BS 8611:2016

BSI Standards Publication

Robots and robotic devices

Guide to the ethical design and application of robots and robotic systems

bsi.

...making excellence a habit.™

Publishing and copyright information

Published by BSI Standards Limited 2016

ISBN 978 0 580 89530 2

ICS 25.040.30

The following BSI references relate to the work on this document:
Committee reference AMT/-/2
Draft for comment 15/30320088 DC

Publication history

First published April 2016

Amendments issued since publication

Date	Text affected

Contents

Summary of pages

This document comprises a front cover, an inside front cover, pages i to ii, pages 1 to 22, an inside back cover and a back cover.

Foreword

Publishing information

This British Standard is published by BSI Standards Limited, under licence from The British Standards Institution, and came into effect on 30 April 2016. This British Standard has been prepared by Technical Subcommittee AMT/-/2, *Robots and robotic devices*. A list of organizations represented on this committee can be obtained on request to its secretary.

Information about this document

This is the first edition of a new standard. It addresses issues in a field of technology that is rapidly changing and influencing society. Although efforts have been made to incorporate the anticipated effects of future developments in the standard, it is not possible for all future circumstances to have been considered or covered, for example non-embodied autonomous systems.

Use of this document

As a guide, this British Standard takes the form of guidance and recommendations. It should not be quoted as if it were a specification or a code of practice and claims of compliance cannot be made to it.

Presentational conventions

The guidance in this standard is presented in roman (i.e. upright) type. Any recommendations are expressed in sentences in which the principal auxiliary verb is "should".

Commentary, explanation and general informative material is presented in smaller italic type, and does not constitute a normative element.

Contractual and legal considerations

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Compliance with a British Standard cannot confer immunity from legal obligations.

1 Scope

This British Standard gives guidance on the identification of potential ethical harm and provides guidelines on safe design, protective measures and information for the design and application of robots. It builds on existing safety requirements for different types of robots; industrial, personal care and medical.

This British Standard describes ethical hazards associated with the use of robots and provides guidance to eliminate or reduce the risks associated with these ethical hazards. Significant ethical hazards are presented and guidance given on how they are to be dealt with for various robot applications.

Ethical hazards are broader than physical hazards. Most physical hazards have associated psychological hazards due to fear and stress. Thus, physical hazards imply ethical hazards and safety design features are part of ethical design. Safety elements are covered by safety standards; this British Standard is concerned with ethical elements.

This British Standard is intended for use by designers and managers, amongst others.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

BS EN ISO 12100:2010, *Safety of machinery – General principles for design – Risk assessment and risk reduction (ISO 12100:2010)*

BS ISO 8373, *Robots and robotic devices -– Vocabulary*

BS ISO 31000, *Risk management – Principles and guidelines*

3 Terms and definitions

For the purposes of this British Standard, the terms and definitions given in BS ISO 8373 and the following apply.

3.1 autonomous system

system which has the ability to perform intended tasks based on current state, knowledge and sensing, without human intervention

3.2 ethical harm

anything likely to compromise psychological and/or societal and environmental well-being

NOTE Examples of ethical harm include stress, embarrassment, anxiety, addiction, discomfort, deception, humiliation, being disregarded. This might be experienced in relation to a person's gender, race, religion, age, disability, poverty or many other factors.

3.3 ethical hazard

potential source of ethical harm

3.4 ethical risk

probability of ethical harm occurring from the frequency and severity of exposure to a hazard

3.5 ethics

common understanding of principles that constrain and guide human behaviour

3.6 harm
injury or damage to health

3.7 responsible research and innovation (RRI)
process that seeks to promote creativity and opportunities for science and innovation that are socially desirable and undertaken in the public interest

[SOURCE: EPSRC framework for responsible innovation [1)]]

3.8 roboticist
person actively engaged in a process leading to the development, production, installation, deployment or maintenance of a robot

3.9 system
set of parts which, when interoperating, has behaviour that is not present in any of the parts themselves

NOTE This definition includes systems of fleeting duration, such as cars passing each other and behaving appropriately during this moment, to long-lived systems such as Voyager II.

3.10 vulnerable people
people having reduced physical, sensory or mental capabilities (e.g. partially disabled, elderly having some reduction in their physical and mental capabilities), or lack of experience and knowledge (e.g. children)

[SOURCE: BS EN 60335-1:2012+A11:2014, **3.Z.5**]

4 Ethical risk assessment

4.1 General

Ethical issues can be divided into societal, application, commercial/financial and environmental, as described in Clause **5** to Clause **8**. Table 1 describes examples of the ethical hazards and risks associated with these issues. See also BS ISO 31000 for further guidance.

1) See <www.epsrc.ac.uk/research/framework> [last viewed 4 April 2016].

