	ARTIFICIAL INTELLIGENCE (AI) OVERALL SAFETY MANAGEMENT MANUAL	编号 Code	MUNIK-AP-X-0X
		版本号/修订次 Version/Revision	A/0
		首次发布日期 First Issue Date	2026-09-11
		修订日期 Revision Date	/
	Template	页数 Page	第 1 页 共 249 页
		编写 Author(s)	Gerhard M. Rieger
		批准 Approved by	Jianguo Wang

Table of Contents

1 Purpose 16
2 Scope 19
2.1 Organizational Scope 19
2.2 Technical Scope 20
2.3 Product and Application Scope 21
2.4 AI Lifecycle Scope 21
2.5 Safety and Security Interface Scope 22
2.6 Applicability and Classification 22
2.7 Exclusions 23
2.8 Scope Management 23
3 Applicable Standards and Frameworks 24
3.1 General 24
3.2 AI-Specific Safety Standards 24
3.3 Functional Safety 25
3.4 Safety of the Intended Functionality 25
3.5 Cybersecurity 26
3.6 Software Update Engineering 26
3.7 Certification and Conformity Assessment 27
3.8 Regulatory and Type-Approval Requirements 27
3.9 Classification of External Requirements 28
3.10 Applicable Requirements List 29
3.11 Determination of Applicability 30
3.12 Management of Changes to External Requirements 30
3.13 Conflicts and Overlapping Requirements 31
3.14 Compliance and Objective Evidence 31
3.15 Responsibility 32
3.16 Reference to the Project-Specific Compliance Framework 33
4 AI Safety Policy 33
4.1 Management Commitment 33
4.2 AI Safety Principles 34
4.2.1 Safety by Design 34
4.2.2 Risk-Based Development 35
4.2.3 Lifecycle Responsibility 35
4.2.4 Traceability and Reproducibility 35
4.2.5 Appropriate Independence 35
4.2.6 Evidence-Based Decision Making 35

4.2.7 Controlled Changes 36
4.2.8 Continuous Monitoring and Assurance 36
4.2.9 Defined Accountability and Authority 36
4.2.10 Controlled Deployment and Release 36
4.3 Management Responsibility 37
4.4 AI Safety Objectives 38
4.5 Communication of the AI Safety Policy 38
4.6 Continual Improvement 39
4.7 Policy Statement 39
5 AI Safety Organization 40
5.1 Organizational Structure 40
5.2 AI Safety Management Function 41
5.3 Relevant Functional Areas and Specialist Interfaces 42
5.4 Responsibility and Authority 43
5.5 Combination of Roles and Organizational Size 44
5.6 Escalation and Safety Authority 44
5.7 External Organizations and Suppliers 45
5.8 Organizational Competence and Resources 46
5.9 Organizational Information and Records 46
5.10 Organizational Change 46
5.11 Organizational Review and Continual Improvement 47
6 Roles and Responsibilities 48
6.1 General 48
6.2 AI Safety Manager 48
6.3 ML Development Engineer 50
6.4 Data Management Responsible / Data Engineer 51
6.5 Verification and Validation Engineer 52
6.6 Functional Safety Manager 53
6.7 SOTIF Responsible 54
6.8 Cybersecurity Responsible 55
6.9 Quality Manager 56
6.10 Configuration and Change Manager 56
6.11 System Engineer 56
6.12 Management and Release Responsibility 57
6.13 Responsibility Interfaces 57
7 Competence and Qualification 58

7.1 General 58
7.2 Competence Requirements 58
7.3 Role-Specific Competence 59
7.4 Competence Demonstration 60
7.5 Qualification and Authorization 60
7.6 Initial and Continuing Competence 61
7.7 Training and Development 62
7.8 Competence Records 62
7.9 External Personnel and Consultants 63
7.10 Competence Gaps 63
7.11 Independence and Competence 63
7.12 Management Responsibility 64
7.13 Competence Effectiveness 64
8 Authority and Responsibility Matrix 65
8.1 General 65
8.2 Authority and Responsibility Matrix 65
8.3 Authority Principles 67
8.4 Safety Release Authority 67
8.5 Separation of Conflicting Responsibilities 68
8.6 Delegation of Authority 69
8.7 Project-Specific Responsibility Matrix 69
8.8 Evidence of Authority 69
8.9 Review and Maintenance 70
9 AI Safety Lifecycle 71
9.1 General Lifecycle 71
9.2 Strategy and Governance 72
9.3 AI Use-Case Definition 72
9.4 Safety Classification 73
9.5 AI Safety Concept 74
9.6 Data Definition and Management 74
9.7 AI/ML Development 75
9.8 Training 75
9.9 Verification 76
9.10 Validation 76
9.11 System Integration 77

9.12 Safety Assessment 78
9.13 Release Approval 79
9.14 Deployment 79
9.15 Operation and Monitoring 80
9.16 Change and Update 81
9.17 Re-Verification, Re-Validation and Re-Assessment 81
9.18 Retirement 82
9.19 Lifecycle Feedback and Continual Improvement 82
10 Lifecycle Entry and Exit Criteria 83
10.1 General 83
10.2 Lifecycle Gate Concept 83
10.3 Example Entry and Exit Criteria 84
10.4 Gate Review and Approval 85
10.5 Open Issues and Deviations 86
10.6 Traceability of Lifecycle Gates 87
10.7 Repetition of Lifecycle Phases 87
11 Interface with ISO 26262 Functional Safety 87
11.1 General Interface Principles 88
11.2 Item Definition and AI Component Context 88
11.3 Safety Goals and Safety Requirements 89
11.4 Allocation of Requirements to AI Components 89
11.5 Assumptions of Use 90
11.6 Safety Mechanisms and Interfaces 90
11.7 Verification and Validation Evidence 91
11.8 Integration Evidence 92
11.9 Confirmation Measures and Independence 93
11.10 Safety Case and Traceability 93
11.11 AI Component as SEooC 93
11.12 Changes and Impact Analysis 94
11.13 Responsibility and Interface Management 94
11.14 Objective Evidence and Records 95
11.15 Interface Review 95
12 Interface with ISO/SAE 21434 Cybersecurity 96
12.1 General Interface Principles 96
12.2 Safety-Relevant AI Assets 97

12.3 Cybersecurity Threats Affecting AI Safety 97
12.4 Interface with Cybersecurity Risk Management 98
12.5 Protection of AI Development and Training 99
12.6 Model and Data Integrity 99
12.7 Adversarial Inputs and AI Runtime Security 100
12.8 Secure Deployment and Updates 100
12.9 Access Control and Protection of Safety-Relevant Assets 101
12.10 Cybersecurity and Safety Verification 101
12.11 Cybersecurity Events and Safety Impact 102
12.12 Interface with AI Safety Risk Management 103
12.13 Responsibility and Interface Management 103
12.14 Objective Evidence and Records 104
12.15 Interface Review 104
13 Interface with ISO 21448 SOTIF 105
13.1 General Interface Principles 105
13.2 Intended Functionality 105
13.3 Functional Insufficiencies 106
13.4 AI Perception and Performance Limitations 106
13.5 Foreseeable Misuse 107
13.6 Operational Design Domain 108
13.7 Scenario Identification and Coverage 108
13.8 Unknown Unsafe Scenarios 109
13.9 AI Validation 110
13.10 Simulation, Testing and Scenario-Based Evidence 110
13.11 Relationship to AI Safety Risk Management 111
13.12 Monitoring and Operational Feedback 111
13.13 Changes and Re-Validation 112
13.14 Responsibility and Interface Management 112
13.15 Objective Evidence and Records 113
13.16 Interface Review 114
14 Integrated Safety and Security Interface 114
14.1 Interface Principles 115
14.2 Shared Information and Dependencies 115
14.3 Interaction of Safety and Security Risks 116
14.4 Conflicting Requirements and Trade-Offs 117

14.5 Integrated Evidence 117
14.6 Integrated Reviews 118
14.7 Change Notification and Impact Analysis 118
14.8 Incident and Operational Feedback 119
14.9 Responsibility and Escalation 120
14.10 Integrated Interface Records 120
14.11 Continual Review of the Integrated Interface 121
15 Data and AI Model Management 121
15.1 Data Management 121
15.1.1 Data Identification and Classification 121
15.1.2 Data Acquisition and Provenance 122
15.1.3 Data Requirements 123
15.1.4 Data Quality 123
15.1.5 Data Preparation and Processing 124
15.1.6 Data Labeling 125
15.1.7 Dataset Composition and Coverage 125
15.1.8 Dataset Partitioning 126
15.1.9 Dataset Versioning and Baseline 126
15.1.10 Dataset Verification and Validation 127
15.1.11 Dataset Approval and Release 127
15.1.12 Data Access, Protection and Integrity 128
15.1.13 Data Changes and Impact Analysis 128
15.1.14 Data Monitoring and Operational Feedback 129
15.1.15 Data Retention and Disposal 130
15.1.16 Data Management Records 131
15.1.17 Data Management Review 131
15.2 AI Model Management 132
15.2.1 Model Identification and Configuration 132
15.2.2 Model Baseline 133
15.2.3 Training and Dataset Traceability 133
15.2.4 Model Verification and Validation 134
15.2.5 Model Performance and Safety Limitations 134
15.2.6 Model Approval and Release 135
15.2.7 Model Deployment 135
15.2.8 Model Changes and Impact Analysis 136
15.2.9 Model Versioning and Traceability 137
15.2.10 Model Retirement and Replacement 137
15.2.11 Model Management Records 137
15.2.12 Model Management Review 138
16 Verification and Validation 138

16.1 Verification and Validation Planning 139
16.2 Verification Activities 140
16.3 Validation Activities 140
16.4 AI-Specific Verification and Validation 141
16.5 Scenario-Based and Boundary-Condition Testing 142
16.6 Robustness and Stress Testing 142
16.7 Fault Injection and Adversarial Testing 143
16.8 Test Environment and Configuration Control 144
16.9 Independence 144
16.10 Acceptance Criteria and Results 144
16.11 Anomalies, Deviations and Open Findings 145
16.12 Objective Evidence and Traceability 146
16.13 Re-Verification and Re-Validation 146
16.14 V&V Records 147
16.15 V&V Review and Continual Improvement 147
17 Continuous Assurance During Operation 148
17.1 Operational Monitoring 148
17.2 Performance Monitoring 149
17.3 Safety-Event Monitoring 149
17.4 Data Distribution Monitoring 150
17.5 Model Degradation and Drift 150
17.6 Anomaly Detection 151
17.7 Incident Management 152
17.8 Customer and Field Feedback 152
17.9 Operational Data and Evidence 153
17.10 Continuous Safety Assessment 153
17.11 Trigger Conditions for Re-Assessment 154
17.12 Feedback into the AI Safety Lifecycle 155
17.13 Change and Update Interface 155
17.14 Responsibilities and Escalation 156
17.15 Objective Evidence and Records 156
17.16 Review of Continuous Assurance Effectiveness 157
18 Change Management 158
18.1 General 158
18.2 Types of Changes 158

18.3 Change Request and Initial Screening 159
18.4 Change Impact Analysis 159
18.5 Change Classification 160
18.6 Change Planning and Approval 160
18.7 Implementation and Configuration Control 160
18.8 Verification and Validation of Changes 161
18.9 Regression Testing 161
18.10 Re-Verification, Re-Validation and Re-Assessment 162
18.11 Interface with Functional Safety, SOTIF and Cybersecurity 162
18.12 Emergency Changes 162
18.13 Change Rejection and Rollback 163
18.14 Change Records and Traceability 163
18.15 Change Management Responsibilities 164
18.16 Review and Continual Improvement 164
19 Deviation and Waiver Management 164
19.1 General 164
19.2 Deviation Identification 165
19.3 Deviation Record 165
19.4 Safety Impact and Risk Assessment 166
19.5 Compensating Measures 166
19.6 Approval of Deviations 166
19.7 Waivers 167
19.8 Deviations Affecting Safety Evidence 167
19.9 Deviations After Release 168
19.10 Deviation Closure 168
19.11 Deviation Records and Traceability 168
19.12 Responsibilities and Escalation 169
19.13 Review and Continual Improvement 169
20 Configuration Management 169
20.1 General 169
20.2 Configuration Items 170
20.3 Configuration Identification 170
20.4 Version and Revision Control 171
20.5 Baseline Management 171
20.6 Configuration Status and Change History 171

20.7 Configuration Control and Change Management 172
20.8 Access Control and Protection 172
20.9 Traceability and Reproducibility 173
20.10 Configuration and Safety Evidence 173
20.11 Configuration Audit and Verification 173
20.12 Release Configuration 174
20.13 Configuration Records 174
20.14 Responsibilities 174
20.15 Review and Continual Improvement 175
21 Toolchain Management 175
21.1 General 175
21.2 Tool Identification and Intended Use 176
21.3 Tool Safety Relevance 176
21.4 Tool Evaluation 176
21.5 Tool Version and Configuration 177
21.6 Toolchain Dependencies 177
21.7 Confidence in Tool Outputs 177
21.8 Tool Anomalies and Limitations 178
21.9 Tool Changes 178
21.10 Tool Reproducibility 178
21.11 Tool Output and Safety Evidence 179
21.12 Toolchain Integrity and Access Control 179
21.13 Tool Records 179
21.14 Responsibilities 180
21.15 Review and Continual Improvement 180
22 Traceability 180
22.1 General 180
22.2 Traceability Chain 181
22.3 Bidirectional Traceability 182
22.4 Requirements Traceability 182
22.5 Data and Model Traceability 182
22.6 Verification and Validation Traceability 183
22.7 Safety Evidence Traceability 183
22.8 Traceability and Safety Case 183
22.9 Traceability and Change Impact Analysis 184

22.10 Traceability and Configuration Management 184
22.11 Traceability Gaps 184
22.12 Traceability Records and Tools 185
22.13 Traceability Review 185
22.14 Responsibilities 185
22.15 Continual Improvement 186
23 Documentation and Evidence Management 186
23.1 General 186
23.2 Documentation and Evidence Structure 187
23.3 Evidence Identification 187
23.4 Evidence Review and Approval 187
23.5 Evidence and Configuration 188
23.6 Evidence Traceability 188
23.7 Evidence Integrity and Protection 188
23.8 Evidence Retrieval and Availability 189
23.9 Evidence Completeness 189
23.10 Evidence from AI-Specific Activities 189
23.11 Interface Evidence 190
23.12 Safety Assessment and Release Evidence 190
23.13 Evidence Retention and Archiving 190
23.14 Evidence Management Responsibilities 191
23.15 Review and Continual Improvement 191
24 Internal Audits 192
24.1 General 192
24.2 Audit Planning 192
24.3 Audit Scope and Criteria 192
24.4 Audit Evaluation 193
24.5 Auditor Competence and Independence 193
24.6 Audit Evidence and Records 193
24.7 Audit Findings and Classification 194
24.8 Corrective Actions and Closure 194
24.9 Follow-Up Audits 194
24.10 Relationship to Safety Assessments and External Assessments 195
24.11 Audit Responsibilities 195
24.12 Audit Records and Continual Improvement 195

25 External Assessment and Audit 196
25.1 General 196
25.2 Assessment Scope and Criteria 196
25.3 Assessment Preparation 197
25.4 Communication with the Assessment Body 197
25.5 External Assessment Activities 197
25.6 External Findings and Nonconformities 198
25.7 Corrective Actions 198
25.8 Changes Affecting the Assessment Scope 198
25.9 External Assessment Evidence 199
25.10 Responsibilities 199
25.11 Review and Continual Improvement 199
26 Safety Review and Release 200
26.1 General 200
26.2 Safety Review Inputs 200
26.3 Review of Safety Requirements and Objectives 201
26.4 Review of Risk and Residual Risk 201
26.5 Review of Verification and Validation 201
26.6 Review of Open Issues 202
26.7 Review of Deviations and Waivers 202
26.8 Review of Interfaces 202
26.9 Review of the Evidence Package 203
26.10 Safety Review Independence 203
26.11 Release Decision 203
26.12 Release Configuration and Status 204
26.13 Release Records 204
26.14 Post-Release Feedback 204
27 AI Safety Case 205
27.1 General 205
27.2 Safety Case Structure 205
27.3 Intended Use and Operating Conditions 206
27.4 Hazards and Safety Risks 206
27.5 Safety Requirements and Safety Objectives 206
27.6 Controlled AI Development 206
27.7 Data as a Safety Argument Element 207

27.8 AI Model and Behavior 207
27.9 Functional Safety, SOTIF and Cybersecurity Interfaces 207
27.10 Operational Limitations and Continuous Assurance 208
27.11 Change and Configuration 208
27.12 Residual Risk and Safety Conclusion 209
27.13 Evidence Supporting the Safety Case 209
27.14 Traceability and Consistency 209
27.15 Review and Approval of the Safety Case 209
27.16 Maintenance of the Safety Case 210
27.17 Safety Case Records 210
28 Training and Awareness 211
28.1 General 211
28.2 Training Needs 211
28.3 AI Safety Training Programme 211
28.4 Role-Specific Training 212
28.5 Awareness 212
28.6 Training Methods 213
28.7 Training Effectiveness 213
28.8 Initial and Continuing Training 213
28.9 Training Records 214
28.10 External Personnel and Suppliers 214
28.11 Training and Organizational Change 214
28.12 Training Records and Continual Improvement 214
29 Incident and Problem Management 215
29.1 General 215
29.2 Identification and Reporting 215
29.3 Examples of AI Safety Incidents 216
29.4 Initial Assessment and Classification 216
29.5 Containment 216
29.6 Incident Investigation 217
29.7 Safety Impact Assessment 217
29.8 Root Cause and Problem Management 217
29.9 Corrective and Preventive Action 218
29.10 Re-Verification, Re-Validation and Re-Assessment 218
29.11 Cybersecurity-Related Incidents 219

29.12 Data and Model Integrity Incidents 219
29.13 Communication and Escalation 219
29.14 Closure 220
29.15 Incident Records and Traceability 220
29.16 Interface with Continuous Assurance 220
29.17 Continual Improvement 221
30 Continuous Improvement 221
30.1 General 221
30.2 Inputs to Continuous Improvement 222
30.3 Evaluation of Improvement Needs 222
30.4 Lessons Learned 222
30.5 Improvement of AI Safety Processes 223
30.6 Technology and AI Method Changes 223
30.7 Standards and Regulatory Developments 223
30.8 Effectiveness of Improvement Actions 223
30.9 Management Review and Continuous Improvement 224
30.10 Improvement Records and Traceability 224
30.11 Continual Improvement of the Safety Case 224
30.12 Organizational Learning 225
31 Management Review 225
31.1 General 225
31.2 Management Review Inputs 225
31.3 Safety Performance 226
31.4 Audit and Assessment Results 226
31.5 Incidents, Problems and Operational Experience 226
31.6 Risks, Deviations and Changes 227
31.7 Resources and Competence 227
31.8 Customer, Regulatory and External Requirements 227
31.9 Improvement Actions 228
31.10 Management Review Outputs 228
31.11 Safety Escalation and Management Decisions 228
31.12 Management Review Records 228
31.13 Relationship to Continuous Improvement 229
31.14 Frequency and Extraordinary Review 229
32 Minimum Mandatory AI Safety Work Products 230

32.1 General 230
32.2 Minimum Work Product Set 230
32.3 Management-System and Project Work Products 232
32.4 Applicability and Tailoring 232
32.5 Work Product Development and Approval 233
32.6 Relationship to the AI Safety Lifecycle 233
32.7 Work Product Traceability 233
32.8 Work Product Completeness 234
32.9 Work Product Control and Retention 234
32.10 Relationship to Assessment 234
32.11 Responsibilities 234
32.12 Continual Improvement of Work Products 235
33 AI Safety Gate Model 235
33.1 General 235
33.2 Gate Decision Principles 236
33.3 Gate 0 – AI Use-Case Approval 236
33.4 Gate 1 – Safety Classification 237
33.5 Gate 2 – Safety Concept Approval 237
33.6 Gate 3 – Development Readiness 237
33.7 Gate 4 – Verification Readiness 238
33.8 Gate 5 – Validation Approval 238
33.9 Gate 6 – Safety Release 239
33.10 Gate 7 – Operational Monitoring 239
33.11 Gate Evidence and Records 240
33.12 Gate Responsibilities and Independence 240
33.13 Gate Repetition and Re-Opening 241
33.14 Relationship to the AI Safety Lifecycle 241
33.15 Gate Model and Minimum Work Products 241
33.16 Continual Improvement of the Gate Model 242
34 Records 242
34.1 General 242
34.2 Minimum Records 243
34.3 Record Identification and Traceability 244
34.4 Record Integrity and Protection 244
34.5 Configuration of Records 244

34.6 Record Retention 245
34.7 Record Retrieval and Availability 245
34.8 Record Disposal 246
34.9 Responsibilities 246
34.10 Review and Continual Improvement 246
35. Compliance Statement 247
35.1 General 247
35.2 Compliance with this Manual 247
35.3 Relationship to Applicable Standards 248
35.4 Product Safety and Compliance 248
35.5 Management Commitment to Compliance 248
35.6 Continuing Validity 249
35.7 Compliance Declaration 249

Table 1: Standards and frameworks considered to the AI Safety Management System. 29
Table 2: Authority and Responsibility Matrix 66
Table 3: Example Entry and Exit Criteria 85
Table 4: Minimum Workproduct Set 232
Table 5: Minimum Records 243

1 Purpose

This **AI Overall Safety Management Manual** defines the organizational framework, governance principles, responsibilities, processes and controls established by **[Company Name]** to ensure the systematic and controlled management of safety throughout the complete lifecycle of Artificial Intelligence (AI) systems and AI-enabled products.

The purpose of this manual is to establish a consistent and organization-wide framework for the **planning, development, verification, validation, deployment, operation, monitoring, maintenance, modification and retirement** of AI systems whose behavior may have an impact on safety.

The manual defines the principles and management controls necessary to ensure that AI-related safety activities are:

- systematically planned and integrated into the overall product lifecycle;
- based on identified safety objectives, requirements, risks and assumptions;
- performed by competent and appropriately authorized personnel;
- supported by defined responsibilities and decision-making authorities;
- subject to appropriate verification, validation and review;
- traceable from safety objectives and requirements through implementation to objective evidence;
- controlled with respect to changes, deviations, configurations and releases;
- continuously monitored and reassessed where operational experience or changes may affect safety.

The manual establishes the organizational and process framework for ensuring that **AI safety is addressed as a lifecycle responsibility and not solely as a testing or verification activity**.

It defines how AI safety activities are **planned, performed, monitored, verified, validated, documented, reviewed, assessed and approved**, and how the resulting objective evidence is maintained to support safety-related decisions and, where applicable, safety claims and safety cases.

Interfaces to Other Disciplines

Because AI functionality may contribute to, interact with or affect several safety and security domains, this manual establishes the required interfaces between AI safety and other relevant engineering and management disciplines.

In particular, the manual defines the relationship between AI safety and:

- **Functional Safety according to ISO 26262**, including the allocation and traceability of safety requirements where AI is part of a safety-related item or function;
- **Safety of the Intended Functionality (SOTIF) according to ISO 21448**, including functional insufficiencies, foreseeable misuse, limitations of intended functionality, scenario coverage and unknown unsafe scenarios;
- **Road Vehicle Cybersecurity according to ISO/SAE 21434**, including cybersecurity threats and vulnerabilities that may compromise AI data, models, software, configuration or safety-relevant behavior;
- **Software Update Engineering according to ISO 24089**, where AI models, software, datasets or other safety-relevant elements are updated during the product lifecycle;
- **AI-specific safety standards, technical reports, guidelines and regulatory requirements**, including applicable requirements derived from **ISO/PAS 8800**, **ISO/IEC TR 5469**, and other applicable standards and regulations;
- **system, hardware, software, data and quality management processes**, where these processes provide inputs to or depend on AI safety activities.

The interfaces shall ensure that safety-related assumptions, requirements, risks, dependencies, changes and evidence are appropriately communicated and consistently managed across the respective disciplines.

Safety Assurance and Objective Evidence

The AI Safety Management System shall provide a structured basis for demonstrating that identified AI-related safety risks have been adequately addressed and that applicable safety objectives and requirements have been fulfilled.
For this purpose, **objective evidence** shall be generated and maintained throughout the lifecycle. Such evidence may include, as applicable:

- AI safety plans and policies;
- safety objectives and requirements;
- hazard and risk analyses;
- AI system and model specifications;
- data and dataset documentation;
- training and model-development records;
- verification and validation results;
- robustness and performance evidence;

- system integration evidence;
- operational monitoring results;
- change and configuration records;
- deviation and waiver records;
- review and audit results;
- safety assessments;
- safety arguments or safety cases;
- release and approval records.

The organization shall ensure that safety-relevant evidence is **identifiable, traceable, controlled, reproducible where applicable, and protected against unauthorized modification**.

Management and Product Responsibility

This manual defines the **management-system requirements and organizational controls** for AI safety. Compliance with this manual alone does not constitute proof that a specific AI system or product is safe.

The safety of an individual AI system or product shall be demonstrated through the applicable **product-specific engineering, verification, validation, assessment and approval activities** and through objective evidence appropriate to its intended use, safety relevance, operating conditions and applicable regulatory or standardization framework.

The overall objective of this manual is therefore to establish a controlled and auditable framework through which **AI-related safety risks are systematically identified, analyzed, evaluated, controlled, verified, validated, monitored and continuously reassessed throughout the complete AI lifecycle**.

2 Scope

2.1 Organizational Scope

This AI Overall Safety Management Manual applies to all organizational units, functions, locations and personnel of **[Company Name]** that are responsible for, contribute to, support, or exercise authority over the development, integration, deployment, operation, maintenance, modification or retirement of Artificial Intelligence (AI) systems and AI-enabled products.

The scope therefore includes, as applicable, organizational activities related to **AI strategy and governance, AI safety management, research and development, system and software engineering, machine-learning development, data acquisition and data engineering, model training, verification and validation, system integration, deployment, operation, monitoring, maintenance, updating, change management, configuration management, incident management, safety assessment and release approval**.

The manual also applies to the management functions responsible for establishing policies, allocating resources, defining responsibilities and authorities, ensuring competence, monitoring process performance and conducting management reviews. AI safety is considered a cross-functional responsibility; consequently, the implementation of this manual shall not be limited to an organizational unit formally designated as the AI or AI Safety function.

Where AI development or safety-relevant activities are performed by external organizations, suppliers, contractors, development partners or other third parties, **[Company Name]** shall determine the extent to which the requirements of this manual are applicable to those activities. The organization shall ensure that externally performed activities that may affect the safety of an AI system are appropriately specified, controlled, monitored and supported by objective evidence.

The organizational scope shall also cover interfaces with other relevant management and engineering disciplines, including **Functional Safety, SOTIF, Cybersecurity, Software Update Management, Quality Management, Configuration Management, System Engineering and Product Compliance**, where these disciplines interact with AI safety.

The responsibility for determining and maintaining the organizational applicability of this manual shall reside with the **AI Safety Manager**, or another formally designated responsible function, subject to approval by the appropriate management authority.

2.2 Technical Scope

This manual applies to Artificial Intelligence and Machine Learning technologies that are developed, integrated or used by **[Company Name]** where their behavior, limitations, failure modes or interactions may have an impact on the safety of a system, product, user, environment or other affected entity.

The technical scope includes, but is not limited to, **machine-learning models, neural networks, deep-learning systems, computer-vision and perception systems, classification and prediction algorithms, decision-making systems, AI-based control functions, AI-supported software and AI-enabled hardware/software systems**.

The scope also includes the supporting technical elements that may influence the safety behavior of an AI system. These elements may comprise **training, validation and test datasets, data-processing and labeling pipelines, model parameters, model architectures, inference software, AI-related algorithms, configuration data, model deployment mechanisms, simulation environments, development and verification tools, and other elements required to develop, verify, validate or operate the AI system**.

Generative AI and other emerging AI technologies are included where they are used within a product or process and may have a relevant impact on safety. The applicability of the requirements shall be determined according to the intended use, system architecture, safety relevance and potential consequences of incorrect, unpredictable or insufficiently controlled AI behavior.

The scope is technology-neutral with respect to the specific AI implementation. It applies whether the AI functionality is implemented as embedded software, hardware-accelerated AI, an AI accelerator, an SoC-based solution, cloud-supported functionality, an edge-AI implementation or a combination of these technologies.

The use of AI technology alone does not automatically establish that the technology is safety-relevant. The safety relevance shall be determined based on the intended function, operating context, system architecture, reasonably foreseeable use and misuse, potential hazards, dependencies on other functions and the consequences of erroneous or insufficient AI behavior.

2.3 Product and Application Scope

This manual applies to AI systems and AI-enabled products developed, supplied, integrated or operated by **[Company Name]** in safety-relevant or potentially safety-relevant applications.

The product and application scope includes, where applicable, **road vehicles, vehicle components, Advanced Driver Assistance Systems (ADAS), automated and highly automated driving systems, automotive semiconductor and System-on-Chip (SoC) products, embedded systems, industrial automation systems, robotics and other systems in which AI functionality may influence safety-related behavior**.

The scope is not restricted to complete AI systems. It also includes AI components, AI-enabled subsystems and AI-related elements that are incorporated into a larger system or supplied to another organization for integration into a safety-relevant product.

For component and semiconductor suppliers, the scope may therefore include AI-enabled **MCUs, MPUs, SoCs, AI accelerators, IP components, perception modules, software libraries, embedded machine-learning functions and other AI-enabled hardware or software elements** where their intended use may contribute to the safety of the final system.

Where an AI component is developed as a **Safety Element out of Context (SEooC)** or for another predefined integration context, the applicable assumptions of use, safety requirements, interfaces, limitations and evidence shall be defined and maintained as part of the applicable safety lifecycle.

For products intended for integration into vehicles or other regulated systems, the scope shall additionally consider applicable market-specific regulatory requirements and type-approval or product-compliance interfaces where these may affect the acceptance or deployment of the AI-enabled product.

2.4 AI Lifecycle Scope

The scope of this manual extends across the complete lifecycle of an AI system, beginning with the identification and classification of the intended AI use case and continuing through concept development, requirements definition, data management, AI/ML development, training, verification, validation, system integration, release, deployment, operation, monitoring, maintenance, updating and eventual retirement.

