relate to the internal felt-states of the individuals being measured. While these technologies can, in some cases, make accurate measurements of physiological features or movements, the further a system travels from the physiological toward the psychological, the more it must buy into one of these often-competing theoretical views.

[12] Various descriptions of the utilitarian ethics and utilitarianism are available online, such as at https://en.wikipedia.org/wiki/Utilitarianism

Some theories describe different types of empathy, particularly "affective" and "cognitive" empathy. Essentially, "affective empathy" involves experiencing an emotional response to other's emotions (e.g., "I feel your pain"), while "cognitive empathy" involves interpreting and understanding other people's emotions without necessarily feeling anything. Humans experience a complex blend of these different kinds of empathy. For more background on emotions and empathy, see B.1, B.2.

On the other hand, the "empathy" of an EA/IS can only be cognitive in nature. No constructed system can have innate, intuitive, affective empathy. EA/IS are only capable of simulating an understanding of emotion or subjectivity. Moreover, such systems should only be used to advance individual and collective human well-being.

Considering the contentious nature of affect and empathy in the scientific community, and the primacy of the self-report in the assessment of these phenomena, it should be considered that neither phenomenon can ever have a wholly true or accurate ground truth. In A/IS, models are based on ground truths that are set by the developers, and are usually based on some agreed measure of objective fact. In the case of EA/IS this can never be the case. Thus, some degree of bias and incompleteness is always introduced in the system.

These newly enabled approaches to measuring human behavior create a situation in which responsibility is placed on the shoulders of the developer of any system that purports to interpret emotional state from behavioral measurement.

See Annex B for more background on the science of emotion and empathy.

2. Normative references

This standard has no normative references.

3. Definitions, acronyms, and abbreviations

3.1 Definitions

For the purposes of this document, the following terms *and definitions apply. The IEEE Standards Dictionary Online* should be consulted for terms not defined in this clause.[13]

affect: Any internal state experienced or expressed by a subject, such as emotion, mood, or cognition.

affective computing: The study and development of systems and devices that can recognize, interpret, process, and simulate human affect.

affective rights: Those normative rights attributed to all humanity which directly relate to communication(s), thought(s), conviction(s), and emotion(s), and the privacy generally afforded to their development and expression.

affective states: *See:* **affect**.

agency: In reference to users, data subjects and other stakeholders who are affected by the system, agency is the nature and extent of their autonomy, freedom, access, and decision-making capacity with respect to the system.

NOTE—Where specified, agency may alternatively refer to the power of a given system to affect individual or groups of users, data subjects, or other stakeholders.

[13] *IEEE Standards Dictionary Online* is available at: http://dictionary.ieee.org. An IEEE account is required for access to the dictionary, and one can be created at no charge on the dictionary *sign-in page.*

algorithmic drift: *See:***drift**.

auditability: Ability to access, open, inspect, and verify content, data, and EA/IS implementations.

biometrics: Any data relating to a human subject, or group of humans, for use in empathic technology, particularly physiological data (e.g., heart rate, facial movement, or gait).

NOTE—Biometrics, in the context of EA/IS, also includes the data traditionally associated only with identification of individuals (see IEC JTC 1/SC 37 [B20]).

concept drift: *See:***drift**.

data drift: *See:***drift**.

developer: The entity responsible for development, deployment, or decommission of the system.

NOTE 1—Developer includes any entity responsible for conforming to the standard.

NOTE 2—Categories of developers for empathic devices include: a) original equipment manufacturers, including both individual employees (e.g., programmer, analyst, quality control personnel) and groups (e.g., corporate departments, external contractors) or the overall organization; b) third-party actors who deploy, modify, or licenses the system for commercial gain or non-profit support; and c) third-party providers (including open-source providers) of sub-components that may be employed in the empathic device at any time during its life cycle.

drift: The expression of deviations between training data and collected data inputs.

NOTE 1—**Data drift** occurs when data changes after the model is deployed. Each data point exists in a high-dimensional space, which is defined by the number of input features. Data drift occurs when new data points arise from a region of the input space that was less populated by training data.

NOTE 2—**Concept drift** occurs when the decision boundary changes, and is often observed in time-series or streaming data. Concept drift exhibits as deviations to or from the desired system behavior based on some agreed measure of objective fact.

NOTE 3—**Algorithmic drift** occurs when a change in the system's processing of the same data results during learning or technical enhancements. This phenomenon leads the system to provide different results when fed the same pre-drift data.

empathic AI: *See:***empathic autonomous and intelligent system(s)**.

empathic autonomous and intelligent system(s) (EA/IS): Affect-sensitive technologies employed to algorithmically infer, model, simulate, or stimulate understanding of emotions, feelings, moods, perspective, attention or intention. Data insights or actions taken in response to those automated inferences typically, but not always, inform future interactions between a person or group and system (or between systems).

NOTE—EA/IS is a general term for empathic technologies. These include similar technologies with alternative names, including "affective computing" systems, "emotion(al) AI," "emotion-detection" systems, and "empathic/empathetic AI," as well as AI that uses processes such as sentiment analysis, biometrics, or natural language processing to analyze, infer, or simulate affect. The abbreviation EA/IS adds "empathic" to the A/IS acronym ("autonomous and intelligent systems") used in other IEEE standards terms.

empathic technologies: *See:***empathic autonomous and intelligent system(s)**.

empathy: To share or understand another's feelings or experience. Any empathy performed by a machine is considered to be an emulation of human empathy rather than the same process.

NOTE—See Annex B for a background on emotion, affect and empathy.

explainability: The ability to describe how A/IS make decisions. Explanations can be produced regarding both the procedures followed by the A/IS (i.e., its inputs, methods, models, and outputs) and the specific decisions that are made.

fairness: Equanimity, balanced and unbiased processes or outcomes.

fitness for purpose: Assurance regarding a system or service that the algorithm will meet its agreed requirements, and perform reliably.

ground truth: Data that is chosen to indicate the actual facts of the matter, e.g., for scientific analysis or for training machine-learning models. For measuring human emotion, affect or empathy, ground truth is elusive and subjective. See 1.13.

modality: *See:* **mode**.

mode: Any category of device or data used in EA/IS (such as facial movement used to infer emotional expression, or room temperature used to infer contextual information about the subject's affect) which may include sensor type, sensor data, or affective feature(s). *Syn*: **modality**.

model: **(A)** A computer program trained on certain data to detect patterns in that data. **(B)** A theoretical representation of a real-world process, concept or phenomenon (e.g., a psychological model of affect).

sentiment: The value (e.g., positive or negative) of affect inferred from the system's data source, such as sentiment from text analysis using natural language processing.

stakeholder: Any person or group potentially affected by the system (regardless of whether that person is a subject or not) or any person or group who may potentially affect the system (e.g., influence its design, deployment or decommission).

NOTE—See 4.3.1 for more on stakeholders.

subject: A person that (voluntarily or involuntarily, or with or without awareness) interacts with the system or whose personal data has been, is, and/or will be acquired for engaging with and/or training the system.

NOTE—"Subject" may include "data subject", which is specifically a person who provides personal data that is used by the system. The General Data Protection Regulation (GDPR) provides further context for defining "data subject". Text of the GDPR states: "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

transparency: The extent of disclosure, explainability, interpretability, or accessibility of the EA/IS, such that appropriate stakeholders can assess or understand the system.

3.2 Acronyms and abbreviations

AI	artificial intelligence
A/IS	autonomous and intelligent system(s)
BOM	bill of materials
EA/IS	empathic autonomous and intelligent system(s)
GDPR	General Data Protection Regulation
WIA	well-being impact assessment

4. System life cycle processes

4.1 Organizational project-enabling processes

4.1.1 Acquisition of required skills, learning, and knowledge

4.1.1.1 Purpose

Acquiring skills and education for EA/IS requires a multidisciplinary approach, and continuous learning.

4.1.1.2 Activities and tasks

The developer *shall* do the following:

a) Publish evidence of the extent to which they have acquired appropriate knowledge and skill (e.g., peer-reviewed learning, certified qualifications) in the underlying nature of EA/IS. Examples of relevant knowledge and skills could include the following:

   1) Emotion science and the contentions within the field (as outlined in 1.15, 1.16, and Annex B)

   2) Abnormal psychology and psychological vulnerability and how they relate to EA/IS

   3) Familiarity with the types and structure of affective data (e.g., modes of biometric data) and the typical uses and issues associated with them

   4) Familiarity with ethics as they apply to EA/IS

b) Publish evidence of the extent to which they are maintaining a continuous process of learning and skill development to keep up with empathic technology and the issues arising from it as it develops, insofar as is relevant to EA/IS.

The developer *should* do the following:

c) Continuously maintain a learning system for education on the wider context surrounding empathic AI, such as: the basics of AI, human behavior, human psychology, and emotions.

d) Create a culture of continuous learning and collaboration, both within the organization and with others in relevant fields, including both specialists (e.g., for technical insight) and affected stakeholders (e.g., for user feedback).

