

ISO/IEC TS 8200:2024-04 (E)

Information technology - Artificial intelligence - Controllability of automated artificial intelligence systems

Contents

Page

- Foreword..... v
- Introduction..... vi
- 1 Scope..... 1**
- 2 Normative references..... 1**
- 3 Terms and definitions..... 1**
- 4 Abbreviations..... 5**
- 5 Overview..... 5**
 - 5.1 Concept of controllability of an AI system..... 5
 - 5.2 System state..... 6
 - 5.3 System state transition..... 7
 - 5.3.1 Target of system state transition..... 7
 - 5.3.2 Criteria of system state transition..... 7
 - 5.3.3 Process of system state transition..... 7
 - 5.3.4 Effects..... 8
 - 5.3.5 Side effects..... 8
 - 5.4 Closed-loop and open-loop systems..... 8
- 6 Characteristics of AI system controllability..... 9**
 - 6.1 Control over an AI system..... 9
 - 6.2 Process of control..... 11
 - 6.3 Control points..... 12
 - 6.4 Span of control..... 13
 - 6.5 Transfer of control..... 13
 - 6.6 Engagement of control..... 15
 - 6.7 Disengagement of control..... 16
 - 6.8 Uncertainty during control transfer..... 17
 - 6.9 Cost of control..... 17
 - 6.9.1 Consequences of control..... 17
 - 6.9.2 Cost estimation for a control..... 18
 - 6.10 Cost of control transfer..... 18
 - 6.10.1 Consequences of control transfer..... 18
 - 6.10.2 Cost estimation for a control transfer..... 18
 - 6.11 Collaborative control..... 18
- 7 Controllability of AI system..... 19**
 - 7.1 Considerations..... 19
 - 7.2 Requirements on controllability of AI systems..... 20
 - 7.2.1 General requirements..... 20
 - 7.2.2 Requirements on controllability of continuous learning systems..... 21
 - 7.3 Controllability levels of AI systems..... 21
- 8 Design and implementation of controllability of AI systems..... 22**
 - 8.1 Principles..... 22
 - 8.2 Inception stage..... 23
 - 8.3 Design stage..... 24
 - 8.3.1 General..... 24
 - 8.3.2 Approach aspect..... 24
 - 8.3.3 Architecture aspect..... 25
 - 8.3.4 Training data aspect..... 25

8.3.5	Risk management aspect	25
8.3.6	Safety-critical AI system design considerations.....	25
8.4	Suggestions for the development stage	25
9	Verification and validation of AI system controllability.....	26
9.1	Verification.....	26
9.1.1	Verification process.....	26
9.1.2	Output of verification.....	26
9.1.3	Functional testing for controllability	26
9.1.4	Non-functional testing for controllability	27
9.2	Validation	28
9.2.1	Validation process.....	28
9.2.2	Output of validation.....	28
9.2.3	Retrospective validation.....	28
	Annex A (informative) Example verification output documentation.....	30
	Annex B (informative) Example validation output documentation.....	32
	Bibliography.....	34