AI safety activities shall therefore not cease when an AI model has successfully completed development or initial validation. Safety-relevant assumptions, limitations, performance

characteristics and operational behavior shall remain subject to appropriate monitoring throughout the operational lifecycle.

Where an AI model, dataset, algorithm, software component, configuration, operating environment or intended use is changed, the organization shall determine whether the change may affect the existing safety argument or safety evidence. Where necessary, the affected activities shall be repeated or extended, including verification, validation, risk analysis and safety assessment.

The lifecycle scope also includes the management of safety-relevant information and evidence necessary to demonstrate continued conformity with applicable requirements.

2.5 Safety and Security Interface Scope

AI safety shall be managed in conjunction with other applicable safety and security disciplines where the AI functionality interacts with or contributes to those disciplines.

For road-vehicle applications, particular attention shall be given to the interfaces between AI safety, **Functional Safety according to ISO 26262, Safety of the Intended Functionality (SOTIF) according to ISO 21448, and Cybersecurity according to ISO/SAE 21434**.

Where software or AI-related content is updated during the product lifecycle, the relationship with **Software Update Engineering according to ISO 24089** shall also be considered.

The organization shall identify and document relevant interfaces, dependencies, shared assumptions and responsibilities. AI safety activities shall complement applicable safety and security processes and shall not be interpreted as replacing requirements established by other applicable standards.

Where an AI function is part of a safety-related system, the organization shall determine which requirements and evidence belong to the AI safety domain and which are addressed through Functional Safety, SOTIF, Cybersecurity or other applicable engineering disciplines. The resulting allocation shall be documented and traceable.

2.6 Applicability and Classification

The applicability of this manual shall be determined for each AI project, product or system through a documented **AI Safety Planning and Project Classification process**.

The classification shall consider, as appropriate, the intended function of the AI system, its operational environment, its interaction with other system functions, the degree of autonomy, the

potential consequences of erroneous or insufficient AI behavior, applicable safety requirements, regulatory requirements, customer requirements and the relationship to Functional Safety, SOTIF and Cybersecurity.

Based on this classification, the organization shall determine the extent and rigor of the AI safety activities required for the specific project.

Not every requirement of this manual will necessarily apply with the same level of rigor to every AI application. Any tailoring or exclusion shall be **documented, technically justified, risk-based and formally approved** by the responsible authority. Requirements that are mandatory due to applicable legislation, regulation, contractual obligations or an applicable certification scheme shall not be excluded through organizational tailoring.

The result of the applicability determination shall be recorded in the applicable **AI Safety Management Plan**, **AI Project Classification Record** or equivalent controlled work product.

2.7 Exclusions

An AI system or AI-related activity may be excluded from specific requirements of this manual only where a documented analysis demonstrates that the excluded activity has no relevant impact on the applicable AI safety objectives and that the exclusion does not compromise the integrity of the overall safety management system.

Exclusions shall be explicitly documented together with their technical justification, responsible authority and approval.

An exclusion from the scope of a particular AI safety activity shall not be interpreted as an exclusion from other applicable safety, cybersecurity, quality, regulatory or legal requirements.

2.8 Scope Management

The defined scope shall be maintained as a controlled element of the AI Safety Management System. Changes to the organizational structure, AI technology, product portfolio, intended use, operating environment, applicable standards, regulatory framework, development approach or system architecture shall be evaluated to determine whether they affect the established scope.

Where a change affects the scope or applicability of this manual, the relevant documentation shall be updated and the change shall be subject to the organization's defined change-management process.

The **AI Safety Manager** shall be responsible for coordinating the review of the scope, while the designated management authority shall retain overall responsibility for approving significant changes to the AI Safety Management System.
The scope of the AI Overall Safety Management System shall be reviewed periodically as part of the **Management Review** process and whenever significant changes occur.

3 Applicable Standards and Frameworks

3.1 General

The AI Safety Management System of **[Company Name]** shall be established and operated with due consideration of all standards, regulations, statutory requirements, contractual obligations, customer requirements and other applicable external requirements relevant to the AI system or product under consideration.
The applicable requirements shall be determined based on the **intended use, product type, safety relevance, technical architecture, development lifecycle, target market, operating environment and intended deployment context** of the AI system.
The organization shall not assume that a single AI standard or regulatory framework is sufficient to address all relevant safety aspects. Depending on the application, AI safety may require the coordinated application of AI-specific requirements together with Functional Safety, SOTIF, Cybersecurity, Software Update Engineering, system engineering, quality management and applicable regulatory requirements.
The applicable requirements shall be identified at an early stage of the project and shall be maintained throughout the complete lifecycle. Changes to the product, intended use, technology, target market, applicable standards or regulatory environment shall trigger a review of the applicable requirements.

3.2 AI-Specific Safety Standards

For AI systems used in road vehicles or other safety-relevant applications, the organization shall identify and apply the AI-specific standards and technical guidance relevant to the intended use.

Particular consideration shall be given to **ISO/PAS 8800**, where applicable to AI-related safety in road vehicles, and **ISO/IEC TR 5469**, which addresses the relationship between Artificial Intelligence and functional safety.

These documents provide an important basis for addressing AI-specific characteristics that may require additional consideration beyond conventional software development and safety engineering approaches. Depending on the application, these characteristics may include data-related limitations, AI model behavior, uncertainty, robustness, performance limitations, verification and validation challenges, operational limitations and changes in AI behavior over the lifecycle.

The organization shall determine which provisions are applicable to the specific AI system and shall document the resulting applicability and implementation strategy.

3.3 Functional Safety

Where an AI system or AI-enabled component contributes to a safety-related function in a road vehicle, the requirements and processes of **ISO 26262** shall be considered and applied where applicable.

The organization shall determine the relationship between AI safety activities and the Functional Safety lifecycle. This includes, as applicable, the relationship between safety goals, functional and technical safety requirements, AI-related requirements, assumptions of use, safety mechanisms, verification activities, integration activities and safety evidence.

AI safety activities shall complement the Functional Safety process and shall not be used to replace applicable ISO 26262 requirements.

Where AI is implemented as part of a safety-related item or as a safety element within a larger system, the allocation of responsibilities and evidence between the AI safety process and the Functional Safety process shall be explicitly defined and maintained.

3.4 Safety of the Intended Functionality

Where safety depends on the intended functionality of an AI-enabled system, the organization shall consider the requirements and methods of **ISO 21448 (SOTIF)**.

This is particularly relevant where hazards may arise without a conventional hardware or software fault, for example as a consequence of limitations of perception, insufficient functionality, inadequate scenario coverage, foreseeable misuse or insufficient understanding of operating conditions.

For AI-based perception, classification, prediction or decision-making functions, the organization shall determine whether SOTIF activities are required and shall define the appropriate interfaces between SOTIF and AI safety activities.
Where both AI safety and SOTIF apply, common assumptions, scenarios, limitations, verification results and validation evidence shall be appropriately coordinated and traceable.

3.5 Cybersecurity

Where AI systems are implemented in road vehicles or other connected systems, cybersecurity requirements shall be considered in accordance with **ISO/SAE 21434** and other applicable cybersecurity requirements.
AI safety and cybersecurity shall be treated as related but distinct disciplines. A cybersecurity compromise may affect the safety behavior of an AI system, for example through manipulation of training data, modification of AI models, unauthorized software or model replacement, manipulation of configuration data or attacks against AI inputs.
The organization shall therefore identify cybersecurity threats that may have a potential impact on AI safety and shall establish appropriate interfaces between the AI Safety Management System and the Cybersecurity Management System.
The allocation of cybersecurity activities and AI safety activities shall be documented to ensure that relevant risks are not omitted or duplicated.

3.6 Software Update Engineering

Where AI software, models, datasets, parameters, configuration or other safety-relevant software elements are updated during the product lifecycle, the organization shall consider **ISO 24089, Software Update Engineering**, where applicable.
The update process shall ensure that safety-relevant changes are identified, assessed and appropriately verified and validated before deployment.
Particular attention shall be given to the potential impact of an update on:

- the validated AI model or algorithm;
- training, validation or operational data;
- safety requirements and assumptions;

- system interfaces;
- cybersecurity;
- SOTIF-related scenarios and limitations;
- Functional Safety assumptions and evidence;
- the existing safety argument or safety case.

Where an update affects safety-relevant characteristics, the organization shall determine the necessary scope of regression verification, validation and safety assessment before release.

3.7 Certification and Conformity Assessment

Where **[Company Name]** performs conformity assessment or certification activities, the applicable certification scheme and accreditation requirements shall be identified and applied.

ISO/IEC 17065 may be applicable where the organization performs certification of products, processes or services within the scope of a certification body. ISO/IEC 17065 itself does not define the technical safety requirements for AI; rather, it establishes requirements for the competence, consistency, impartiality and operation of bodies performing applicable conformity assessment activities.

Where AI safety certification or assessment is performed under an accredited scheme, the organization shall clearly distinguish between:

1. the **technical requirements and standards against which the AI system or product is assessed**; and
2. the **conformity-assessment and certification requirements governing the assessment and certification process**.

The applicable certification scheme, assessment criteria, decision-making process and accreditation scope shall be documented and controlled.

3.8 Regulatory and Type-Approval Requirements

For products intended for regulated markets, the organization shall identify all applicable statutory and regulatory requirements in addition to voluntary standards.

For road-vehicle applications, this may include applicable **UNECE Regulations**, type-approval requirements and market-specific regulatory requirements relevant to the vehicle, system or component.

The organization shall determine whether and how AI functionality contributes to compliance with applicable regulatory requirements. Where AI-related evidence forms part of a vehicle manufacturer's or supplier's regulatory or type-approval evidence package, the relevant interfaces, assumptions and responsibilities shall be documented.

Regulatory compliance shall not be assumed solely on the basis of compliance with an engineering standard. Conversely, compliance with an engineering standard shall not automatically be interpreted as evidence of regulatory approval unless this is explicitly established within the applicable regulatory framework.

The applicable regulatory requirements shall be determined for each target market.

3.9 Classification of External Requirements

The following standards and frameworks are considered particularly relevant to the AI Safety Management System. Their applicability shall be determined for each project rather than being assumed universally.

Standard / Framework	Principal Area of Application	Typical Relevance to AI Safety
ISO/PAS 8800	AI-related safety for road vehicles	AI safety lifecycle, safety management and AI-specific safety considerations
ISO/IEC TR 5469	Artificial Intelligence and Functional Safety	AI-related aspects of functional safety and safety assurance
ISO 26262	Functional Safety of road vehicles	Safety lifecycle, HARA, safety requirements, implementation, verification and validation
ISO 21448	Safety of the Intended Functionality (SOTIF)	Functional insufficiencies, foreseeable misuse, scenario coverage and unknown unsafe scenarios
ISO/SAE 21434	Road Vehicle Cybersecurity Engineering	Cybersecurity risks and interfaces affecting AI safety

ISO 24089	Software Update Engineering	Safety-relevant software, model and configuration updates
ISO/IEC 17065	Conformity Assessment – Certification of Products, Processes and Services	Certification-body requirements where applicable
Applicable UNECE Regulations	Vehicle regulatory compliance and type approval	Regulatory requirements relevant to AI-enabled vehicle functions
Applicable national/regional legislation	Legal and regulatory compliance	Mandatory requirements applicable to the target market
Customer-specific requirements	Contractual / customer compliance	Additional safety, security, quality and evidence requirements
[Other applicable standard]	[Area]	[Relevance]

Table 1: Standards and frameworks considered to the AI Safety Management System.

The table shall be maintained as a controlled reference and shall be reviewed whenever relevant standards, regulations or customer requirements are revised.

3.10 Applicable Requirements List

For each AI project or product, an **Applicable Requirements List (ARL)** shall be established and maintained as a controlled work product.
The ARL shall identify the standards, regulations, contractual requirements, customer requirements and other external obligations that are applicable to the specific project.
The ARL shall, as appropriate, identify:

- the applicable standard or regulation;
- edition or revision;
- applicable clauses or requirements;
- applicability rationale;
- responsible organizational function;
- affected lifecycle phase;

- required work products or evidence;
- relationship to other applicable requirements;
- compliance status;
- identified gaps or deviations.

The ARL shall provide the baseline for determining the applicable AI safety processes and assessment activities.

3.11 Determination of Applicability

The determination of applicability shall be performed in a systematic and documented manner.
The organization shall consider, at minimum, the **intended use and function of the AI system, safety relevance, product category, system architecture, operating environment, target market, regulatory classification, customer requirements, interfaces to other safety and security disciplines, and the consequences of incorrect or insufficient AI behavior**.
The applicability assessment shall distinguish between requirements that are:

- mandatory due to legislation or regulation;
- mandatory due to contractual or customer requirements;
- applicable through the selected certification or assessment scheme;
- applicable engineering standards adopted by the organization;
- recommended guidance or best-practice documents.

Where a standard is not applicable, the rationale shall be documented.
Where only selected requirements of a standard are applicable, the applicable clauses and the rationale shall be identified.

3.12 Management of Changes to External Requirements

The organization shall maintain a process for monitoring relevant changes to standards, regulations and other external requirements.
Changes shall be evaluated for their potential impact on:

- the AI Safety Management System;
- existing projects;
- released products;
- certification or assessment status;
- safety requirements;
- safety evidence;
- customer commitments;
- regulatory compliance.

Where a change is determined to affect an existing product or project, the organization shall perform an appropriate impact assessment and shall determine whether updates to requirements, processes, analyses, verification, validation or safety evidence are necessary.
The outcome of the review shall be documented and incorporated into the applicable change-management process.

3.13 Conflicts and Overlapping Requirements

Where requirements from different standards, regulations or customer specifications overlap or appear to conflict, the organization shall identify and resolve the conflict through a documented requirements analysis.
The organization shall not resolve conflicting requirements by simply selecting the less stringent requirement without technical justification and appropriate approval.
Particular attention shall be given to overlapping requirements between **AI safety, Functional Safety, SOTIF and Cybersecurity**, since the same system behavior, assumption, scenario or technical measure may be relevant to more than one discipline.
Where appropriate, a common requirement or evidence item may be used to support multiple standards, provided that the evidence demonstrably satisfies the respective requirements.

3.14 Compliance and Objective Evidence

Compliance with applicable standards and regulatory requirements shall be demonstrated through objective evidence appropriate to the nature and safety relevance of the AI system.

The organization shall maintain traceability between the applicable requirements and the evidence demonstrating their fulfillment.
Objective evidence may include, as applicable:

- requirements specifications;
- analyses and assessments;
- AI safety plans;
- data and model documentation;
- verification and validation results;
- test reports;
- review records;
- safety cases;
- audit reports;
- conformity assessment records;
- certificates;
- regulatory submissions;
- approval records.

The existence of a document or procedure shall not, by itself, be considered sufficient evidence of implementation. Where applicable, the organization shall demonstrate that defined processes have been effectively implemented and that the resulting technical and safety objectives have been achieved.

3.15 Responsibility

The **AI Safety Manager** shall coordinate the identification and maintenance of AI safety-related external requirements. The relevant Functional Safety, SOTIF, Cybersecurity, Quality, Regulatory and Certification functions shall be responsible for the requirements within their respective areas of competence.
Overall responsibility for ensuring that applicable legal, regulatory, contractual and standardization requirements are identified and addressed shall remain with the designated management authority.
The applicable requirements and their implementation status shall be reviewed at defined lifecycle gates and as part of the management review process.

3.16 Reference to the Project-Specific Compliance Framework

The requirements identified in this chapter establish the overall framework for determining the applicable standards and regulatory requirements. They shall be translated into project-specific requirements through the **AI Safety Management Plan, Applicable Requirements List, Compliance Matrix and relevant AI safety procedures and checklists**.
The project-specific compliance framework shall provide a clear and traceable relationship between:

Applicable Standard / Regulation
→ **Applicable Requirement**
→ **AI Safety / Engineering Requirement**
→ **Implementation**
→ **Verification / Validation**
→ **Objective Evidence**
→ **Assessment / Approval**

This structure shall ensure that the AI Safety Management System provides a controlled basis for demonstrating compliance throughout the complete lifecycle of the AI system or product.

4 AI Safety Policy

4.1 Management Commitment

[Company Name] recognizes that Artificial Intelligence and Machine Learning may introduce safety characteristics and risks that differ from those associated with conventional deterministic software and hardware. AI safety is therefore considered a fundamental element of responsible product development and lifecycle management.
[Company Name] is committed to establishing, implementing, maintaining and continually improving an effective **AI Safety Management System** that provides a systematic framework for identifying, analysing, evaluating, controlling and monitoring AI-related safety risks throughout the complete lifecycle of AI systems and AI-enabled products.

The organization is committed to developing, integrating and delivering AI systems that are appropriate for their defined intended use and operating conditions and whose safety-related characteristics are supported by sufficient objective evidence. AI systems shall be subject to appropriate verification and validation, and their known limitations, assumptions and residual risks shall be identified, documented and appropriately controlled.

AI safety shall be addressed from the earliest stages of concept and requirements definition and shall remain an active responsibility through development, integration, verification, validation, release, deployment, operation, monitoring, maintenance, modification, updating and retirement.

The organization shall ensure that safety-relevant AI activities are performed by personnel with appropriate competence and authority and that sufficient organizational, technical and financial resources are made available to achieve the defined safety objectives.

AI safety decisions shall be based on **objective evidence, documented analysis and sound engineering judgment**. Commercial, schedule or other non-technical considerations shall not justify the release or continued operation of an AI system where identified safety risks have not been adequately addressed or formally accepted by an authorized decision-maker.

Where applicable, AI safety shall be coordinated with **Functional Safety, SOTIF, Cybersecurity, Software Update Engineering, Quality Management and Regulatory Compliance** to ensure that safety-relevant risks and dependencies are addressed consistently across the organization.

The organization is committed to complying with applicable **laws, regulations, standards, certification requirements, contractual obligations and customer requirements** relevant to the AI system or product.

AI safety shall not be regarded solely as a testing, validation or final-gate activity. It shall be treated as a **lifecycle-wide engineering and management responsibility**.

4.2 AI Safety Principles

The AI Safety Management System shall be based on the following fundamental principles.

4.2.1 Safety by Design

Safety shall be considered from the beginning of the AI lifecycle and shall be incorporated into the concept, architecture, requirements, data strategy, model development, system integration and operational design.

Safety shall not depend solely on detecting unsafe behavior after development has been completed.

4.2.2 Risk-Based Development

AI safety activities shall be proportionate to the potential safety impact, complexity, intended use and operating environment of the AI system.
Safety risks shall be systematically identified, analyzed, evaluated and controlled. The level of verification, validation, independence, documentation and assessment shall be appropriate to the identified risk.

4.2.3 Lifecycle Responsibility

AI safety responsibility extends across the complete lifecycle.
The organization shall consider not only initial development and release but also deployment, operational monitoring, maintenance, software and model updates, changes in data or operating conditions, incident management and retirement.

4.2.4 Traceability and Reproducibility

Safety-relevant requirements, assumptions, decisions, data, models, configurations, verification and validation results, changes and approvals shall be appropriately traceable.
Where applicable, the organization shall maintain sufficient information to reproduce or reconstruct safety-relevant development and verification activities.

4.2.5 Appropriate Independence

Verification, validation, review and assessment activities shall be performed with an appropriate degree of independence based on the safety relevance and organizational requirements of the AI system.
Personnel shall not be placed in a position where commercial, schedule or personal interests improperly influence safety-related decisions.
Where independent assessment or confirmation measures are required, these shall be formally defined and implemented.

4.2.6 Evidence-Based Decision Making

Safety-related decisions shall be supported by objective and sufficient evidence.

Claims concerning the safety, robustness, performance or suitability of an AI system shall be supported by appropriate analyses, verification, validation, testing, reviews or other objective evidence.
The absence of evidence shall not be interpreted as evidence of safety.

4.2.7 Controlled Changes

Safety-relevant changes to AI models, datasets, algorithms, software, hardware, configuration, tools, operating conditions or intended use shall be subject to controlled change management.
The potential impact of a change on existing safety requirements, assumptions, risks, verification and validation results and safety evidence shall be evaluated before the change is released.

4.2.8 Continuous Monitoring and Assurance

AI safety shall continue to be assured after deployment.
Where appropriate, operational performance, safety-relevant events, data characteristics, model behavior, environmental conditions and other relevant indicators shall be monitored.
New information, incidents, field experience or changes in operating conditions shall be evaluated to determine whether the existing safety argument and evidence remain valid.

4.2.9 Defined Accountability and Authority

Safety responsibilities, authorities and interfaces shall be clearly defined and communicated.
Personnel responsible for AI safety shall have sufficient authority to escalate safety concerns and, where necessary, prevent or stop a release or deployment until safety-relevant issues have been appropriately resolved.
Accountability for safety-related decisions shall be clearly assigned and documented.

4.2.10 Controlled Deployment and Release

No AI system shall be deployed or released for a defined safety-relevant use without completion of the applicable release criteria and authorization by the designated responsible authority.
Known limitations, open issues, residual risks and approved deviations shall be formally evaluated before release.
Uncontrolled deployment, including deployment of unapproved AI models, datasets, configurations or software versions, shall not be permitted.

4.3 Management Responsibility

Top management retains overall responsibility for ensuring that the AI Safety Management System is established, implemented and maintained effectively within **[Company Name]**.

Top management shall provide the leadership, resources and organizational conditions necessary to achieve the organization's AI safety objectives. This includes ensuring that sufficient qualified personnel, appropriate infrastructure, suitable tools and adequate financial and organizational resources are available.

Management shall ensure that AI safety responsibilities and authorities are clearly defined and that the **AI Safety Manager** and other safety-related functions have sufficient authority, competence and organizational access to perform their responsibilities effectively.

Management shall establish and communicate the organization's AI safety objectives and shall ensure that these objectives are consistent with the organization's overall business, quality, safety and compliance objectives.

Management shall ensure that safety-related decisions are documented and that unresolved safety risks are appropriately escalated to the level of authority required for their resolution or formal acceptance.

Where a safety risk exceeds the defined acceptance criteria, management shall ensure that the risk is not accepted implicitly through schedule pressure, commercial considerations or organizational inaction.

Management shall ensure that personnel are encouraged and authorized to report safety concerns without inappropriate negative consequences. Safety concerns shall be evaluated objectively and shall be escalated where required.

Management shall also ensure that the organization maintains appropriate independence for safety-critical verification, validation, review and assessment activities.

In particular, **commercial interests, project schedules, customer pressure or management targets shall not override defined safety requirements or release criteria**. Any deviation from an established safety requirement or process shall be handled through the formally defined deviation and waiver process and shall require appropriate technical justification and authorization.

Top management shall periodically review the effectiveness and suitability of the AI Safety Management System through formal **Management Review** activities. The review shall consider, as

applicable, safety performance, audit results, incidents, significant changes, open risks, deviations, resource requirements, competence, regulatory developments and opportunities for improvement.

4.4 AI Safety Objectives

Based on this policy, **[Company Name]** shall establish measurable and, where appropriate, project-specific AI safety objectives.
The objectives shall support the organization's commitment to:

- preventing unacceptable AI-related safety risks;
- ensuring that applicable safety requirements are identified and addressed;
- ensuring appropriate verification and validation;
- maintaining traceable and controlled safety evidence;
- ensuring effective management of AI-related changes;
- maintaining appropriate competence and independence;
- ensuring controlled release and deployment;
- monitoring safety throughout the operational lifecycle;
- identifying and addressing lessons learned;
- continually improving the AI Safety Management System.

The specific objectives, targets and associated performance indicators shall be defined and monitored through the applicable **AI Safety Management Plan** and management processes.

4.5 Communication of the AI Safety Policy

The AI Safety Policy shall be communicated to all personnel whose activities may affect the safety of AI systems or AI-enabled products.
Personnel shall understand:

- the organization's commitment to AI safety;
- their individual safety responsibilities;
- their authority to raise safety concerns;
- the applicable escalation mechanisms;

- the consequences of uncontrolled changes or deployment.

Where relevant, the policy and associated safety expectations shall also be communicated to external organizations, suppliers and development partners whose activities may affect AI safety.
The policy shall be reviewed periodically and whenever significant changes occur to the organization's AI activities, applicable requirements or safety management framework.

4.6 Continual Improvement

[Company Name] shall continually improve the suitability, adequacy and effectiveness of its AI Safety Management System.
Improvement shall be based on objective information, including audit findings, assessment results, safety incidents, operational experience, customer feedback, lessons learned, changes in technology, changes in applicable standards and regulations, and identified process weaknesses.
Corrective and preventive actions shall be tracked to completion and their effectiveness shall be evaluated where appropriate.
The organization shall ensure that lessons learned from one AI project are considered for application to other relevant projects and products, thereby preventing the recurrence of identified safety issues.

4.7 Policy Statement

Through this AI Safety Policy, **[Company Name]** commits to managing AI safety as a fundamental and lifecycle-wide responsibility.
The organization shall provide the governance, competence, resources, processes, independence and objective evidence necessary to support safe and controlled development and deployment of AI systems.
1. The organization shall ensure that **AI systems are not released merely because they function as intended, but only when the applicable safety requirements, risks, limitations and evidence have been appropriately addressed and the defined release criteria have been satisfied**.

5 AI Safety Organization

5.1 Organizational Structure

The organization shall establish and maintain an **AI Safety Organization** appropriate to the size, organizational complexity, technologies, products, intended use and safety relevance of its AI systems.

The organizational structure shall ensure that AI safety responsibilities, authorities, competencies, interfaces and escalation paths are clearly defined throughout the applicable AI lifecycle.

The organization shall provide sufficient competent personnel, resources and organizational authority to perform AI safety activities effectively. Where required by the safety relevance or applicable requirements, appropriate independence between development, verification, assessment and approval activities shall be maintained.

The AI Safety Organization shall comprise the management function responsible for AI safety and shall have defined interfaces to all organizational and technical functions that contribute to or may affect AI safety.

A generic organizational model may be established as follows:

Top Management

↓

AI Safety Manager / AI Safety Responsible

↓

AI Safety Management System

↙ ↓ ↘

Relevant Functional Areas and Specialist Interfaces

The functional areas relevant to AI safety may include:

- **AI / Machine Learning Development**
- **Data Engineering and Data Management**

- **Verification and Validation**
- **System Engineering**
- **Functional Safety**
- **SOTIF**
- **Cybersecurity**
- **Software Update Engineering**
- **Quality Management**
- **Configuration and Change Management**
- **Regulatory and Product Compliance**, where applicable

The above functional areas represent the competencies and organizational interfaces required to manage AI safety. They do **not necessarily constitute reporting units within the AI Safety Organization** and shall not be interpreted as a mandatory reporting hierarchy.

The actual organizational structure, reporting lines and allocation of functions may differ according to the organization's size, product portfolio, business model and operating model. Irrespective of the organizational arrangement, the responsibilities and authorities associated with each relevant safety function shall be clearly defined.

5.2 AI Safety Management Function

The organization shall designate an **AI Safety Manager or AI Safety Responsible** with appropriate competence, authority and organizational access to coordinate the AI Safety Management System.

The AI Safety Manager shall be responsible for ensuring that the AI safety framework is established, implemented, maintained and continually improved in accordance with this manual.

The AI Safety Manager shall coordinate the interfaces between the relevant functional areas and shall ensure that AI safety activities are appropriately planned, assigned, performed, reviewed and documented.

The AI Safety Manager shall not automatically assume the technical responsibility of the specialist functions contributing to AI safety. Technical responsibility shall remain with the competent function or personnel assigned to the respective activity.

The AI Safety Manager shall have sufficient authority to escalate unresolved or significant safety concerns to the appropriate level of management and shall have access to the information and resources necessary to perform the assigned responsibilities.

5.3 Relevant Functional Areas and Specialist Interfaces

The organization shall identify the functional areas that are relevant to the safety of its AI systems and shall define the interfaces between those areas and the AI Safety Management System.

Depending on the AI system and its application, different functional areas may have different levels of involvement. Not every AI project necessarily requires active participation from every functional area listed in Section 5.1.

For example, the **AI / Machine Learning Development** function is responsible for the development and technical implementation of AI models and associated software, while **Data Engineering and Data Management** is responsible for the controlled management of safety-relevant data and datasets.

Verification and Validation is responsible for the applicable verification and validation activities, including the definition and execution of appropriate methods and the generation of objective evidence.

System Engineering ensures that the AI component is considered within its system context and that relevant interfaces, requirements, assumptions and system-level behavior are appropriately addressed.

Where applicable, the **Functional Safety**, **SOTIF** and **Cybersecurity** functions shall provide their respective specialist expertise and shall ensure that the relevant interfaces to ISO 26262, ISO 21448 and ISO/SAE 21434 are appropriately managed.

Software Update Engineering shall be involved where AI software, models, datasets, parameters or other safety-relevant elements are subject to controlled updates.

Quality Management shall support the effective implementation and monitoring of the applicable management-system processes.

Configuration and Change Management shall ensure that safety-relevant configurations, changes and baselines are appropriately controlled.

Regulatory and Product Compliance shall be involved where regulatory, type-approval, certification or other external compliance requirements apply.

The organization shall ensure that these functional areas cooperate where their activities are interdependent. Safety-relevant information, requirements, assumptions, risks, changes and evidence shall be communicated to the functions affected by them.

5.4 Responsibility and Authority

The organization shall define responsibility and authority for all safety-relevant activities.

Responsibilities shall include, as applicable:

- performing safety-related activities;
- providing specialist input;
- reviewing work products and results;
- verifying compliance with defined requirements;
- performing safety assessments;
- approving safety-relevant decisions;
- escalating unresolved safety concerns; and
- authorizing release of the applicable AI configuration.

Responsibility for performing an activity and authority for approving its result shall be clearly distinguished where required by the safety classification or applicable organizational and external requirements.

The detailed allocation of responsibilities and authorities shall be defined in the **Authority and Responsibility Matrix** specified in Chapter 8 and shall be consistent with the role definitions in Chapter 6.

No person shall exercise safety-relevant approval authority without the required competence and formally assigned authority.

5.5 Combination of Roles and Organizational Size

The organization may combine functions or assign several roles to the same individual where this is appropriate to its size, organizational structure and safety relevance.

Such combination shall only be permitted where the organization can demonstrate that the required competence, authority, objectivity, independence and effectiveness of the respective activities are maintained.