4.2 Technical management processes

4.2.1 Contribution to human flourishing

4.2.1.1 Purpose

Recognizing that EA/IS present unique and potentially high risks of harm, it is insufficient only to plan for remedy of potential impacts after they occur. Rather, prediction and prevention of harm is required as far as feasible. However, this standard also requires that the developer goes further than harm prevention, by demonstrating how the system contributes to the well-being of any relevant stakeholders.

It is not sufficient that an EA/IS is published (e.g., as a “beta” version) after solely evaluating risks and issues. It is also not sufficient that the developer demonstrates that empathic AI can improve the service provided to the subject by the system. The developer is required to publish evidence of testing that demonstrates a strongly positive balance of well-being outcomes over risks and issues.

4.2.1.2 Activities and tasks

The developer *shall* execute and publish a well-being impact assessment (WIA) demonstrating how the system contributes to the well-being of the subject, other affected stakeholders, and, where relevant, the well-being and flourishing of the wider human populace.

The developer *should* do the following:

a) Conduct a WIA using a recognized standard process for the assessment, such as IEEE Std 7010 [B26].

b) Include in the WIA an assessment of stakeholder hierarchy conflicts that details the risks and impacts of prioritizing one stakeholder type or group over another.

c) Use a reputable third party to conduct the WIA or gain peer review of the WIA.

4.2.2 Risk, issue and impact management

4.2.2.1 Purpose

Risk, issue and impact management ("risk management") processes are used to identify potential risks and avert potential harm from the EA/IS. Risk management in the context of EA/IS requires oversight of the risk process to determine whether the system's inputs, activities, and outputs are likely to pose a risk of harm.

While this standard considers EA/IS in general to be high-risk systems, the developer is nevertheless required to assess the system's relative levels of potential risks, issues and impacts.

This standard requires the developer to conduct a risk, issue, and impact assessment, as described in the normative statements in this section. As a general guide, characteristics of a high-risk EA/IS can include one that has the potential to substantially influence the emotions, well-being, or psychological state of individuals or groups. Alternatively, a high-risk EA/IS may be involved in critical areas, such as healthcare, mental health support, education, or social interactions, whereby the system can raise significant ethical concerns due to its potential for emotional manipulation, privacy infringement, algorithmic biases, or other potentially harmful consequences. As such, empathic systems can significantly impact individuals' autonomy, privacy, or psychological welfare.

NOTE 1—One useful definition of "high risk" in relation to AI systems is provided by the European Commission [B15].

NOTE 2—IEEE/ISO/IEC 16085-2021 [B32] provides guidance on risk management.

4.2.2.2 Activities and tasks

The developer *shall* do the following:

a) Not release the system (e.g., publicly) without publishing evidence of its safety.

b) Publish risk, issue and impact assessment(s) that include the unique risks, issues and impacts inherent to EA/IS (e.g., as outlined in 1.4).

c) Treat the system and its components as high risk prior to demonstrating through the risk, issue and impact assessment(s) that they are low risk.

d) Provide approaches to proactively monitor the EA/IS to identify the emergence of unintended risks or issues, and approaches to effectively manage these unanticipated risks and issues.

e) Include strategies and process strategies for mitigation of projected risks, issues and impacts.

f) Include assessment of risks, issues, and impacts relating to ethics, culture, and diversity.

g) Address all potentially affected stakeholders (see 4.3.1).

h) Document a clear definition of the person or persons responsible for the risks, issues and impacts of the system.

i) Renew risk, issue, and impact assessments at each feasible stage of the system life cycle (e.g., development, deployment, and decommission), and at any appropriate time intervals, or each year.

The developer *should* do the following:

j) Use a qualified third party for risk, issue, and impact assessment, where possible.

k) Personalize the system's behavior based on individual subjects, taking into account their unique circumstances, goals, and risk tolerance. This can lead to more effective risk management strategies that are tailored to the needs of the individual.

l) Attempt to detect cognitive biases in subject and help correct them. For example, if a subject is prone to overconfidence bias or loss aversion, an EA/IS can provide personalized feedback and recommendations that help the subject make more rational decisions.

m) Attempt to communicate with subjects in a human-like way where possible (e.g., using natural language and tone of voice) with the aim to manage risk more effectively.

4.2.3 Transparency, interpretability, and explainability

4.2.3.1 Purpose

Considering the potentially high risks, issues and impacts of A/IS, and those further heightened by EA/IS (see 1.4), this standard requires the developer to adopt a strong stance with respect to making the system transparent, interpretable, and explainable ("transparent") for all relevant stakeholders.

Transparent systems increase the potential for positive agency by stakeholders, rigorous assessment by outside parties, and reduced risk through making the systems understandable to others. The following statements require the developer to publish explanations of the system that go beyond technical explainability (e.g., listing system features) to express the rationale for each part of the system's design and expected use, as well as for the developer's ethical stance with respect to the purpose of, and approach to, the system's development, deployment and decommission ("ethical explainability").

The statements in this section describe how to make available to all relevant stakeholders, in formats that are clear and accessible to those stakeholders, information on the following:

a) *What* decisions the EA/IS makes (e.g., generating a score for a particular emotion).

b) *How* it makes those decisions (while the decision-making of some forms of EA/IS, e.g., those with unsupervised self-learning, may be too complex to explain, the developer can still provide transparency with respect to the processes and mechanisms on which the system operates to make decisions, and their probabilistic nature). See B.4.2.

c) *Why* those decisions have been chosen by the developers (i.e., what theoretical or practical frameworks were chosen for the EA/IS's workflows, and how were they implemented in the system).

Furthermore, transparency documentation is required to be accessible to people with varying degrees of experience, capabilities, knowledge, and resources, and published in forms and locations that are readily accessible and understandable to all relevant stakeholders (i.e., not just technically sophisticated readers). "Relevant" here goes beyond all those who are potentially affected by the system to include others who may have a reasonable interest in the system (e.g., auditors, government, journalists). Additionally, as far as possible, explainability documentation should support replicability of system outputs for testing by appropriate parties.

As outlined in 1.15, 1.16, and Annex B, there are several contentious, subjective, and poorly understood aspects to human affect and EA/IS (such as how to accurately measure an emotion). This entails a need for EA/IS to be accompanied by transparency processes and assets that examine these aspects.

The heightened risks (see 1.4) and contentious nature (see 1.15 and 1.16) of EA/IS entail a heightened need for stakeholders to be able to interrogate and understand the system. Well-designed transparency assets should empower stakeholders to be able to easily understand how the system works, what potential impacts it can have, and what their choices are with respect to it.

Transparency "assets" here can be any kind of system feature (e.g., an automated notification), content (e.g., published text), or action (e.g., human intervention).

Appropriate modes of transparency depend on the context of use. For instance, an app may provide a notice that pops up on a screen at appropriate times, whereas a small physical device may have lights to indicate certain data processing is happening. Where appropriate, longer statements may be provided elsewhere (e.g., online).

The provision of transparency features and content does not override the requirement for the subject to opt-in to the system or for appropriate agency and controls to be granted to the subject, as described in the following subclauses.

In general, greater transparency is preferred. However, this standard urges caution with respect to cases where excessive transparency can lead to negative outcomes. For instance, "automation bias" describes how subjects can have inappropriately high levels of trust in the system (e.g., by being unaware of when empathic AI is in use, being poorly informed of the contentious nature of the underlying theory, or believing that the system's outputs are more accurate or correct than they are). As such, increasing transparency in the system can in some cases decrease subjects' safety, trust, and, ultimately, their well-being.

4.2.3.2 Activities and tasks

The developer *shall* do the following:

a) Publish appropriate materials wherever this standard states that an activity (e.g., a risk assessment) shall be "published." Furthermore, the following shall apply:
    1) The developer shall publish the outcomes of the activity (e.g., risk assessment scores) to an appropriate extent such that it is readily accessible to all relevant stakeholders.
    2) If the developer asserts that publishing to an "appropriate extent" would not include making the content readily accessible to the public, the developer shall publish their justification for limiting the audience of their publications.
    3) Where possible, publish any required materials prior to deployment of the system.
    4) The developer shall update all required materials at a reasonable frequency throughout the system life cycle.
b) Publish a signed statement of conformity to this standard, on completion.
c) Declare unambiguously in any appropriate location (e.g., notification, label, accompanying documentation) and at a reasonable timeframe and frequency, including, where possible, notification at the point of usage of the system the following:
    1) That EA/IS is in use in the system or if it has been developed using affective data or technology (e.g., emotion data sets)
    2) The nature of the EA/IS technology in use in the system

3) Where an empathic system is in use and not a human

4) If the system is designed to operate on groups rather than individuals, the developer should provide notice to the group

5) Provide notice of the probabilistic and subjective nature of EA/IS (see B.4.2), clarifying that the system does not claim to be "accurate" or "correct" with respect to human affect

d) Examine and publish an explanation of their ethical stance with regard to the purpose of and approach to the system's development, deployment, and decommission (e.g., if the system is designed to protect children, state why children, rather than competent adults or others, need protection).

e) Publish explanations of the following:

1) Intended purpose of the system (such as via a software bill of materials (sBOM) (ISO/IEC 5962:2021 [B28])

2) The expected scope of use of the system and the conditions under which the system can be expected to function as intended

f) Publish documentation of safety tests conducted on the system, detailing methodologies such as response to additive random noise or targeted adversarial attacks and demonstrating the system's robustness and resilience against potential vulnerabilities. This shall include a comprehensive report on the outcomes of each test, ensuring the system's robustness and resilience against potential vulnerabilities, and highlighting the ethical importance of maintaining system integrity against external threats.

g) Establish and publish a document of components and dependencies of the system. A bill of materials (BOM) shall be conducted and published prior to system deployment and maintained throughout the life cycle of the system. It shall include hardware, software, and data.

h) Publish a data management plan.