Table 1 **Ethical issues, hazards and risks**

Ethical issue	Ethical hazard	Ethical risk	Mitigation	Comment	Verification/ Validation
Societal	Loss of trust (human robot)	Robot no longer used or is misused, abused	Design to ensure reliability in behaviour	If unexpected behaviour occurs, ensure traceability to help explain what happened	User validation
	Deception (intentional or unintentional)	Confusion, unintended (perhaps delayed) consequences, eventual loss of trust	Avoid deception due to the behaviour and/or appearance of the robot and ensure transparency of robotic nature	–	Software verification; user validation; expert guidance
	Anthropo-morphization	Misinterpretation	Avoid unnecessary anthropomorphization Clarification of intent to simulate human or not, or intended or expected behaviour	See deception (above) Use anthropomorphization only for well-defined, limited and socially-accepted purposes	User validation; expert guidance
	Privacy and confidentiality	Unauthorized access, collection and/or distribution of data, e.g. coming into the public domain or to unauthorized, unwarranted entities	Clarity of function Control of data, justification of data collection and distribution Ensure user awareness of data management and obtain informed consent in appropriate contexts	Privacy by design Data encryption, storage location, adherence to legislation	Software verification
	Lack of respect for cultural diversity and pluralism	Loss of trust in the device, embarrassment, shame, offence	Awareness of cultural norms incorporated into programming	Organizational, professional, regional	Software verification; user validation
	Robot addiction	Loss of human capability, dependency, reduction in willingness to engage with others, isolation	Raise awareness of dependency	A difficult area, particularly in relation to vulnerable people Careful evaluation of potential applications is needed	User validation; expert guidance

Table 1 **Ethical issues, hazards and risks**

Ethical issue	Ethical hazard	Ethical risk	Mitigation	Comment	Verification/ Validation
Societal	Employment	Social dislocation, job replacement, unemployment, loss of skills/need to re-skill	Targeted deployment of robots to unsocial, hazardous activities, organizational; design	–	Economic and social assessment
Application	Misuse	All in this column	Minimize potential for unethical use via ethical risk assessment Adherence to standards and regulations	Innovate responsibly	Software verification
	Unsuitable divergent use	Unsuitable or inappropriate outcomes	Consider balance between useful/positive divergent use and inappropriate usage, inherent design	–	Software verification
	Dehumanization of humans in the relationship with robots	Inappropriate control exercised by the robot Loss of respect for human dignity, dehumanization and human rights	Inbuilt limits, design and implementation principles and practice	–	User validation; expert guidance
	Inappropriate "trust" of a human by robot	Malign or inadequate human control	Model of appropriate human control	Design robot to reach safe state with respect to any other tasks the robot is executing	Software verification; user validation
	Self-learning system exceeding its remit	Inappropriate use of resources	Limits on extent of self-learning without human monitoring	Keeping the human in the loop is usually advisable	Software verification
Commercial/ Financial	Approbation of legal responsibility and authority	Failure to meet fair contract conditions leading to illegality, inappropriate actions, avoidance of responsibilities	Attention to contracts, recognizing the characteristics of the robot, its intended domain of application and its operators	An issue involving corporate social responsibility, entraining legal and social considerations	Software verification; expert guidance

Table 1 **Ethical issues, hazards and risks**

Ethical issue	Ethical hazard	Ethical risk	Mitigation	Comment	Verification/ Validation
Commercial/ Financial	Employment issues	Job replacement, job change, unemployment, loss of tax revenue	Appropriate support networks, appropriate taxation, retraining opportunities	Note literature on robot economics	Economic and social assessment
	Equality of access	Propagation of the "digital divide", isolation of minorities, non-compliance with human rights legislation	Inclusive design of robot behaviour to conform with Corporate Social Responsibility, and recognition of characteristics of intended application domain Support networks to minimize risks	–	Legal assessment; software verification
	Learning by robots that have some degree of behavioural autonomy	Robot might develop new or amended action plans, or omit steps in processes, that could have unforeseen consequences for safety and/or quality of outcomes	Designers can enable robots to inform their operators when a new form of behaviour has been developed	Important consideration for human-robot trust	Software verification; user validation
	Informed consent	Unaware operators causing accidents, unwanted consequences, unfair and inequitable responsibilities placed upon consenter, inability to respond to situations	Design contracts to make explicit, in plain language, what the consenter is agreeing to, as far as the robot is concerned, including risks Also, contracts to make plain what personal or private information could be passed on to the network(s) to which the robot is connected	In principle, the actions of sets of robots, especially when networked, is nearly unlimited There are privacy and security issues that might arise as robots working in close co-ordination with humans become more informed about human behaviour	User validation

Table 1 **Ethical issues, hazards and risks**

Ethical issue	Ethical hazard	Ethical risk	Mitigation	Comment	Verification/ Validation
Commercial/ Financial	Informed command	User is unaware of extent or legality or social acceptance of the tasks given to the robot, consequences of the tasks might ramify in unexpected ways and extents	As a principle, robots ought to only act on the basis of a properly-constructed command. If the robot cannot construe such a command from the human's communication, it ought to seek further guidance	A generally accepted principle is that "the last person to issue a command is responsible for its outcome". For equity, safety and security, it is necessary to ensure that operators are as informed as possible about the capabilities and limitations of the robots that they utilize [A)]	Software verification
Environmental	Environmental awareness (robot and appliances)	Cause concerns about wastage and destruction of the environment Failure to conform to regulations and/or codes of practice designed to protect the environment resulting in harm to the society	Adopt good/best management practice in design (see BS EN ISO 9000 series) and support for recycling and other circular economy activities	Likely to be of steadily-increasing importance	Lifecycle compliance tests
	Environmental awareness (operations and applications)	Execution of non-sustainable actions, harm to local situation, reputational harm	Good/best management practice (see BS EN ISO 14000 series) built into contract terms, design of robot sensing to ensure recognition and labelling of physical environment, and, particularly in relation to pick-and-place actions, sources and sinks are known	–	Compliance tests

[A)] For examples of "properly-constructed commands", designers and operators can consult NATO's C-BML (Coalition Battle Management Language). See **5.1.12**.