The combination of roles shall not result in the omission of required activities or in the uncontrolled approval of work performed by the same individual.

Where a person performs both development activities and review, verification, assessment or approval activities, the organization shall evaluate the potential conflict of interest and shall establish appropriate measures to maintain objective decision-making.

Depending on the safety relevance, such measures may include independent review, separate approval authority, additional verification, peer review, independent safety assessment or external assessment.

The applicable arrangement shall be documented where the combination of roles could affect the independence or objectivity of a safety-relevant decision.

5.6 Escalation and Safety Authority

The organization shall establish clear escalation paths for unresolved, significant or potentially unacceptable AI safety concerns.

Personnel involved in AI safety activities shall have the authority and responsibility to report and escalate safety concerns without being prevented from doing so by commercial, organizational or schedule-related considerations.

Safety concerns that cannot be resolved at the responsible technical level shall be escalated to the AI Safety Manager and, where necessary, to Top Management or the designated safety decision authority.

The organization shall ensure that escalation does not depend solely on the hierarchical reporting line of the person identifying the concern.

Significant safety concerns shall remain controlled until an authorized decision has been made regarding their resolution, acceptance, restriction, further investigation or impact on the applicable release.

5.7 External Organizations and Suppliers

Where external organizations, suppliers, consultants or service providers perform activities that may affect AI safety, the organization shall define the applicable responsibilities, interfaces and expectations.

The organization shall ensure that externally performed safety-relevant activities are subject to appropriate competence, contractual, technical, configuration, evidence and review controls.

External personnel shall not automatically receive safety-relevant decision authority solely by virtue of performing technical activities. Where external personnel participate in verification, assessment or approval activities, their competence, independence and authority shall be defined and controlled.

Relevant supplier and external-party interfaces shall be considered in the AI Safety Management Plan and, where applicable, in the Authority and Responsibility Matrix.

5.8 Organizational Competence and Resources

The organization shall ensure that the AI Safety Organization has sufficient competent personnel and resources to perform the activities required by the applicable AI safety lifecycle.

Resource planning shall consider the safety relevance and complexity of the AI system, required technical disciplines, lifecycle phases, V&V activities, independent reviews, assessments, operational monitoring and expected changes.

Where a required competence is not available internally, the organization shall establish appropriate measures, such as training, recruitment, competent external personnel or independent specialist support.

Resource limitations shall not be used as justification for omitting mandatory safety activities.

Competence requirements and qualification of personnel shall be managed in accordance with Chapter 7, **Competence and Qualification**, and Chapter 28, **Training and Awareness**.

5.9 Organizational Information and Records

The organizational structure, relevant functional areas, responsibilities, authorities and reporting or escalation relationships shall be documented as controlled organizational information.

The organization shall maintain sufficient records to demonstrate the assignment of safety-relevant responsibilities and authorities.

Changes to organizational responsibilities shall be reflected in the applicable organizational documentation, Authority and Responsibility Matrix and, where relevant, project-specific safety plans.

5.10 Organizational Change

Organizational changes that may affect AI safety responsibilities, competence, independence, resources, interfaces or decision authority shall be subject to an appropriate impact assessment.

Such changes may include organizational restructuring, changes in management, transfer of responsibilities, loss of key competence, outsourcing or insourcing of safety activities, introduction of new AI technologies or significant changes to the organization's AI product portfolio.

Where necessary, the organization shall implement corresponding measures for competence transfer, training, authorization, independent review, resource allocation or transitional control before the changed organization is relied upon for safety-relevant activities.

The effectiveness of significant organizational changes shall be reviewed after implementation.

5.11 Organizational Review and Continual Improvement

The effectiveness of the AI Safety Organization shall be periodically reviewed as part of the AI Safety Management System.

The review shall consider, as applicable, audit findings, assessment results, incidents, recurring safety issues, competence gaps, resource constraints, escalation effectiveness, organizational changes and experience from AI projects.

Where weaknesses in the organizational structure or interfaces are identified, appropriate corrective or improvement actions shall be initiated.

The organizational structure shall therefore be treated as a controlled element of the AI Safety Management System rather than as a static organizational chart.

The objective of the AI Safety Organization is to ensure that the organization has clearly defined and appropriately authorized responsibilities, competent resources and effective interfaces for managing AI safety throughout the complete AI lifecycle, while maintaining the required objectivity and independence of safety-relevant decisions.

6 Roles and Responsibilities

6.1 General

The organization shall define and maintain clear roles, responsibilities, authorities and interfaces for all personnel performing or supporting AI safety-related activities.

The assignment of a role under this chapter shall not necessarily require the creation of a separate organizational position. Depending on the size, structure and competence of **[Company Name]**, one person may perform more than one role, provided that the required competence, authority, objectivity and independence are maintained.

Where roles are combined, the organization shall assess potential conflicts of responsibility and interest, particularly with respect to independent verification, validation, assessment and release decisions. Appropriate compensating measures shall be implemented where required.

The responsibilities defined in this chapter represent the generic minimum framework. Project-specific responsibilities, deliverables and decision authorities shall be further defined in the applicable **AI Safety Management Plan**, responsibility matrix, project plan or equivalent controlled work product.

6.2 AI Safety Manager

The **AI Safety Manager** shall be responsible for the overall coordination, planning and monitoring of AI safety activities within the defined organizational or project scope.

The AI Safety Manager shall ensure that AI safety is managed as a lifecycle-wide activity and that the required AI safety processes, responsibilities, work products and interfaces are established and implemented in accordance with this manual and the applicable project requirements.

The AI Safety Manager shall coordinate the development and maintenance of the **AI Safety Management Plan** and shall ensure that the plan defines the applicable lifecycle activities, responsibilities, standards, methods, work products, reviews, independence requirements, milestones and release criteria.

The responsibilities of the AI Safety Manager shall include, as applicable, the following:

1. establishing and maintaining the AI Safety Management Plan;
2. coordinating the identification and implementation of applicable AI safety processes and requirements;

3. supporting or coordinating AI project classification and applicability determination;
4. ensuring that AI safety activities are planned throughout the complete lifecycle;
5. coordinating the management of AI-related safety risks, assumptions, limitations and dependencies;
6. monitoring the status, completeness and consistency of required AI safety work products and objective evidence;
7. coordinating interfaces with Functional Safety, SOTIF, Cybersecurity, System Engineering, Software Update Engineering, Quality Management and other relevant disciplines;
8. ensuring that safety-relevant interfaces, responsibilities and dependencies are appropriately documented;
9. reviewing or coordinating the review of safety-relevant deviations, waivers and unresolved issues;
10. monitoring changes that may affect established safety requirements, assumptions or evidence;
11. coordinating or supporting safety reviews and assessments;
12. ensuring that unresolved safety concerns are escalated through the defined escalation process;
13. supporting or recommending AI safety release decisions within the limits of the assigned authority;
14. ensuring that the AI safety evidence structure is established, maintained and appropriately controlled; and
15. supporting continual improvement through the evaluation of audits, assessments, incidents, field experience and lessons learned.

The AI Safety Manager shall have sufficient organizational authority and direct access to the appropriate management level to raise and escalate safety concerns.

The AI Safety Manager shall have the authority to recommend that a release or deployment be withheld where applicable safety requirements or release criteria have not been adequately demonstrated. Where the organization's governance model assigns release-blocking authority directly to the AI Safety Manager, this authority shall be formally documented.

No person, including the AI Safety Manager, shall approve an unsafe release merely on the basis of commercial, schedule or other non-technical considerations.

The AI Safety Manager shall not replace the technical responsibility of the responsible engineering functions. Technical responsibility for AI architecture, data, model development, verification, validation and other engineering decisions shall remain with the designated competent role.

6.3 ML Development Engineer

The **ML Development Engineer** shall be responsible for the technical development and implementation of AI and Machine Learning functionality within the defined project scope.

The ML Development Engineer shall develop the AI model and associated implementation in accordance with approved requirements, architectural constraints, safety-related assumptions and applicable development processes.

Responsibilities shall include, as applicable, the definition and implementation of AI/ML architectures, model development, algorithm selection, training configuration, parameterization, model optimization and implementation of approved safety-related requirements.

The ML Development Engineer shall ensure that relevant technical characteristics of the developed model are sufficiently documented to support verification, validation, integration, reproducibility, configuration management, change impact analysis and safety assessment.

The ML Development Engineer shall, as applicable:

- implement the approved AI/ML architecture and requirements;
- develop and maintain the AI model and associated software;
- define and document relevant model configurations and parameters;
- implement controlled training configurations;
- support reproducibility of training and model generation;
- document model versions and baselines;
- implement applicable safety mechanisms, constraints, monitors or fallback-related measures assigned to the AI function;
- identify and communicate known technical limitations and assumptions;
- support analysis of model behavior and performance limitations;
- participate in reviews, verification and validation activities;
- correct identified defects within the assigned responsibility;
- assess the impact of technical changes on the AI model and existing evidence; and
- provide the technical documentation required to support safety assessment and release.

The ML Development Engineer shall not independently modify a released or baselined AI model outside the defined configuration and change-management process.

6.4 Data Management Responsible / Data Engineer

The **Data Engineer** shall be responsible for the controlled acquisition, preparation, management and maintenance of data used for AI development, training, verification, validation and, where applicable, operational monitoring.

Because the behavior and safety characteristics of an AI system may be strongly influenced by its data, data management shall be treated as a safety-relevant activity where the data contributes to the performance or behavior of a safety-relevant AI function.

The Data Engineer shall ensure that relevant datasets are appropriately identified, documented, version-controlled and traceable.

Responsibilities shall include, as applicable:

- defining or implementing approved data requirements;
- acquiring data from identified and controlled sources;
- maintaining information on data origin and provenance;
- preparing, preprocessing and transforming data using controlled methods;
- supporting the assessment of data quality, completeness and suitability;
- supporting labeling and annotation activities where applicable;
- maintaining the separation and controlled use of training, validation and test datasets;
- maintaining dataset identification, versioning and configuration baselines;
- ensuring traceability between datasets, model versions and relevant development activities;
- supporting data distribution analysis and identification of relevant gaps or limitations;
- documenting relevant assumptions, limitations and known weaknesses associated with the data;
- controlling access to safety-relevant datasets;
- preventing unauthorized, undocumented or uncontrolled changes;
- supporting verification and validation activities involving datasets; and
- participating in change impact assessments where data-related changes may affect AI safety.

The Data Engineer shall ensure that changes to a safety-relevant dataset are identifiable and subject to the applicable configuration and change-management processes.

Where data acquisition, labeling or preparation is performed by an external organization, the Data Engineer or designated responsible function shall ensure that the applicable requirements, responsibilities and evidence are appropriately defined and controlled.

6.5 Verification and Validation Engineer

The **Verification and Validation Engineer** shall be responsible for planning, performing, coordinating or overseeing verification and validation activities necessary to provide objective evidence that the AI system satisfies its defined requirements and acceptance criteria within the applicable scope.

Verification and validation activities shall be appropriate to the safety relevance, intended use, operating environment, technical complexity and identified risks of the AI system.

The Verification and Validation Engineer shall define or support the definition of appropriate verification and validation strategies, methods, acceptance criteria and evidence requirements.

Responsibilities shall include, as applicable:

- defining verification and validation strategies and plans;
- establishing test objectives, methods and acceptance criteria;
- planning and controlling test activities;
- ensuring appropriate identification and configuration of test items and test environments;
- executing or coordinating verification and validation activities;
- evaluating functional performance and safety-related behavior;
- evaluating robustness, boundary conditions and relevant operating conditions;
- supporting the assessment of scenario coverage;
- identifying unknown, insufficiently covered or inadequately represented scenarios;
- evaluating relevant model limitations and performance degradation;
- performing or supporting regression testing following safety-relevant changes;
- documenting test procedures, results, anomalies and conclusions;
- ensuring traceability between requirements, test activities and results;
- reviewing evidence for completeness and consistency;
- supporting independent review where the required level of independence applies; and
- communicating failures, anomalies and safety-relevant findings to the responsible functions and the AI Safety Manager.

The Verification and Validation Engineer shall ensure that verification or validation results are not represented as successful where applicable acceptance criteria have not been fulfilled.
Open anomalies, limitations and residual risks shall be clearly documented and communicated for evaluation prior to release.

6.6 Functional Safety Manager

Where Functional Safety is applicable, the **Functional Safety Manager** shall be responsible for coordinating the interface between the AI Safety Management System and the applicable Functional Safety lifecycle, including activities performed in accordance with **ISO 26262** or another applicable Functional Safety framework.
The Functional Safety Manager shall ensure that the role and contribution of AI functionality within a safety-related item or system are appropriately understood and reflected in the Functional Safety activities.
Responsibilities shall include, as applicable:

- coordinating the interface between AI safety and Functional Safety activities;
- ensuring that relevant safety goals and safety requirements are communicated to the AI development and AI safety functions;
- coordinating the allocation of safety requirements to AI-related elements where applicable;
- defining or reviewing assumptions of use and dependencies relevant to AI functionality;
- ensuring appropriate traceability between Functional Safety requirements and AI-related requirements;
- coordinating safety-related interfaces between AI functions and conventional hardware and software elements;
- supporting the definition and assessment of safety mechanisms and related architectural measures;
- coordinating verification and validation activities where AI behavior contributes to the fulfillment of safety requirements;
- assessing the impact of AI-related changes on Functional Safety assumptions and evidence;
- supporting relevant confirmation measures, reviews and safety assessments; and
- ensuring that AI safety activities are not incorrectly used as a substitute for applicable Functional Safety requirements.

The Functional Safety Manager shall coordinate closely with the AI Safety Manager where AI-specific characteristics create additional uncertainty, limitations or evidence requirements not sufficiently addressed through conventional Functional Safety activities.

6.7 SOTIF Responsible

Where SOTIF is applicable, the **SOTIF Responsible** shall be responsible for coordinating the interface between AI safety and the applicable SOTIF activities.
The SOTIF Responsible shall address safety-relevant behavior resulting from limitations of the intended functionality rather than only from failures or malfunctions in the conventional sense.
This role is particularly relevant for AI-based functions involving perception, classification, prediction, interpretation or decision-making.
Responsibilities shall include, as applicable:

- coordinating the interface between AI safety and SOTIF activities;
- identifying functional insufficiencies and performance limitations relevant to safety;
- identifying limitations of the intended functionality;
- considering reasonably foreseeable misuse where applicable;
- supporting the identification of known and unknown potentially unsafe scenarios;
- coordinating scenario identification and coverage analysis;
- evaluating limitations associated with perception, classification, prediction or decision-making;
- coordinating assumptions relating to the intended operating environment;
- reviewing whether validation evidence adequately addresses relevant scenario classes and limitations;
- ensuring that identified SOTIF-related limitations are communicated to the AI Safety Manager and relevant engineering functions;
- assessing the impact of changes to the AI model, data, system or operating conditions on SOTIF assumptions and evidence; and
- supporting the definition of appropriate measures to reduce or control identified risks.

The SOTIF Responsible and AI Safety Manager shall ensure that AI-specific limitations and SOTIF-related limitations are consistently addressed and that common evidence is appropriately shared without creating gaps or duplicated responsibilities.

6.8 Cybersecurity Responsible

The **Cybersecurity Responsible** shall be responsible for ensuring that cybersecurity risks and controls with a potential impact on AI safety are identified and appropriately addressed.

AI safety and cybersecurity are distinct disciplines; however, a cybersecurity compromise may cause or contribute to unsafe AI behavior. The Cybersecurity Responsible shall therefore maintain an effective interface with the AI Safety Manager and relevant engineering functions.

Responsibilities shall include, as applicable:

- identifying cybersecurity assets relevant to AI safety;
- coordinating the assessment of threats that may affect the integrity, availability or authenticity of safety-relevant AI assets;
- identifying risks associated with adversarial attacks and malicious manipulation where applicable;
- considering manipulation, poisoning or unauthorized modification of training, validation or operational data;
- considering unauthorized modification, replacement or extraction of AI models where relevant;
- addressing unauthorized changes to safety-relevant AI software, parameters or configurations;
- supporting the protection and integrity of AI model deployment and update mechanisms;
- considering software update integrity and authenticity;
- supporting the protection of AI development and operational assets against unauthorized access;
- coordinating cybersecurity-related change impact analysis where a cybersecurity measure may affect AI safety behavior;
- communicating cybersecurity incidents that may invalidate AI safety assumptions or evidence; and
- ensuring that relevant cybersecurity risks are addressed through the applicable cybersecurity processes.

Where an AI system is within the scope of **ISO/SAE 21434**, the Cybersecurity Responsible shall ensure that the required interfaces between AI safety activities and cybersecurity engineering are clearly defined.

6.9 Quality Manager

The **Quality Manager**, where applicable, shall support the effective implementation and continual improvement of the processes used within the AI Safety Management System.

The Quality Manager shall provide appropriate interfaces to document control, internal audits, nonconformity management, corrective actions, management review and process improvement.

The Quality Manager shall support the AI Safety Manager in ensuring that documented AI safety processes are implemented consistently and that identified process deficiencies are appropriately addressed.

Quality assurance activities shall not replace the technical verification, validation or safety assessment required for an individual AI system. The respective responsibilities shall remain clearly separated.

6.10 Configuration and Change Manager

The **Configuration and Change Manager**, where applicable, shall ensure that safety-relevant AI configuration items and changes are uniquely identified, controlled and traceable throughout their lifecycle.

Configuration items may include requirements, datasets, data-processing configurations, AI models, model parameters, source code, software versions, tools, tool configurations, test cases, test results, deployment configurations and release packages.

The Configuration and Change Manager shall ensure that baselines are established where required and that it is possible to determine which controlled configuration was used to generate specific verification, validation or safety evidence.

Safety-relevant changes shall be processed through the defined change-management process, including appropriate impact analysis and required review, verification and validation activities.

6.11 System Engineer

The **System Engineer** shall be responsible for ensuring that AI functionality is appropriately defined, integrated and managed within the overall system context.

The System Engineer shall establish and maintain the relevant system-level requirements, architecture, interfaces, assumptions and constraints affecting the AI system.

Where AI functionality contributes to a safety-relevant function, the System Engineer shall ensure that the relationship between system requirements, safety requirements, AI requirements and system-level verification and validation is appropriately defined and traceable.
The System Engineer shall coordinate with the AI Safety Manager and the relevant safety and security functions to ensure that AI-related assumptions and limitations are correctly reflected at system level.

6.12 Management and Release Responsibility

The designated **Management and Release Authority** shall be responsible for approving safety-relevant release decisions within the limits of its formally assigned authority.
Before approval, the responsible authority shall ensure that the applicable release process has been completed and that the required evidence, open issues, residual risks, deviations and limitations have been appropriately evaluated.
A release approval shall not imply that all conceivable risks have been eliminated. It shall represent a documented decision that the applicable requirements and release criteria have been addressed and that any remaining risks or limitations have been handled in accordance with the defined acceptance and approval process.
Where the available evidence is incomplete or significant safety concerns remain unresolved, the responsible authority shall withhold approval, request additional evidence or initiate the applicable escalation or deviation process.

6.13 Responsibility Interfaces

The responsibilities defined in this chapter shall be implemented through defined interfaces between the relevant roles.
The organization shall ensure that no safety-relevant activity is left without a clearly assigned responsible function and that overlapping responsibilities do not result in assumptions that another function is addressing the activity.
Where multiple roles contribute to the same activity, the organization shall define the respective responsibility, authority, consultation and information requirements through an appropriate **Responsibility and Authority Matrix**, such as a RACI matrix.

The project-specific allocation of responsibilities shall be documented in the applicable AI Safety Management Plan and shall remain consistent with the organizational structure defined in Chapter 5.

7 Competence and Qualification

7.1 General

Personnel performing activities that may affect AI safety shall possess and maintain the competence necessary to perform their assigned responsibilities effectively.
The organization shall determine competence requirements based on the role, responsibility, authority, safety relevance and complexity of the activities performed. Competence requirements shall be proportionate to the potential impact of the individual's activities on the safety of the AI system.
Competence shall not be considered solely as completion of formal training. The organization shall consider the combination of relevant education, technical knowledge, practical experience, training, demonstrated capability and, where required, formal qualification or authorization.
Where personnel perform multiple roles, the organization shall ensure that the combined competence requirements of those roles are fulfilled and that the combination does not compromise the required independence or objectivity.

7.2 Competence Requirements

Competence requirements for AI safety-related roles shall be defined and maintained by the organization.
Depending on the assigned role, competence requirements shall consider, as applicable:

- AI and Machine Learning principles and technologies;
- AI model development and lifecycle management;
- data engineering and data management;
- system engineering and architecture;
- software and hardware engineering;
- Functional Safety and applicable safety engineering methods;

- SOTIF and intended-functionality safety;
- cybersecurity and security engineering;
- verification and validation;
- risk management and safety analysis;
- applicable AI safety standards and regulations;
- applicable automotive, industrial or other domain-specific requirements;
- AI development, verification and assessment tools;
- configuration and change management;
- statistical and performance evaluation methods, where relevant;
- safety case development and safety argumentation, where applicable; and
- conformity assessment and certification requirements, where relevant to the assigned role.

Not every individual shall be required to possess competence in all of these areas. The organization shall define the competence profile appropriate to each role and shall ensure that the collective competence of the project team covers all areas necessary to perform the required AI safety activities.

7.3 Role-Specific Competence

The organization shall define role-specific competence requirements for personnel performing AI safety-related activities.

For example, the competence profile of an **AI Safety Manager** should include knowledge of AI safety principles, lifecycle management, risk management, applicable standards, safety governance, assessment methods and the interfaces to Functional Safety, SOTIF and Cybersecurity.

An **ML Development Engineer** shall have appropriate competence in AI/ML development, model architecture, training methods, model configuration, implementation and the technical limitations relevant to the assigned AI application.

A **Data Engineer** shall have appropriate competence in data acquisition, preprocessing, quality assessment, dataset management, data provenance, versioning and the identification of data-related limitations.

A **Verification and Validation Engineer** shall have appropriate competence in test strategy, verification and validation methods, performance evaluation, robustness evaluation, scenario analysis and interpretation of AI-specific test results.

Personnel responsible for Functional Safety, SOTIF or Cybersecurity interfaces shall possess the applicable domain competence in addition to sufficient understanding of the AI characteristics relevant to their responsibilities.
The organization shall determine the required competence profiles for other roles according to their assigned responsibilities.

7.4 Competence Demonstration

Competence shall be demonstrated through an appropriate combination of objective evidence. Depending on the role and activity, acceptable evidence may include:

- formal education or academic qualifications;
- professional training;
- specialized AI or AI safety training;
- training in applicable standards;
- relevant professional experience;
- documented participation in comparable projects;
- examinations or knowledge assessments;
- practical demonstrations;
- supervised performance of assigned activities;
- peer or technical review;
- formal professional qualification or certification, where applicable; and
- documented authorization to perform specific activities.

The organization shall determine which forms of evidence are appropriate for each role.
Completion of a training course alone shall not automatically be considered sufficient evidence of competence for safety-critical activities where practical experience or demonstrated capability is additionally required.

7.5 Qualification and Authorization

Where an activity requires formal qualification or authorization, personnel shall be qualified and authorized before performing the activity independently.

The organization shall define the conditions under which personnel may:

- perform AI safety activities independently;
- perform activities under supervision;
- review or approve safety-related work products;
- perform independent verification or validation;
- conduct safety assessments or audits; and
- make safety-related release or acceptance decisions.

Authorization shall be based on demonstrated competence and shall be appropriate to the assigned responsibility and authority.
Where competence is not yet sufficient, personnel may perform defined activities under appropriate supervision. The required supervision and limitations shall be documented.

7.6 Initial and Continuing Competence

The organization shall ensure that personnel maintain their competence throughout the period in which they perform AI safety-related activities.
Continuing competence shall be considered particularly when:

- new AI technologies or methods are introduced;
- significant changes are made to AI development or deployment processes;
- new standards or regulatory requirements become applicable;
- personnel take on additional responsibilities;
- new tools or development environments are introduced;
- significant safety incidents or findings identify a competence gap; or
- personnel have not performed the relevant activity for an extended period.

Where necessary, additional training, supervised work, reassessment or requalification shall be performed.
The organization shall periodically review whether the competence of personnel remains appropriate for their assigned responsibilities.

7.7 Training and Development

The organization shall establish appropriate training and development activities to maintain and improve AI safety competence.
Training needs shall be identified based on role requirements, changes in technology and standards, identified competence gaps, audit findings, incidents, lessons learned and organizational changes.
Training may include internal training, external courses, workshops, technical reviews, practical exercises, mentoring, supervised project work or other appropriate methods.
Where training concerns a specific standard or technical method, the organization shall ensure that the training is sufficiently relevant to the activities actually performed by the personnel.

7.8 Competence Records

The organization shall maintain documented information demonstrating the competence, qualification and authorization of personnel performing AI safety-related activities.
Records shall include, as applicable:

- education and professional qualifications;
- relevant professional experience;
- completed training;
- examinations and assessments;
- practical competence evaluations;
- supervised work;
- role-specific qualifications;
- authorization records;
- requalification or competence reviews; and
- identified competence gaps and corresponding actions.

Records shall be maintained in accordance with the organization's document and record-control processes.
The records shall provide sufficient evidence to demonstrate that personnel performing safety-relevant activities were appropriately competent at the time the activity was performed.

7.9 External Personnel and Consultants

Where external personnel, consultants, contractors or suppliers perform AI safety-related activities, the organization shall ensure that their competence is appropriate to the assigned responsibilities.
Applicable competence and qualification requirements shall be defined in the relevant contractual or technical arrangements.
The organization shall obtain and retain appropriate evidence of competence where required to demonstrate that externally performed activities were conducted by suitably qualified personnel.
Use of external expertise shall not remove the organization's responsibility for ensuring that AI safety activities are adequately controlled and that the resulting evidence is suitable for its intended purpose.

7.10 Competence Gaps

Identified competence gaps shall be documented and appropriately addressed.
Where a competence gap could affect the validity or reliability of a safety-relevant activity, the organization shall determine whether additional training, supervision, independent review, reassessment or re-performance of the activity is necessary.
Competence gaps affecting safety-critical activities shall not be implicitly accepted through project schedule or resource constraints.
Where insufficient competence is identified after an activity has already been performed, the organization shall assess the potential impact on the validity of the resulting work product and associated safety evidence.

7.11 Independence and Competence

Competence and independence shall be considered together for activities requiring objective or independent judgment.
Personnel performing independent verification, validation, assessment, review or approval activities shall possess the competence necessary for the activity and shall have an appropriate degree of independence from the work being reviewed.
Assignment of an independent role to a person who does not possess the required technical competence shall not be considered sufficient to satisfy independence requirements.

Where organizational constraints require combined roles, the organization shall define appropriate compensating measures to maintain both competence and objectivity.

7.12 Management Responsibility

Management shall ensure that sufficient competent personnel are available to perform the AI safety activities required within the organization's scope.
Resource planning shall consider not only the number of personnel but also the required technical competence, experience, independence and availability at the relevant lifecycle stages.
Management shall periodically review competence availability as part of the AI Safety Management System and Management Review process.
Where required competence is unavailable internally, the organization may obtain appropriate external expertise, provided that the associated responsibilities, competence, interfaces and evidence are adequately controlled.

7.13 Competence Effectiveness

The effectiveness of competence development shall be evaluated where appropriate.
The organization shall not assess training effectiveness solely on the basis of attendance or completion. Where justified by the safety relevance of the activity, effectiveness may be demonstrated through examinations, practical exercises, supervised work, technical reviews, performance evaluation or other objective methods.
Lessons learned from audits, assessments, incidents, nonconformities and project experience shall be used to identify potential improvements to competence requirements and training programs.
The organization shall therefore maintain competence as a **controlled lifecycle capability**, rather than treating qualification as a one-time administrative activity.

8 Authority and Responsibility Matrix

8.1 General

The organization shall establish and maintain a role-based **Authority and Responsibility Matrix** defining the responsibilities, authorities and required involvement of the functions participating in AI safety activities.

The matrix shall provide a consistent basis for determining who is responsible for performing an activity, who has approval or decision authority, and which functions shall be consulted before a safety-relevant decision is made.

The matrix shall be established at organizational level and, where necessary, refined for individual AI projects through the applicable **AI Safety Management Plan** or project-specific responsibility matrix.

The assignment of authority shall be consistent with the organizational structure defined in Chapter 5 and the roles and responsibilities defined in Chapter 6.

No individual shall approve or release work for which the individual does not possess the required competence or formally assigned authority.

8.2 Authority and Responsibility Matrix

The following represents the generic MUNIK AI Safety authority and responsibility model. The actual assignment may be adapted to the organization's structure and project characteristics, provided that responsibilities, authority, competence and required independence remain effective.

Activity / Work Product	AI Safety Manager	ML Engineer	Data Engineer	V&V Engineer	FS Manager	SOTIF Responsible	Cybersecurity Responsible	System Engineer	Quality / Change
AI Safety Management Plan	**A**	C	C	C	C	C	C	C	C
AI Project Classification	**A/R**	C	C	C	C	C	C	R/C	C
Dataset Release	**A**	C	**R**	C	C	C	C	C	C
Model Release	**A**	**R**	C	**V**	C	C	C	C	C
V&V Strategy / Plan	**A**	C	C	**R**	C	C	C	C	C

Activity / Work Product	AI Safety Manager	ML Engineer	Data Engineer	V&V Engineer	FS Manager	SOTIF Responsible	Cybersecurity Responsible	System Engineer	Quality / Change
Safety Assessment	**A**	C	C	**R**	C	C	C	C	C
SOTIF Assessment	A/C	C	C	R/C	C	**R/A***	C	C	C
Cybersecurity Assessment Interface	A/C	C	C	C	C	C	**R/A***	C	C
Safety-Relevant Change Assessment	**A**	**R**	**R**	C	C	C	C	R	C
Change Approval	**A**	R	R	C	C	C	C	R	**C/R***
Safety Deviation / Waiver	**A**	C	C	C	C	C	C	C	C
Final Safety Release Recommendation	**A/R**	C	C	C	C	C	C	C	C
Final Product / AI Release Approval	**A***	C	C	C	C	C	C	C	C

Table 2: Authority and Responsibility Matrix

* Where formally assigned by the organization's governance model and applicable project requirements.