The developer *should* do the following:

i) Publish explanations of the following:

1) Expected capabilities and limitations of the system (e.g., with respect to the accuracy of its affective modeling)

2) The system's training and testing data sets, and how they were obtained

3) The system's inputs, describing the sensor data captured (or acquired from an external source), specific affective modalities, and the resulting extracted features

4) The system's outputs, describing the labels being inferred by the system, and the scale of those labels if appropriate

5) How inputs and outputs are transformed into features consumed by the model

6) How the system inputs and outputs are appropriate to the purpose of the system, providing evidence where possible (e.g., peer-reviewed publications)

7) The theoretical framework(s) on which the system's affective modeling is based, if appropriate (e.g., Discrete Emotion theory) and how the framework(s) is/are represented in the system

8) The affective model(s) (e.g., Facial Action Coding system[14]) used to generate empathic features, with evidence (e.g., peer-reviewed publications) to justify the efficacy and appropriateness of the chosen model(s)

9) Rationale for the estimates and decisions made by the system (system outputs).

[14]Facial Action Coding System is a system for categorizing facial movements, such as for estimating emotional expression. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7264164/

j) Include in the ethical explainability documentation, including the following:

1) The main approaches taken to the design, build and operation of the system (e.g., what kind of data was acquired, which affective models were used), and a balanced analysis of the possible implications of those approaches.

2) Explanation of why each approach was seen to be the ethical approach (e.g., if children do indeed need protection, explain why the current approach addresses the problem, such as it being grounded in the understanding of the user group and other stakeholders such as guardians).

NOTE—Guidance on transparency practices can be found in IEEE Std 7001-2021 [B23].

4.2.4 Agency and autonomy

4.2.4.1 Purpose

Several of the heightened risks that are accompanied by EA/IS can impact the agency and autonomy of stakeholders (e.g., by coercing, manipulating, reducing decision-making capability).

In particular, EA/IS are trained on data sets that are affective in nature and, thus, inherently contain biases, inaccuracies, and subjective information that could result in unfair or discriminatory outcomes. It is important to identify and reduce such biases so that the system can be more fair and unbiased and can avoid potential harms in its outputs (e.g., recommendations, decisions, and actions).

Furthermore, EA/IS entails a heightened level of care required in designing and controlling the system (e.g., human oversight, additional controls, transparency).

4.2.4.2 Activities and tasks

The developer *shall* do the following:

a) Demonstrate that all affective data was ethically obtained (e.g., with prior informed consent) and the subjects who provided the data are adequately protected (e.g., anonymous).

b) Validate all affective data so that it represents the diversity of the population it is intended to interact with, as far as possible.

c) Cleanse the system's affective data of any identifiable data issues frequently and thoroughly.

The developer *should* do the following:

d) Collect affective data from a wide range of sources (e.g., culturally diverse) to improve accuracy and reliability of system outputs.

e) Label affective data in a way that accurately reflects the different perspectives and experiences of the people who provided the data.

f) Provide subjects with controls to manage how their data is processed and intervene in the function of the system (e.g., shutting it off) when required.

g) Provide subjects with the facility to access and securely share information about the inferences made about them and the system decisions made based on those inferences.

h) Provide qualified human oversight of, and intervention in, the system.

i) Provide a method for subjects to calibrate the system (e.g., fidelity of outputs).

j) Provide a competent representative (e.g., assigned family member or carer) when the subject may lack adequate mental capacity (e.g., of making safe and healthy decisions about their use of the system).

4.2.4.3 EA/IS-specific particularities

EA/IS can enable building trust between humans and technology. As users interact with the system and receive personalized and empathic support, they may feel more comfortable and confident in using the technology to make data-driven decisions. This can produce both positive and negative outcomes that can be both foreseeable and unforeseeable.

EA/IS collect sensitive personal data, including emotional and psychological information about subjects. This data is highly sensitive and thus requires that it is acquired with prior informed consent and stored and processed securely with access limited to authorized personnel. To the greatest extent possible, the developer is required to provide control of data to the subject and keep them informed as to how it is being used.

EA/IS tend to use more complex algorithms to make recommendations and decisions than other forms of A/IS or traditional systems. Thus, this standard requires a heightened requirement for these algorithms to be made transparent and explainable. This way, subjects can understand how the system arrived at its outputs, have confidence in the system's decision-making, and be able to make informed decisions about the system.

4.2.5 Context and diversity

4.2.5.1 Purpose

The efficacy of EA/IS, as well as their attendant ethical factors, can be greatly affected by context. For example, the context (e.g., geographic, environmental) in which affective data is recorded can impact on the performance of affective models and can produce different results depending on the context in which it is deployed.

Various factors affect the capability of developers to measure or emulate affect accurately and reliability (e.g., camera systems have been found to analyze lighter skin tones more effectively than darker tones). The analysis of affective data can reinforce erroneous algorithm loops. This can lead to negative consequences that are potentially increased for vulnerable parties, such as the young, people with cognitive disabilities, and marginalized groups. Additional care is required for such vulnerable parties.

With respect to EA/IS, contextual considerations can include many factors, such as the following:

— Cultural aspects
— Physical environment (e.g., location, weather, time, speed)
— Formal versus informal settings
— Social status and relationships

In particular, there are many variable aspects of human diversity that can impact on subjects' emotional expressions, and the analysis of affective data by systems and their developers. Relevant diversity variables include but are not limited to the following:

— Age
— Race
— Assigned sex
— Gender identity and gender expression
— Physical ability and disability
— Language

— Ethnicity
— Cultural background, and current cultural context
— Religious and spiritual beliefs
— Personality, and neurodivergence
— Health condition
— Morphological variation

4.2.5.2 Activities and tasks

The developer *shall* do the following:

a) Publish an explanation of how the system adheres to human affective rights (see definition in 3.1, norms and regulations across all relevant vectors of diversity. For more on affective rights, see Ferraro [B16]).
b) Publish documentation of any potential biases in the system, and the steps taken to manage them.
c) Identify, enumerate, and publish the contextual scenarios in which the system is designed to be deployed.

The developer *should* do the following:

d) Make all reasonable effort to identify, record, and reduce any potential bias that may develop within the system.
e) Include methods or tools for checking and mitigating for biases.
f) Include contextual factors in the functioning of the system (e.g., to design interactions that are as natural as possible).
g) Provide the system with the capability to consider and adapt to variations in context (e.g., time, location, speed, time of day, changing societal mores) that can all impact affect.

4.2.6 Quality assurance

4.2.6.1 Purpose

In consideration of the heightened risks, issues, and impacts associated with EA/IS (see 1.4), such technologies entail heightened requirements for quality attributes, such as accuracy, performance, and reliability.

At the same time, EA/IS raises a number of ethical issues related to quality assurance aimed at achieving responsible and ethical operation of the system. Some key concepts for quality assurance include the following:

a) In recognition of the probabilistic and subjective nature of affect and EA/IS (as outlined in B.4.2), claims of system "accuracy," such as to provide a precise measurement of affect, can never represent objective truth. Accuracy claims can only be based on assumptions or simplifications of the subject's affect relative to a chosen measure, such as the subject's self-report.
b) The concept of "high accuracy" does not necessarily correlate with a more ethical system, as claims of high accuracy can increase risks for subjects such as raising believability of, and reliance on, the system's capabilities, as well as a heightened sense of the system's invasiveness experienced by the subject.

c) Regardless of the learning techniques applied (e.g., supervised or unsupervised), the effectiveness and reliability of an EA/IS's algorithms can vary as they continue to learn, thus changing the risk levels inherent in the system.

d) Compared to other A/IS or traditional systems, EA/IS tend to exhibit low "reproducibility" (i.e., the ability of an independent team to replicate, in an equivalent environment, domain, or area, the same experiences or results using the same methods, data, software, codes, algorithms, models, and documentation, and to reach the same conclusions as the original research or activity).

e) Compared to other A/IS or traditional systems, EA/IS tend to be more greatly affected by context, both in their design (e.g., context of data acquisition) and deployment (e.g., context of use). Thus, "robustness" (i.e., the stability, resilience, and performance of the system in dealing with changing environments and contexts) is paramount to quality. However, compared to other A/IS or traditional systems, EA/IS tend to exhibit low "robustness" and, thus, their fitness for purpose.

f) Compared to other A/IS or traditional systems, once some EA/IS are operationalized they can adapt through learning and training to a heightened degree. Quality assurance processes should recognize this potential and monitor the EA/IS after it has been put into production.