4.2 Ethical hazard identification

The concept for a new robot or application of an existing robot should be reviewed to identify which groups of humans or animals are likely to be affected by it. There might be ethical harm to more groups than the immediate users. The range of ethical hazards might vary depending on the intended use of the robot and the perception and background of the subjects. New applications for existing robots can result in new and different ethical hazards.

The principles and techniques for responsible research and innovation (RRI) as defined in BS ISO 31000 should be adapted and followed.

BS EN ISO 12100:2010 and ISO/TR 14121-2 provide requirements and guidance on performing risk assessments, including risk analysis which focuses on hazard identification. The guidance contained in BS EN ISO 12100:2010, Clause **5** has been derived from the iterative process of applying safeguarding measures (in accordance with the general principles for designing robots given in that standard) to hazards identified as ergonomic hazards. When considering the overlap with ethical hazards, the same procedure should be applied. The source of an ergonomic hazard could be, for example, the design or location of indicators and visual display units with the consequence of discomfort due to misunderstanding of the robot.

Ethical hazards could be considered as part of the ergonomic hazards that can be present in mobile robots (including self-driven vehicles and personal care robots). A risk analysis should be carried out to identify any further hazards that might be present.

BS EN ISO 14971 provides requirements and guidance in performing risk assessment including risk analysis for medical devices through hazard identification and management of risks. The benefit to the user and expected experience of intended design should be balanced against the physical and emotional hazards. The probability of occurrence against benefit in a quantifiable manner guides users' education and training, as well as providing guidelines for inspection, instructions for use, warning, tractability, design modifications for safety and ergonomics principle of good design. Hazards are also identified in other medical and non-medical standards and overlap with ethical hazards, where the same procedure should be applied. The designer's ethical education should be complementary to user-centred design and validation of users' experiences.

NOTE As an example, the benefit to users such as elderly people might be more important than the potential risk of deception, whereas deformation of building character in education and experience of children might not be justifiable.

4.3 Ethical risk assessment

The data collected as part of ethical hazard identification should be analysed qualitatively and/or quantitatively to determine the ethical risk associated with the identified hazards.

A risk assessment should be carried out on those hazards identified in **4.2** with careful attention paid to various human–robot contact situations such as normalcy. This risk assessment should take into account:

a) unauthorized use/reasonably foreseeable misuse such as providing access to personal data;

b) uncertainty of situations to be dealt with, where robots and individuals might cohabit, to minimize stress and fear;

c) the psychological effect of failure in the control system;

d) possible reconfiguration of programs/data by users and associated changes in responsibilities; and

e) where necessary, the ethical hazards associated with the specific robot application.

In performing a risk evaluation, the decision on whether an ethical risk is acceptable or not depends on the application.

As a general principal, the ethical risk of a robot should not be higher than the risk of a human operator performing the same action.

4.4 Learning robots

Learning can be categorized into three classes, depending on the degree of autonomy granted to the robot:

a) *Environmental*: the robot is able to learn about its operational environment and classify the objects within it, whether stationary or transient, fixed or amorphous, so that it can act safely and efficiently;

b) *Performance enhancement*: the robot is able to develop incremental improvements to the processes involved in task execution for efficiency or effectiveness; and

c) *Strategy:* the robot can decide whether or not to commence execution, change the task or process, or to refuse a command. It is important that operators and other humans who could be affected by such changes to expected behaviour are informed of them in some way, and that they can be countermanded if they affect safety, privacy, security, etc.

If robots are able to learn, they might differentiate themselves from other identical robots in their class. This raises further risks for safety, and designers and operators should have the means to ensure that such changes in robots are easily identifiable individually.

5 Ethical guidelines and measures

5.1 Societal ethical guidelines and measures

5.1.1 General societal ethical guidelines

What society considers to be ethical issues should be identified and defined by engaging with end users, specific stakeholders and the public. The following principles should be taken into account:

a) robots should not be designed solely or primarily to kill or harm humans;

b) humans, not robots, are the responsible agents;

c) it should be possible to find out who is responsible for any robot and its behaviour;

d) robots as products should be designed to be safe, secure and fit for purpose, as other products;

e) robots as manufactured artefacts should not be designed to be deceptive and likely to cause ethical harm;

f) the Precautionary Principle (see Note 2);

g) privacy by design;

NOTE 1 For guidance on privacy by design, see Guidelines on regulating robotics *[1].*

h) robots able to learn can distance themselves from the intentions of their designers and operators;

i) potential users should not be discriminated against or forced to acquire and use a robot.