The matrix shall be interpreted together with the applicable project-specific responsibilities. An **A** designation does not necessarily mean that the person performs the technical activity. It identifies the role with the authority to accept, approve or make the corresponding organizational decision. An **R** designation identifies the role performing or leading the activity. **C** identifies functions whose technical expertise or review is required. **V** identifies an explicitly assigned verification role. Where a role is designated as both **R** and **A**, the organization shall verify that this combination does not compromise required independence or objective review.

8.3 Authority Principles

Authority shall be assigned in a manner that ensures that safety-relevant decisions are made by personnel having appropriate competence, responsibility and organizational authority.
In particular:

- responsibility for performing an activity shall be clearly assigned;
- approval authority shall be explicitly assigned;
- technical decisions shall be made by appropriately competent personnel;
- safety-relevant approval shall not be based solely on project management authority;
- independent verification or validation shall be maintained where required;
- unresolved safety concerns shall have a defined escalation path;
- safety deviations shall require documented evaluation and approval;
- changes affecting established safety evidence shall require appropriate impact assessment; and
- final release shall be based on defined release criteria and documented objective evidence.

The authority to approve a work product shall not automatically imply technical responsibility for producing that work product.

8.4 Safety Release Authority

The organization shall define the authority responsible for the final safety release decision.
The final safety release authority shall ensure that, before release:

- applicable safety requirements have been addressed;
- required verification and validation activities have been completed;
- relevant safety assessments have been performed;
- open safety issues have been evaluated;
- residual risks and known limitations have been documented and accepted in accordance with the applicable process;
- approved deviations or waivers are available where required;
- the released configuration is clearly identified;

- relevant Functional Safety, SOTIF and Cybersecurity interfaces have been addressed; and
- the required objective evidence is complete and traceable.

Where the AI Safety Manager is not the formal release authority, the AI Safety Manager shall nevertheless have the authority to escalate unresolved safety concerns to the designated release authority and management.
A release decision shall not be considered valid merely because the responsible technical team has completed development activities. The defined safety release criteria and approval process shall be fulfilled.

8.5 Separation of Conflicting Responsibilities

The organization shall identify and appropriately control conflicts between responsibilities.
Particular attention shall be given to combinations involving:

- development and independent verification;
- development and independent validation;
- assessment and approval;
- commercial responsibility and safety acceptance;
- change implementation and independent change review; and
- safety assessment and final release approval.

Where the organization permits combined roles because of its size or organizational structure, the combination shall be justified and appropriate compensating measures shall be established.
Compensating measures may include independent technical review, peer review, management review, review by another qualified organizational unit or involvement of an appropriately qualified external person.
The selected measures shall be documented where they are required to demonstrate the necessary level of independence and objectivity.

8.6 Delegation of Authority

Authority may be delegated only where the delegation is formally defined and the delegated person possesses the required competence.
Delegation shall not remove the organization's responsibility for ensuring that the applicable AI safety governance requirements are fulfilled.
Delegated authority shall define, as applicable:

- the scope of the authority;
- the applicable activities or decisions;
- the duration of the delegation;
- competence requirements;
- applicable limitations; and
- required records.

Temporary delegation shall be reviewed when the underlying organizational or project conditions change.

8.7 Project-Specific Responsibility Matrix

The generic matrix in this chapter shall be tailored where necessary to reflect the actual AI project organization.
The project-specific matrix shall identify, at minimum, the responsible and approval functions for safety-relevant activities and work products.
The project-specific allocation shall be documented in the AI Safety Management Plan or equivalent controlled project documentation.
Changes to project responsibilities, personnel or organizational structure shall be evaluated for their potential impact on competence, independence, authority and continuity of AI safety activities.

8.8 Evidence of Authority

The organization shall maintain objective evidence that the assigned responsibilities and authorities are formally established.
Evidence may include:

- organizational charts;
- role descriptions;
- appointment letters;
- responsibility matrices;
- project-specific AI Safety Management Plans;
- approval records;
- delegation records;
- review and assessment records; and
- release authorization records.

The evidence shall permit an assessor to determine who was responsible for an activity, who reviewed it, who approved it and under which authority the decision was made.

8.9 Review and Maintenance

The Authority and Responsibility Matrix shall be reviewed periodically and whenever significant organizational or project changes occur.
The review shall consider changes to:

- organizational structure;
- personnel or key role holders;
- AI technology or architecture;
- product scope;
- safety relevance;
- applicable standards or regulations;
- project lifecycle phase;
- external suppliers or service providers; and
- required independence.

The matrix shall remain consistent with the actual organization. Where a mismatch is identified, the organization shall correct the documented responsibilities and, where necessary, assess whether previously performed safety activities or decisions may have been affected.

9 AI Safety Lifecycle

9.1 General Lifecycle

AI safety activities shall cover the complete lifecycle of the AI system or AI-enabled product, from initial strategy and use-case definition through development, deployment, operation, change and eventual retirement.

The AI safety lifecycle shall be established as an integrated part of the overall product and engineering lifecycle. Safety activities shall be planned according to the characteristics and safety relevance of the AI system and shall be performed at the appropriate lifecycle stages.

The generic AI safety lifecycle is:

Strategy & Governance
↓
AI Use-Case Definition
↓
Safety Classification
↓
AI Safety Concept
↓
Data Definition & Management
↓
AI/ML Development
↓
Training
↓
Verification
↓
Validation
↓
System Integration
↓
Safety Assessment
↓
Release Approval
↓
Deployment
↓
Operation & Monitoring
↓
Change / Update
↓
Re-Verification / Re-Validation
↓
Re-Assessment / Re-Approval
↓
Deployment / Continued Operation
↓
Retirement

The lifecycle shall not be interpreted as a strictly linear sequence. Results obtained during verification, validation, operation, monitoring, incident analysis or safety assessment may require activities to return to an earlier lifecycle stage.
For example, identified limitations may require changes to the AI use case, data, model, architecture, requirements or safety concept. Such feedback shall be controlled through the applicable change-management process.

9.2 Strategy and Governance

AI safety activities shall begin during the strategy and governance phase.
The organization shall establish the governance framework necessary to determine how AI may be developed, acquired, integrated, deployed and operated within the organization's products and services.
At this stage, the organization shall determine, as applicable:

- the intended business and technical purpose of the AI system;
- applicable products and markets;
- relevant regulatory and contractual requirements;
- applicable safety, cybersecurity and quality frameworks;
- organizational responsibilities and authorities;
- required competence and resources;
- applicable AI safety processes; and
- initial safety-related constraints and assumptions.

The governance framework shall provide the basis for determining whether a proposed AI use case requires formal AI safety activities and to what extent.

9.3 AI Use-Case Definition

The intended AI use case shall be defined before safety requirements and risks can be adequately assessed.
The use-case definition shall describe, as applicable:

- intended functionality;
- intended users and operators;
- intended operating environment;
- relevant inputs and outputs;
- interfaces to other system functions;
- operating conditions and constraints;
- expected AI behavior;
- foreseeable misuse;
- dependencies on external systems or data;
- applicable performance expectations; and
- limitations known at the concept stage.

The intended use shall be sufficiently precise to support safety classification, risk analysis, requirements definition and subsequent verification and validation.
Where the intended use changes during development or operation, the change shall be assessed for its potential impact on AI safety.

9.4 Safety Classification

The organization shall determine the safety relevance of the AI system or AI-enabled function.
The classification shall consider, as applicable:

- the intended function;
- potential consequences of incorrect or unsafe AI behavior;
- degree of autonomy;
- interaction with safety-related functions;
- operating environment;
- foreseeable misuse;
- AI model characteristics and limitations;
- interfaces with Functional Safety, SOTIF and Cybersecurity;
- applicable regulatory requirements; and
- customer or contractual requirements.

The classification shall determine the rigor and extent of the AI safety activities, evidence, verification, validation, independence and assessment required.
The classification shall be documented and shall be reviewed when relevant system, product, intended-use, technology or operating conditions change.

9.5 AI Safety Concept

Based on the use case and safety classification, the organization shall establish an appropriate AI Safety Concept.
The AI Safety Concept shall define the high-level safety objectives, principles, constraints, assumptions and measures used to control AI-related safety risks.
Where applicable, the AI Safety Concept shall establish interfaces with:

- Functional Safety;
- SOTIF;
- Cybersecurity;
- Software Update Engineering;
- system engineering;
- operational safety; and
- regulatory or product compliance.

The AI Safety Concept shall provide a basis for deriving detailed safety-related requirements and for demonstrating that the implemented AI system remains consistent with the intended safety objectives.

9.6 Data Definition and Management

Safety-relevant data activities shall be planned and controlled as part of the AI lifecycle.
The organization shall define appropriate requirements for data acquisition, selection, preparation, labeling, quality, representativeness, provenance, versioning, storage and access.
Training, validation and test datasets shall be appropriately identified and controlled.
Data-related assumptions and limitations that may influence AI safety shall be documented and considered during verification, validation and safety assessment.

Changes to safety-relevant datasets shall be subject to configuration and change management and shall be assessed for their potential impact on previously established safety evidence.

9.7 AI/ML Development

AI/ML development shall be performed according to defined requirements, architecture, development methods and safety constraints.
The development process shall provide appropriate traceability between requirements, model architecture, implementation, configuration and resulting evidence.
Relevant model characteristics, assumptions, limitations, parameters and configurations shall be documented sufficiently to support reproducibility, verification, validation, change impact analysis and safety assessment.
Development activities shall remain under configuration and change control.

9.8 Training

Where machine learning involves training, the training process shall be controlled and reproducible to the extent necessary for the applicable safety objectives.
The organization shall control, as applicable:

- training datasets;
- preprocessing and data transformation;
- training algorithms;
- hyperparameters;
- model architecture;
- initialization and randomization;
- training environment;
- relevant software and tools;
- training configuration;
- resulting model version; and
- training results.

The organization shall maintain sufficient information to establish which training configuration and dataset resulted in a particular model version.
Where training variability can affect safety-relevant behavior, the organization shall define appropriate controls to assess and manage that variability.

9.9 Verification

Verification activities shall provide objective evidence that defined requirements have been correctly implemented.
Verification shall be planned according to the safety relevance and technical characteristics of the AI system.
Verification may include, as applicable:

- requirements verification;
- architecture verification;
- implementation verification;
- model verification;
- data and dataset verification;
- interface verification;
- robustness analysis;
- performance analysis;
- configuration verification;
- regression testing; and
- verification of safety mechanisms.

Verification results shall be documented and traceable to the applicable requirements, configuration and test environment.

9.10 Validation

Validation shall provide evidence that the AI system is suitable for its intended use under the defined operating conditions.

Validation shall consider the complete operational context rather than only the internal correctness of the AI model.
Depending on the application, validation may include:

- intended-use scenarios;
- relevant operational environments;
- scenario coverage;
- boundary conditions;
- foreseeable misuse;
- performance limitations;
- robustness;
- interaction with other system functions;
- human interaction;
- degraded or abnormal conditions; and
- relevant safety-related operating situations.

Validation shall consider limitations and uncertainties that may not be identified through conventional verification activities.

9.11 System Integration

The AI system shall be integrated into the target system using a controlled integration process.
Integration activities shall verify that:

- interfaces are correctly implemented;
- assumptions of use remain valid;
- system-level requirements are fulfilled;
- AI behavior remains within defined constraints;
- safety mechanisms operate as intended;
- interactions with other system elements do not introduce unacceptable risks; and
- the integrated configuration is correctly identified.

AI safety evidence shall be reviewed following integration to determine whether evidence generated at component or model level remains applicable at system level.

9.12 Safety Assessment

Before release, the organization shall perform an appropriate AI safety assessment.
The assessment shall determine whether the available objective evidence is sufficient to support the defined safety claims and whether applicable safety requirements and acceptance criteria have been fulfilled.
The assessment shall consider, as applicable:

- safety requirements;
- risk analysis;
- safety concept;
- AI model and data evidence;
- verification and validation results;
- system integration evidence;
- known limitations;
- assumptions;
- residual risks;
- deviations and waivers;
- Functional Safety interfaces;
- SOTIF interfaces;
- Cybersecurity interfaces;
- configuration status; and
- applicable regulatory or certification requirements.

The assessment shall result in a documented conclusion regarding the safety status of the AI system within the defined scope.

9.13 Release Approval

AI systems shall not be released for their intended operational use until the applicable release criteria have been fulfilled and the required approval has been obtained.
The release decision shall be based on the approved configuration and the available objective evidence.
The release package shall identify, as applicable:

- approved AI model version;
- relevant dataset versions;
- software and hardware configuration;
- applicable requirements;
- verification and validation status;
- safety assessment status;
- known limitations and assumptions;
- approved deviations or residual risks;
- applicable operating restrictions; and
- required monitoring or post-release measures.

The release decision shall be documented and traceable to the responsible approval authority.

9.14 Deployment

Deployment shall be controlled to ensure that the released AI configuration is transferred to the intended operational environment without unauthorized or uncontrolled changes.
The organization shall verify, as applicable:

- identity and integrity of the released configuration;
- deployment configuration;
- applicable interfaces;
- operating environment;
- security controls;
- safety-relevant parameters;

- monitoring functions; and
- applicable operational restrictions.

Where deployment conditions differ materially from those considered during validation, the impact shall be assessed before deployment.

9.15 Operation and Monitoring

AI safety responsibilities shall continue after deployment.
The organization shall establish appropriate monitoring and operational assurance activities based on the safety relevance and characteristics of the AI system.
Monitoring may include, as applicable:

- AI performance;
- safety-relevant behavior;
- operational boundary conditions;
- data drift;
- performance degradation;
- unexpected behavior;
- safety incidents;
- near misses;
- customer or operator reports;
- cybersecurity events;
- changes in the operational environment; and
- effectiveness of safety mechanisms.

Operational information shall be evaluated to determine whether existing safety assumptions, requirements, validation evidence or risk assessments remain valid.
Safety-relevant incidents or observations shall be subject to the applicable incident and corrective-action processes.

9.16 Change and Update

Changes to the AI system shall be controlled throughout its operational lifecycle.
Safety-relevant changes may include changes to:

- AI models;
- algorithms;
- training data;
- validation or test data;
- software;
- hardware;
- model parameters;
- configuration;
- interfaces;
- operating conditions;
- deployment environment;
- intended use; or
- applicable safety requirements.

Before implementing a safety-relevant change, the organization shall perform an appropriate change impact analysis.
The analysis shall determine whether existing verification, validation, safety assessment and approval evidence remains valid or whether additional activities are required.

9.17 Re-Verification, Re-Validation and Re-Assessment

Where a change may affect AI safety, the organization shall perform the level of re-verification and re-validation necessary to demonstrate continued compliance with applicable requirements.
The extent of re-verification and re-validation shall be based on the change impact and safety relevance. A change shall not automatically require complete repetition of all lifecycle activities where a documented analysis demonstrates that existing evidence remains valid.
Where the change affects the safety argument, assumptions, residual risks or applicable requirements, an updated safety assessment shall be performed.

Where required, the changed system shall undergo a new release approval before being deployed.

9.18 Retirement

The organization shall define controlled retirement activities for AI systems that are removed from operational use.
Retirement shall consider, as applicable:

- controlled removal or deactivation;
- termination of operational interfaces;
- retention of required safety records;
- preservation of relevant configuration and evidence;
- handling of remaining safety risks;
- cybersecurity implications;
- contractual or regulatory retention requirements; and
- lessons learned.

Retirement records shall provide sufficient traceability to establish the final operational status of the AI system.

9.19 Lifecycle Feedback and Continual Improvement

Information generated at any stage of the AI lifecycle shall be capable of influencing earlier lifecycle activities where necessary.
Examples include:

- validation findings leading to changes in requirements;
- operational monitoring identifying previously unknown limitations;
- incidents leading to changes in the safety concept;
- data drift leading to additional data requirements;
- cybersecurity events leading to changes in AI safety controls;
- model changes leading to renewed SOTIF analysis; and
- new standards or regulatory requirements leading to reassessment.

Such feedback shall be managed through the applicable change, risk and continual-improvement processes.

The AI Safety Lifecycle shall therefore operate as a **controlled lifecycle with feedback loops**, rather than as a one-directional development sequence.

10 Lifecycle Entry and Exit Criteria

Lifecycle entry and exit criteria define the technical and process conditions for completion or initiation of a lifecycle phase. Formal AI Safety Gates defined in Chapter 33 provide the management decision mechanism for authorizing progression between defined lifecycle stages.

10.1 General

Each AI safety lifecycle phase shall have defined entry and exit criteria appropriate to the safety relevance, complexity and maturity of the AI system.

Entry criteria shall define the conditions that shall be fulfilled before an activity or lifecycle phase may formally begin. Exit criteria shall define the conditions that shall be fulfilled before the phase may be considered complete and the AI system may proceed to the next applicable lifecycle phase.

The criteria shall be defined in sufficient detail to provide objective evidence that the required activities have been completed and that the relevant outputs are available, reviewed and approved where required.

The applicable entry and exit criteria shall be documented in the **AI Safety Management Plan**, project plan, lifecycle procedure or equivalent controlled documentation.

No lifecycle phase shall be formally closed where mandatory exit criteria remain unfulfilled unless the unresolved item has been evaluated and an explicitly approved deviation or waiver permits continuation.

An approved deviation shall not automatically mean that the underlying requirement has been fulfilled. The deviation shall define the applicable justification, conditions, responsibilities, compensating measures and acceptance authority.

10.2 Lifecycle Gate Concept

Lifecycle entry and exit criteria shall be used as formal control gates.

A lifecycle gate shall provide sufficient evidence that:

- the required preceding activities have been completed;
- required work products are available and under configuration control;
- applicable reviews have been performed;
- identified safety risks have been evaluated;
- open issues have been assessed;
- required verification and validation activities have been completed or appropriately planned;
- applicable assumptions remain valid;
- required interfaces with Functional Safety, SOTIF and Cybersecurity have been addressed; and
- the responsible authority has approved continuation where approval is required.

The rigor of a lifecycle gate shall be proportionate to the safety classification and project-specific requirements.

10.3 Example Entry and Exit Criteria

The following table provides a generic example. The actual criteria shall be tailored to the AI system and documented in the applicable project-level safety plan.

Phase	Entry Criteria	Exit Criteria
AI Concept	AI use case and intended purpose identified	Use case defined; initial safety relevance determined; applicable stakeholders identified
Safety Classification	Use case and system context available	AI safety classification completed and approved
AI Safety Concept	Safety classification available	Safety objectives, safety strategy, assumptions and high-level safety measures defined and approved
Data Development	Data requirements and intended use defined	Applicable dataset requirements fulfilled; dataset baseline established; quality and limitations assessed
Model Development	Approved requirements and applicable data baseline available	Model baseline established; relevant configuration documented; development evidence available

Phase	Entry Criteria	Exit Criteria
Verification	Controlled model baseline available; verification criteria defined	Required verification activities completed; acceptance criteria fulfilled; verification evidence reviewed
Validation	Verified AI baseline available; validation criteria and operational context defined	Validation acceptance criteria fulfilled; limitations and relevant scenarios evaluated
System Integration	Validated AI component available; integration requirements defined	Integration completed; interface and system-level evidence available; integrated configuration controlled
Safety Assessment	Required safety evidence available	Safety assessment completed; open issues and residual risks evaluated; assessment conclusion documented
Release Approval	Safety assessment completed; open issues evaluated	Release criteria fulfilled; required approvals obtained; released configuration uniquely identified
Deployment	Approved release available	Deployment completed using controlled configuration; deployment integrity and applicable operating conditions confirmed
Operation & Monitoring	Approved deployment available; monitoring measures established	Monitoring operational; safety-relevant observations evaluated and managed
Change / Update	Approved change request and applicable impact analysis initiated	Required change activities completed; re-verification/re-validation and re-assessment performed as applicable; updated release approved
Retirement	Retirement decision authorized	AI system removed or deactivated as defined; required records retained; residual obligations addressed

Table 3: Example Entry and Exit Criteria

The table shall not be interpreted as requiring identical activities for every AI system. Activities may be combined, repeated or tailored where justified by the project-specific safety classification and documented requirements.

10.4 Gate Review and Approval

Where a formal lifecycle gate is required, the organization shall perform a documented gate review.

The review shall determine whether the defined entry and exit criteria have been fulfilled and whether the AI system is authorized to proceed to the next lifecycle phase.
The gate review shall identify, as applicable:

- completed activities;
- outstanding activities;
- open issues;
- deviations and waivers;
- residual risks;
- assumptions requiring confirmation;
- required corrective actions; and
- approval status.

Gate decisions shall be documented and traceable to the responsible authority.

10.5 Open Issues and Deviations

Open issues shall be classified according to their potential impact on AI safety.
Issues that prevent fulfillment of mandatory exit criteria shall not be treated as closed merely because the project is required to proceed.
Where continuation is permitted through a formal deviation or waiver, the deviation shall identify:

- the affected requirement or criterion;
- the reason for the deviation;
- the associated safety impact;
- applicable compensating measures;
- responsible owner;
- acceptance authority;
- validity or expiration conditions; and
- required follow-up actions.

Deviations shall remain subject to monitoring until formally closed or otherwise dispositioned.

10.6 Traceability of Lifecycle Gates

The organization shall maintain traceability between lifecycle phases, entry and exit criteria, work products, reviews and approval decisions.
The evidence shall allow an assessor to determine:

- which criteria were applicable;
- which evidence demonstrated fulfillment;
- who performed the activity;
- who reviewed or verified the result;
- who approved continuation; and
- which deviations or open issues were accepted.

Lifecycle gate records shall be maintained as part of the AI safety evidence structure.

10.7 Repetition of Lifecycle Phases

AI safety lifecycle phases may be repeated when new information, changes, incidents, validation findings or operational experience indicate that previously completed activities are no longer sufficient.
A return to an earlier lifecycle phase shall be controlled through the applicable change-management and safety-management processes.
The organization shall determine the scope of repeated activities based on documented impact analysis rather than assuming that all previous evidence remains valid or, conversely, that all activities must automatically be repeated in full.

11 Interface with ISO 26262 Functional Safety

Where an AI component contributes to a safety-related function, the organization shall define and maintain its interface to the applicable ISO 26262 functional safety lifecycle.
The purpose of this interface is to ensure that the contribution of the AI component to functional safety is clearly identified, appropriately integrated into the functional safety lifecycle, and supported by sufficient objective evidence. The interface shall define the relationship between AI safety

activities and the applicable functional safety activities and shall ensure that responsibilities, requirements, assumptions, interfaces and evidence remain traceable.

11.1 General Interface Principles

The organization shall determine, for each applicable project, how the AI component interacts with the ISO 26262 safety lifecycle. The extent and depth of the interface shall be appropriate to the role of the AI component, its safety relevance, its architectural position, its intended use and the applicable safety requirements.

The interface shall be established early enough to ensure that AI-related safety considerations are incorporated into the functional safety concept and subsequent lifecycle activities.

AI safety activities shall complement, and shall not automatically replace, applicable ISO 26262 activities. Compliance with an AI-specific safety process or standard shall not, by itself, constitute evidence of compliance with applicable ISO 26262 requirements.

Where an AI component is not itself safety-related but contributes to a function that has safety relevance, the organization shall nevertheless determine whether its behavior, limitations, interfaces or failure modes can affect the fulfillment of functional safety requirements.

11.2 Item Definition and AI Component Context

Where applicable, the AI component shall be identified within the item definition and its role within the overall vehicle or system architecture shall be documented.

The organization shall determine, as applicable:

- the intended function of the AI component;
- its relationship to the item and the relevant vehicle or system functions;
- its inputs, outputs and interfaces;
- its operational environment and operating conditions;
- its dependencies on other hardware, software, data or system elements;
- its safety-related responsibilities;
- its limitations and assumptions of use.

Where the AI component is supplied as a component, library, IP, semiconductor function or SEooC, the information required for its integration into the receiving system shall be identified and documented.

11.3 Safety Goals and Safety Requirements

The organization shall establish the relationship between applicable ISO 26262 safety goals and the AI-related safety requirements.
Where safety requirements are allocated to an AI component, the allocation shall be documented and shall identify the responsibility of the AI component for contributing to the fulfillment of the respective requirement.
The relationship between:
Safety Goal → Functional Safety Requirement → Technical Safety Requirement → AI Safety Requirement → AI implementation or safety measure → Verification/Validation Evidence
shall be traceable where applicable.
AI-specific requirements may be derived where the characteristics of AI technology require additional controls or constraints. Such requirements shall be managed consistently with the applicable safety lifecycle.

11.4 Allocation of Requirements to AI Components

The organization shall define which functional and technical safety requirements are allocated to the AI component and which requirements remain the responsibility of other system elements.
The allocation shall consider the complete AI-related contribution, including, where applicable:

- AI model behavior;
- training and inference software;
- data and datasets;
- model parameters and configuration;
- AI hardware or acceleration functions;
- input processing;
- output processing;
- monitoring and diagnostic mechanisms;

- fallback or degradation mechanisms;
- interfaces to conventional safety mechanisms.

The allocation shall avoid ambiguity concerning responsibility for the fulfillment of safety requirements.

11.5 Assumptions of Use

Where the safety contribution of the AI component depends on assumptions concerning its operating environment or its integration into the receiving system, such assumptions shall be explicitly identified and documented.
Such assumptions may concern, as applicable:

- sensor characteristics and input quality;
- operating conditions;
- environmental conditions;
- expected input ranges;
- system timing and resource availability;
- surrounding software and hardware;
- monitoring mechanisms;
- fallback behavior;
- human interaction;
- cybersecurity protection;
- system integration constraints.

Assumptions that are necessary for the safe use of the AI component shall be communicated to the relevant receiving organization or system development function and shall be subject to appropriate verification during integration.

11.6 Safety Mechanisms and Interfaces

The organization shall identify safety mechanisms that contribute to controlling risks associated with the AI component.

Depending on the architecture and safety concept, such mechanisms may include:

- plausibility checks;
- input validation;
- output monitoring;
- confidence or uncertainty monitoring;
- independent monitoring functions;
- redundancy or diversity;
- fault detection;
- degraded operation;
- fallback mechanisms;
- runtime supervision;
- safety-related constraints on AI outputs.

The interaction between AI-specific safety mechanisms and conventional ISO 26262 safety mechanisms shall be defined and verified as applicable.
Hardware, software, communication and system interfaces relevant to the safety contribution of the AI component shall be controlled and traceable.

11.7 Verification and Validation Evidence

The organization shall define the verification and validation evidence required to demonstrate that the AI component fulfills its allocated safety-related requirements.
The evidence shall be appropriate to the characteristics and risks of the AI component and may include, as applicable:

- requirements verification;
- model verification;
- software verification;
- data-related verification;
- robustness testing;
- scenario-based testing;
- performance evaluation;

- fault injection;
- interface verification;
- integration testing;
- validation under representative operating conditions.

AI-specific verification and validation evidence shall be linked to the applicable safety requirements and shall be maintained as controlled objective evidence.

11.8 Integration Evidence

Where the AI component is integrated into a larger safety-related system, the organization shall ensure that evidence is available demonstrating that the integrated AI component performs its allocated safety-related function as intended.
Integration activities shall consider, as applicable:

- interfaces;
- timing;
- data flow;
- system assumptions;
- resource constraints;
- safety mechanisms;
- monitoring;
- fallback behavior;
- interactions with other system functions;
- configuration and version compatibility.

Evidence generated at the AI component level shall not automatically be considered sufficient evidence for system-level functional safety. The organization shall determine what additional integration and system-level evidence is required.

11.9 Confirmation Measures and Independence

Where ISO 26262 confirmation measures are applicable, the organization shall determine how AI-related activities and evidence are included within those measures.
The required degree of independence shall be determined according to the applicable safety lifecycle, safety relevance and organizational responsibilities.
AI development, verification, assessment and approval responsibilities shall be assigned such that conflicts of interest are appropriately controlled and the required independence is maintained.

11.10 Safety Case and Traceability

Where a safety case is established, the contribution of the AI component shall be incorporated into the applicable safety argument and supported by objective evidence.
The organization shall maintain traceability between relevant:

- safety goals;
- functional and technical safety requirements;
- AI safety requirements;
- assumptions and dependencies;
- AI design and implementation;
- safety mechanisms;
- verification and validation activities;
- integration evidence;
- assessment results;
- residual risks and limitations.

Traceability shall be maintained throughout the applicable lifecycle and shall support assessment, release and subsequent change activities.

11.11 AI Component as SEooC

Where an AI component is developed as a Safety Element out of Context (SEooC), the organization shall define and document the assumptions, requirements, safety measures, interfaces and evidence necessary for its intended integration.

The SEooC safety evidence shall clearly identify the conditions under which the evidence is valid.
The receiving organization shall determine whether the assumptions and evidence of the AI component are fulfilled in the target application. Where assumptions are not fulfilled, an appropriate impact analysis and further safety activities shall be performed.

11.12 Changes and Impact Analysis

Changes to an AI component may affect functional safety even where conventional software or hardware interfaces remain unchanged.
Therefore, changes to the following shall be subject to appropriate impact analysis where applicable:

- AI model;
- model architecture;
- model parameters;
- training data;
- data preprocessing;
- training method;
- AI software;
- inference environment;
- configuration;
- safety mechanisms;
- operating conditions;
- system interfaces.

The impact analysis shall determine whether additional verification, validation, integration testing, safety assessment or confirmation measures are required.
Where the change affects the validity of existing safety evidence, the affected evidence shall be re-established or supplemented before the changed AI component is released for safety-related use.

11.13 Responsibility and Interface Management

Responsibilities for the interface between AI safety and functional safety shall be clearly assigned.

The AI Safety Manager, Functional Safety Manager, System Engineering, AI/ML Development, V&V and other relevant functions shall establish appropriate interfaces for planning, requirement allocation, evidence generation, review, assessment and approval.
Responsibilities shall be documented in the applicable AI Safety Management Plan, Functional Safety Plan, project responsibility matrix or equivalent controlled documentation.

11.14 Objective Evidence and Records

The organization shall maintain objective evidence demonstrating that the defined interface with ISO 26262 has been implemented.
Records shall include, as applicable:

- applicable safety requirements and their allocation;
- AI safety requirements;
- assumptions of use;
- interface specifications;
- safety mechanisms;
- verification and validation evidence;
- integration evidence;
- confirmation measure results;
- safety assessment results;
- safety case contributions;
- change-impact analyses;
- residual risks and limitations;
- relevant approvals and release decisions.