4.2.6.2 Activities and tasks

The developer *shall* do the following:

a) Publish information on the methods and results of measuring the system's quality, performance, effectiveness, and fitness-for-purpose, using appropriate standard methods and metrics.

b) Publish evidence of the degree to which the EA/IS has been shown to perform at its stated purpose, in real or realistic contexts.

c) Reduce claims of "accuracy" or other attributes that could be understood as accuracy. Instead, the developer may publish values for "confidence," "certainty," "error," or other probabilistic measurements. See B.4.2.

d) Publish information as to how the system's claimed "accuracy" (if any such claims are made) is proportional to the system's risks, issues, and impacts (e.g., a casual game may not carry the same need for "accuracy" as a medical app).

e) Include a method or methods for subjects to challenge the results of the system that is proportional to the risk level of the system, such as the following:

   1) For a low-risk system, a user feedback form may suffice.

   2) For a high-risk system, an interactive feature for the users to disagree with results directly within the system interface may suffice.

f) Sign off on product authentication and be accountable for accuracy of claims before deployment.

g) Perform regular reviews of the affective data and carry out procedures to correct it if necessary.

4.3 Technical processes

4.3.1 Stakeholder needs and requirements definition

4.3.1.1 Purpose

For the ethical development, deployment, and decommission of EA/IS, this standard requires continuous, active engagement with a diverse set of stakeholders, including but not limited to the following:

a) Those responsible for development, business strategy, and senior leadership, designers, producers, suppliers, and marketers who design, set requirements, or use/operate the system

b) Regulatory bodies, government, compliance, auditing and risk agencies, industry bodies, and others who may assess the system

c) Affected parties such as users, data subjects, affected non-users, and other interested parties

The developer is required to define all relevant classes of stakeholders and their needs. These needs can then be used to create system stakeholder requirements. It is important to engage actively with stakeholders throughout the entire system life cycle in order to create more effective, usable, and ethical systems. This process aims to help the system meet the needs of all parties involved, to contribute to human flourishing, and reduce potential negative outcomes.

Stakeholder needs and requirements definition can help the following:

d) Identifying potential risks and ethical considerations associated with the system. This can help developers proactively address these issues and manage potential negative impacts

e) Helping the system meet legal and regulatory requirements, including applicable data protection and privacy laws

f) Building trust and credibility with users and other stakeholders by demonstrating that their needs and concerns have been taken into account

g) Holding developers and other stakeholders accountable and responsible for the system and its outcomes

It is a general attribute of EA/IS that there will always be at least one affective data subject, whose data is collected or analyzed for the system to function. The processing of this intimate class of data creates an obligation for the developer to afford such subjects adequate protections (e.g., data security) and ethical treatment (e.g., prior informed consent).

NOTE—Disambiguation: The term “stakeholder(s)” is used throughout this standard to describe those parties that are potentially affected by, or have an interest in, the system (e.g., developers, policymakers, subjects, end users). These stakeholders are addressed in this section. However, to be clear, the primary “stakeholder” of this standard, whom the standard primarily addresses, is different. This stakeholder is the developer (including third-party resellers, etc.) See definition of “developer” in 3.1.

4.3.1.2 Activities and tasks

The developer *shall* do the following:

a) Identify and analyze the different stakeholders who could interact with, or be affected by, the system.

b) Conduct research (e.g., interviews, surveys, focus groups) to gather needs and preferences of stakeholders. This involves capturing their expectations, desired functionalities, and the level(s) and mode(s) of empathy they expect from the system.

c) Publish the findings of stakeholder analysis and how it influences the development, deployment and decommission of the system.

The developer *should* do the following:

d) Treat all stakeholders as subservient to the subject (e.g., user, data subject), who should have the final say where possible.

e) Treat the system's default audience as to include vulnerable parties such as children, and otherwise formally declare if it is not the default audience, as well as declaring how the system is restricted to other stakeholder classes (e.g., adults).

f) Apply a human-centered approach to system development, deployment, and decommission by prioritizing stakeholder needs and requirements at all stages.

4.3.2 Security, privacy, and consent

4.3.2.1 Purpose

EA/IS generally operate with particularly sensitive and intimate types of data (e.g., physiological metrics), and interactions (e.g., estimating the subject's mood) that can lead to negative outcomes, such as psychological manipulation or violations of the subject's right to freedom of thought and feeling, and, thus, human autonomy and free agency. Therefore, EA/IS, and the empathic interactions that subjects experience, heighten safety, security and privacy concerns, as well as wider ethical concerns. This further produces heightened levels of responsibility and accountability for the developer that are somewhat proportional to the risk and sensitivity levels of the data and interaction modes of the system.

An increasing number of professionally adopted frameworks [e.g., the General Data Protection Regulation (GDPR) in the European Union] address special classes of data, with requirements for special treatment thereof. Examples include data on political opinion, membership, and religion, which are considered to relate to possibly private or intimate knowledge, experience, and identity. The affective data that is broadly used in EA/IS includes similarly "intimate" data, such as that relating to physiology, behavior, thought, affect, and opinion. Thus, this standard considers EA/IS to generally contain data that requires special care and consideration.

Furthermore, EA/IS exposes stakeholders to a novel concern in relation to its ability to "feel into" or expose information that the subject would have otherwise been able to disguise, control, or keep private. Humans have several ways in which they can keep knowledge about their affect private, such as by masking our expressions (e.g., trying not to show pain for the sake of a grandchild so they won't get worried) or displaying alternative expressions in order to comply to cultural norms (e.g., giving a polite smile when a colleague says something tactless). In some cases, EA/IS reduce or alter the subject's ability to manage their affect, and its expression and communication, and as such degrades the subject's agency over their affective rights.

Official frameworks such as the GDPR give particular attention to biometric data. Biometric data (See IEC JTC 1/SC 37 [B20]) can be considered to comprise two separate classes of "hard" and "soft" biometrics. "Hard" biometrics are typically connected with identification of a person (e.g., facial recognition), while soft"

biometrics concern the features of a person (such as behavior or facial expression). In general, EA/IS are developed using "soft" biometrics, which can engender heightened levels of care and consideration. This standard requires the developer to treat affective data (including "soft" biometrics) as having at least the same protections as personal health data does under current law.

4.3.2.2 Activities and tasks

The developer *shall* do the following:

a) Publish details of the system's safety, security, and privacy measures (e.g., encryption methods).

b) Obtain records of informed consent from all subjects or other sources of data acquisition. Consent shall be freely given, as well as being informed, convenient, appropriate, and meaningful, consistent with all applicable laws and regulations.

c) Inform subjects of known potential harms, prior to obtaining informed consent.

d) Obtain informed consent actively (e.g., via opt-in checkbox), never passively (e.g., via opt-out checkbox).

e) Set spatial (e.g., geographic) and temporal (e.g., time limit and frequency) limits to the informed consent licenses.

f) Maintain a record of the subject's informed consent, which is made easily available to the subject when required, for the entire system life cycle.

g) Provide subjects with the facility to reclaim their license at any time in the system life cycle. This is not required after the subject's affective data has been generalized or anonymized to the extent that it would be practically infeasible to remove it and, at the same time, the developer can provide assurance that the subject is no longer vulnerable to further acquisition, processing, or sharing of their affective data.

h) Provide subjects with the facility to retrieve their affective data with a log of how that data has been used, unless the developer can demonstrate that this is infeasible or inappropriate.

i) Not transfer a subject's affective data to other stakeholders, including internal (e.g., in another department of the developer's organization) or external, without having prior informed consent or obtaining revised informed consent.

j) Not share affective data beyond the scope of the original informed consent.

k) Not make use of the system conditional on granting access to subject's affective data to third-parties (e.g., ad networks).

l) Provide subjects with information regarding any third-parties (including but not limited to ad networks and content intermediaries) that can be party to affective data or influence system behavior.

m) Not use deceptive design patterns (e,g., coercive) to influence subjects' behavior (e.g., to opt in or out of any part of the system).

For high-risk systems, the developer *shall* do the following:

n) Include a capable human "in the loop" to provide oversight and intervention in the system.

o) Include an "emergency stop" mechanism for subjects to halt empathic functioning of the system.

p) Include secondary "watchdog" systems (including EA/IS where relevant) to monitor the system and protect subjects from harm.