NOTE 2 a) to e) are the EPSRC Principles of robotics *[2] and are in accordance with responsible research and innovation (see* ***3.7****). f) and g) come from the European Commission (for information on f), see* COM (2000) 1, Communication from the commission on the precautionary principle *[3]).*

Roboticists, particularly those who conduct research, should work responsibly and encourage others to do likewise; this includes (but is not limited to):

1) engagement with the public and taking responsibility for the public image of robots;

2) addressing public/consumer concerns to help all make progress to advance robotics in a holistically acceptable manner;

3) demonstrating commitment to the best possible codes of practice;

4) working with experts from other disciplines including social sciences, law, philosophy and the arts;

5) taking the time to contact or correct reporting journalists in the case of noting erroneous accounts in the press; and

6) providing clear instructions to users.

NOTE 3 For further information, see the Rome Declaration [4]. More detailed information can be found in Responsible research and innovation – Europe's ability to respond to societal challenges *[5].*

5.1.2 Industry, research and public engagement

There is a need for all sectors of society to be included in research and public engagement so that ethically accepted robots are developed and deployed. The following organizations are examples of organizations that can help with potential engagement:

- AAAI (Association for the Advancement of Artificial Intelligence);
- BARA (British Automation and Robot Association);
- BMA (British Medical Association);
- Chartered Institute of Ergonomics and Human Factors (CIEHF);
- CLAWAR Association Ltd;
- Consumer and Public Interest Network;
- Consumer Association;
- IET (Institution of Engineering and Technology);
- IMechE (Institution of Mechanical Engineers);
- ILO (International Labour Organization);
- IPEM (Institute of Physics and Engineering in Medicine);
- Royal Academy of Engineering;
- Science parliamentary committee;
- universities;
- WMA (World Medical Association); and
- WHO (World Health Organization).

5.1.3 Privacy and confidentiality

Robots are expected to take and store personal information. This can be beneficial for connectivity and linkages, e.g. the NHS, social care, family members and robot manufacturers. However, there are also issues around the collection and use of information concerning robot users, control of data storage and use of information by third parties.

The information storage issues to be resolved include the type of information the robot is allowed to record, who should have access to the information, who is intended to be using the information, how long to keep the data stored and whether it is necessary to obtain informed consent from the user.

Robots should follow the principle of "privacy by design".

NOTE Attention is drawn to existing privacy laws and other regulations on data collection and protection, e.g. those applicable to NHS, CCTV, IT. Further information on "privacy by design" can be found at the UK Information Commissioner's Office. [2)]

5.1.4 Respect for human dignity and human rights

Robots should be designed, built and operated in such a way that they do not violate human dignity and human rights (for example, there could be a facility for the user to turn off the robot temporarily so that it is not "witness" to very private activities – where the need for privacy is determined by the user). Robots should promote human dignity (for example, through self-sufficiency).

NOTE Attention is drawn to the UN Convention on the rights of persons with disabilities [6], the UN Convention on the rights of the child [7], the Human Rights Act 1998 [8] and other applicable international humanitarian law.

5.1.5 Respect for cultural diversity and pluralism

Robot applications should take account of different cultural norms, including respect for language, religion, age and gender by formal interaction with representatives of these groups.

5.1.6 Dehumanization of humans in the relationship with robots

Robots and robotic systems should be designed to avoid inappropriate control over human choice, for example forcing the speed of repetitive tasks on an assembly line. The ultimate authority should stay with the human.

5.1.7 Legal issues

The roles, responsibilities and legal liabilities should be clearly identified for all stages of the robot's life cycle. It should always be possible to easily discover the person(s) legally responsible for the robot and its behaviour during all stages of the life cycle.

5.1.8 Benefit and risk balance

Robots should be designed and deployed in light of proper judgement of acceptable residual (economic, social, legal, psychological, physical, etc.) ethical risks.

NOTE Norms for acceptable residual ethical risks are as yet undefined.

---

2) See <https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-by-design> [last viewed 4 April 2016].

5.1.9 Individual and organizational responsibility

There should be defined levels of responsibilities agreed with the individual user, the organization deploying the robot and the organization manufacturing the robot.

5.1.10 Social responsibility

Appropriate deployment of the robot should be taken into account. Robots have the potential to perform many different tasks that are presently carried out by people. However, the fact that a task can be performed by a robot does not mean that it ought to be. Other considerations should be taken into account, including but not limited to: employment and environmental considerations, special expertise and capacities (e.g. first aid training), and the resilience provided by general human intelligence and experience.

5.1.11 Informed consent

For a robot with autonomy used in, for example, medical and domestic applications, the user or else the guardian/appropriate legal entity should be informed of risks, benefits and constraints/limitations of use, and informed consent should be obtained. This is particularly important in the case of children and other vulnerable people.

NOTE Attention is drawn to the UN Convention of the rights of the child [7] and the Family Reform Act 1969 [9].

5.1.12 Informed command

On receipt of a command, the robot should be able to construe it into a properly-constructed command, and check it for coherence and compatibility with its internal constraints (including any ethical constraints). If the command does not appear to be properly constructed, the robot should be programmed to pause and query the command.

NOTE For examples of commands that are properly constructed, see NATO's C-BML. For further information, see Applying a formal language of command and control for interoperability between systems *[10].*

5.1.13 Robot addiction

The human potential to be behaviourally conditioned and to become addicted to using the robot should not be negatively exploited.