The records shall be maintained under the applicable configuration and document control processes.

11.15 Interface Review

The interface between AI safety and functional safety shall be reviewed at appropriate lifecycle gates and following significant changes to the AI component, system architecture, intended use, operating environment, safety requirements or applicable standards.

The review shall determine whether the defined interface, assumptions, requirements, evidence and responsibilities remain valid.

Any identified gap shall be documented, evaluated and addressed through the applicable corrective action, deviation, change or risk management process.

12 Interface with ISO/SAE 21434 Cybersecurity

AI safety and cybersecurity shall be coordinated where cybersecurity threats or compromises may affect the safety of an AI-enabled system.

The organization shall define and maintain the interface between the AI Safety Management System and the applicable ISO/SAE 21434 Cybersecurity Management System (CSMS). The interface shall ensure that cybersecurity risks capable of affecting AI safety are identified, communicated, assessed and appropriately controlled throughout the applicable lifecycle.

12.1 General Interface Principles

The organization shall determine how cybersecurity activities interact with AI safety activities based on the intended use, architecture, safety relevance, connectivity, threat exposure and potential consequences of compromise.

Cybersecurity activities shall address threats that could result in unsafe AI behavior, degradation of safety functions, loss of safety-related evidence or unauthorized changes to safety-relevant AI assets.

AI safety and cybersecurity responsibilities shall remain clearly defined. Cybersecurity risk management shall not automatically replace AI safety risk management, and AI safety activities shall not automatically replace cybersecurity activities.

Where a cybersecurity threat can affect a safety-related function, the relationship between the cybersecurity risk and the potential safety impact shall be documented and managed through the appropriate interfaces.

12.2 Safety-Relevant AI Assets

Relevant AI assets shall be identified and protected according to their cybersecurity and safety relevance.
Such assets may include, as applicable:

- training data;
- datasets and data repositories;
- data-labeling information;
- AI models;
- model parameters and weights;
- AI software and libraries;
- AI configuration;
- model and software artifacts;
- deployment packages;
- update packages;
- AI infrastructure;
- development and training environments;
- test and validation environments;
- model evaluation results;
- safety-relevant AI documentation and evidence;
- credentials, keys and access mechanisms protecting AI assets.

The organization shall determine which assets are safety-relevant and shall ensure that the applicable cybersecurity controls are consistent with their protection requirements.

12.3 Cybersecurity Threats Affecting AI Safety

The organization shall consider cybersecurity threats that could cause or contribute to unsafe AI behavior.
Examples include, as applicable:

- malicious manipulation of training data;

- unauthorized modification of datasets;
- model manipulation;
- unauthorized model replacement;
- malicious modification of model parameters;
- adversarial inputs;
- manipulation of AI inputs or sensor-related data;
- compromised development or training environments;
- compromised deployment mechanisms;
- compromised update mechanisms;
- unauthorized access to safety-relevant AI assets;
- unauthorized modification of AI configuration;
- insertion or modification of malicious software or AI components;
- compromise of interfaces between AI components and other system elements.

The organization shall consider both direct and indirect effects of such threats on safety-related functions.

12.4 Interface with Cybersecurity Risk Management

Where applicable, cybersecurity risks identified through the ISO/SAE 21434 cybersecurity lifecycle shall be reviewed for their potential impact on AI safety.
The interface shall provide for the communication and traceability of relevant information, including:

- cybersecurity threats;
- attack paths and attack scenarios;
- affected AI assets;
- cybersecurity controls;
- residual cybersecurity risks;
- potential safety consequences;
- safety-related assumptions and dependencies;
- required monitoring or detection mechanisms.

Where a cybersecurity risk may result in an unsafe condition, the corresponding safety implications shall be incorporated into the applicable AI safety risk management process.
Similarly, safety-relevant AI assets and safety assumptions shall be communicated to the cybersecurity function where they are relevant to cybersecurity risk assessment and treatment.

12.5 Protection of AI Development and Training

Where AI safety depends on the integrity of development, training or validation activities, cybersecurity controls shall protect the relevant environments and information.
This may include protection against:

- unauthorized access;
- unauthorized modification;
- malicious software;
- compromised development tools;
- compromised training infrastructure;
- unauthorized changes to training data;
- unauthorized changes to model configurations;
- unauthorized replacement of approved models or software;
- loss or corruption of safety-relevant evidence.

The organization shall ensure that the integrity and provenance of safety-relevant AI artifacts can be demonstrated where required.

12.6 Model and Data Integrity

The integrity of AI models, datasets, model parameters and configuration shall be protected according to their safety and cybersecurity relevance.
Where appropriate, mechanisms shall be established to detect unauthorized or unintended modification.
The organization shall maintain sufficient version and configuration information to determine which model, dataset, software version and configuration were used for a particular safety-relevant release.

Where model or data integrity cannot be demonstrated, the organization shall determine the potential safety impact and whether the affected AI component remains suitable for safety-related use.

12.7 Adversarial Inputs and AI Runtime Security

Where applicable, the organization shall consider cybersecurity threats involving manipulated or intentionally crafted inputs that may cause unsafe AI behavior.
The assessment shall consider, as appropriate:

- manipulation of sensor or input data;
- adversarial examples;
- spoofing;
- abnormal or malicious input patterns;
- exploitation of AI interfaces;
- denial or degradation of AI functions;
- manipulation of supporting information used by the AI system.

Where such threats may affect safety, appropriate prevention, detection, mitigation or fallback mechanisms shall be defined and verified as applicable.

12.8 Secure Deployment and Updates

Deployment and update mechanisms for safety-relevant AI components shall be subject to appropriate cybersecurity controls.
The organization shall ensure, as applicable, that:

- only authorized AI artifacts are deployed;
- the integrity and authenticity of deployment packages can be verified;
- unauthorized model replacement is prevented or detected;
- update packages are appropriately protected;
- the identity and version of deployed AI artifacts can be established;
- failed or compromised updates are appropriately detected and managed;

- safety-relevant assumptions and validation status are not invalidated by unauthorized changes.

The interface with the applicable software update lifecycle shall be maintained where AI models, data, software or configuration can be updated after deployment.

12.9 Access Control and Protection of Safety-Relevant Assets

Access to safety-relevant AI assets shall be controlled according to defined authorization requirements.
The organization shall determine appropriate controls for access to, as applicable:

- training and validation datasets;
- AI models and model parameters;
- development and training environments;
- source code;
- deployment infrastructure;
- update mechanisms;
- configuration information;
- safety-relevant evidence.

Unauthorized access, modification or deletion of safety-relevant AI assets shall be treated as a potential safety-relevant event where such actions may affect the validity or behavior of the AI component.

12.10 Cybersecurity and Safety Verification

Where cybersecurity controls are intended to prevent or mitigate safety-relevant AI threats, their effectiveness shall be appropriately verified.
Verification may address, as applicable:

- protection of AI assets;
- access controls;

- model integrity;
- data integrity;
- secure deployment;
- secure updates;
- detection of unauthorized changes;
- adversarial robustness;
- monitoring and detection mechanisms;
- fallback or mitigation mechanisms.

The resulting evidence shall be linked to the applicable cybersecurity requirements and, where relevant, to AI safety requirements and safety risks.

12.11 Cybersecurity Events and Safety Impact

Cybersecurity incidents affecting AI assets shall be evaluated for potential safety consequences. Examples include:

- unauthorized model modification;
- compromised training data;
- unauthorized software or configuration changes;
- compromised update packages;
- compromise of AI infrastructure;
- successful attacks affecting AI inputs or outputs;
- loss of integrity of safety-relevant evidence.

Where an incident may affect the safety of an AI-enabled system, the applicable incident management, AI safety risk management, change management and release processes shall be initiated.

The organization shall determine whether re-verification, re-validation, safety assessment, cybersecurity assessment or re-approval is required before the affected AI component is returned to safety-related operation.

12.12 Interface with AI Safety Risk Management

Cybersecurity threats with potential safety consequences shall be incorporated into the applicable AI safety risk management process.
The relationship shall be traceable, where applicable, between:
Cybersecurity Threat → Affected AI Asset → Cybersecurity Risk → Potential Safety Impact → Safety Risk → Control Measures → Verification Evidence → Residual Risk
The organization shall ensure that cybersecurity controls do not introduce new unacceptable safety risks and that safety-related assumptions concerning cybersecurity protection are explicitly documented.

12.13 Responsibility and Interface Management

Responsibilities for the interface between AI safety and cybersecurity shall be clearly assigned.
The AI Safety Manager and Cybersecurity Responsible function shall establish appropriate interfaces for:

- identification of safety-relevant AI assets;
- cybersecurity threat and risk information;
- assessment of potential safety consequences;
- definition and allocation of controls;
- verification and validation;
- incident handling;
- change and update management;
- safety and cybersecurity assessments;
- release and re-approval decisions.

Responsibilities shall be documented in the applicable AI Safety Management Plan, Cybersecurity Plan, project responsibility matrix or equivalent controlled documentation.

12.14 Objective Evidence and Records

The organization shall maintain objective evidence demonstrating that the defined interface between AI safety and cybersecurity has been implemented.
Records shall include, as applicable:

- identification of safety-relevant AI assets;
- cybersecurity threats and risks;
- identified safety impacts;
- cybersecurity controls;
- AI safety risk assessments;
- assumptions and dependencies;
- verification and validation evidence;
- incident records;
- change-impact analyses;
- assessment results;
- residual risks;
- relevant approvals and release decisions.

Records shall be maintained under the applicable configuration, document and information security controls.

12.15 Interface Review

The interface between AI safety and cybersecurity shall be reviewed at appropriate lifecycle gates and following significant changes to the AI component, system architecture, cybersecurity environment, threat landscape, intended use or operating conditions.
The review shall determine whether the identified AI assets, cybersecurity threats, safety impacts, controls, assumptions and evidence remain valid.
Identified gaps shall be documented, evaluated and addressed through the applicable risk management, corrective action, deviation, change or improvement process.

13 Interface with ISO 21448 SOTIF

Where AI functionality contributes to the intended functionality of a system, the organization shall define and maintain its interface to the applicable ISO 21448 SOTIF lifecycle.
The purpose of this interface is to ensure that safety risks arising from limitations of the intended functionality, including limitations of AI-based perception, decision-making and performance, are systematically identified, analysed, controlled and validated.

13.1 General Interface Principles

The organization shall determine how AI safety activities interact with SOTIF activities based on the intended functionality, the role of the AI component, its safety relevance, the operational environment and the potential consequences of insufficient performance.
Particular attention shall be given to safety risks that cannot be adequately addressed through conventional fault-based safety analysis.
AI safety activities shall complement, and shall not automatically replace, applicable SOTIF activities. Similarly, SOTIF activities shall not automatically replace AI-specific analysis of model behavior, data limitations, uncertainty, robustness or scenario coverage.
The relationship between AI safety requirements, SOTIF objectives, identified limitations, mitigation measures and validation evidence shall be appropriately traceable.

13.2 Intended Functionality

The organization shall establish a clear understanding of the intended functionality supported by the AI component and its role within the overall system.
The intended functionality shall address, as applicable:

- the purpose and expected behavior of the AI function;
- relevant inputs and outputs;
- expected operating conditions;
- environmental assumptions;
- applicable performance expectations;
- limitations of the AI functionality;

- interactions with other system functions;
- conditions under which the intended functionality may become insufficient.

The intended functionality shall provide an appropriate basis for identifying potential SOTIF-related hazards and limitations.

13.3 Functional Insufficiencies

The organization shall identify and evaluate functional insufficiencies that may arise from the intended functionality of the AI component, including situations in which the AI system operates according to its design but its performance is insufficient for the intended use.
Functional insufficiencies may relate, as applicable, to:

- perception accuracy;
- object or event classification;
- localization;
- prediction;
- decision-making;
- planning;
- control;
- environmental interpretation;
- confidence or uncertainty;
- temporal behavior;
- interaction with other functions.

The organization shall determine whether identified insufficiencies can contribute to hazardous behavior under relevant operating conditions.

13.4 AI Perception and Performance Limitations

Where AI functionality depends on perception or classification, the organization shall consider limitations associated with the AI model, input data and operating environment.
Such limitations may include:

- insufficient recognition performance;
- ambiguous or incomplete inputs;
- unusual environmental conditions;
- rare or poorly represented objects or events;
- occlusion;
- adverse weather or lighting;
- sensor limitations;
- distribution changes;
- insufficient generalization;
- model uncertainty;
- performance degradation outside representative conditions.

Relevant limitations shall be documented and considered in the applicable SOTIF analysis and validation activities.

13.5 Foreseeable Misuse

The organization shall consider reasonably foreseeable misuse of the AI-enabled functionality where such misuse may contribute to unsafe behavior.
Foreseeable misuse may include, as applicable:

- use outside intended operating conditions;
- inappropriate user interaction;
- incorrect interpretation of system capabilities;
- use under conditions not adequately represented during development;
- inappropriate reliance on AI outputs;
- interaction with other system functions in an unintended manner.

Where foreseeable misuse is relevant to safety, appropriate preventive, mitigating or monitoring measures shall be defined and evaluated.

13.6 Operational Design Domain

Where an Operational Design Domain (ODD) is applicable, the organization shall define the conditions within which the AI-enabled functionality is intended to operate.
The ODD shall be considered in relation to, as applicable:

- geographic conditions;
- road or infrastructure characteristics;
- traffic conditions;
- environmental conditions;
- weather and lighting;
- vehicle operating conditions;
- speed ranges;
- sensor availability and performance;
- connectivity or infrastructure dependencies;
- system configuration.

The organization shall establish appropriate mechanisms for determining whether the current operating conditions remain within the applicable ODD.
Where the AI system approaches or exceeds relevant ODD boundaries, appropriate warnings, degradation, fallback or transition mechanisms shall be considered as applicable.

13.7 Scenario Identification and Coverage

The organization shall establish a systematic approach for identifying and evaluating scenarios relevant to the safety of the AI-enabled functionality.
Scenario analysis shall consider, as applicable:

- normal operating scenarios;
- challenging scenarios;
- rare scenarios;
- boundary conditions;
- combinations of environmental conditions;

- interactions with other road users or system functions;
- foreseeable misuse scenarios;
- scenarios associated with functional insufficiencies;
- scenarios identified through incidents, field data or operational monitoring;
- scenarios associated with changes in data distribution or operating conditions.

Scenario coverage shall be evaluated against the intended functionality, identified risks and applicable safety objectives.
High test coverage or statistical performance alone shall not automatically be considered sufficient evidence that unknown or insufficiently covered unsafe scenarios have been adequately addressed.

13.8 Unknown Unsafe Scenarios

The organization shall consider the potential for unsafe scenarios that were not identified during the initial development and analysis activities.
Where applicable, activities shall support the identification of previously unknown or insufficiently understood unsafe scenarios through:

- scenario generation;
- simulation;
- testing;
- edge-case analysis;
- operational monitoring;
- incident analysis;
- field data;
- expert review;
- analysis of model limitations;
- analysis of changes in the operational environment.

Newly identified unsafe scenarios shall be evaluated and, where necessary, incorporated into the applicable risk management, validation, monitoring and change processes.

13.9 AI Validation

AI validation shall provide evidence that the AI-enabled functionality performs safely within its intended operational context.
Validation shall consider, as applicable:

- representative operating conditions;
- relevant scenarios;
- boundary conditions;
- rare and challenging scenarios;
- functional insufficiencies;
- foreseeable misuse;
- ODD limitations;
- interactions with other system functions;
- AI model limitations;
- uncertainty and robustness;
- performance degradation;
- safety-related acceptance criteria.

Validation shall be based on defined acceptance criteria and controlled configurations of the AI model, data, software and system environment.

13.10 Simulation, Testing and Scenario-Based Evidence

Where simulation or scenario-based testing is used to support SOTIF activities, the organization shall establish appropriate controls for the validity and representativeness of the scenarios and simulation environment.
The organization shall consider, as applicable:

- scenario relevance;
- parameter ranges;
- environmental representation;
- model assumptions;

- simulation limitations;
- correlation between simulated and real-world behavior;
- scenario completeness;
- configuration and version control.

The limitations of simulation and scenario-based evidence shall be documented and considered when determining the overall sufficiency of the validation evidence.

13.11 Relationship to AI Safety Risk Management

SOTIF-related findings shall be incorporated into the applicable AI safety risk management process where they may affect safety.
The relationship shall be traceable, where applicable, between:
Functional Insufficiency / Scenario → Hazardous Behavior → Safety Risk → Mitigation Measure → Verification/Validation → Residual Risk
The organization shall ensure that mitigation measures addressing SOTIF-related risks are appropriately implemented and that their effectiveness is verified or validated.

13.12 Monitoring and Operational Feedback

Where the AI-enabled functionality is deployed in operation, relevant operational information shall be used to identify previously unknown or insufficiently understood safety-relevant scenarios.
Monitoring may consider, as applicable:

- performance degradation;
- new environmental conditions;
- changes in operating scenarios;
- newly observed edge cases;
- user interactions;
- incidents and near misses;
- changes in traffic or infrastructure conditions;
- changes in data distribution;
- unexpected AI behavior.

Relevant findings shall be fed back into the SOTIF, AI safety risk management, validation and change-management processes.

13.13 Changes and Re-Validation

Changes to the AI model, training data, software, configuration, operating conditions, ODD or system architecture may affect the validity of existing SOTIF evidence.
Such changes shall therefore be subject to an appropriate impact analysis.
The impact analysis shall determine whether additional:

- scenario analysis;
- testing;
- simulation;
- verification;
- validation;
- SOTIF analysis;
- safety assessment;
- monitoring

is required.
Where existing validation evidence is no longer considered representative or sufficient, the affected evidence shall be supplemented or re-established before the changed AI functionality is released for safety-related operation.

13.14 Responsibility and Interface Management

Responsibilities for the interface between AI safety and SOTIF shall be clearly assigned.
The AI Safety Manager, SOTIF Responsible, System Engineering, AI/ML Development, V&V and other relevant functions shall establish appropriate interfaces for:

- intended-function definition;
- identification of functional insufficiencies;

- scenario analysis;
- ODD definition;
- risk analysis;
- mitigation measures;
- verification and validation;
- operational monitoring;
- change management;
- SOTIF assessment;
- release and re-approval decisions.

Responsibilities shall be documented in the applicable AI Safety Management Plan, SOTIF Plan, project responsibility matrix or equivalent controlled documentation.

13.15 Objective Evidence and Records

The organization shall maintain objective evidence demonstrating that the defined interface between AI safety and SOTIF has been implemented.
Records shall include, as applicable:

- intended-function definition;
- functional insufficiency analysis;
- SOTIF-related risk analysis;
- ODD definition and assumptions;
- scenario definitions and coverage analysis;
- foreseeable misuse analysis;
- validation plans and results;
- simulation and testing evidence;
- identified limitations;
- unknown or newly identified unsafe scenarios;
- mitigation measures;
- monitoring results;
- change-impact analyses;
- SOTIF assessment results;

- residual risks;
- relevant approvals and release decisions.

Records shall be maintained under the applicable configuration, document and information control processes.

13.16 Interface Review

The interface between AI safety and SOTIF shall be reviewed at appropriate lifecycle gates and following significant changes to the intended functionality, AI model, training data, system architecture, ODD, operating environment or relevant safety assumptions.
The review shall determine whether the identified functional insufficiencies, scenarios, limitations, assumptions, mitigation measures and validation evidence remain valid and sufficient.
Identified gaps shall be documented, evaluated and addressed through the applicable risk management, corrective action, deviation, change or continual improvement process.

14 Integrated Safety and Security Interface

The organization shall establish and maintain an integrated interface between AI Safety and the applicable Functional Safety, SOTIF and Cybersecurity processes.
The integrated interface shall ensure that safety and security considerations are coordinated throughout the applicable AI lifecycle and that relevant information, assumptions, requirements, risks and evidence are exchanged between the responsible functions.
The integrated relationship shall be established between:

AI Safety
↕
Functional Safety – ISO 26262
↕
SOTIF – ISO 21448
↕
Cybersecurity – ISO/SAE 21434

The organization shall ensure that the interfaces between these disciplines are defined, maintained and reviewed throughout the lifecycle.

14.1 Interface Principles

The integrated interface shall address, as applicable:

- responsibilities and authorities;
- common inputs and outputs;
- shared assumptions;
- dependencies between safety and security activities;
- allocation and interaction of requirements;
- conflicting or potentially conflicting requirements;
- risk dependencies;
- evidence exchange;
- review and approval responsibilities;
- change notification mechanisms;
- escalation mechanisms;
- interfaces to release and re-approval activities.

The interfaces shall be defined in a manner appropriate to the complexity, safety relevance, cybersecurity exposure and intended use of the AI-enabled system.

14.2 Shared Information and Dependencies

Relevant information generated by one discipline shall be made available to the other disciplines where it may affect their respective activities.
This may include, as applicable:

- safety goals;
- AI safety requirements;
- functional safety requirements;
- SOTIF objectives and findings;

- cybersecurity requirements;
- cybersecurity threats and risks;
- safety risks;
- assumptions of use;
- ODD definitions;
- system and AI interfaces;
- safety mechanisms;
- cybersecurity controls;
- identified limitations;
- residual risks;
- verification and validation results;
- assessment results;
- incidents and operational feedback;
- planned changes.

Dependencies between safety and security assumptions shall be explicitly identified where the validity of one activity depends on controls or results established by another discipline.

14.3 Interaction of Safety and Security Risks

The organization shall consider situations in which cybersecurity threats may affect functional safety or SOTIF-related behavior.
Examples include, as applicable:

- manipulation of AI inputs;
- modification or replacement of AI models;
- manipulation of training or operational data;
- unauthorized modification of safety mechanisms;
- compromise of monitoring functions;
- unauthorized changes to AI configuration;
- compromised software or model updates;
- attacks affecting system availability or performance.

Where a cybersecurity threat can contribute to a safety-relevant condition, the cybersecurity risk and its potential safety impact shall be communicated to the relevant safety functions.
Similarly, safety-relevant assets, assumptions and functions shall be communicated to the cybersecurity function where they influence cybersecurity risk assessment or treatment.

14.4 Conflicting Requirements and Trade-Offs

The organization shall identify and manage potential conflicts between safety, SOTIF and cybersecurity requirements.
Examples may include situations where:

- a cybersecurity control affects system timing or resource availability;
- security monitoring affects AI performance;
- security mechanisms modify or restrict AI inputs or outputs;
- safety mechanisms introduce cybersecurity exposure;
- update requirements affect previously validated AI behavior;
- operational restrictions introduced for cybersecurity reasons affect the intended functionality.

Conflicts shall be evaluated by the responsible functions and resolved based on documented technical and safety considerations.
No discipline shall unilaterally override a safety- or security-relevant requirement without appropriate analysis, authorization and documentation.

14.5 Integrated Evidence

Where evidence generated by one discipline supports or affects another discipline, the relationship shall be documented and traceable.
The organization shall determine, as applicable:

- which evidence may be reused;
- which evidence requires independent review;
- which assumptions apply to the evidence;
- which limitations affect its applicability;

- which additional verification or validation is required.

Evidence shall not be considered transferable between disciplines solely because it addresses a related topic. Its applicability to the receiving safety or security activity shall be assessed and documented.

14.6 Integrated Reviews

The organization shall establish appropriate integrated reviews at defined lifecycle gates.
Integrated reviews shall consider, as applicable:

- changes to safety or security requirements;
- changes to assumptions;
- new or changed risks;
- identified interactions between safety and security;
- unresolved findings;
- changes to AI models, data or configuration;
- changes to the operating environment;
- changes to the threat landscape;
- changes to the ODD;
- changes to applicable standards or regulations.

The frequency and depth of integrated reviews shall be appropriate to the safety and security relevance of the AI-enabled system.

14.7 Change Notification and Impact Analysis

Changes that may affect more than one discipline shall be communicated to the affected responsible functions.
This shall include, as applicable, changes to:

- AI models;

- training or operational data;
- AI software;
- system architecture;
- safety mechanisms;
- cybersecurity controls;
- interfaces;
- configuration;
- ODD;
- operating conditions;
- update mechanisms;
- applicable requirements.

The responsible functions shall determine whether the change requires additional verification, validation, safety analysis, SOTIF analysis, cybersecurity analysis, assessment or approval.
The results of the impact analysis shall be documented and incorporated into the applicable change-management process.

14.8 Incident and Operational Feedback

Safety-relevant incidents, cybersecurity incidents, unexpected AI behavior and relevant operational observations shall be communicated across the affected disciplines.
The organization shall determine whether an event requires:

- safety risk reassessment;
- SOTIF reassessment;
- cybersecurity reassessment;
- additional verification or validation;
- corrective action;
- change-impact analysis;
- software or model update;
- re-assessment or re-approval.

Operational feedback shall be used as an input to continual improvement of the integrated safety and security management processes.

14.9 Responsibility and Escalation

Responsibilities and authorities for the integrated interface shall be clearly defined.
The AI Safety Manager shall coordinate the AI safety perspective, while the Functional Safety, SOTIF and Cybersecurity responsible functions shall retain responsibility for their respective technical and management domains.
Where an issue crosses organizational or disciplinary boundaries, the responsible functions shall jointly evaluate the issue and determine the required actions.
Unresolved safety or security conflicts shall be escalated according to the defined management and authority structure.

14.10 Integrated Interface Records

The organization shall maintain objective evidence demonstrating that the integrated interface has been established and maintained.
Records shall include, as applicable:

- responsibility and interface definitions;
- shared assumptions;
- interface requirements;
- risk dependencies;
- integrated review results;
- evidence exchange and applicability decisions;
- conflict resolutions;
- change notifications;
- impact analyses;
- incident and escalation records;
- integrated assessment results;
- relevant approvals and release decisions.

The records shall be maintained under the applicable document, configuration, change and information-control processes.

14.11 Continual Review of the Integrated Interface

The integrated safety and security interface shall be reviewed periodically and following significant changes to the AI-enabled system, its architecture, intended functionality, operating environment, threat landscape or applicable requirements.
The organization shall ensure that the interface remains effective in identifying and managing interactions between AI safety, functional safety, SOTIF and cybersecurity.
Identified weaknesses shall be addressed through the applicable risk management, corrective action, change-management or continual-improvement processes.

15 Data and AI Model Management

15.1 Data Management

AI data shall be managed as a controlled lifecycle asset where it may affect the safety of an AI-enabled system.
The organization shall establish and maintain a controlled data-management process covering the identification, acquisition, qualification, preparation, use, modification, release, monitoring, retention and disposal of safety-relevant data.
The data-management process shall ensure that data used for AI development and operation is appropriately defined, controlled and traceable to its intended purpose and applicable safety requirements.

15.1.1 Data Identification and Classification

Safety-relevant data shall be identified and classified according to its intended use, safety relevance, origin, content, quality characteristics and potential impact on the AI-enabled system.
The organization shall determine, as applicable:

- data category and intended purpose;

- data source and origin;
- data ownership and responsibility;
- safety relevance;
- applicable usage restrictions;
- applicable quality requirements;
- applicable protection requirements;
- intended lifecycle stage;
- relationship to the AI model or system;
- applicable retention requirements.

Data classification shall be used to determine the level of control, review, protection and traceability required.

15.1.2 Data Acquisition and Provenance

Safety-relevant data shall be acquired through controlled processes appropriate to its intended use. The organization shall maintain information necessary to establish data provenance, including, as applicable:

- data source;
- acquisition method;
- acquisition date or relevant time period;
- acquisition conditions;
- source configuration;
- applicable licenses or usage rights;
- processing applied before use;
- responsible organization or function.

Where data is obtained from external organizations, the organization shall evaluate and control the relevant external data source according to its safety relevance and applicable supplier-management processes.
Data provenance shall remain traceable throughout the applicable data lifecycle.

15.1.3 Data Requirements

Data requirements shall be established where data characteristics may affect the safety or intended functionality of the AI-enabled system.
Data requirements may address, as applicable:

- representativeness;
- completeness;
- accuracy;
- consistency;
- relevance;
- timeliness;
- coverage of relevant operating conditions;
- coverage of relevant scenarios;
- class distribution;
- resolution and sensor characteristics;
- labeling accuracy;
- permitted data sources;
- known limitations and assumptions.

Data requirements shall be consistent with the intended use, applicable AI safety requirements, system assumptions and identified safety risks.

15.1.4 Data Quality

The organization shall establish criteria for evaluating the quality of safety-relevant data.
Data-quality evaluation shall consider characteristics relevant to the intended use and identified risks.
Where applicable, data-quality evaluation shall address:

- completeness;
- correctness;
- consistency;
- accuracy;

- representativeness;
- distribution;
- duplication;
- missing or corrupted data;
- outliers;
- class imbalance;
- labeling quality;
- temporal or environmental coverage;
- known sources of bias or systematic limitations.

Data-quality limitations that may affect safety shall be documented and considered in the applicable risk assessment, model development, verification and validation activities.

15.1.5 Data Preparation and Processing

Data preparation and processing shall be controlled where such activities may affect model behavior or safety evidence.
The organization shall identify and control, as applicable:

- filtering;
- cleaning;
- normalization;
- transformation;
- augmentation;
- feature extraction;
- preprocessing;
- labeling and relabeling;
- data selection;
- removal or exclusion criteria.

Data-processing activities shall be traceable to the applicable dataset version.
Where automated or tool-supported data-processing activities are safety-relevant, the applicable tools, configurations and processing procedures shall be controlled in accordance with Chapter 21.

15.1.6 Data Labeling

Where supervised or otherwise labeled data is used, the labeling process shall be appropriately defined and controlled.
The organization shall establish, as applicable:

- labeling criteria;
- labeling rules and definitions;
- labeling responsibility;
- qualification or competence requirements;
- review and verification of labels;
- handling of ambiguous or uncertain labels;
- labeling quality criteria;
- correction and change processes.

Significant labeling errors or systematic labeling limitations that may affect safety shall be identified, evaluated and addressed.