Where possible, the developer *should* not unduly withhold delivery of service where a subject has chosen or failed to provide effective informed consent.

4.3.3 Data acquisition and management

4.3.3.1 Purpose

For the purposes of this standard, "data acquisition" refers to the process of acquiring and categorizing/labeling information that the system can use to perform EA/IS functions (e.g., estimate emotion, make affect-based decisions). The normative statements in this section apply whether the developer acquires this data directly or indirectly (e.g., via an external source).

This standard requires that the processes of affective data acquisition are designed to protect subjects' privacy and rights, promote high-quality data acquisition and management, and promote ethical use of the data.

Anonymous data is not considered sensitive or at risk.

Prior informed consent to data acquisition and processing is a key aspect of ethical system design.

4.3.3.2 Activities and tasks

The developer *shall* do the following:

a) Publish an explanation of the source(s) and method(s) of acquisition of affective data.

b) Publish an explanation of the volume, type(s), and range of affective data acquired (e.g., facial images from 1000 adult European male subjects).

c) Publish a data retention policy and plan in accordance with appropriate frameworks (e.g., GDPR) for treatment of sensitive data such as personal health data or personally identifiable information (PII). This shall include the following:

   1) High levels of data protection and security (e.g., multifactor authentication).

   2) Privacy or anonymization of affective data, with access restricted to authorized personnel.

   3) Automatic deletion of affective data based on a preset time period at the request of the subject.

   4) Definition of the intended purpose of use of affective data.

d) Only collect the data necessary to achieve their specific and limited purpose (as published by the developer).

e) Not retain affective data for longer than necessary. Once the data is no longer needed, it shall be securely deleted or anonymized.

f) State the intended use of the data, and only use it for the limited use for which they have obtained informed consent. If the data is to be reused, additional permission shall be sought. General-use permission is not acceptable.

g) Not acquire or use affective data for purposes that can foreseeably harm subjects, discriminate against them, or violate their privacy or rights.

4.3.4 System training

4.3.4.1 Purpose

EA/IS system training typically exposes the system to various types of affective data (e.g., text, images, or audio). In some training approaches, this data is labeled to indicate affect (e.g., emotion labels such as happiness, sadness, anger, or fear). This labeled data is then used to learn patterns and develop algorithms that enable the system to recognize, respond to, or emulate affect. In other training scenarios, (e.g., unsupervised learning) system training is done without labeled data and the system infers its own meaning to the data.

To develop empathy in an EA/IS the system is typically trained on how to interpret and respond appropriately to affective cues (e.g., changes in tone of voice, facial expressions, internal physiological signals, or changing text sentiment). Training the system with empathy involves developing an understanding of the subject's affect, as well as the surrounding context, such as cultural and individual differences that can influence emotional expression.

Effective system training also typically involves incorporating feedback from stakeholders to review and improve the system's performance over time. A process of continuous learning and adaptation is required to maintain an effective system as the context changes over time (e.g., changing attitudes toward affect, emergence of new technologies).

This standard requires the developer to apply heightened levels of care and quality to system training to promote human flourishing, as well as remove or mitigate risks, issues, and impacts.

4.3.4.2 Activities and tasks

The developer *shall* do the following:

a) Publish an explanation of the methods of system training and justification for the approach. See 4.2.3.

b) Train the system on affective data and models that are as diverse and culturally sensitive as possible. See 4.2.5.

c) Publish their approach to continuous incorporation of stakeholder feedback into system training, to improve system quality and robustness, and to promote ethical design and use.

The developer *should* do the following:

d) Use high-quality and recognized affective data sets (e.g., peer reviewed) if acquiring data from third parties.

e) Train the system on a mix of different modes of data (e.g., text, audio, image, various physiological metrics) where this is likely to improve the quality and robustness of the system.

4.3.4.3 EA/IS-specific particularities

Training EA/IS requires a diverse, multidisciplinary approach that combines expertise in machine learning, psychology, human-centered design, and ethical principles. By taking into account these unique considerations, developers can create EA/IS that are able to better understand and respond to the needs and emotions of their subjects.

4.3.5 Modeling of affect or empathy

4.3.5.1 Purpose

EA/IS typically include models of affect or empathy, usually based on machine learning methods. Any attempt to model such human, probabilistic (see B.4.2), and subjective phenomena entails ethical considerations. See 1.13.

This standard requires the developer to apply special care and consideration in modeling with the aim to develop, deploy, and decommission the system ethically.

4.3.5.2 Activities and tasks

The developer *shall* do the following:

a) Publish an explanation of the methods and frameworks used for model design and use. See 4.2.3. This shall include the following:
   1) Enumeration of the types of affect modeled, or influenced by, the model(s)
   2) Supporting validation (e.g., publications) for the approach
   3) Explanation of the ranges of validity of affective data and models (e.g., where and when they are expected to function effectively)

b) Provide information to all potentially affected stakeholders that clearly labels any affective inferences made by the system as such (e.g., labeled as "estimate" not as "truth").

c) If the system simulates affect (e.g., generates a smiley face, or says "I'm sorry"), information shall be provided to potentially affected stakeholders that clearly states that any expressions of affect are purely trained, logical responses and not the result of actual emotional expressions—or words to that effect.

d) Not apply any affective modeling to subjects' affective data (except that which is anonymous) without prior informed consent. See 4.3.2.

e) Analyze how consistent the system's model(s) is/are across all foreseeable contexts of use.

f) Analyze the effectiveness of the EA/IS model(s) with a representative sample of all relevant stakeholders.

Where possible, the developer *should* provide subjects with the facility to be able to calibrate the degree of intimacy and obtrusiveness of the system's behavior.

NOTE—For more on how to transparently identify human and machine agency, see IEEE P3152 (Draft 12, February 2024), Draft Standard for Transparent Agency Identification of Humans and Machines [B21].[15]

4.3.6 Ongoing monitoring and validation

4.3.6.1 Purpose

Monitoring is the process of collecting, analyzing, and using information to track applications and infrastructure to guide the development, deployment and decommission of the system or to guide decisions (e.g., business decisions). Monitoring provides developers with feedback to help them quickly find and fix problems with the systems. The methods for operating and monitoring EA/IS are generally similar to those of other AI/S, but the EA/IS can introduce novel complexity, challenges, and risks and, thus, require heightened care.

To address these challenges, monitoring the deployment of EA/IS requires a multi-faceted approach that involves ongoing evaluation, testing, and subject feedback. This may involve developing new metrics to assess the effectiveness and ethical implications of the system as well as conducting regular user studies and surveys to gather feedback and identify areas for improvement. Additionally, it may be necessary to establish clear policies and guidelines for the use of EA/IS (e.g., within the developer's organization or for system users), both to promote their ethical use and to reduce the risk of unintended consequences.

Overall, monitoring an EA/IS is an important part of ensuring that the system operates effectively and ethically. Monitoring requires a combination of methods and strategies tailored to the specific system and its context. In particular, monitoring is required for the inputs and outputs of the affective model(s) in the system in which they operate. Each model requires a different approach and different metrics that require publication prior to deployment.

[15]Numbers preceded by P are IEEE authorized standards projects that were not approved by The IEEE SA Standards Board at the time this publication went to Sponsor ballot/press. For information about obtaining drafts, contact the IEEE.

A particular challenge for the ongoing monitoring of EA/IS is that machine learning models are known to decay over time, and, as such, need to be deployed again and again. The affective models that are used in EA/IS have potential to decay faster than other machine learning models and drift further from their intended behavior. Thus, there is a requirement for close monitoring throughout the use of the system to determine if the statistical properties of the target variable(s) change in unforeseen ways.

The factors outlined in this subclause, including the uniquely rapid drift of fitness-for-purpose in the empathic context, entail a requirement for close, high-frequency monitoring, validation, re-evaluation, and modification throughout the system life cycle. This may include some form of AI-enabled monitoring able to automatically shut down or pause the system and/or refer it to a human for review and decision.

4.3.6.2 Activities and tasks

The developer *shall* do the following:

a) Perform ongoing monitoring of the EA/IS. This may include, but is not limited to, system drift from its original goal, performance, accuracy metrics, response time, and user satisfaction.

b) Publish an explanation of the system's monitoring plan, and ongoing results. This shall include the follwoing:

   1) Explanation of the monitoring process and features (e.g., System alerts on critical thresholds, if relevant).

   2) Performance metrics for each model (e.g., metrics monitored include R2, root mean squared error, mean absolute error, explained variance).

   3) Coverage of all key areas of the system and a gap analysis of what is being monitored versus what is not (since monitoring everything is not possible, and ensuring that the right data is used is important).

   4) Explanation of the features and methods for models to remain production-ready and perform as intended.