5.1.14 Dependence on robots

Circumstances where the human might become unnecessarily reliant on the robot should be taken into account.

This can be an individual or global issue.

At an individual level it is necessary to balance the benefits of using robotics with the risks of dependency. For example, dependence on a robotic wheelchair might be positive for a person with a permanent disability but damaging to someone who has the potential (if exercised) to recover from a disabling condition.

The risk of general dependence on robots can result in the loss of important human capability that is learned, maintained and developed through practice, e.g. robots used in surgery.

5.1.15 Anthropomorphization of the robots

The degree of anthropomorphization and humanization, particularly with children and other vulnerable people, should be taken into account.

Where there is a degree of anthropomorphization/humanization, a written document explaining how to introduce a robot to children and other vulnerable individuals should be provided with robot products.

Such a document should provide:

a) a description of the robot;

b) references to the robot's level of autonomy; and

c) advice on how to operate the robot, including how privacy concerns should be addressed.

NOTE For some applications, anthropomorphization and humanization might be necessary as part of the functionality of the robot. However, there is a known issue of the "uncanny valley" (see Mori, 1970 [11]) where the closer a robot resembles a human being but lacks human emotions and social behaviours, the more alien and untrustworthy it can appear to be. There is also a risk that children and other vulnerable people might respond by becoming emotionally attached to the robot.

5.1.16 Robots and employment

The economic, psychological and social consequences of the introduction of robots on employment should be assessed and concerns addressed (see **5.1.10**).

5.2 Application of ethical recommendations and measures

5.2.1 General

Where ethical frameworks exist in application domains (e.g. medical, societal, financial, legal) these should be incorporated as far as possible in requirements as design constraints.

NOTE Design principles are largely qualitative and hence not easy to quantify.

5.2.2 Rehabilitation

Where robots are used for rehabilitation purposes, the potential ethical issues should be taken into account. For example, physiotherapy augmentation of a human, in which case control by a human should be a consideration.

5.2.3 Medical applications

Use of a robot can, in many circumstances, lead to improved or enhanced outcomes, sometimes better than those that can be achieved by humans. For example, certain medical treatments or diagnoses might be carried out better with a robot.

However, the consequences of the wide-spread application of robots should be carefully evaluated as this could lead to a loss of expertise amongst humans that in the long term might be detrimental.

5.2.4 Military use

The use of robots in military applications should not remove responsibility and accountability from a human.

The deployment of robots should be in accordance with international humanitarian law and laws governing armed conflict.

5.3 Commercial/financial ethical guidelines and measures

There is a need to ensure that only ethically acceptable products are manufactured.

5.4 Environmental ethical guidelines, measures and sustainable development

Studies have considered sustainable development in technology design, but quantification has been a problem. The following should be taken into account:

a) use of scarce materials;

b) minimal waste (e.g. graceful and harmless degradation, recoverable materials, zero-waste); and

c) repairability.

NOTE Attention is drawn to 2009/125/EC Ecodesign requirements for energy-related products *[12] and various environment and waste management legislation.*

6 Ethics-related system design recommendations

6.1 General principles for ethical design

6.1.1 General

A methodology for the safety-directed design of machines is described in BS EN ISO 12100:2010, Clause **5**. A machine type robot should be designed in accordance with the principles of BS EN ISO 12100:2010 for all hazards, including the ethical hazards identified in this British Standard.

A robot user should be protected from any identified ethical hazards to ensure that the continuous use and operation of the robot can be carried out safely.

The robot should be designed so as to ensure that the ethical risks of identified hazards are as low as reasonably practicable.

In addition, measures should be taken into account for ethical hazards that might affect the robot user's psychological well-being.

6.1.2 Inherently ethical design

NOTE Inherently ethical design refers to the integrity of the robot, not to artifacts that might be attached to it.

Inherently ethical design measures are the first and most important step in reducing ethical risk because protective measures inherent to the characteristics of the machine are likely to remain effective.

Inherently ethical design measures are achieved by avoiding ethical hazards or reducing ethical risks by a suitable choice of design features of the machine itself and/or interaction between the exposed person and the robot.

6.1.3 Ethically-related control system performance

Where the implementation of inherently ethical design measures is not practicable, the use of safeguards and protective measures should be taken into account in order to protect robot users from harm caused by significant ethical hazards.

6.2 Protection against the perception of harm

An ethical risk assessment should be made to determine the perception of harm, including:

a) the effects of close encounters/near collisions between a robot (e.g. personal care robot, self-driven vehicle, etc.) and people or animals, causing fear (see Walters et al., 2009 [13] and Koay et al., 2014 [14]);

b) unexpected noise or vibrations; and

c) unanticipated, unfamiliar or erratic motion.

Sudden changes in position and orientation or unintended contacts or noise and vibrations could trigger fear and stress that are ethical hazards. These should be avoided or limited as far as reasonably possible.

7 Verification and validation

*COMMENTARY ON CLAUSE **7***

In general terms, "verification" checks that a system does what its specification requires it to do, whereas "validation" checks that a system does what its users expect. Precise specifications are needed in order to carry out verification, while user engagement is needed in order to carry out validation.

7.1 General

A failure of either verification or validation might result in a lack of trust between the system and the user(s) or, in extreme cases, it could result in a system that is unsafe.