15.1.7 Dataset Composition and Coverage

The composition of safety-relevant datasets shall be defined and controlled.
The organization shall consider whether the dataset provides appropriate coverage of the conditions and scenarios relevant to the intended use.
As applicable, the organization shall consider:

- operational design domain;
- environmental conditions;
- relevant system states;
- expected input characteristics;
- relevant scenario classes;
- nominal and non-nominal conditions;
- boundary conditions;
- rare but safety-relevant situations;

- foreseeable misuse;
- known limitations of available data;
- potential unknown or insufficiently represented conditions.

Dataset limitations shall be considered when defining verification and validation activities and when establishing the limitations and assumptions of the AI system.

15.1.8 Dataset Partitioning

Datasets used for training, validation and testing shall be appropriately separated and controlled to prevent unintended information leakage and to preserve the validity of the applicable evaluation results.
The organization shall define appropriate partitioning criteria based on the characteristics of the data and the intended evaluation.
Where relevant, the organization shall consider dependencies between data samples, such as common sources, subjects, vehicles, routes, environments, time periods or scenarios.
The basis for dataset partitioning shall be documented and traceable.

15.1.9 Dataset Versioning and Baseline

Each safety-relevant dataset shall have a unique identification and version.
The organization shall maintain sufficient information to determine the content and configuration of each controlled dataset version.
A dataset baseline shall identify, as applicable:

- dataset identity and version;
- included data;
- excluded data;
- labeling status;
- preprocessing status;
- partitioning;
- applicable quality evaluation;
- known limitations;
- approval or release status.

Changes to a controlled dataset baseline shall be managed through the applicable change-management process.

15.1.10 Dataset Verification and Validation

Safety-relevant datasets shall be subject to appropriate verification and validation activities before being used for safety-relevant model development or evaluation.
The extent of evaluation shall be based on the intended use, safety relevance, data characteristics and identified risks.
Evaluation may address, as applicable:

- data quality;
- dataset composition;
- scenario and condition coverage;
- labeling quality;
- partitioning;
- data integrity;
- representativeness;
- distribution characteristics;
- known limitations;
- suitability for the intended model development or evaluation activity.

The results shall be associated with the specific dataset version evaluated.

15.1.11 Dataset Approval and Release

A safety-relevant dataset shall have an appropriate approval or release status before it is used for activities requiring controlled data.
The release decision shall consider, as applicable:

- fulfillment of applicable data requirements;
- data-quality results;
- dataset composition and coverage;

- labeling results;
- partitioning;
- identified limitations;
- applicable risk assessments;
- traceability and configuration completeness;
- required reviews and approvals.

Development, evaluation and released datasets shall be clearly distinguishable where applicable.

15.1.12 Data Access, Protection and Integrity

Access to safety-relevant data shall be controlled according to its safety relevance and protection requirements.
The organization shall establish controls to prevent unauthorized:

- access;
- modification;
- deletion;
- copying;
- replacement;
- disclosure.

Appropriate mechanisms shall be established to protect the integrity and availability of safety-relevant datasets.
Cybersecurity controls relevant to data protection shall be coordinated with the cybersecurity processes defined in Chapter 12.

15.1.13 Data Changes and Impact Analysis

Changes to safety-relevant data or controlled datasets shall be managed in accordance with Chapter 18.
Changes may include, as applicable:

- addition or removal of data;
- changes to labels;
- correction of data;
- changes to preprocessing;
- changes to dataset composition;
- changes to dataset partitioning;
- changes to data sources;
- changes to data-generation methods;
- changes to data-selection criteria.

An appropriate impact analysis shall determine whether the change may affect:

- safety requirements;
- model behavior;
- model performance;
- safety assumptions;
- SOTIF assumptions;
- verification results;
- validation results;
- cybersecurity controls;
- functional safety evidence;
- safety case arguments;
- approved release status.

The impact analysis shall determine the need for additional verification, validation, assessment or approval.

15.1.14 Data Monitoring and Operational Feedback

Where safety-relevant AI systems depend on operational or continuously collected data, the organization shall establish appropriate controls for monitoring relevant data characteristics. Monitoring may include, as applicable:

- changes in data distribution;
- changes in operating conditions;
- changes in input characteristics;
- data-quality degradation;
- unexpected data patterns;
- newly identified scenarios;
- changes affecting the validity of model assumptions.

Relevant operational findings shall be fed back into the AI Safety Lifecycle and shall be evaluated for potential impact on the model, safety case, validation evidence and approved release.

15.1.15 Data Retention and Disposal

Safety-relevant datasets and associated information shall be retained for the period necessary to support safety, traceability, investigation, assessment and applicable contractual or regulatory obligations.
Retention shall consider, as applicable:

- released dataset versions;
- datasets supporting released models;
- associated metadata;
- provenance information;
- quality and validation results;
- labeling information;
- change history;
- approval records.

Disposal of safety-relevant data shall be controlled where disposal could affect safety evidence, traceability or the ability to investigate incidents or reproduce relevant results.

15.1.16 Data Management Records

The organization shall maintain controlled records sufficient to demonstrate that safety-relevant data has been appropriately identified, acquired, evaluated, controlled, released and maintained. Records shall include, as applicable:

- data requirements;
- data source and provenance;
- data classification;
- data-quality evaluations;
- labeling information;
- dataset composition;
- dataset versions and baselines;
- dataset partitioning;
- preprocessing and transformation information;
- verification and validation results;
- known limitations;
- approval and release records;
- access and protection information;
- change records;
- impact analyses;
- retention and disposal records.

The records shall be maintained under the applicable configuration, documentation, change and record-management processes.

15.1.17 Data Management Review

Data management shall be reviewed at appropriate lifecycle gates and following significant changes, incidents, unexpected model behavior, identified data-quality issues or changes in intended use or operating conditions.

The review shall determine whether safety-relevant data remains appropriately identified, controlled, traceable, suitable for its intended use and consistent with the assumptions and evidence supporting the AI system.

Any identified gap shall be addressed through the applicable risk-management, corrective-action, change-management or assessment process.

Objective

The objective of Data Management is to ensure that safety-relevant AI data is appropriately defined, qualified, controlled, protected and traceable throughout its applicable lifecycle, and that the data used to develop, evaluate and operate an AI system remains suitable for its intended safety-related purpose.

15.2 AI Model Management

The organization shall establish and maintain a controlled process for the identification, configuration, evaluation, approval, release, deployment, monitoring and change management of AI models that may affect the safety of an AI-enabled system.

Each safety-relevant AI model shall have a unique identification and shall be traceable to its applicable development, training, verification, validation, configuration and release information.

15.2.1 Model Identification and Configuration

Each AI model intended for safety-relevant use shall have a unique model identification and version.

The organization shall maintain, as applicable:

- model ID;
- model version;
- model architecture;
- model type and intended purpose;
- model configuration;
- training configuration;
- training dataset and dataset version;
- validation dataset and dataset version;
- test dataset and dataset version;
- hyperparameters;
- training software and environment;
- tool versions;

- libraries and dependencies;
- hardware or accelerator configuration, where relevant;
- model parameters or weights;
- preprocessing and post-processing configuration;
- inference environment;
- performance results;
- safety-related limitations;
- known weaknesses;
- applicable assumptions of use;
- approval and release status.

The configuration information shall be sufficient to identify the applicable model baseline and, where required, to reproduce or appropriately reconstruct the relevant development and evaluation environment.

15.2.2 Model Baseline

An AI model intended for safety-related use shall be established as a controlled baseline before formal verification, validation or release activities are performed.

The baseline shall uniquely identify the model and the configuration information necessary to establish its identity and evaluation status.

Changes to a controlled model baseline shall be subject to the applicable change-management process.

15.2.3 Training and Dataset Traceability

The organization shall maintain traceability between the model and the datasets and training configuration used to develop and evaluate it.

Where applicable, traceability shall include:

- dataset identity and version;
- data source and provenance;
- data preparation and preprocessing;
- labeling information;
- dataset partitioning;
- training configuration;
- training parameters;
- training environment;

- model generation process.

The organization shall ensure that the datasets used for training, validation and testing are appropriately identified, controlled and distinguishable.

15.2.4 Model Verification and Validation

AI models intended for safety-relevant use shall be subject to applicable verification and validation activities.

The organization shall define appropriate acceptance criteria based on the intended use, safety relevance and identified risks.

Evaluation may include, as applicable:

- functional performance;
- robustness;
- accuracy and relevant error characteristics;
- scenario coverage;
- edge-case behavior;
- uncertainty or confidence behavior;
- performance under relevant operating conditions;
- degradation behavior;
- safety-related limitations;
- interaction with safety mechanisms.

The results shall be associated with the specific model version and configuration evaluated.

15.2.5 Model Performance and Safety Limitations

Model performance shall not be evaluated solely in terms of general statistical performance.

The organization shall identify and document limitations that may affect safety, including, as applicable:

- known weaknesses;
- insufficient scenario coverage;
- sensitivity to environmental conditions;
- distribution limitations;
- uncertainty;
- performance degradation;

- known failure modes;
- limitations outside the intended operating conditions;
- dependencies on input quality or system configuration.

Safety-relevant limitations shall be communicated to the functions responsible for system integration, safety assessment, validation and release.

15.2.6 Model Approval and Release

An AI model shall not be deployed for safety-related use without an approved release status.

The release decision shall consider, as applicable:

- fulfillment of applicable requirements;
- verification and validation results;
- identified limitations;
- open anomalies and deviations;
- residual risks;
- applicable safety and cybersecurity assessments;
- configuration completeness;
- traceability of the model and applicable datasets;
- integration status;
- required approvals.

The approved release status shall be associated with the specific model version and configuration.

Development, test and released model versions shall be clearly distinguishable.

15.2.7 Model Deployment

The organization shall ensure that the model deployed in a safety-relevant environment corresponds to the approved model release.

Where appropriate, mechanisms shall be established to verify the identity and integrity of the deployed model.

The deployed model shall remain traceable to the applicable:

- model version;
- software version;
- configuration;

- dataset and training baseline;
- verification and validation evidence;
- release approval.

Unauthorized model replacement or modification shall be prevented or detected in accordance with the applicable cybersecurity controls.

15.2.8 Model Changes and Impact Analysis

Changes to an AI model shall be controlled.

Changes may include, as applicable:

- model architecture;
- model parameters or weights;
- training data;
- data preprocessing;
- training method;
- hyperparameters;
- software environment;
- dependencies;
- inference software;
- model configuration;
- deployment environment.

The organization shall perform an appropriate impact analysis to determine whether the change may affect:

- safety requirements;
- model performance;
- known limitations;
- safety risks;
- SOTIF assumptions;
- cybersecurity controls;
- functional safety evidence;
- validation results;
- integration evidence;
- release status.

The impact analysis shall determine the need for additional verification, validation, safety assessment, cybersecurity assessment or re-approval.

15.2.9 Model Versioning and Traceability

Model versions shall be uniquely identifiable and shall provide traceability throughout the applicable lifecycle.

The organization shall be able to determine, where required:

- which model was developed;
- which model was evaluated;
- which model was approved;
- which model was deployed;
- which datasets and configurations were used;
- which verification and validation results apply;
- which changes occurred between versions;
- who approved the applicable release.

This traceability shall support incident investigation, safety assessment, change management and rollback where required.

15.2.10 Model Retirement and Replacement

When an AI model is replaced or retired, the organization shall ensure that the transition is controlled.

The organization shall determine, as applicable:

- whether the replacement model has been appropriately verified and validated;
- whether the replacement model fulfills the applicable requirements;
- whether integration assumptions remain valid;
- whether existing safety evidence remains applicable;
- whether additional assessment or approval is required;
- whether the previous model must remain available for investigation or rollback.

Retired model versions and their associated safety-relevant records shall be retained according to applicable record-retention requirements.

15.2.11 Model Management Records

The organization shall maintain controlled records sufficient to demonstrate that safety-relevant AI models have been appropriately identified, evaluated, approved, released and maintained.

Records shall include, as applicable:

- model identification and version;
- model configuration;
- training and dataset information;
- verification and validation results;
- performance evaluation;
- known limitations and weaknesses;
- assumptions of use;
- change history;
- impact analyses;
- approval status;
- release records;
- deployment information;
- replacement or retirement records.

The records shall be maintained under the applicable configuration, document, change and information-control processes.

15.2.12 Model Management Review

AI model management shall be reviewed at appropriate lifecycle gates and following significant changes, incidents, unexpected model behavior, or changes in intended use or operating conditions.

The review shall determine whether the model remains appropriately identified, controlled, traceable, evaluated and approved for its intended safety-related use.

Any identified gap shall be addressed through the applicable risk-management, corrective-action, change-management or assessment process.

Objective

The objective of Data and AI Model Management is to ensure that safety-relevant datasets and AI models remain uniquely identifiable, appropriately controlled, traceable and reproducible throughout their applicable lifecycle, and that changes to data or models are assessed before they can affect an approved safety-related configuration.

16 Verification and Validation

The organization shall establish and maintain a systematic verification and validation process for AI systems and AI components that may affect safety.

Verification and validation activities shall be planned according to the safety relevance, intended use, architecture, AI characteristics, identified risks and applicable requirements.
Verification shall provide evidence that specified requirements and defined technical or safety properties have been fulfilled.
Validation shall provide evidence that the AI-enabled system is suitable for its intended use and performs acceptably under the relevant operating conditions and scenarios.

16.1 Verification and Validation Planning

Verification and validation activities shall be planned early enough to support the applicable AI safety lifecycle and shall be documented in the relevant project plans or equivalent controlled information.
The planning shall define, as applicable:

- objectives and scope;
- applicable requirements;
- verification and validation criteria;
- test methods;
- scenarios and operating conditions;
- datasets;
- test environment;
- simulation environment;
- required tools and equipment;
- responsibilities and required independence;
- acceptance criteria;
- required evidence and records;
- handling of anomalies and deviations.

The extent and rigor of verification and validation shall be proportionate to the safety relevance and identified risks of the AI system.

16.2 Verification Activities

Verification activities shall determine whether the AI system and its constituent elements fulfill their specified requirements.
Activities may include, as applicable:

- requirements verification;
- data verification;
- model verification;
- software verification;
- interface verification;
- configuration verification;
- performance testing;
- robustness testing;
- scenario-based testing;
- boundary-condition testing;
- stress testing;
- fault injection;
- simulation;
- hardware-in-the-loop testing;
- integration testing;
- static or analytical verification;
- verification of safety mechanisms.

The organization shall determine which verification methods are appropriate for the characteristics and risks of the AI system.

16.3 Validation Activities

Validation shall evaluate whether the AI-enabled functionality is suitable for its intended use and operating conditions.
Validation activities may include, as applicable:

- system-level validation;
- scenario-based validation;
- representative operational testing;
- boundary-condition validation;
- environmental testing;
- robustness evaluation;
- field validation;
- simulation-supported validation;
- hardware-in-the-loop validation;
- validation of human interaction;
- validation of fallback or degraded operation;
- validation of safety-relevant performance limitations.

Validation shall consider relevant scenarios identified through AI safety risk management, SOTIF analysis and operational experience.

16.4 AI-Specific Verification and Validation

Verification and validation shall consider characteristics specific to AI systems where relevant. These may include:

- model uncertainty;
- generalization capability;
- data distribution;
- dataset limitations;
- insufficient scenario coverage;
- rare or edge-case behavior;
- robustness against variations in inputs;
- sensitivity to environmental conditions;
- performance degradation;
- unexpected model behavior;
- interaction between AI and conventional software or hardware;

- changes in model or data characteristics.

Statistical performance alone shall not automatically be considered sufficient evidence of safety where additional safety-related characteristics or limitations require evaluation.

16.5 Scenario-Based and Boundary-Condition Testing

Where AI behavior depends on the operating environment or scenario, verification and validation shall include relevant scenarios and boundary conditions.
The organization shall consider, as applicable:

- normal operating conditions;
- challenging conditions;
- rare events;
- edge cases;
- combinations of relevant conditions;
- transitions between operating states;
- ODD boundaries;
- foreseeable misuse;
- degraded conditions;
- conditions identified through incidents or operational monitoring.

Scenario selection and coverage shall be justified with respect to the intended use and identified safety risks.

16.6 Robustness and Stress Testing

Where appropriate, the organization shall evaluate the robustness of AI behavior against variations and disturbances that may occur during operation.
Testing may consider, as applicable:

- input variations;
- sensor noise;

- environmental changes;
- degraded inputs;
- incomplete or ambiguous information;
- data distribution changes;
- computational limitations;
- timing variations;
- resource constraints;
- unexpected combinations of conditions.

The organization shall determine whether observed degradation remains within acceptable safety limits.

16.7 Fault Injection and Adversarial Testing

Where applicable, fault injection shall be used to evaluate the behavior of AI-related functions and safety mechanisms under defined fault conditions.
Where cybersecurity threats may affect AI safety, adversarial testing or equivalent techniques shall be considered as appropriate.
Such testing may address, as applicable:

- manipulated inputs;
- corrupted data;
- abnormal inputs;
- model-related faults;
- interface faults;
- communication faults;
- degraded sensors;
- compromised or unauthorized AI artifacts;
- failures of monitoring or safety mechanisms.

The applicability and extent of such testing shall be determined based on the identified risks and system characteristics.

16.8 Test Environment and Configuration Control

Verification and validation shall be performed using controlled and documented configurations.
The organization shall identify, as applicable:

- AI model version;
- dataset version;
- software version;
- hardware configuration;
- AI accelerator configuration;
- tool versions;
- test environment;
- simulation environment;
- test parameters;
- system configuration.

The results shall remain traceable to the configuration under which the verification or validation was performed.

16.9 Independence

The organization shall determine the required degree of independence for verification and validation activities based on the safety relevance, complexity and applicable requirements.
Where independent verification or validation is required, the organization shall ensure that the assigned personnel have appropriate competence and sufficient independence from the development activities being evaluated.
Independence shall be maintained without preventing the necessary technical cooperation between development and V&V functions.

16.10 Acceptance Criteria and Results

Verification and validation activities shall be performed against defined acceptance criteria.
Results shall identify, as applicable:

- activity performed;
- configuration evaluated;
- test or analysis method;
- expected result;
- actual result;
- pass/fail or equivalent evaluation;
- identified anomalies;
- deviations;
- limitations;
- conclusions;
- responsible reviewer or approver.

Results shall be reviewed by appropriately competent personnel.

16.11 Anomalies, Deviations and Open Findings

Anomalies and unexpected results identified during verification and validation shall be documented and evaluated.
The evaluation shall determine, as applicable:

- safety impact;
- root cause or contributing factors;
- affected requirements;
- affected scenarios;
- required corrective action;
- residual risk;
- need for additional verification or validation;
- effect on release status.

Open findings shall not be disregarded solely because overall statistical or functional performance remains acceptable.

Where a deviation from an acceptance criterion is accepted, the deviation shall be formally documented, justified, approved by authorized personnel and included in the applicable safety assessment and release decision.

16.12 Objective Evidence and Traceability

Verification and validation shall provide objective evidence supporting the applicable safety decision.
Evidence shall be traceable, as applicable, to:
Requirement → Verification/Validation Method → Configuration → Test or Analysis → Result → Finding → Conclusion → Safety Decision
The evidence shall be controlled and retained according to the applicable document, configuration and record-management processes.

16.13 Re-Verification and Re-Validation

Changes to the AI model, data, software, configuration, system architecture, operating conditions or intended use shall be evaluated to determine whether existing verification and validation evidence remains valid.
Where necessary, the affected verification and validation activities shall be repeated or supplemented.
The organization shall consider, as applicable:

- model changes;
- training-data changes;
- software updates;
- configuration changes;
- changes to the ODD;
- changes to system interfaces;
- changes to safety mechanisms;
- cybersecurity-related changes;
- newly identified scenarios;
- incidents or unexpected behavior.

Successful re-verification or re-validation shall not automatically constitute approval for renewed deployment. Where applicable, the results shall be subject to the defined safety assessment and release approval process.

16.14 V&V Records

The organization shall maintain records sufficient to demonstrate that the required verification and validation activities have been performed and that the results have been appropriately evaluated. Records shall include, as applicable:

- V&V plans;
- requirements and acceptance criteria;
- test specifications;
- test scenarios;
- datasets;
- test configurations;
- test and analysis results;
- simulation results;
- identified anomalies;
- deviations;
- corrective actions;
- conclusions;
- approvals;
- evidence of independence where applicable.

16.15 V&V Review and Continual Improvement

Verification and validation results shall be reviewed at appropriate lifecycle gates and shall provide input to safety assessment, risk management, release decisions and continual improvement.
Lessons learned from testing, incidents, field experience and newly identified scenarios shall be considered when updating V&V strategies, scenarios, acceptance criteria and methods.
The organization shall periodically evaluate whether its verification and validation methods remain appropriate for the evolving AI technology, intended use, safety risks and applicable requirements.

17 Continuous Assurance During Operation

AI safety shall continue to be managed after product release and throughout the applicable operational lifecycle.

The organization shall establish and maintain mechanisms for continuous monitoring, evaluation and assurance of AI safety during operation. The extent and rigor of operational assurance shall be appropriate to the safety relevance, intended use, operational environment, AI characteristics and identified risks.

Operational assurance shall provide mechanisms for detecting changes in AI behavior, operating conditions, data characteristics, safety assumptions and other factors that may affect the continued validity of the safety case or release decision.

17.1 Operational Monitoring

The organization shall establish appropriate operational monitoring for safety-relevant AI functionality.

Monitoring may include, as applicable:

- operational conditions;
- system status;
- AI performance;
- AI outputs;
- safety-related indicators;
- confidence or uncertainty indicators;
- safety mechanisms;
- fallback and degraded operation;
- environmental conditions;
- ODD compliance;
- system resource conditions;
- relevant cybersecurity events.

Monitoring activities shall be defined according to the identified safety risks and operational requirements.

17.2 Performance Monitoring

The organization shall monitor relevant AI performance characteristics during operation where changes in performance may affect safety.
Performance monitoring may consider, as applicable:

- prediction or classification performance;
- false-positive and false-negative behavior;
- detection performance;
- robustness indicators;
- confidence or uncertainty;
- response time;
- degradation of performance;
- unexpected behavior;
- performance under relevant environmental conditions.

The organization shall define appropriate thresholds, indicators or evaluation criteria where necessary to identify potentially safety-relevant degradation.

17.3 Safety-Event Monitoring

Safety-relevant events and observations shall be systematically identified, recorded and evaluated. Such events may include:

- hazardous or potentially hazardous behavior;
- unexpected AI behavior;
- safety mechanism activations;
- fallback events;
- loss of intended functionality;
- near misses;

- incidents;
- repeated anomalies;
- violations of safety assumptions;
- operation outside relevant ODD conditions.

Safety events shall be evaluated to determine whether they indicate a previously unknown risk, an existing risk becoming unacceptable, or a limitation in the existing safety measures.

17.4 Data Distribution Monitoring

Where AI performance depends on the characteristics of operational data, the organization shall consider monitoring for relevant changes in data distribution.
Such changes may include:

- changes in environmental conditions;
- new object or scenario characteristics;
- changes in sensor characteristics;
- changes in operating conditions;
- changes in user behavior;
- changes in traffic or infrastructure;
- changes in geographic conditions;
- changes in data quality;
- significant differences between development and operational data.

Where relevant distribution changes may affect safety, their potential impact shall be assessed.

17.5 Model Degradation and Drift

The organization shall consider the potential for model degradation or changes in AI behavior during operation.
Model degradation may result from, as applicable:

- changes in operational data;

- changes in the operating environment;
- changes in input characteristics;
- changes in system configuration;
- aging or degradation of supporting sensors or hardware;
- changes in software dependencies;
- interactions with updated system components.

Where evidence indicates that model performance or safety-related behavior may have degraded, the organization shall initiate an appropriate investigation and determine whether additional verification, validation, assessment or corrective action is required.

17.6 Anomaly Detection

The organization shall establish appropriate mechanisms for identifying and evaluating abnormal or unexpected AI behavior.
Anomaly detection may be based on:

- runtime monitoring;
- safety monitors;
- statistical indicators;
- threshold violations;
- unexpected outputs;
- inconsistent system behavior;
- operational data;
- customer reports;
- incident information;
- cybersecurity monitoring.

The organization shall define how detected anomalies are classified, recorded, investigated and escalated according to their potential safety impact.

17.7 Incident Management

Safety-relevant AI incidents shall be managed through a controlled incident-management process.
The process shall provide for, as applicable:

- event identification;
- immediate containment;
- preservation of relevant evidence;
- technical investigation;
- safety impact assessment;
- cybersecurity assessment where applicable;
- root-cause or contributing-factor analysis;
- corrective action;
- communication and escalation;
- decision on continued operation;
- verification of corrective measures;
- re-assessment and re-approval where required.

Incidents shall provide input to the applicable risk management, SOTIF, cybersecurity, V&V and change-management processes.

17.8 Customer and Field Feedback

Relevant customer feedback and field information shall be systematically evaluated where it may provide evidence concerning AI safety.
The organization shall consider, as applicable:

- customer complaints;
- reported unexpected behavior;
- operational observations;
- field incidents;
- near misses;
- performance concerns;

- newly identified scenarios;
- reports of operation outside expected conditions.

Feedback shall not be dismissed solely because the reported behavior has not previously been observed during development or validation.

17.9 Operational Data and Evidence

Operational data used for safety assurance shall be appropriately controlled and traceable.
The organization shall determine, as applicable:

- data source;
- time period;
- system and model version;
- operating conditions;
- data quality;
- applicable privacy and cybersecurity requirements;
- analysis methods;
- limitations of the data;
- conclusions derived from the analysis.

Where operational data is used to support safety decisions, the organization shall ensure that its relevance and limitations are understood.

17.10 Continuous Safety Assessment

Operational monitoring results shall be periodically evaluated to determine whether the assumptions and evidence supporting the safety decision remain valid.
The evaluation shall consider, as applicable:

- newly identified risks;
- changes in model performance;
- data distribution changes;

- new scenarios;
- incidents;
- changes in the operating environment;
- changes to the ODD;
- cybersecurity events;
- changes to system configuration;
- changes to applicable requirements.

Where operational evidence indicates that existing safety assumptions are no longer valid, the organization shall initiate an appropriate investigation, change, corrective action or re-assessment.

17.11 Trigger Conditions for Re-Assessment

The organization shall define conditions that may trigger additional safety activities.
Triggers may include:

- significant performance degradation;
- repeated safety-relevant anomalies;
- newly identified unsafe scenarios;
- operation outside validated conditions;
- significant data-distribution changes;
- model degradation or drift;
- safety-related incidents;
- cybersecurity compromise;
- significant system changes;
- changes to intended use;
- changes to the ODD;
- invalidated safety assumptions.

The organization shall determine whether such conditions require additional verification, validation, risk analysis, SOTIF analysis, cybersecurity assessment, safety assessment, corrective action or release re-approval.

17.12 Feedback into the AI Safety Lifecycle

Operational assurance shall provide feedback into the AI safety lifecycle.
Relevant operational findings shall be used, as applicable, to update:

- safety risks;
- safety requirements;
- SOTIF scenarios;
- validation scenarios;
- datasets;
- model requirements;
- monitoring mechanisms;
- safety mechanisms;
- cybersecurity controls;
- assumptions of use;
- change-management activities.

This feedback shall support continual improvement and shall ensure that experience gained during operation is incorporated into subsequent AI development and safety activities.

17.13 Change and Update Interface

Where operational findings result in changes to the AI model, data, software, configuration or operating conditions, the applicable change-management process shall be initiated.
The organization shall determine the impact of the change on:

- safety requirements;
- functional safety;
- SOTIF;
- cybersecurity;
- validation evidence;
- safety assumptions;
- residual risks;

- release status.

Changes shall not be deployed for safety-related use until the applicable verification, validation, assessment and approval activities have been completed.

17.14 Responsibilities and Escalation

Responsibilities for continuous assurance shall be clearly assigned.
Relevant functions may include:

- AI Safety Management;
- AI/ML Development;
- V&V;
- Functional Safety;
- SOTIF;
- Cybersecurity;
- System Engineering;
- Quality;
- Operations;
- Customer Support;
- Incident Management.

Escalation criteria shall be defined for safety-relevant events, unexpected AI behavior, significant performance degradation and other conditions that may affect continued safe operation.

17.15 Objective Evidence and Records

The organization shall maintain objective evidence demonstrating that continuous AI safety assurance is performed.
Records shall include, as applicable:

- monitoring plans and results;
- performance indicators;

- safety-event records;
- anomaly records;
- field-data analyses;
- customer feedback;
- incident records;
- data-distribution analyses;
- model degradation analyses;
- investigation results;
- corrective actions;
- re-assessment decisions;
- change-impact analyses;
- updated validation evidence;
- release or re-approval decisions.

Records shall be maintained under the applicable document, configuration, information and record-control processes.

17.16 Review of Continuous Assurance Effectiveness

The effectiveness of continuous assurance shall be periodically reviewed.
The review shall determine whether the established monitoring and assurance mechanisms remain capable of detecting relevant changes in AI behavior, operating conditions, safety risks and assumptions.
The organization shall consider lessons learned from operational experience, incidents, customer feedback, cybersecurity events and newly identified scenarios when improving its continuous assurance processes.

18 Change Management

18.1 General

All safety-relevant changes affecting the AI system, its development environment, its operating conditions, or its supporting processes shall be systematically identified, controlled, assessed and approved before implementation or release.

Change Management shall ensure that changes do not unintentionally invalidate existing safety requirements, safety assumptions, risk controls, verification and validation results, safety arguments, cybersecurity controls, SOTIF assumptions, or operational limitations.

Change Management shall apply throughout the complete AI lifecycle, including development, integration, deployment, operation, maintenance and update activities. The organization shall establish and maintain a controlled process covering the initiation, analysis, classification, planning, implementation, verification, validation, assessment, approval and release of changes.

No safety-relevant change shall be introduced into an approved release without completion of the applicable change-management activities and the required approval.

18.2 Types of Changes

Changes subject to Change Management may affect any element that can influence the safety of the AI system. This includes, for example, changes to the AI model or its architecture, model weights and parameters, training and validation data, data sources, labeling and preprocessing, algorithms, software, hardware, interfaces, system architecture and configuration.

Changes to the Operating Design Domain (ODD), intended use, operating conditions, safety requirements, safety mechanisms, fallback mechanisms, cybersecurity controls, monitoring mechanisms or other safety-related assumptions shall also be subject to Change Management.

