

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.
AVA Documents™
Document Sharing Hub

Publishing and copyright information

The BSI copyright notice displayed in this document indicates when the document was last issued.

© The British Standards Institution 2016

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
------	---------------

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
	Anthropomorphization	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
	Misuse	All in this column	Minimize potential for unethical use via ethical risk assessment Adherence to standards and regulations	Innovate responsibly	Software verification
Application	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
	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
Commercial/ Financial					

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 consenters, inability to respond to situations	Design contracts to make explicit, in plain language, what the consenters are 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