As with any system to be deployed, an appropriate risk assessment should be carried out beforehand. The nature of a robot's actions could mean that a failure in verification or validation might result in more physical risk than for non-physical autonomous systems. Any risks should be mitigated by a mixture of designing to avoid risk, risk analysis, comprehensive testing and software verification.

Ethical risk can, in principle, be treated in the same way, although to perform an ethical risk assessment for an intended use application, a precise ethical specification of the robot is needed. This might be difficult to obtain, especially since some ethical principles are not universal and validation of a robot's ethics could well produce different outcomes with different users. The following issues should be taken into account.

a) Designing to avoid risk means developing the system to reduce the risk of problems occurring. For example, techniques to mitigate physical risk include having soft contact areas and utilizing low energies. Mitigating ethical risk is less obvious, but the issues listed in Clause **4** and Clause **5** should provide a starting point.

b) Risk analysis (or assessment) means reducing the risk of serious incident to an acceptable level (e.g. less than 10^{-9} risk of a catastrophic aircraft failure per hour). Such analysis is expected by insurers (despite the results often lacking public acceptance), although the paucity of data regarding ethical risk makes this type of risk assessment difficult.

c) Testing is traditionally used to increase confidence in built systems. Even if testing only covers a very small proportion of real situations, a case can be made for the sample tested being representative of many of the real situations. As with formal verification, a precise specification of the required ethical behaviour is recommended.

d) Software verification is a relatively new discipline that attempts to verify the software component of a system. One major problem in verifying ethical aspects is that of extracting the ethical decisions, so they can be matched against a specification; the internal complexity of most robot programs makes fully formal software verification impractical.

7.2 Suggested approaches

The techniques for verification and validation developed in other industries (such as the aircraft, automotive and machine tool industries) can also be applied to robots. These techniques include:

a) redundancy -– building up a system with multiple independent control pathways to greatly reduce the risk of simultaneous failure in all of them; and

b) independent safety systems that mitigate the effect of failure.

What makes physical autonomous systems (such as robots) unusual is the mixture of physical interaction with complex behaviour. Listing what behaviour is acceptable is normally impractical; it is better to specify by predicates detailing what is safe and/or desired. Safety requirements should also take into account historical points of failure in mechanical systems that capture “best practice”; for example, BS EN ISO 13482 gives safety requirements for personal care robots and lists an eight-point plan consisting of the steps to follow: inspection, practical tests, measurement, observation during operation, examination of circuit diagrams, examination of software, review of task-based risk assessment, and examination of layout drawings and relevant documents.

Software engineers have developed techniques that attempt to verify automatically whether a system precisely meets its specification:

1) Algorithmic verification techniques intelligently verify each possible computational pathway in a program. They work well for systems with large but finite state spaces, but themselves fail when the state-space is infinite, as often happens with physical interactions and where a position can take any of an infinite number of values. Such formal verification systems are often called model checkers, as they check the health of a model of the system’s behaviours across its range of possible states. MAGIC is an example of such a tool. 3)

2) Stochastic verification techniques use probabilistic measures to check the most likely pathways through a program. If the probabilistic models are chosen wisely the result might be suitable as a risk analysis of failure. PRISM is an example of such a tool. 4)

Both algorithmic and stochastic verification techniques suffer from difficulties when tackling large state spaces, and neither of them tend to capture one essential aspect of many physical systems, in that the borderline between success and failure can be very fine. For example, a difference in position of a few microns could be all that distinguishes a successful motion from a catastrophic collision. An expert analysis of the system’s behaviour might be required to identify the critical issues in the design of a robotic program – that is, some form of abstraction from the infinity of real-world values to a smaller set of concepts. For example, a care robot might determine whether it has (or has not) an accurate idea of whether the patient in its care is physically safe or not; or whether the patient is happy (or not).

3) See <www.cs.cmu.edu/~chaki/magic> for further information [last viewed 4 April 2016].

4) See <www.prismmodelchecker.org> for further information [last viewed 4 April 2016].

7.3 Suggested methods

A number of techniques can be applied in accordance with the guidelines in Clause **4**.

a) User validation – the system is validated using either the end user or a representative sample of the end users.

b) Software verification – formal and semi-formal verification of the software against its specification.

c) Expert guidance – call upon subject matter experts (e.g. psychologists, human factors engineers, social scientists, legal experts, religious leaders, economists).

d) Economic and social assessment – statistical evaluation of social outcomes.

e) Legal assessment – evaluation of legal implications.

f) Compliance tests – as defined by other standards and directives.

These techniques are given in the final column of Table 1.

An added complication for verification and validation of autonomous systems is that they are – almost by definition – prone to do the unexpected. In addition, autonomous systems can now make their own decisions and so often what has to be considered is not just what the system does, but "why" it chose to do one thing rather than another.

One route to formal verification that has relevance to ethical risk is to separate out the software making the high level decisions from the software control interacting with the environment (see Fisher et al., 2013 [15]). In this way, formal verification techniques can be applied to the robot's decision-making software, potentially retaining tractability. The traditional control software is then verified by other means, such as testing or control systems analysis. This allows formal verification to be applied to the high-level, often ethical, decision making, rather than the detailed environmental interaction (see, for example, Dennis et al. [16]).