Changes to development, training, simulation or test environments, toolchains, software libraries, dependencies, operating systems or other supporting technologies shall be considered where they may affect the validity of safety-related results or evidence.

Changes may also result from operational experience, field incidents, newly identified scenarios or safety risks, or changes to applicable standards, regulations or customer requirements. Changes to supporting processes shall be controlled where they may affect safety-related activities, evidence or decisions.

18.3 Change Request and Initial Screening

Safety-relevant changes shall be initiated through a controlled change request or an equivalent documented mechanism. The change request shall describe the proposed change, its purpose and the affected configuration.

The initial screening shall determine whether the proposed change may have an impact on safety. This determination shall consider the affected AI model, data, software, hardware, requirements, assumptions, safety functions and existing safety evidence, as well as relevant Functional Safety, SOTIF and Cybersecurity interfaces.

Changes initially classified as non-safety-relevant shall be supported by sufficient documented justification to demonstrate why no safety impact is expected. The classification and its rationale shall remain traceable to the change record.

18.4 Change Impact Analysis

Before implementation of a safety-relevant change, an appropriate impact analysis shall be performed. The analysis shall determine whether the change may affect safety requirements, risk controls, AI behaviour, training or validation data, assumptions of use, operating conditions, functional safety, SOTIF, cybersecurity, system interfaces, safety mechanisms, monitoring, previously established V&V evidence, safety arguments, residual risks or regulatory and customer requirements.

The impact analysis shall consider both direct and indirect effects.

For example, a change to a training dataset may not modify the model architecture or software implementation, but may nevertheless modify model behaviour and invalidate existing verification or validation results. Similarly, a change to an inference library, compiler, accelerator, operating system or hardware platform may affect the behaviour or performance of an already validated AI model.

The results of the impact analysis shall be documented and traceable to the affected configuration items and safety evidence.

18.5 Change Classification

Each safety-relevant change shall be classified according to its potential safety impact. The classification method shall be defined by the organization and shall be appropriate to the safety relevance, complexity and potential consequences of the affected system or function.

The classification shall consider, as applicable, the safety relevance of the affected function, the potential consequences of a changed behaviour, the extent of the affected configuration, changes to AI behaviour or data, changes to operating conditions or safety assumptions, and potential effects on Functional Safety, SOTIF, Cybersecurity, V&V evidence and the safety case.

The resulting classification shall determine the level of analysis, review, verification, validation, regression testing, safety assessment and approval required.

The classification shall not be based solely on the apparent technical size of the change. A technically small modification may have a significant safety impact if it changes the behaviour of the AI system or invalidates an existing safety assumption.

18.6 Change Planning and Approval

Each safety-relevant change shall be planned before implementation. The change plan shall define the affected configuration, responsible functions, required analyses and reviews, applicable V&V activities, regression scope, safety assessment activities and required approval authorities.

Where a change affects several engineering disciplines, the relevant responsible functions shall participate in the change evaluation. This shall include Functional Safety, SOTIF, Cybersecurity, System Engineering or other functions where their responsibilities are affected.

The planned activities shall be proportionate to the classification and impact of the change. No change shall proceed to release where required safety activities or approvals remain incomplete, unless an explicitly controlled deviation or emergency-change process is applicable.

18.7 Implementation and Configuration Control

Approved changes shall be implemented under configuration management. The organization shall ensure that the approved change is uniquely identifiable, that affected configuration items are controlled and that the implemented configuration corresponds to the approved change request.

For AI systems, configuration control shall include, as applicable, the AI model and model weights, datasets, preprocessing and postprocessing, parameters, software, libraries, hardware, toolchain and deployment configuration.
The resulting configuration shall be uniquely identifiable so that the exact configuration subjected to verification, validation and safety assessment can be demonstrated.

18.8 Verification and Validation of Changes

Changes shall be subject to verification and validation activities appropriate to their classification and impact.
Verification shall determine whether the implemented change satisfies the applicable requirements and defined change objectives. Validation shall determine whether the changed AI system remains suitable for its intended use and relevant operating conditions.
The selected methods may include targeted verification, regression testing, model-performance testing, robustness testing, scenario-based testing, simulation, integration testing, hardware-in-the-loop testing, cybersecurity testing or system-level validation. The extent and combination of these activities shall be justified by the change impact analysis.
Completion of testing alone shall not constitute approval for release. The results shall be reviewed and incorporated into the applicable safety assessment and release decision.

18.9 Regression Testing

Where a change may affect previously verified or validated functionality, appropriate regression testing shall be performed.
The regression strategy shall be based on the identified dependencies and potential impact of the change. It shall consider not only the changed function but also other safety-related functions and interfaces that may be affected.
For AI systems, regression testing may need to address changes in model performance, scenario coverage, false-positive and false-negative behaviour, robustness, uncertainty, boundary conditions, safety mechanisms, interactions with conventional software or other AI components, computational performance and runtime behaviour.
The selected regression scope and its rationale shall be documented.

18.10 Re-Verification, Re-Validation and Re-Assessment

Where the impact analysis identifies an effect on previously established safety evidence, the affected evidence shall be re-established to the extent necessary.

The applicable change process shall therefore follow the principle:

Change → Impact Analysis → Classification → Re-Verification → Re-Validation → Re-Assessment → Re-Approval → Release

Not every change requires complete repetition of all lifecycle activities. The required extent shall be determined by the documented impact analysis and change classification.

Re-verification or re-validation shall not automatically restore the previous safety approval status. Where applicable, the safety assessment shall determine whether previous conclusions remain valid, whether additional evidence is required, whether residual risks or assumptions have changed, and whether the safety case, operational restrictions or monitoring measures require updating.

18.11 Interface with Functional Safety, SOTIF and Cybersecurity

Changes shall be evaluated for their potential cross-domain effects. Where a change affects Functional Safety, SOTIF or Cybersecurity, the responsible functions shall participate in the change analysis and assessment.

A cybersecurity-related change, for example, may affect AI behaviour or system performance and therefore require a safety impact assessment. Conversely, a model or dataset change introduced to improve safety performance may create cybersecurity or SOTIF implications.

The Change Management process shall therefore provide an effective interface between the AI Safety Management System and the applicable Functional Safety, SOTIF, Cybersecurity, Software Update, System Engineering and regulatory processes.

18.12 Emergency Changes

Where an immediate change is required to address a significant safety, security or operational issue, the organization shall apply a controlled emergency-change process.

The emergency process may allow an accelerated implementation route where justified, but shall not eliminate the requirement for safety control. The change shall have a documented justification,

defined authority, controlled implementation and appropriate verification before deployment where practicable.

Following deployment, the changed configuration shall be subject to appropriate monitoring and the required verification, validation and safety assessment activities shall be completed retrospectively where they could not reasonably be completed beforehand.

18.13 Change Rejection and Rollback

A change shall not be released where the applicable acceptance criteria have not been fulfilled or where identified residual risks have not been appropriately addressed and accepted through the defined process.

Where a deployed change results in unacceptable safety behaviour, unexpected degradation or invalidated safety assumptions, appropriate containment shall be initiated. Depending on the circumstances, this may include restricting the affected function, increasing monitoring, initiating corrective action, suspending deployment or reverting to a previously approved configuration.

Rollback configurations shall themselves be controlled, uniquely identifiable and demonstrably approved for the intended operational use.

18.14 Change Records and Traceability

Safety-relevant changes shall generate objective evidence demonstrating that the change has been controlled throughout its lifecycle.

The change record shall provide traceability between the proposed change, affected configuration items, impact analysis, applicable requirements and assumptions, verification and validation results, safety assessment, approval and final released configuration.

Where applicable, the record shall also include deviations, open issues, cybersecurity or SOTIF assessments, updated safety evidence, release records and deployment or rollback information.

The organization shall be able to demonstrate which exact configuration was assessed and approved and how the corresponding safety evidence relates to that configuration.

18.15 Change Management Responsibilities

The **AI Safety Manager** shall ensure that the Change Management process is established and effectively implemented. The relevant technical functions shall be responsible for analysing and implementing changes within their areas of competence.

The **Functional Safety Manager**, **SOTIF Responsible**, **Cybersecurity Responsible**, **System Engineer**, **V&V Responsible** and other affected functions shall participate where the change impacts their respective responsibilities.

The responsible release authority shall ensure that the required change activities, safety assessments and approvals have been completed before the changed configuration is released for operational use.

18.16 Review and Continual Improvement

The effectiveness of Change Management shall be periodically reviewed as part of the AI Safety Management System.

The review shall consider experience from implemented changes, unexpected safety effects, incidents, rollback events, recurring findings and the effectiveness of impact analysis, regression testing and safety assessment.

Lessons learned shall be used to improve change classification criteria, V&V strategies, safety assessment methods and other elements of the AI Safety Management System where appropriate.

The objective of Change Management is not to prevent change. It is to ensure that change remains controlled, traceable and appropriately assessed before the changed configuration is released.

19 Deviation and Waiver Management

19.1 General

Deviations from approved requirements, processes, procedures, acceptance criteria or other defined safety-related provisions shall be formally identified, assessed, documented and approved.

Deviation and Waiver Management shall ensure that deviations are transparent, justified and controlled and that their potential impact on AI safety is understood before the deviation is accepted.

A deviation shall not be treated as an informal agreement or undocumented exception. Any deviation that may affect safety-related activities, evidence, assumptions, requirements or decisions shall be subject to the defined deviation-management process.
The organization shall ensure that deviations do not become a means of circumventing mandatory safety requirements or releasing a product without the required safety evidence.

19.2 Deviation Identification

A deviation may arise when an approved requirement, process, specification, acceptance criterion or planned activity cannot be fulfilled as defined.
Deviations may be identified during development, verification and validation, safety assessment, audits, reviews, production, deployment or operation. They may also result from technical constraints, newly identified conditions, supplier limitations, unexpected test results or changes in applicable requirements.
The deviation shall be identified as soon as it becomes known and shall be recorded in a controlled deviation record.
Where a deviation is identified after release, its potential impact on the released configuration and previously established safety conclusions shall be assessed without undue delay.

19.3 Deviation Record

Each safety-relevant deviation shall be uniquely identified and documented.
The deviation record shall describe the affected requirement or process, the reason for the deviation and the technical justification for the proposed departure from the approved provision. It shall also document the affected product or configuration, the potential safety impact and the associated risk assessment.
Where the deviation is accepted, the record shall define any compensating measures, restrictions or additional controls required to maintain an acceptable level of safety.
The record shall identify the responsible person and the authority responsible for approval. For temporary deviations or waivers, the applicable validity period or defined closure criterion shall also be specified.

The deviation record shall contain sufficient information to allow an independent reviewer to understand why the deviation was necessary, what safety consequences were considered and on what basis the deviation was accepted.

19.4 Safety Impact and Risk Assessment

Every safety-relevant deviation shall be evaluated with respect to its potential impact on the safety of the affected AI system.

The assessment shall consider whether the deviation may affect safety requirements, safety objectives, assumptions, risk controls, V&V evidence, SOTIF considerations, cybersecurity controls, operating limitations or the safety argument.

The associated risk shall be assessed using the applicable risk-management process. Where the deviation changes an existing safety risk or introduces a new risk, the affected risk records shall be updated accordingly.

The acceptance of a deviation shall therefore be based on documented technical and safety considerations rather than on schedule, cost, commercial pressure or convenience alone.

19.5 Compensating Measures

Where a deviation results in an increased or otherwise changed safety risk, appropriate compensating measures shall be identified and implemented where necessary.

Compensating measures may include additional verification or validation, increased monitoring, operational restrictions, additional safety mechanisms, restricted deployment, additional user or operator information, or other technical or organizational controls.

The effectiveness of compensating measures shall be verified to the extent necessary for the associated safety risk.

Compensating measures shall be controlled as part of the affected product or project configuration and shall remain effective for the duration of the deviation.

19.6 Approval of Deviations

A safety-relevant deviation shall be approved by an authority with appropriate competence and formal responsibility.

The approval authority shall review the technical justification, safety impact, risk assessment and proposed compensating measures before accepting the deviation.
The approval shall be traceable to the specific deviation record and affected configuration.
Where a deviation affects several safety disciplines, the relevant responsible functions shall participate in the assessment. This may include Functional Safety, SOTIF, Cybersecurity, V&V, System Engineering or other affected functions.
Approval of a deviation shall not automatically imply acceptance of the affected product for release. The release authority shall determine whether the deviation is compatible with the applicable release criteria and overall safety assessment.

19.7 Waivers

A waiver may be used where a formally approved temporary exception from an established requirement or process is necessary and technically justified.
A waiver shall be explicitly limited in scope and shall define either a fixed expiration date or a clear closure criterion. The waiver shall identify the affected requirement, product or configuration and shall specify any applicable restrictions or compensating measures.
A waiver shall not be open-ended. Before its expiration or closure criterion is reached, the responsible organization shall either eliminate the underlying deviation, formally renew the waiver based on a new assessment, or take another approved action.
Repeated renewal of a waiver shall not be used to avoid permanently resolving a known safety-related issue.

19.8 Deviations Affecting Safety Evidence

Where a deviation affects the basis on which verification, validation or safety assessment conclusions were established, the affected evidence shall be reviewed.
The organization shall determine whether additional analysis, verification, validation or safety assessment is necessary and whether the existing safety argument remains valid.
A deviation shall not invalidate previously established evidence without appropriate assessment. Where existing evidence is no longer applicable, it shall be identified and replaced or supplemented by appropriate objective evidence.

19.9 Deviations After Release

Where a deviation is identified in an already released configuration, the organization shall assess its potential impact on operational safety without undue delay.
Depending on the result of the assessment, appropriate actions may include additional monitoring, corrective action, a software or model update, operational restrictions, customer communication, incident management, re-verification, re-validation, safety re-assessment or withdrawal of the affected release.
The decision and its technical justification shall be documented.

19.10 Deviation Closure

A deviation shall remain open until the defined closure conditions have been fulfilled.
Closure may require implementation of a corrective change, completion of additional V&V, removal of the affected condition, expiry of the applicable waiver, or confirmation that the deviation is no longer applicable.
The closure shall be reviewed and approved by the responsible authority and shall be traceable to the corresponding deviation record.
Open deviations shall be periodically reviewed to ensure that they remain valid, controlled and appropriately addressed.

19.11 Deviation Records and Traceability

Deviation and waiver records shall be maintained as controlled safety records.
The records shall provide traceability from the original requirement or process through the deviation, technical justification, safety and risk assessment, compensating measures, approval, validity period and final closure.
The organization shall maintain sufficient objective evidence to demonstrate that deviations were identified, assessed and controlled throughout their lifecycle.

19.12 Responsibilities and Escalation

The person identifying a deviation shall ensure that it is reported through the defined process. The responsible technical function shall perform or coordinate the technical analysis, while the AI Safety Manager shall ensure that safety-relevant deviations are appropriately assessed and controlled.
The applicable approval authority shall have the competence and authority required to accept the deviation. Deviations with significant or unresolved safety implications shall be escalated to the appropriate management or safety authority.
No individual shall approve a deviation solely on the basis of their involvement in creating or implementing the deviation where this would compromise the required independence of the decision.

19.13 Review and Continual Improvement

Deviation and Waiver Management shall be periodically reviewed as part of the AI Safety Management System.
The review shall consider recurring deviations, repeated waiver requests, overdue closures, deviations associated with incidents or findings, and the effectiveness of compensating measures.
Recurring or systematic deviations shall be evaluated to determine whether they indicate weaknesses in requirements, processes, competence, tools, resources or the overall AI Safety Management System.
A deviation is an exception that must be controlled; it is not an alternative to fulfilling the applicable safety requirements. A waiver provides only a formally justified and time- or criterion-limited exception and shall never be used to circumvent mandatory safety requirements.

20 Configuration Management

20.1 General

All safety-relevant AI work products, configurations and supporting artefacts shall be subject to configuration management throughout their applicable lifecycle.

Configuration Management shall ensure that the organization can uniquely identify, control, reproduce and trace the configuration on which safety-related development, verification, validation, assessment and release decisions are based.

Configuration Management shall apply to development, verification and validation, safety assessment, release, deployment, operation, maintenance and change activities. The level of configuration control shall be appropriate to the safety relevance and complexity of the affected item.

The configuration management system shall ensure that the organization can determine at any time which versions of relevant work products, models, datasets, software, tools, parameters and environments belong to a particular assessed or released configuration.

20.2 Configuration Items

Configuration items shall be identified according to their relevance to the AI Safety Management System and the applicable project.

For an AI system, configuration items may include requirements and safety plans, datasets and data-processing configurations, AI models and model weights, source code and scripts, software libraries, tools and tool configurations, model and inference parameters, test cases and test results, simulation and test environments, assessment reports, safety cases and release packages.

Configuration control shall also be applied to other work products where changes to those work products could affect safety-related decisions or the validity of objective evidence.

The organization shall define which configuration items are subject to mandatory control and shall maintain this definition throughout the project lifecycle.

20.3 Configuration Identification

Each configuration item shall have a unique and unambiguous identification.

Identification shall allow the applicable version, status and relationship to other configuration items to be determined.

For AI models, identification shall distinguish at least the released model version from development or experimental versions. Where relevant, the identification shall also establish the relationship between the model, its training and validation datasets, parameters, software environment and deployment configuration.

The configuration identification scheme shall prevent ambiguity between different versions or configurations that could otherwise result in the use of incorrect or non-approved safety evidence.

20.4 Version and Revision Control

Changes to configuration items shall be managed through controlled version and revision mechanisms.

The organization shall ensure that previous versions remain identifiable where required for traceability, investigation, reproducibility or regulatory purposes. Current and obsolete versions shall be clearly distinguishable.

Version control shall apply not only to source code and documents but also to AI-specific configuration elements such as datasets, models, model weights, parameters, preprocessing and postprocessing configurations, and relevant software and tool environments.

A released configuration shall therefore be reproducible from uniquely identified configuration items rather than from informal descriptions or uncontrolled copies.

20.5 Baseline Management

Configuration baselines shall be established at defined lifecycle stages where a stable and controlled configuration is required.

A baseline shall represent an explicitly identified set of configuration items that provides the approved basis for subsequent activities.

Baselines may be established for requirements, AI models, datasets, verification and validation, safety assessment and product release, as appropriate to the project.

Once established, a baseline shall only be changed through the defined Change Management process. The relationship between a baseline and its associated verification, validation and safety assessment evidence shall remain traceable.

20.6 Configuration Status and Change History

The status of safety-relevant configuration items shall be controlled throughout their lifecycle.

The organization shall maintain sufficient change history to determine what was changed, when the change occurred, why the change was made, who authorized it and which resulting configuration was affected.

Configuration status shall distinguish, as applicable, development, review, verification, validation, approved, released, superseded and obsolete configurations.
This status information shall support the determination of whether a particular configuration item is permitted to be used for safety-related activities or operational deployment.

20.7 Configuration Control and Change Management

Changes to controlled configuration items shall be performed in accordance with the established Change Management process defined in Chapter 18.
A configuration change shall not be introduced into an approved baseline without appropriate impact analysis, classification and approval.
Where a change affects previously established safety evidence, the relationship between the changed configuration and the affected verification, validation or safety assessment shall be evaluated.
Configuration Management and Change Management shall therefore operate as closely linked processes:
Change → Configuration Identification → Impact Analysis → Implementation → V&V → Safety Assessment → Approved Baseline → Release

20.8 Access Control and Protection

Access to safety-relevant configuration items shall be controlled according to defined responsibilities and authorization.
The organization shall protect configuration items against unauthorized modification, deletion, substitution or corruption.
Particular protection shall be applied to safety-relevant AI assets such as approved models, model weights, datasets, safety evidence and released configurations.
Where configuration items are maintained in distributed or externally hosted environments, the organization shall ensure that applicable access, integrity, version and traceability controls remain effective.

20.9 Traceability and Reproducibility

Configuration Management shall provide traceability between related configuration items and safety evidence.

For an AI system, this shall allow the organization to establish, as applicable, which requirements were applicable to a particular model, which datasets and training configuration were used, which software and tools supported development and verification, which test environment was used, and which verification, validation and safety assessment results supported the corresponding release.

The organization shall maintain sufficient information to reproduce the relevant configuration and, where technically practicable, the results of safety-relevant activities.

Reproducibility shall take into account not only the model and source code but also relevant datasets, parameters, software dependencies, tool versions, hardware or accelerator configuration, randomization or seed information where relevant, and the applicable execution environment.

20.10 Configuration and Safety Evidence

Safety-related evidence shall be associated with the configuration for which it was generated.

Verification and validation results, assessment reports, safety cases and release decisions shall therefore identify or reference the applicable configuration baseline.

Evidence generated for one configuration shall not automatically be assumed to remain valid for another configuration following a change.

Where configuration changes affect the validity of existing evidence, the organization shall apply the applicable impact analysis and re-verification, re-validation or re-assessment process.

20.11 Configuration Audit and Verification

The organization shall perform configuration checks or audits at appropriate lifecycle stages to confirm that the documented configuration corresponds to the actual configuration.

Configuration verification shall confirm, as applicable, that required configuration items are present, correctly identified, at the appropriate version, and consistent with the approved baseline.

Particular attention shall be given to the configuration used for final verification, validation, safety assessment and release.

Any discrepancy between the documented and actual configuration shall be recorded and assessed for its potential impact on safety evidence and release status.

20.12 Release Configuration

Each approved release shall have a uniquely identifiable configuration baseline.
The release configuration shall identify all configuration items necessary to define the released AI system and its relevant operating environment.
The organization shall ensure that only an approved configuration is released for its intended operational use.
Where applicable, the release configuration shall include the AI model, model weights, software, datasets or dataset references, parameters, hardware configuration, deployment configuration and relevant safety and cybersecurity configuration.
The exact relationship between the released configuration and the supporting safety evidence shall remain traceable.

20.13 Configuration Records

Configuration Management records shall be maintained as controlled records of the AI Safety Management System.
Records shall provide objective evidence of configuration identification, version history, baselines, changes, approvals, status and relationships between configuration items.
The records shall be retained for the applicable period and protected against unauthorized alteration or loss.

20.14 Responsibilities

The **Configuration and Change Manager**, or an appropriately designated responsible person, shall ensure that the Configuration Management process is established and maintained.
Project and technical functions shall be responsible for correctly identifying and controlling the configuration items within their areas of responsibility.
The **AI Safety Manager** shall ensure that configuration control is adequate to support safety-related development, V&V, assessment and release decisions.

All personnel contributing to safety-relevant AI development or assessment shall comply with the applicable configuration-control requirements.

20.15 Review and Continual Improvement

The effectiveness of Configuration Management shall be periodically reviewed as part of the AI Safety Management System.

The review shall consider configuration discrepancies, unauthorized changes, reproducibility problems, incomplete traceability, audit findings and incidents related to incorrect or uncontrolled configurations.

Identified weaknesses shall be addressed through corrective action and, where appropriate, improvements to configuration identification, baseline management, access control, traceability and reproducibility.

The objective of Configuration Management is to ensure that the exact configuration on which safety decisions are based can be uniquely identified, controlled, traced and, where required, reproduced throughout the AI lifecycle.

21 Toolchain Management

21.1 General

Tools used for safety-relevant AI activities shall be identified, evaluated and controlled throughout their applicable lifecycle.

Toolchain Management shall ensure that tools used to generate, transform, analyse, verify, validate, assess or otherwise support safety-relevant AI work products are suitable for their intended use and are used under controlled conditions.

The organization shall determine the level of control required for each tool based on its intended use, safety relevance and potential influence on safety-related results or decisions.

Toolchain Management shall apply to tools used throughout the AI lifecycle, including requirements management, data preparation, model development and training, simulation, verification and validation, analysis, safety assessment, configuration management and release activities.

21.2 Tool Identification and Intended Use

Safety-relevant tools shall be uniquely identified and their intended use shall be documented.
The tool description shall be sufficient to establish the function performed by the tool, the activities in which it is used and the safety-related work products or decisions that may depend on its output.
The organization shall distinguish, where relevant, between tools that directly contribute to a safety-related result and tools that are used for supporting or administrative purposes.
The intended use shall define the applicable operating conditions, relevant inputs and expected outputs of the tool.

21.3 Tool Safety Relevance

The organization shall assess the safety relevance of each tool used for safety-related AI activities.
The assessment shall consider whether an incorrect, incomplete, corrupted or otherwise unreliable tool output could affect a safety requirement, model, dataset, verification or validation result, safety assessment, safety argument or release decision.
The level of control shall be proportionate to the potential consequences of an incorrect tool result.
A tool shall not be considered safety-irrelevant solely because it is used indirectly. Where an indirect tool output contributes to safety-related evidence or decisions, its relevance shall be considered accordingly.

21.4 Tool Evaluation

Safety-relevant tools shall be evaluated with respect to their suitability for the intended use.
The evaluation shall consider, as applicable, the tool's functional correctness, limitations, known anomalies, version, configuration, interfaces, dependencies and reproducibility.
The organization shall determine whether the available information and evidence provide sufficient confidence that the tool can be used for its intended safety-relevant purpose.
Depending on the safety relevance, appropriate confidence measures may include vendor information, documented tool behaviour, comparison with reference results, review of tool outputs, independent checks, testing, qualification or other suitable measures.
Formal tool qualification shall be applied where required by the applicable process, standard, certification scheme or project-specific safety strategy.

21.5 Tool Version and Configuration

The version and relevant configuration of safety-relevant tools shall be controlled.
Tool configurations may include configuration files, scripts, plug-ins, libraries, compiler settings, execution parameters, environment variables, model settings or other settings that can influence the generated result.
The organization shall ensure that the tool configuration used to generate safety-relevant evidence can be identified and, where necessary, reproduced.
A change of tool version or configuration shall not be assumed to be safety-neutral. Its potential impact shall be evaluated according to the applicable Change Management process.

21.6 Toolchain Dependencies

AI safety activities frequently depend on a chain of interconnected tools rather than on a single tool.
The organization shall consider relevant dependencies within the toolchain where the output of one tool becomes the input to another tool or contributes to a subsequent safety-related activity.
Where such dependencies may influence the validity of safety evidence, the relevant tool versions, configurations and interfaces shall be traceable.
This may include, for example, the relationship between data-processing tools, training frameworks, compilers, simulation environments, testing frameworks, analysis tools and reporting systems.
The objective shall be to ensure that the relevant toolchain configuration can be reconstructed for a safety-relevant result.

21.7 Confidence in Tool Outputs

Where the correctness of a tool output can influence a safety-related result or decision, appropriate measures shall be established to provide confidence in that output.
The measures shall be proportionate to the safety relevance and may include independent review, cross-checking, testing, comparison against known or reference results, monitoring of tool behaviour, alternative methods or other appropriate controls.
The organization shall document the rationale for the selected measures where the tool has significant influence on safety-related evidence.

Confidence in a tool output shall not be based solely on the fact that the tool is widely used, commercially available or provided by a recognized supplier.

21.8 Tool Anomalies and Limitations

Known tool anomalies, limitations and restrictions relevant to the intended use shall be identified and considered.
Where a tool limitation may affect the validity of a safety-related result, appropriate compensating measures shall be established or the tool shall not be used for the affected purpose.
Unexpected tool behaviour identified during development, testing, assessment or operation shall be evaluated for its potential impact on previously generated safety evidence.
Where necessary, affected results shall be re-evaluated, re-generated or independently confirmed.

21.9 Tool Changes

Changes to safety-relevant tools, including new versions, configuration changes, updates, patches, replacement tools or changes to dependencies, shall be controlled through Configuration Management and Change Management.
The organization shall determine whether a tool change can affect previously generated results or safety evidence.
Where such an impact is possible, the affected activities shall be identified and additional verification, validation, analysis or assessment shall be performed as necessary.
Particular attention shall be given to changes affecting tools used for AI training, model generation, data transformation, compilation, simulation, verification, validation or safety assessment.

21.10 Tool Reproducibility

The organization shall maintain sufficient information to reproduce safety-relevant tool executions where reproducibility is required.
This shall include, as applicable, the tool version, configuration, relevant dependencies, input data, execution environment and relevant parameters.
For AI development, reproducibility may additionally depend on the training framework, libraries, hardware or accelerator, dataset version, randomization settings and other environmental conditions.

Where exact reproduction is not technically practicable, the organization shall establish an appropriate alternative means of demonstrating the validity and traceability of the resulting safety evidence.

21.11 Tool Output and Safety Evidence

Safety-relevant tool outputs shall be identifiable and traceable to the tool and configuration that generated them.
Where tool outputs are used as objective evidence, the organization shall ensure that their origin, relevant configuration and relationship to the applicable work product can be demonstrated.
Tool-generated results shall be subject to appropriate review where their correctness or interpretation can affect safety-related decisions.
The use of a tool shall therefore not remove the responsibility of competent personnel for evaluating and approving the resulting safety evidence.

21.12 Toolchain Integrity and Access Control

Access to safety-relevant tools and tool configurations shall be controlled according to defined responsibilities and authorization.
The organization shall protect the toolchain against unauthorized modification, replacement or manipulation where such activities could affect safety-related results.
Particular attention shall be given to tools and environments used for model training, dataset processing, model conversion, compilation, testing and release generation.
Where externally provided or cloud-based tools are used, the organization shall assess the applicable dependencies, versioning, access and configuration aspects necessary to maintain confidence in the resulting safety evidence.

21.13 Tool Records

The organization shall maintain appropriate records for safety-relevant tools.
The records shall provide sufficient information to establish the tool's identity, intended use, safety relevance, applicable version and configuration, known limitations or anomalies, evaluation results and any additional confidence measures applied.

Where formal qualification or another specific evaluation has been performed, the associated evidence shall be maintained as part of the tool records.

21.14 Responsibilities

The responsible engineering function shall identify and manage tools used within its safety-relevant activities.
The **AI Safety Manager** shall ensure that the overall Toolchain Management process is established and that safety-relevant tool dependencies are appropriately considered.
Tool users shall comply with the approved tool versions, configurations and usage conditions.
Where a tool has a significant influence on safety evidence, the responsible technical function shall ensure that the required evaluation, confidence measures and configuration controls have been completed before the tool is used for the applicable safety-related activity.