   5) Explanation of how system monitoring addresses aspects that are unique or heightened in the empathic context (e.g., inaccurate or potentially embarrassing claims of the subject's affect, changes in system context). See 1.13.

c) Engage relevant stakeholders (e.g., users) for feedback on the system's ongoing performance and issues.

d) Publish the procedure and expected time for restoration of the system following a service incident or detection of an issue that could impact subjects.

The developer *should* do the following:

e) Use qualified third-party assessment and auditing of the system.

f) Monitor the system in as close to real time as feasible.

g) Perform human oversight of the system, and intervention where required.

4.3.7 Decommission and disposal

4.3.7.1 Purpose

The end of the EA/IS life cycle requires that all system components (e.g., models, data) are effectively and ethically retired, deleted, decommissioned, or disposed of, and cannot be reused or mistreated.

This standard requires the developer to consider the potential impacts of decommission and disposal of the system that are heightened in the context of EA/IS. For instance, subjects may have developed a bond or attachment to the system and could suffer psychological harm from its removal.

4.3.7.2 Activities and tasks

The developer *shall* do the following:

a) Publish a decommission and disposal plan prior to deployment. This shall include the following:
   1) The methods of decommission and disposal, and justification for the approach
   2) Foreseeable risks, issues and impacts of decommission and disposal including, but not limited to the follwoing:
      i) Personal (e.g., psychological) outcomes for subjects
      ii) Societal outcomes
      iii) Environmental outcomes (e.g., waste of hazardous materials in electronic components)
      iv) Security breaches (e.g., data theft)
      v) Policy or compliance breaches
   3) Any reasonable measures to ameliorate harms from decommission and disposal (e.g., provision of counseling or other support for any subjects who may be affected by the loss of the system).
b) Adhere to the expiry date and plan for all affective data, according to the system's data retention policy. See 4.3.3.

NOTE—ISO/IEC 27001 [B29] provides a framework for information security management systems (ISMS), including the secure disposal of electronic devices and data.

Annex A

(informative)

Examples and use cases

A.1 Examples of EA/IS

Some examples of EA/IS include the following:

a) Emotion estimation in natural language processing (NLP) to facilitate naturalistic system interaction, such as a virtual assistant that responds in a more empathic and human-like manner

b) An emotion-sensing wearable that detects health issues via sensors that measure physiological changes (such as heart rate, skin conductance, or neurological outputs) to infer a user's emotional state

c) A music streaming service that uses algorithms to select songs based on estimation of a user's emotional state to create personalized playlists that cater to their mood

d) A social robot designed to interact with humans in a more natural and empathic way by recognizing and responding to emotions through facial recognition, voice analysis, or other emotional analysis modes

e) A chatbot designed to communicate with users in a conversational manner, with a simulation of emotional intelligence, by recognizing and responding to the user's emotional state

A.2 Use cases

This subclause includes examples of EA/IS use cases, showing how positive impact or potential harm can result. These examples are purely illustrative.

A.2.1 Mental health support

Positive: Help medical personnel to quickly respond to mental health issues, such as anxiety or depression. The EA/IS can analyze a user's tone of voice, facial expressions, and other cues to detect signs of distress and offer appropriate support.

Potential harm: EA/IS used for mental health diagnosis without sufficient human oversight could potentially misdiagnose or stigmatize individuals based on their emotional data. This can lead to incorrect or harmful treatment.

A.2.2 Hiring

Positive: To help companies to improve hiring practices by assessing the emotional response of candidates and/or hiring managers (e.g., during interviews, to refine questions, improve interpersonal interaction).

Potential harm: One specific potential harm is biased decision-making by the system. EA/IS can learn biases from the data they are trained on, and if not properly addressed, these biases can lead to discriminatory or harmful decisions. For example, EA/IS used in hiring may inadvertently discriminate against certain groups of people based on their emotional data.

A.2.3 Customer service

Positive: EA/IS could be used to improve customer service experience. For example, if a customer is frustrated or angry, the system can detect this and respond in a more compassionate and understanding way, which can help defuse the situation.

Potential harm: EA/IS may inadvertently reinforce certain biases or stereotypes, as their learning algorithms can inadvertently pick up and amplify existing societal biases in the customer service activities because of the data they are trained on.

A.2.4 Education

Positive: Improve the learning experience for students. For example, by analyzing a student's emotional state, the EA/IS can adjust the pace and difficulty of the material so that the student is not overwhelmed or bored.

Potential harm: The EA/IS can misinterpret the student's emotional state and respond inappropriately, causing psychological harm. It may also incorrectly recognize a student's cognitive ability related to a topic and respond inappropriately, causing psychological harm.

A.2.5 Personal assistants

Positive: To provide more personalized assistance to individuals. For example, the system can detect when a user is feeling stressed or overwhelmed and offer helpful suggestions or activities to help them relax.

Potential harm: By being trained on the personal data and context of an individual, the system can be perceived as becoming uncomfortably intimate with the user or expose them to potential breaches of their security or psychological well-being.

A.2.6 Autonomous vehicles

Positive: To improve safety in autonomous vehicles. For example, by analyzing a passenger's emotions, the EA/IS can adjust the driving style to provide a more comfortable and safer ride.

Potential harm: Increased levels of anxiety due to the continuous monitoring of emotional states and the potential for the autonomous vehicle responding and acting incorrectly as a result.

A.2.7 Gaming

Positive: To provide a more immersive gaming experience. For example, the system can detect when a player is feeling anxious or excited and adjust the game's difficulty or storyline to match the player's emotional state.

Potential harm: Adapting to the user's playing style to provide increasingly strong or frequent emotional reactions can lead to misuse, overuse, overreliance, or desensitizing the user to real-life emotional experiences.

A.2.8 Marketing

Positive: To improve marketing strategies. For example, by analyzing a customer's emotional response to ads or promotions, the EA/IS can adjust its approach to better resonate with the customer's emotional state.

Potential harm: EA/IS that use emotional data to influence individuals' decisions or behavior without their knowledge or prior informed consent can be seen as manipulative or unethical. This could further lead to distrust of the technology.

A.2.9 Legal

Positive: Assisting the prosecution process. For example, by analyzing common trends in psychophysiological signals from criminals.

Potential harm: EA/IS used in the criminal justice system, such as predictive policing or sentencing algorithms, can be biased and perpetuate systemic racism and discrimination. This can result in unfair treatment of certain groups of people based on their emotional data.

A.2.10 Smart cities

Positive: Insights for designing better urban environments. For example, by estimating group mood in urban spaces based on signals such as walking speed, gait, and facial expressions.

Potential harm: EA/IS used for surveillance purposes, such as detecting emotions in public spaces or monitoring social media, can infringe on individuals' privacy and civil liberties. This can lead to a lack of trust in the technology and the entities that use it.

Annex B

(informative)

Background on the science of emotion, affect, and empathy

B.1 Emotion and affect

There have been many treatises and musings on the nature of emotion and how it influences human behavior throughout the centuries of recorded thought. However, the modern scientific exploration of emotions is usually traced back to Charles Darwin and the publication of his book "The Expression of the Emotions in Man and Animals" [B10]. Certainly, this book started a strand of emotion research that has a certain preoccupation with the relationship between emotion, faces, and the contraction of facial muscles that produces expressions. This tradition led onto the work of Paul Ekman and colleagues in the 1960s and 1970s and emphasized the role of evolution, with an argument for the universality of emotion and the existence of a discrete set of basic emotions with concomitant facial expressions. Not long after Darwin's book William James published an article "What is an Emotion?" [B34] in the philosophical journal *Mind*. James argued that physiological changes that precede the felt-state of emotion or physiological changes that occur as a result of a perceived emotion inducing event are the emotion itself. This has become known as the James-Lange theory as it is shared with Carl Lange. Importantly, for current purposes, this represents the start of another strand of emotion research that gives a primacy to the physiological changes that occur when people are said to be in an emotional state. This strand, in turn, led to the Cannon-Bard [B7] and Schachter-Singer [B43] theories of emotion that argued for differing conceptualizations of the role that physiology and the autonomic nervous system in particular play within the human experience of emotion. Schachter and Singer [B43] brought in a cognitive element to the understanding of emotion, and together with the work of Magda Arnold [B2], started a third major strand of emotion theories in psychology that emphasize the cognitive interpretation or appraisal of physiological feelings as part of the emotional experience. Important figures in this tradition are Nico Frijda, who emphasized the role of emotion in preparing people for action [B18] and Richard Lazarus ("Relational Meaning in Discrete Emotions" in *Appraisal Process in Emotion: Theory, Methods, Research*[B1]). A final major theoretical strand emphasizes social and cultural elements in the understanding of emotion. The social constructivist strand arises with James Averill [B3]. Other constructivist approaches were developed by James Russell [B42] and, most recently, by Lisa Feldman Barrett [B5].