A promising approach is to use an ethical verification system that operates independently from the main control system and monitors the actions in terms of their ethical outcomes. This monitor exists purely to determine whether a proposed (or actual) action is ethical (for example, the "consequence engine" described in Dennis et al., 2015 [17]). In this way, multiple ethical tests could also be implemented as parallel monitors with, say, one checking for environmental issues, another for privacy, etc. (This has the advantage of separating concerns, but says nothing in itself about how to manage disagreements between monitors.)

8 Information for use

*COMMENTARY ON CLAUSE **8***

This Clause relates only to ethical considerations in terms of information for use.

8.1 General

Information for use consists of information for the ethical use of a robot. It is intended for use by the user and other stakeholders.

Instructions and other text should be written in an official language of the region in which the robot is expected to be sold or used.

Markings, symbols and written warnings should be readily understandable and unambiguous, especially as regards the part of the function(s) of the robot to which they are related. Readily understandable signs (pictograms) should be used in preference to written warnings. However, signs and pictograms should only be used if they are understood in the culture of the region in which the robot is to be sold.

Attention is drawn to the fact that, in a typical environment for robots, not all users are able to read the instruction handbook or to notice and understand acoustic or visual warning signs. This includes, but is not limited to, the following situations and user groups:

a) children, elderly people, mentally impaired people;

b) animals;

c) guests/visitors in private areas; and

d) third parties near the robot in public areas.

Where it is foreseeable that the information for use is not likely to be available for certain groups of people, this should be anticipated and provision made so that it does not lead to additional risks.

The markings should be clearly legible and durable. In considering the durability of the marking, the effect of normal use should be taken into account. For example, marking by means of paint or enamel, other than vitreous enamel, on containers that are likely to be cleaned frequently is not deemed to be durable.

Except for the information described in **8.2**, information for use can be supplied not only with printed material but any electronic media as long as they are easily available in any region where the robot is to be sold.

8.2 Markings or indications

Markings on the robot should be clearly discernible from the outside of the robot or, if necessary, after removal of a cover.

As a minimum, the name, trademark or identification mark of the manufacturer or responsible supplier and the model or type reference should be visible when the robot is in normal use. If a robot is integrated into a building or another framework (e.g. furniture), this applies after the robot has been installed in accordance with the instructions provided with the robot.

Switches and controls should be clearly marked to cause no confusion.

In addition to safety data and any other information that might be required by other standards, the following ethical information should be marked or indicated on the robot:

a) current status of connectivity to the outside world;

b) current status of sensors gathering and recording data (e.g. camera, audio equipment);

c) current status of the mode of operation;

d) maximum speed for mobile robots and maximum speed of moving arms;

e) maximum force; and

f) weight handling capability.

Units of physical quantities and their symbols should be in accordance with the International System of Units (SI).

If figures are used for indicating the different positions, the off position should be indicated by the numeric character "0" and the position for a higher value (e.g. output, input, speed or cooling effect) should be indicated by a higher character number.

The character "0" should not be used for any other indication unless it is positioned and associated with other numbers so that it does not give rise to confusion with the indication of the off position.

For signals and warning devices, visual signals (e.g. flashing lights) and audible signals (e.g. sirens) may be used to warn of an impending hazardous event (e.g. robot start-up or over-speed). Such signals may also be used to warn the operator before the triggering of automatic protective measures.

These signals should:

1) be unambiguous and clearly differentiable from all other signals used; and
2) be clearly recognizable to the operator and other people.

The warning devices should be designed and located such that checking is easy. The information for use should prescribe regular checking of the warning devices as appropriate.

The attention of designers is drawn to the possibility of "user overload", which can result from too many signals, leading to confusion that might defeat the effectiveness of the warning devices. Potential users should be consulted regarding this.

8.3 User manual

In addition to operator's instructions, safety data and other information that might be required by other standards (where applicable), the following ethical information should be provided with the robot:

a) a statement that the robot has been designed to meet the highest ethical standards but that it is the responsibility of the user to use the robot in accordance with accepted ethical standards;
b) a detailed description of information collected and recorded by the robot, including:
    1) how and when the information can be accessed and interpreted (e.g. by the user and/or third parties);
    2) who can access the information;
    3) how access can be controlled; and
    4) the degree to which the information can be managed;
c) a description of the default settings and how these can be changed;
d) a warning of the potential risk of addiction as a result of the functionality of the robot or people's vulnerability;
e) a detailed description of the self-learning capabilities and characteristics (e.g. learning about the environment and/or user);
f) a description of the limits of reprogramming and a warning of potential consequences;
g) a description of the robot's autonomous capabilities;
h) guidance on how to issue proper instructions and the consequences of improperly given instructions.

8.4 Service manual

The service manual should include sufficient information to allow the user to maintain or enhance the "as designed" degree of ethical compliance.