B.2 Empathy

Where conversations about the nature of emotion stretch back into antiquity, the word empathy is thought to have been coined by psychologist Edward Titchener in the early 20th century from the Greek word "empatheia." Conceptually it is derived from the German "Einfühlung," which translates as "feeling-into," drawn from German aesthetics and formulated into something people would recognize as modern empathy by Theodor Lipps (Wispe [B47]). In a similar manner to ideas around emotion, the meaning of the word empathy remains ill-defined. Benjamin M.P. Cuff conducted a review in 2016 noting at least 43 different ways to define empathy [B8]. While there was a lot of overlap in these definitions, some were contradictory. One commonly used division of the concept is into *affective* empathy, which captures a shared emotional experience, and *cognitive* empathy, which is oriented toward understanding and making inferences about others' internal states. The role of these components of empathy is a subject of much debate, as is how they may interact with one another. Another important aspect of empathy is its interpersonal nature. Where emotions can be experienced as individual experiences, empathy has an inherent interpersonal nature. A caveat is that empathy is often measured using self-report psychometric scales that treat it as an individual difference. Davis's Interpersonal Reactivity Index (IRI) [B11] is a popular measure as is Bagby, Parker and Taylor's Toronto Alexithymia Scale (TAS) [B4].

B.3 Theoretical foundations and their limitations

Presented here is a brief discussion on the theoretical foundations of emotion science, affective computing, and some key limitations in how they influence the design and use of EA/IS. The design of any EA/IS is generally founded on one or more psychological theories, but there are issues with any theory underlying EA/IS. The picture painted here suggests that much debate exists concerning emotions and empathy. In particular, there is debate concerning the nature of the internal felt-states at the core of these phenomena and how these felt-states are outwardly expressed as measurable phenomena. From the relatively new science of Affective Computing has arisen a suite of technologies that measure visual, acoustic, physiological, and biomechanical human behavior, and subsequently pertain to infer interior states. This is becoming both inexpensive and pervasive. The various strands and theories of emotion and empathy research each recommend differing interpretations of how these measurements relate to the internal felt-states of the individuals being measured.

While these technologies can, in some cases, make accurate measurements of physiological features or movements, the further a system travels from the physiological toward the psychological, the more it must buy into one of these often-competing theoretical views. These newly enabled approaches to measuring human behavior create a situation in which responsibility is placed on the shoulders of the developers and users of any system that purports to interpret emotional state from behavioral measurement.

To summarize some of the leading theories of emotion, Ekman describes six basic/discrete emotions [B12]; Mehrabian and Russell prefer a dimensional description [B38]; Scherer and predecessors prefer an appraisal view (*Appraisal Process in Emotion: Theory, Methods, Research*[B1]); Averill [B3] and Barrett [B5] advocate social constructionist views; Fridlund thinks the word emotion is defined so differently by so many people that it is effectively useless [B17]. As is common in an academic evaluation, there is not a consensus on what emotion means.

There is also a temporal element to the study of affect that is argued about in the scientific community. Some see emotions are very short lived, while others argue that they are longer.

Furthermore, emotional experience itself is multidimensional in nature. A complete assessment of the emotional experience would require an understanding of the subjective experience of how the subjects themselves would use language, words, or other expressions to describe their experience, as well as the different physiological responses related to the emotion, including autonomic, brain, and motor, as well as behavior, which is what an observer would usually be able to "read."

The correlation between the different dimensions of an emotion experience is not necessarily perfect. For an anecdotal example, consider when people mask their expressions—especially their facial expressions. Technically, the choice of how they express or control their experience is also part of that emotion experience. For instance, a person might feel guilty but not let it show because of related feelings of shame, when, at other times, they might feel guilty and want the person they are interacting with to clearly see how guilty they feel. In the scientific community there are several reasons for disagreements between dimensions of affect, and this entails that any attempt by an EA/IS to emulate affect is only ever one part of the story, even if that system includes multimodal recognition (i.e., using combinations of multiple physiological and/or behavioral signals).

Considering these issues, any EA/IS might perform well in some circumstances but not in others, and it can be impractical to predict all such circumstances. The implications of this limitation can include a user disagreeing with an inference of the system. In the scientific community, the gold standard for assessing affect is the self-report, in which the subject describes how they feel. Generally speaking, it would be inappropriate for an EA/IS to disagree with a subject's self-reported affect, although there can be allowances in special circumstances, such as when a qualified medical professional chooses to "override" the subject's claim with a qualified observation.

B.4 Key ethical considerations

B.4.1 The assumption of affect can impact on the rights of subject(s)

Any attempt by an EA/IS to estimate or simulate (infer) affect entails assumptions about how the subject feels, which may be incorrect, misleading, intrusive or harmful. As described in further detail in the definition of affective rights (see 3.1) and Angelo Ferraro's paper on Affective Rights [B16] the subject is granted fundamental rights to their freedom of feeling, thought, and expression; the right to their privacy (e.g., for their feelings to remain private); and the right to protest.

Sensitivity is required for how the subject is presented with the affective inferences made by the system, as they cannot be objective fact. For instance, when an inferred state is presented to a system user, it can be interpreted by the user as being told how they "truly feel" or are "supposed to feel." Presentation of the inferences should not bend the subject to those inferences. Such categorizing of people (e.g., labeling them as "happy" or "sad") engenders a need for heightened agency for affected stakeholders (e.g., system users).

These issues are potentially increased for vulnerable parties, such as the young, the elderly, people with cognitive disabilities, and marginalized groups. Additional care is required for such vulnerable parties, and consideration is required pertaining to the expression of emotions and their meanings within different cultural contexts.

With these issues in mind, it may be preferable for some systems not to present inferences of affect in the first place.

Developers are required to consider how the design and action of their system could infringe on the subject's rights. Guidance is provided in this standard for developers to manage this potential issue, but caution is always prudent.

B.4.2 Accuracy is contentious, probabilistic, and subjective

Any attempt by an EA/IS to estimate or simulate affect entails assumptions about how accurately that affect is estimated or simulated relative to reality. Since human affect is subjective in the first place—both from the perspective of the person or group having a feeling and from the outside perspective of an observing person or system—accuracy is always in question.

Furthermore, there is evidence that people trust systems and assume accuracy or truth. They believe what they are being told, even though the system is only able to make estimates based on contentious models.

This accuracy challenge entails ethical considerations. Guidance is provided in this standard for developers to address issues of accuracy, but caution is always prudent.

B.5 Climate and energy

With empathic and affect-sensitive technologies often making use of AI and deep-learning technologies, questions of energy consumption are raised. "Energy mindfulness" is advisable, given that CO_2 and other environmental/pollution costs of a) training deep learning systems and b) subsequent usage of these trained systems is high. Mindfulness in this setting means factoring for how accurate an AI system needs to be and the complexity of the models generated therein. Accuracy in this context refers to object and feature detection, such as a facial expression rather than "accuracy of measuring emotion."

"Energy justification" is also advisable. This is a principle that echoes "purpose limitation" in that any system deployed should be mindful of the energy required to fulfill a given organizational or system goal. Thought should also be given to where energy is sourced, demands made by devices in edge-based systems, and physical materials such as metals involved in production.

Annex C

(informative)

Required materials

Table C.1 provides a list of all materials required to be published to conform to the standard, listed in one place.

Table C.1—Required materials

Material (e.g., document, notification)	Relevant section in this standard
Well-being impact assessment.	4.2.1
Risk, impact and issue assessment.	4.2.2
Notes on the system's purpose.	4.2.3
Notes on the system's intended use.	4.2.3
Signed statement of conformance to the standard	4.2.3
Notification that EA/IS is in use.	4.2.3
Notification of the nature of EA/IS in use.	4.2.3
Cautionary notification of the probabilistic and subjective nature of EA/IS.	4.2.3
A bill or materials (BOM) for hardware, software and data.	4.2.3
Data management plan.	4.2.3
Quality and performance measurement.	4.2.6
Explanation of "accuracy" claims (if any are made), proportional to the risks, issues and impacts	4.2.6
Evidence of the effectiveness and fitness-for-purpose of the system.	4.2.6
Stakeholder analysis.	4.3.1
Safety, security and privacy measures.	4.3.2
Records of informed consent (for acquisition and use of affective data).	4.3.2
Explanation of the source(s) and method(s) of acquisition of affective data.	4.3.3
Data retention policy and plan.	4.3.3
System training methods and approach, including use of Stakeholder feedback for System improvement.	4.3.4
Explanation of the methods and frameworks used for model design and use, including supporting validation (e.g., publications) for the approach.	4.3.5
Monitoring plan and results.	4.3.6
Service restoration procedure and expected time (in case of system incident).	4.3.6
Decommission and disposal plan.	4.3.7

Annex D

(informative)

Bibliography

Bibliographical references are resources that provide additional or helpful material but do not need to be understood or used to implement this standard. Reference to these resources is made for informational use only.