Bibliography

Standards publications

For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

BS EN 60335-1:2012+A1:2014, *Household and similar electrical appliances – Safety – Part 1: General requirements*

BS EN ISO 9000 (series), *Quality management systems*

BS EN ISO 13482, *Robots and robotic devices – Safety requirements for personal care robots*

BS EN ISO 14000 (series), *Environmental management*

BS EN ISO 14971, *Medical devices – Application of risk management to medical devices*

ISO/TR 14121-2, *Safety of machinery – Risk assessment – Part 2: Practical guidance and examples of methods*

Other publications

[1] PALMERINI, E., AZZARRI, F., BATTAGLIA, A., BERTOLINI, A., CARNEVALE, A., CARPANETO, J., CAVALLO, F., DI CARLO, A., CEMPINI, M., CONTROZZI, M., KOOPS, B.J., LUCIVERO, F., MUKERJI, N., NOCCO, L., PIRNI, A., SHAH, H., SALVINI, P., SCHELLEKENS, M. and WARWICK, K. *D6.2 – Guidelines on regulating robotics*. Pisa, Italy: Robolaw project, 2014. 5)

[2] BODEN, M., BRYSON, J., CALDWELL, D., DAUTENHAHN, K., EDWARDS, L., KEMBER, S., NEWMAN, P., PARRY, V., PEGMAN, G., RODDEN, T., SORRELL, T., WALLIS, M., WHITBY, B. and WINFIELD, A. *Principles of robotics: regulating robots in the real world*. EPSRC, May 2011. 6)

[3] COMMISSION OF THE EUROPEAN COMMUNITIES. *COM (2000) 1. Communication from the commission on the precautionary principle.* Brussels, 02.02.2000. 7)

[4] PRESIDENCY OF THE COUNCIL OF THE EUROPEAN UNION. Rome declaration on responsible research and innovation in Europe, 21 November 2014. 8)

[5] COMMISSION OF THE EUROPEAN COMMUNITIES. *Responsible research and innovation – Europe's ability to respond to societal challenges*. 2012. 9)

[6] UNITED NATIONS. Convention on the rights of persons with disabilities.

[7] UNITED NATIONS. Convention on the rights of the child.

[8] GREAT BRITAIN. Human Rights Act 1998. London: The Stationary Office.

[9] GREAT BRITAIN. Family Reform Act 1969. London: HMSO.

5) Available from <http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf> [last viewed 4 April 2016].

6) Available from <https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics> [last viewed 4 April 2016].

7) Available from <http://www.precaution.org/lib/06/eu_communication_on_pp.pdf> [last viewed 4 April 2016].

8) Available from <https://ec.europa.eu/research/swafs/pdf/rome_declaration_RRI_final_21_November.pdf> [last viewed 4 April 2016].

9) Available from <http://ec.europa.eu/research/science-society/document_library/pdf_06/responsible-research-and-innovation-leaflet_en.pdf> [last viewed 4 April 2016].

[10] HIEB, M.R. and SCHADE, U. *Applying a formal language of command and control for interoperability between systems*. Fairfax, VA: Center of Excellence in C4I, George Mason University, 2008. [10]

[11] MORI, M. The Uncanny Valley. *In: Energy*, vol 7, no 4, PP.33-35, 1970. [11]

[12] EUROPEAN COMMUNITIES. 2009/125/EC. Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for the setting of ecodesign requirements for energy-related products. Luxembourg: Office for Official Publications of the European Communities, 2009.

[13] WALTERS, M.L., DAUTENHAHN, K., TE BOEKHORST, R., KOAY, K.L., SYRDAL, D.S. and NEHANIV, C.L. An empirical framework for human-robot proxemics. *In: Procs of New Frontiers in Human-Robot Interaction*, 2009.

[14] KOAY, K.L., SYRDAL, D.S., OSKOEI, M.A., WALTERS, M.L. and DAUTENHAHN, K. Social roles and baseline proxemic preferences for a domestic service robot. *In: I. J. Social Robotics*, 6(4), 2014, PP.469-488.

[15] FISHER, M., DENNIS, L. and WEBSTER, M. Verifying autonomous systems. *In: Communications of the ACM*, September 2013, 56(9), PP.84-93.

[16] DENNIS, L., FISHER, M., LINCOLN, N.K., LISITSA, A. and VERES, S.M. Practical verification of decision making in agent-based autonomous systems. *In: Journal of Automated Software Engineering*.

[17] DENNIS, L.A., FISHER, M. and WINFIELD, A.F.T. Towards verifiably ethical robot behaviour. *In: Workshop on Artificial Intelligence and Ethics*. Association for the Advancement of Artificial Intelligence, 2015. [12]

[10] Available from <http://c4i.gmu.edu/eventsInfo/reviews/2008/papers/27_Hieb.pdf> [last viewed 4 April 2016].

[11] Available from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6213238> [last viewed 4 April 2016].

[12] Available from <http://www.aaai.org/ocs/index.php/WS/AAAIW15/paper/viewFile/10119/10131> [last viewed 4 April 2016].

 This page deliberately left blank

This page deliberately left blank

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards -based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

Useful Contacts:

Customer Services
Tel: +44 845 086 9001
Email (orders): orders@bsigroup.com
Email (enquiries): cservices@bsigroup.com

Subscriptions
Tel: +44 845 086 9001
Email: subscriptions@bsigroup.com

Knowledge Centre
Tel: +44 20 8996 7004
Email: knowledgecentre@bsigroup.com

Copyright & Licensing
Tel: +44 20 8996 7070
Email: copyright@bsigroup.com

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

bsi.