[B1] Appraisal Process in Emotion: Theory, Methods, Research (Series in Affective Science), Scherer, K.R., A. Schorr, and T. Johnstone, eds. Oxford University Press, 2001.

[B2] Arnold, M. B., Emotion and Personality: Volume 1: Psychological Aspects. New York: Columbia University Press, 1960.

[B3] Averill, J.R., "Chapter 12—A Contructionist View of Emotion," in Theories of Emotion, Plutchik, R. and H. Kellerman, eds. Academic Press, 1980, pp. 305–339.

[B4] Bagby, R. M., J. D. Parker, and G. J. Taylor, "The twenty-item Toronto Alexithymia Scale I—Item selection and cross-validation of the factor structure," Journal of Psychosomatic Research, vol. 38, no. 1, pp. 23–32, January 1994, http://dx.doi.org/10.1016/0022-3999(94)90005-1.

[B5] Barrett, L. F., How Emotions Are Made: The Secret Life of the Brain. New York: Houghton Mifflin Harcourt, 2017.

[B6] Barrett, L. F., B. Mesquita, and M. Gendron, "Context in Emotion Perception," Current Directions in Psychological Science, vol. 20, no. 5, pp. 286–290, October 2011, http://dx.doi.org/10.1177/0963721411422522.

[B7] Cannon, W. B., "The James-Lange theory of emotions: A critical examination and an alternative theory," American Journal of Psychology, vol. 39, no. 1/4, pp. 106–124, 1927, http://dx.doi.org/10.2307/1415404.

[B8] Cuff, B. M. P, S. J. Brown, and D. J. Howat, "Empathy: A Review of the Concept," Emotion Review, vol. 8, no. 2, 2016, https://doi.org/10.1177/1754073914558466.

[B9] "Convention on the Rights of the Child," United Nations Treaty Collection, Status of Treaties, Chapter IV, Section 11.

[B10] Darwin, C., 1872. The Expression of the Emotions in Man and Animals. London: John Murray, 1927.

[B11] Davis, M. H., "Measuring individual differences in empathy: Evidence for a multidimensional approach," Journal of Personality and Social Psychology, vol. 44, no. 1, pp. 113–126, 1983, http://dx.doi.org/10.1037/0022-3514.44.1.113.

[B12] Ekman, P., "An argument for basic emotions," Cognition and Emotion, vol. 6, no. 3-4, 2008, https://doi.org/10.1080/02699939208411068.

[B13] Elfenbein, H. A. and N. Ambady, "On the universality and cultural specificity of emotion recognition: A meta-analysis," Psychological Bulletin, vol. 128, no. 2, pp. 203–235, March 2002, http://dx.doi.org/10.1037/0033-2909.128.2.203.

[B14] Ethically Aligned Design, IEEE, https://standards.ieee.org/industry-connections/ec/.

[B15] European Commission, Proposal for a Regulation laying down harmonized rules on artificial intelligence, April 21, 2021.[16]

[B16] Ferraro, A., "Affective Rights: A Foundation for Ethical Standards," 2020 IEEE International Symposium on Technology and Society (ISTAS), pp. 1–11.http://dx.doi.org/10.1109/ISTAS50296.2020.9462172[17,18]

[B17] Fridlund, A. J., Human Facial Expression: An Evolutionary View. Academic Press, 1994.

[B18] Frijda, N.H., *The Emotions*. Cambridge University Press, 1987.

[B19] Hollis, V., A. Pekurovsky, E. Wu, and S. Whittaker, "On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vo. 2, no. 3 Article No. 114, p. 1–31, Sept. 2018, http://dx.doi.org/10.1145/3264924.

[B20] IEC JTC 1/SC 37, Biometrics.[19]

[B21] IEEE P3152 (Draft 12, February 2024), Draft Standard for Transparent Agency Identification of Humans and Machines.[20]

[B22] IEEE Std 7000™-2021, IEEE Standard Model Process for Addressing Ethical Concerns during Systems Design.

[B23] IEEE Std 7001™-2021, IEEE Standard for Transparency of Autonomous Systems.

[B24] IEEE Std 7002™-2022, IEEE Standard for Data Privacy Process.

[B25] IEEE Std 7007™-2021, IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems.

[B26] IEEE Std 7010™-2020, IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being.

[B27] ISO/IEC 5338, Information technology—Artificial intelligence—AI system life cycle processes.[21]

[B28] ISO/IEC 5962:2021, Information technology—SPDX® Specification.

[B29] ISO/IEC 27001, Information security, cybersecurity and privacy protection—Information security management systems—Requirements.

[B30] ISO/IEC/IEEE 12207:2017, Systems and software engineering—Software life cycle processes.

[B31] ISO/IEC/IEEE 15288:2015, Systems and software engineering—System life cycle processes.

[16]Available at: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence
[17]IEEE standards or products are trademarks owned by The Institute of Electrical and Electronics Engineers, Incorporated.
[18]IEEE publications are available from The Institute of Electrical and Electronics Engineers (https://standards.ieee.org/).
[19]IEC publications are available from the International Electrotechnical Commission (https://www.iec.ch) and the American National Standards Institute (https://www.ansi.org/).
[20]Numbers preceded by P are IEEE authorized standards projects that were not approved by The IEEE SA Standards Board at the time this publication went to Sponsor ballot/press. For information about obtaining drafts, contact the IEEE.
[21]ISO publications are available from the International Organization for Standardization (https://www.iso.org/) and the American National Standards Institute (https://www.ansi.org/).

[B32] ISO//IEC/IEEE 16085-2021, ISO/IEC/IEEE International Standard—Systems and software engineering—Life cycle processes—Risk management.

[B33] ISO/IEC/IEEE 24774:2021, Systems and software engineering—Life cycle management—Specification for process description.

[B34] James, W., "What Is an Emotion?" Mind, vol. 9, no. 34, pp. 188–205, 1884, http://dx.doi.org/10.1093/mind/os-IX.34.188.

[B35] Lanzoni, S., "Empathy in Translation: Movement and Image in the Psychological Laboratory" Science in Context, vol. 25, no. 3, pp. 301–327, September 2012, https://doi.org/10.1017/S0269889712000154.

[B36] Matsumoto, D., "Ethnic differences affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample," Motivation and Emotion, vol. 17, no. 2, pp. 107–123, 1993, http://dx.doi.org/10.1007/BF00995188.

[B37] McStay, A., Emotional AI: The Rise of Empathic Media. Sage Publications, 2018, http://dx.doi.org/10.4135/9781526451293.

[B38] Mehrabian, A. and J. A. Russell, An Approach to Environmental Psychology. The MIT Press, 1980.

[B39] Nielsen, M. D., "Haun, J. Kärtner, and C.H. Legare, "The persistent sampling bias in developmental psychology: A call to action," Journal of Experimental Child Psychology, vol. 162, pp. 31–38, October 2017, http://dx.doi.org/10.1016/j.jecp.2017.04.017.

[B40] Ong, D. C., "An Ethical Framework for Guiding the Development of Affectively-Aware Artificial Intelligence," 2019 9th International Conference on Affective Computing and Intelligent Interaction.

[B41] Proposal for a Regulation of the European Parliament and The Council of Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, Document 52021PC0206.[22]

[B42] Russell, J. A., "Core Affect and the Psychological Construction of Emotion," Psychological Review, vol. 110, no. 1, pp. 145–172, 2003, http://dx.doi.org/10.1037/0033-295X.110.1.145.

[B43] Schachter, S. and J. E. Singer, "Cognitive, social, and physiological determinants of emotional state," Psychological Review, vol. 69, no. 5, pp. 379–399, September 1962, http://dx.doi.org/10.1037/h0046234.

[B44] Shazly, H. A., A. Ferraro, and K. Bennet, "Ethical Concerns: An overview of Artificial Intelligence system Development and Life Cycle," 2020 IEEE International Symposium on Technology and Society (ISTAS), pp. 33–42, http://dx.doi.org/10.1109/ISTAS50296.2020.9462201.

[B45] Shimo, S., "Risks of Bias in AI-Based Emotional Analysis Technology from Diversity Perspectives," 2020 IEEE International Symposium on Technology and Society (ISTAS), pp.66–68, http://dx.doi.org/10.1109/ISTAS50296.2020.9462168.

[B46] United Nations, Universal Declaration of Human Rights.

[B47] Wispé, L., "History of the Concept of Empathy" in Empathy and Its Development, Eisenberg, N. and J. Strayer, eds. Cambridge: Cambridge University Press, 1987.

[22]Availabe at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206

IEEE SA
STANDARDS ASSOCIATION

RAISING THE WORLD'S STANDARDS

Connect with us on:

Facebook: facebook.com/ieeesa
LinkedIn: linkedin.com/groups/1791118
Beyond Standards blog: beyondstandards.ieee.org
YouTube: youtube.com/ieeesa

standards.ieee.org
Phone: +1 732 981 0060

IEEE