21.15 Review and Continual Improvement

The effectiveness of Toolchain Management shall be periodically reviewed.
The review shall consider tool anomalies, unexpected results, reproducibility issues, tool changes, audit findings and experience with the effectiveness of the applied confidence measures.
Where a tool or toolchain is found to be unsuitable or insufficiently controlled for its intended safety-relevant use, the organization shall initiate appropriate corrective action and determine whether previously generated safety evidence is affected.
The objective of Toolchain Management is to ensure that tools influencing safety-related AI activities are appropriately evaluated, controlled, traceable and used with a level of confidence proportionate to their potential impact on safety.

22 Traceability

22.1 General

Traceability shall be established and maintained between relevant lifecycle work products, decisions, configurations and safety evidence throughout the AI lifecycle.

The purpose of Traceability is to ensure that the relationship between safety objectives, requirements, AI development activities, data, models, verification and validation results, safety assessments and release decisions can be demonstrated.

Traceability shall be established to a level appropriate to the safety relevance, complexity and architecture of the AI system. It shall support the demonstration that applicable safety requirements have been addressed and that the evidence used for safety decisions relates to the correct configuration.

Traceability shall be maintained throughout development, assessment, release, operation and change activities.

22.2 Traceability Chain

Where applicable, the organization shall establish a traceability chain connecting the relevant safety and engineering work products:

Safety Objective

↓

Safety Requirement

↓

AI Requirement

↓

Data Requirement

↓

Model / Implementation

↓

Verification

↓

Validation

↓

Safety Evidence

↓

Release Decision

The exact structure of the chain shall be adapted to the architecture and lifecycle of the applicable AI system. Not every AI project will require every element to be represented as a separate work product.

Where an AI component contributes to a safety-related function, the traceability chain shall also maintain the relationship to the applicable Functional Safety, SOTIF and Cybersecurity requirements and evidence.

22.3 Bidirectional Traceability

Traceability shall be maintained in both directions where necessary to support safety analysis and assessment.
Forward traceability shall demonstrate how safety objectives and requirements are translated into AI requirements, data, implementation and corresponding verification and validation activities.
Backward traceability shall demonstrate how a model, dataset, test result, safety finding or released configuration relates to the requirements, assumptions and safety objectives for which it was developed.
This bidirectional capability shall allow the organization to determine the potential impact of a change to any relevant configuration item.

22.4 Requirements Traceability

Safety-related requirements shall be traceable to their source, implementation and associated verification or validation evidence.
The organization shall be able to determine which AI requirements originate from safety objectives, system requirements, Functional Safety requirements, SOTIF considerations, Cybersecurity requirements, regulatory requirements or other applicable sources.
Where requirements are derived from assumptions, operating conditions or identified risks, the corresponding relationship shall also be maintained.
Untraced or incompletely traced safety requirements shall be identified and addressed before the applicable lifecycle gate or release decision.

22.5 Data and Model Traceability

For AI systems, traceability shall extend to the data and model elements that influence safety-relevant behaviour.

The organization shall be able to establish, where applicable, the relationship between relevant data requirements, datasets, data-processing activities, training configurations, model versions and resulting verification and validation evidence.
Changes to training or validation data shall therefore be traceable to the affected model versions and to the safety evidence that may depend on those models.
This traceability shall support the determination of whether a change to data, model or processing configuration requires additional verification, validation or safety assessment.

22.6 Verification and Validation Traceability

Verification and validation activities shall be traceable to the requirements, configurations and acceptance criteria against which they are performed.
Test results shall identify the applicable test case or method, configuration, relevant inputs and resulting outcome to the extent necessary to demonstrate their validity.
Where a verification or validation result identifies an anomaly, limitation or finding, the finding shall be traceable to the affected requirement, configuration and subsequent safety decision.
The organization shall be able to demonstrate which verification and validation evidence supports a particular safety conclusion or release.

22.7 Safety Evidence Traceability

Safety evidence shall be traceable to the requirements, analyses, configurations and activities from which it was derived.
This includes, as applicable, verification and validation results, risk assessments, SOTIF analyses, cybersecurity assessments, safety assessments, safety arguments and operational evidence.
Safety evidence shall be associated with the configuration for which it was generated. Evidence from a previous configuration shall not automatically be considered valid for a changed configuration.
Where the applicability of existing evidence is affected by a change, the applicable traceability relationships shall be reviewed and updated.

22.8 Traceability and Safety Case

Traceability shall provide a structured basis for the development and maintenance of the safety case.

The safety case shall be supported by identifiable relationships between safety claims, supporting arguments, assumptions, requirements and objective evidence.
Where a safety claim depends on several engineering disciplines, the relevant evidence and interfaces shall remain traceable.
Traceability shall therefore support not only the collection of evidence but also the demonstration of the logical relationship between the evidence and the corresponding safety conclusion.

22.9 Traceability and Change Impact Analysis

Traceability shall support efficient and systematic change impact analysis.
When a configuration item is changed, the organization shall be able to identify the requirements, assumptions, models, datasets, tests, assessments and release decisions potentially affected by the change.
Conversely, where a requirement, safety objective, assumption or operating condition is changed, the organization shall be able to identify the affected implementation, model, data, V&V activities and safety evidence.
This capability shall form an important input to the Change Management process defined in Chapter 18.

22.10 Traceability and Configuration Management

Traceability shall be linked to Configuration Management.
The organization shall ensure that traceability relationships identify the relevant versions or baselines of configuration items where this is necessary to establish the validity of the associated evidence.
For example, a verification result shall be traceable to the specific model, dataset, software, tool configuration and test environment used to generate the result where these elements are safety-relevant.
This shall ensure that the relationship between evidence and the actual assessed or released configuration remains demonstrable.

22.11 Traceability Gaps

Traceability gaps affecting safety-relevant work products shall be identified and assessed.

Where complete traceability cannot reasonably be established, the organization shall document the limitation and assess its potential impact on safety evidence and the safety decision.
Critical safety conclusions shall not rely on undocumented or unverifiable relationships.
Where necessary, additional analysis, evidence or verification shall be performed to compensate for an identified traceability gap.

22.12 Traceability Records and Tools

Traceability may be implemented using requirements-management systems, configuration-management systems, lifecycle-management tools, model-management systems, test-management systems or controlled documentation, provided that the resulting relationships remain reliable and accessible.
The organization shall ensure that traceability information itself is configuration-controlled where its loss or alteration could affect safety-related decisions.
The integrity of traceability information shall be protected against unauthorized modification.

22.13 Traceability Review

Traceability shall be reviewed at appropriate lifecycle gates and before significant safety decisions or releases.
The review shall confirm that the required relationships have been established, that safety-relevant requirements are adequately covered, and that the supporting evidence corresponds to the applicable configuration.
Particular attention shall be given to traceability completeness following significant changes, integration activities, re-validation or safety re-assessment.

22.14 Responsibilities

The relevant engineering functions shall establish and maintain traceability for the work products under their responsibility.
The **AI Safety Manager** shall ensure that the overall Traceability framework is defined and that sufficient traceability exists to support safety management, assessment and release decisions.

The **System Engineer**, **Functional Safety Manager**, **SOTIF Responsible**, **Cybersecurity Responsible**, **V&V Responsible**, **Data Engineer** and **AI/ML Development functions** shall maintain the traceability relationships relevant to their respective activities.
The responsible release authority shall ensure that the required traceability has been established before approving a safety-relevant release.

22.15 Continual Improvement

The effectiveness of Traceability shall be periodically reviewed.
The review shall consider traceability gaps, audit findings, difficulties encountered during change impact analysis, assessment activities, safety-case development and incident investigations.
Where weaknesses are identified, the organization shall improve the applicable processes, tools, responsibilities or data structures.
The objective of Traceability is to ensure that the organization can demonstrate, in both directions, how safety objectives and requirements are implemented and evidenced, and how every safety-relevant result or released configuration can be related back to its underlying requirements, assumptions and safety decisions.

23 Documentation and Evidence Management

23.1 General

Safety-related documentation and objective evidence shall be established, controlled and maintained throughout the AI lifecycle.
The organization shall establish and maintain a controlled evidence repository or equivalent controlled documentation system in which safety-related evidence can be securely stored, identified, retrieved and associated with the applicable project, product and configuration.
Documentation and Evidence Management shall ensure that evidence used to demonstrate compliance, support safety decisions or substantiate safety claims remains available, traceable, complete and protected against unauthorized modification.

The level of documentation and evidence control shall be appropriate to the safety relevance, complexity and lifecycle stage of the AI system.

23.2 Documentation and Evidence Structure

The evidence structure shall provide an organized representation of the information required to demonstrate effective implementation of the AI Safety Management System and the safety of the applicable AI system.

Depending on the project and applicable processes, the evidence structure shall include management and organizational evidence, competence records, requirements, risk analyses, data and model evidence, verification and validation results, integration evidence and relevant Functional Safety, SOTIF and Cybersecurity interface evidence.

It shall also include applicable change and deviation records, audit results, safety assessment results and release decisions.

The evidence structure shall be defined in a manner that supports systematic retrieval and traceability rather than merely serving as a document archive.

23.3 Evidence Identification

Each safety-relevant document or evidence item shall be uniquely identifiable or otherwise unambiguously associated with the applicable work product.

Identification shall allow the organization to determine, as applicable, the document or evidence type, version, status, responsible function, applicable project or product and relationship to the relevant configuration.

Evidence generated by external organizations or suppliers shall be incorporated into the controlled evidence structure where it is used to support a safety-related conclusion.

23.4 Evidence Review and Approval

Safety-related documentation and evidence shall be subject to appropriate review before it is used as the basis for a formal safety decision.

The required level of review and approval shall be determined according to the safety relevance of the document or evidence.

Where evidence supports a safety assessment, safety case or release decision, its applicability, completeness and relationship to the relevant configuration shall be confirmed.
Approval status shall be distinguishable from draft or working status so that uncontrolled or incomplete information cannot inadvertently be used as approved safety evidence.

23.5 Evidence and Configuration

Safety evidence shall be associated with the configuration for which it was generated.
The evidence management system shall therefore maintain the relationship between evidence and relevant configuration items, including models, datasets, software, hardware, tools, test environments and release configurations where applicable.
A document or test result shall not be considered valid for a different configuration solely because the document itself has not changed.
Where a configuration change may affect the validity of existing evidence, the applicable Change Management and Re-Assessment processes shall be initiated.

23.6 Evidence Traceability

Documentation and evidence shall be traceable to the applicable requirements, risks, assumptions, activities and safety decisions.
The evidence structure shall support the traceability chain established in Chapter 22 and shall allow the organization to demonstrate the relationship between requirements, implementation, V&V results, safety assessment and release decisions.
Where evidence is generated by a tool or automated process, the relevant tool, configuration and input information shall be identifiable where necessary to establish the validity and reproducibility of the result.

23.7 Evidence Integrity and Protection

Safety-related evidence shall be protected against unauthorized modification, deletion, corruption or substitution.
Access shall be controlled according to defined responsibilities and authorization.

The organization shall ensure that approved evidence cannot be altered without detection and that changes to controlled documentation are managed through the applicable configuration and change-control processes.
Where electronic repositories are used, appropriate measures shall be established to protect data integrity, access and availability.

23.8 Evidence Retrieval and Availability

Safety-related evidence shall remain identifiable and retrievable for the period required by the applicable project, contractual, regulatory, certification or organizational requirements.
The evidence repository shall allow authorized personnel to retrieve the information required for safety assessments, audits, change impact analyses, incident investigations and release decisions.
Evidence shall remain accessible in a form that allows its content, status and relationship to the applicable configuration to be understood.

23.9 Evidence Completeness

The organization shall establish criteria for determining whether the required safety evidence is complete for a given lifecycle stage or release.
Evidence completeness shall be assessed against applicable requirements, lifecycle entry and exit criteria, safety plans, assessment criteria and release conditions.
Missing, incomplete, contradictory or unresolved evidence shall be identified and managed through the applicable review, deviation, risk or change-management processes.
The absence of evidence shall not be interpreted as evidence that the corresponding activity or requirement has been fulfilled.

23.10 Evidence from AI-Specific Activities

AI-specific evidence shall be maintained in a manner that supports reproducibility and safety assessment.
Depending on the AI system, this may include evidence related to dataset definition and quality, data preparation, model architecture, model versions and weights, training configuration, model performance, limitations, robustness, scenario coverage, validation results and operational monitoring.

The evidence shall be sufficient to establish which data, model and configuration were used for the relevant safety-related activity.

23.11 Interface Evidence

Evidence supporting the interfaces between AI Safety and other safety-related disciplines shall be maintained as part of the controlled evidence structure.
This includes, where applicable, evidence supporting Functional Safety, SOTIF and Cybersecurity interfaces, including assumptions, requirements, analyses, reviews, assessments and decisions resulting from these interfaces.
Interface evidence shall remain traceable to the responsible disciplines and the applicable system or product configuration.
Evidence generated within one discipline shall not automatically be considered sufficient evidence for another discipline. Its applicability shall be assessed according to the respective safety objective and decision.

23.12 Safety Assessment and Release Evidence

Safety assessment results and release decisions shall be maintained as controlled records.
The evidence shall demonstrate the basis for the safety conclusion and identify the configuration to which the conclusion applies.
Release decisions shall be traceable to the applicable requirements, risk assessments, V&V results, open findings, deviations and safety assessments.
Where a release is subject to restrictions, assumptions, residual risks or compensating measures, these shall be documented together with the release decision.

23.13 Evidence Retention and Archiving

Safety-related documentation and evidence shall be retained for the period defined by applicable legal, regulatory, contractual, certification and organizational requirements.
Archived evidence shall remain identifiable and protected against unauthorized alteration.
Where obsolete or superseded evidence is retained, its status shall be clearly indicated so that it cannot be confused with the currently applicable evidence.

Evidence shall not be deleted solely because it has been superseded where its historical value is required for traceability, safety analysis, incident investigation, conformity assessment or regulatory purposes.

23.14 Evidence Management Responsibilities

The **AI Safety Manager** shall ensure that the overall Documentation and Evidence Management process is established and maintained.

Each responsible function shall ensure that safety-related documentation and evidence generated within its area is complete, correctly identified, reviewed as required and transferred to the controlled evidence repository.

The **Configuration and Change Manager** shall ensure that documentation and evidence are appropriately integrated with Configuration Management and Change Management.

The release authority shall ensure that the evidence required to support a release decision is available, traceable and appropriately approved before release.

23.15 Review and Continual Improvement

The effectiveness of Documentation and Evidence Management shall be periodically reviewed as part of the AI Safety Management System.

The review shall consider evidence gaps, retrieval difficulties, obsolete or uncontrolled documents, configuration mismatches, audit findings, assessment findings and experience from change and incident investigations.

Where weaknesses are identified, corrective actions shall be established to improve the structure, integrity, accessibility and effectiveness of the evidence management process.

The objective of Documentation and Evidence Management is to ensure that safety decisions are supported by controlled, identifiable, traceable and retrievable objective evidence and that the organization can demonstrate the basis for its safety conclusions throughout the AI lifecycle.

24 Internal Audits

24.1 General

The organization shall establish and maintain an internal audit process to determine whether the AI Safety Management System is effectively implemented, maintained and continually improved.
Internal audits shall provide an independent and systematic evaluation of whether defined AI safety processes are being followed, responsibilities and authorities are effectively exercised, required evidence is available, and applicable lifecycle and safety criteria are fulfilled.
Audits shall cover the management system as well as relevant project and product activities. The scope, depth and frequency of audits shall be appropriate to the safety relevance, complexity, organizational structure and lifecycle status of the AI systems concerned.
Internal audits shall not be limited to reviewing the existence of documented processes. They shall also evaluate whether the processes are effectively implemented and whether they produce the intended safety outcomes and objective evidence.

24.2 Audit Planning

The organization shall establish an audit programme covering the relevant elements of the AI Safety Management System and, where applicable, individual AI projects or products.
Audit planning shall consider the safety relevance and complexity of the activities, results of previous audits, identified risks, changes to the organization or technology, significant incidents, deviations, previous assessment findings and areas with known or emerging weaknesses.
Audits shall be performed at appropriate lifecycle stages and shall provide sufficient coverage before significant safety decisions or releases where this is required by the applicable safety plan or management-system arrangements.

24.3 Audit Scope and Criteria

Each audit shall have a defined scope and documented audit criteria.
The criteria may include applicable internal procedures, AI Safety Management System requirements, project-specific safety plans, applicable standards and regulatory or contractual

requirements, lifecycle entry and exit criteria, approved responsibilities and applicable safety objectives.

The audit scope shall identify the organizational functions, processes, projects, products, lifecycle phases, work products or interfaces subject to examination.

24.4 Audit Evaluation

Internal audits shall evaluate, as applicable, whether AI safety processes are implemented as defined and whether assigned responsibilities and authorities are effectively fulfilled.

The audit shall consider the availability and adequacy of required objective evidence, compliance with lifecycle entry and exit criteria, effectiveness of change and configuration management, appropriate management of deviations and waivers, fulfilment of competence requirements and the effectiveness of interfaces with Functional Safety, SOTIF, Cybersecurity and other relevant disciplines.

Particular attention shall be given to whether documented processes are actually reflected in engineering practice and whether safety decisions can be supported by objective evidence.

24.5 Auditor Competence and Independence

Internal audits shall be performed by personnel with appropriate competence in auditing and in the subject matter being audited.

Auditors shall have sufficient independence from the activities they audit to ensure an objective evaluation. A person shall not audit their own work where this would compromise the objectivity or credibility of the audit.

For technically complex or safety-critical activities, the audit team shall collectively possess the technical and safety competence necessary to evaluate the subject matter.

24.6 Audit Evidence and Records

Audit conclusions shall be based on objective evidence.

Audit evidence may include examination of documented information, interviews, reviews of work products, configuration records, traceability information, V&V results, safety assessments, release records and, where appropriate, observation of relevant activities.

Audit records shall identify the audit scope and criteria, audited activities, evidence examined, findings, conclusions and responsible persons.
The audit records shall be maintained as controlled records in accordance with Chapter 23, Documentation and Evidence Management.

24.7 Audit Findings and Classification

Audit findings shall be documented and classified according to their significance and potential impact on the effectiveness of the AI Safety Management System or the safety of the applicable AI system. Findings may identify, for example, nonconformities, weaknesses in process implementation, missing or inadequate evidence, ineffective controls, insufficient competence, traceability gaps or opportunities for improvement.
Where an audit finding may have a direct or indirect impact on product safety, the responsible organization shall determine whether additional risk assessment, containment, change analysis, re-verification, re-validation or safety assessment is required.
Audit findings shall not be closed merely because a corrective action has been planned. Closure shall be based on objective evidence demonstrating that the required action has been effectively implemented and, where appropriate, that its effectiveness has been verified.

24.8 Corrective Actions and Closure

Audit findings shall be assigned to responsible persons or functions and shall be tracked to closure. The responsible function shall determine the cause of the finding and establish appropriate corrective action. Where a finding indicates a systemic weakness, the corrective action shall address the underlying cause rather than only the individual observation.
The status of significant findings shall be monitored by the AI Safety Manager or designated responsible function. Overdue or ineffective corrective actions shall be escalated according to the defined authority and escalation process.

24.9 Follow-Up Audits

Where necessary, follow-up audits or reviews shall be performed to determine whether corrective actions have been effectively implemented.
The extent of follow-up shall depend on the significance of the finding and the potential safety impact.

If corrective actions are found to be ineffective, the finding shall remain open or be re-opened, and additional action shall be initiated.

24.10 Relationship to Safety Assessments and External Assessments

Internal audits are an element of the AI Safety Management System and shall complement, but not replace, project-level safety assessments, independent assessments, certification activities or regulatory conformity assessments.

Audit results may provide relevant input to safety assessments, management reviews, change impact analyses and continual improvement. However, an internal audit finding or audit conclusion shall not by itself constitute evidence of product safety unless the applicable assessment process explicitly establishes such use.

24.11 Audit Responsibilities

Top Management shall provide the resources and authority necessary for the effective implementation of the internal audit process.

The AI Safety Manager shall coordinate the overall audit programme and ensure that significant findings are appropriately addressed and escalated.

Process and project owners shall provide access to relevant information, support the audit activities and ensure that assigned corrective actions are implemented.

Auditors shall conduct audits objectively, document sufficient evidence to support their conclusions and communicate findings through the defined reporting process.

24.12 Audit Records and Continual Improvement

Audit results shall be reviewed for recurring findings, systemic weaknesses and trends that may indicate deficiencies in the AI Safety Management System.

Information from audits shall be considered together with incidents, deviations, change records, assessment results, operational feedback and management reviews to identify opportunities for improvement.

Where audit results indicate that an existing process, responsibility, competence requirement or control is ineffective, the relevant management-system documentation shall be reviewed and updated through the applicable change-control process.

The objective of Internal Audits is to provide systematic and independent evidence that the AI Safety Management System is effectively implemented, that safety-related responsibilities and controls are functioning as intended, and that identified weaknesses are appropriately corrected and prevented from recurring.

25 External Assessment and Audit

25.1 General

Where external assessment, certification, conformity assessment, regulatory assessment or other independent audit activities are applicable, the organization shall establish and maintain arrangements to support the effective and controlled execution of such activities.

External assessments shall be managed as an integral part of the AI Safety Management System while maintaining the required independence of the external assessment or certification body.

The organization shall ensure that the scope, applicable criteria, responsibilities, evidence and communication channels for each external assessment are clearly defined and controlled.

25.2 Assessment Scope and Criteria

Before an external assessment or audit, the organization shall establish and document the applicable assessment scope.

The scope shall identify, as applicable, the organization, locations, processes, AI systems, products, lifecycle activities, management-system elements and technical areas subject to assessment.

The applicable assessment criteria shall also be identified, including relevant standards, regulatory requirements, certification schemes, contractual requirements and defined assessment criteria.

Any limitations, exclusions or boundaries of the assessment scope shall be documented and communicated to the relevant parties.

25.3 Assessment Preparation

The organization shall ensure that the documentation and objective evidence required to support the external assessment are available and controlled.

Relevant evidence shall be identifiable, retrievable and associated with the applicable configuration and assessment scope in accordance with Chapter 23, Documentation and Evidence Management.

Assessment preparation shall include confirmation that relevant personnel are available, responsibilities are understood and required interfaces between management, engineering, quality and assessment functions are established.

Preparation activities shall not result in the creation of evidence solely for the purpose of an assessment if that evidence does not accurately represent the actual implementation of the AI Safety Management System.

25.4 Communication with the Assessment Body

Responsibilities and authority for communication with the external assessment or certification body shall be formally defined.

The designated responsible function shall coordinate assessment planning, scheduling, information exchange, clarification requests and the controlled provision of assessment evidence.

Technical questions shall be addressed by personnel with appropriate competence in the relevant subject matter. Where necessary, Functional Safety, SOTIF, Cybersecurity, AI/ML, V&V, Quality or other responsible functions shall participate directly in the communication.

Communication with the assessment body shall be accurate, complete and consistent with the approved scope and available objective evidence.

25.5 External Assessment Activities

The organization shall provide the external assessment body with appropriate access to the information, personnel, processes and evidence necessary to conduct the agreed assessment.

Assessment activities shall be supported without compromising the independence or impartiality of the external assessment body.

The organization shall distinguish between factual information provided to the assessment body and safety or conformity conclusions that are the responsibility of the assessment body itself.

An organization shall not represent an assessment as certification or regulatory approval unless the applicable certification or regulatory process has actually been completed and the corresponding decision has been formally issued.

25.6 External Findings and Nonconformities

Findings, observations, nonconformities and other results issued by an external assessment body shall be formally recorded and tracked.
Each finding shall be assigned to a responsible function or person, and its significance and potential impact on the AI Safety Management System or applicable AI product shall be evaluated.
Where an external finding may affect safety-related assumptions, requirements, evidence, configurations, released products or certification status, the applicable Change Management, Deviation Management, Risk Management or Safety Assessment processes shall be initiated as appropriate.

25.7 Corrective Actions

Corrective actions resulting from external assessments shall be documented and managed through a controlled process.
The organization shall determine the cause of significant findings and establish corrective actions appropriate to the nature and potential impact of the finding.
Corrective action records shall provide sufficient evidence to demonstrate what was changed, who was responsible, when the action was completed and how its effectiveness was verified.
Where the assessment body requires formal closure or acceptance of corrective actions, the status and communication with the assessment body shall be documented.

25.8 Changes Affecting the Assessment Scope

Changes that may affect an existing assessment, certification, approval or conformity status shall be identified and evaluated.
Such changes may include significant organizational changes, changes to the AI system or product, changes to the intended use or operating conditions, changes to the applicable standards or regulatory framework, changes to the assessed processes or locations, or significant changes to the AI Safety Management System.

The organization shall determine whether the change requires notification to, consultation with or reassessment by the relevant external assessment or certification body.
Changes shall not be represented as covered by an existing assessment or certification where the applicable scope or conditions have been exceeded.

25.9 External Assessment Evidence

Evidence provided to an external assessment body shall remain under the organization's document and configuration control.
The organization shall maintain records of the evidence submitted, assessment scope, assessment results, findings, correspondence where relevant, corrective actions and formal assessment or certification decisions.
Where evidence is subsequently changed, superseded or withdrawn, the organization shall determine whether the change affects the validity of the assessment conclusions or certification status.

25.10 Responsibilities

Top Management shall ensure that adequate resources and authority are available to support applicable external assessments.
The **AI Safety Manager** shall coordinate the AI safety aspects of external assessments and ensure that relevant findings and assessment results are integrated into the AI Safety Management System.
The designated assessment coordinator shall manage formal communication and coordination with the external assessment or certification body.
Process owners and technical functions shall provide accurate evidence and support the assessment activities within their areas of responsibility.

25.11 Review and Continual Improvement

Results from external assessments shall be reviewed together with internal audit results, safety assessments, incidents, deviations, operational feedback and other relevant information.
Recurring findings or significant assessment issues shall be evaluated for systemic causes and may require changes to processes, responsibilities, competence, documentation, tools or controls.

The organization shall use external assessment results as an input to continual improvement of the AI Safety Management System.

The objective of External Assessment and Audit is to ensure that applicable external assessments, certifications and conformity activities are conducted in a controlled, transparent and evidence-based manner, while maintaining clear responsibilities, effective communication and systematic management of assessment findings and changes affecting the assessed scope.

26 Safety Review and Release

26.1 General

Before an AI system or safety-relevant AI configuration is released for its intended use, the organization shall perform a formal Safety Review.

The Safety Review shall provide a structured and documented evaluation of whether the applicable safety requirements, objectives, risks, limitations and evidence have been adequately addressed and whether the defined release criteria have been fulfilled.

The review shall be performed against the specific configuration intended for release. The organization shall ensure that the configuration subject to the Safety Review is uniquely identified and consistent with the configuration for which the relevant verification, validation and safety evidence have been generated.

Completion of individual engineering activities, including verification or validation, shall not by itself constitute release approval.

26.2 Safety Review Inputs

The Safety Review shall consider the information necessary to support a sound release decision.

Depending on the AI system and applicable safety classification, the review shall consider applicable requirements and safety objectives, risk analyses and risk-control measures, assumptions and operating limitations, data and model baselines, verification and validation results, integration results,

Functional Safety, SOTIF and Cybersecurity interface results, open issues, deviations and waivers, safety assessment results and the applicable evidence package.
The review shall also consider relevant changes since the previous approved configuration and the results of the associated impact analysis and re-assessment activities.

26.3 Review of Safety Requirements and Objectives

The Safety Review shall determine whether the applicable safety requirements have been adequately addressed and whether the defined safety objectives have been fulfilled or otherwise formally addressed through an approved and controlled decision.
The review shall confirm that requirements allocated to the AI system or AI component have appropriate implementation and supporting verification or validation evidence.
Where requirements are subject to assumptions, restrictions or conditions of use, these shall be clearly identified and incorporated into the release conditions where applicable.

26.4 Review of Risk and Residual Risk

The Safety Review shall determine whether identified safety risks have been adequately controlled and whether the resulting residual risks have been evaluated in accordance with the applicable risk acceptance criteria.
Known limitations, uncertainties, assumptions and restrictions affecting the safe use of the AI system shall be reviewed.
Residual risks shall not be considered acceptable solely because of commercial, schedule or resource considerations. Where residual risk remains subject to specific conditions, the corresponding controls, monitoring measures or operational restrictions shall be clearly defined.

26.5 Review of Verification and Validation

The Safety Review shall confirm that the required verification and validation activities have been completed to the extent defined by the applicable safety plan, lifecycle criteria and safety classification.
The review shall consider whether the relevant results were obtained using the configuration intended for release and whether identified anomalies, failed criteria or limitations have been appropriately evaluated.

Where re-verification or re-validation was required following a change, the review shall confirm that the applicable activities have been completed and that their results have been incorporated into the release decision.

26.6 Review of Open Issues

All known open issues relevant to the release shall be identified and assessed.
The organization shall determine whether each issue affects safety requirements, assumptions, risks, V&V results, safety evidence, intended use or release conditions.
An open issue shall not prevent release only where its impact has been appropriately assessed and the resulting decision, restrictions, compensating measures or residual risk have been formally accepted by the authorized responsible party.
Unresolved issues with unacceptable safety impact shall prevent release until appropriately resolved or otherwise controlled through an applicable formal process.

26.7 Review of Deviations and Waivers

All deviations and waivers relevant to the release shall be reviewed for validity, scope, safety impact and approval status.
The Safety Review shall confirm that required deviations or waivers have been formally approved by competent authority and that any associated compensating measures, restrictions and validity periods are defined.
An expired, unauthorized or inadequately justified deviation or waiver shall not be used as a basis for release.

26.8 Review of Interfaces

The Safety Review shall confirm that relevant interfaces with other disciplines have been addressed.
Depending on the system and applicable standards, this shall include interfaces with Functional Safety, SOTIF, Cybersecurity, Software Update Engineering, system engineering, hardware/software development and applicable regulatory or conformity requirements.
The review shall consider whether assumptions, requirements, evidence and safety decisions exchanged between disciplines remain consistent for the release configuration.

26.9 Review of the Evidence Package

The Safety Review shall confirm that the evidence package required to support the release decision is complete, controlled and traceable.

Evidence shall be associated with the configuration intended for release and shall provide an adequate basis for the safety conclusions.

Evidence gaps shall be explicitly identified and assessed. The absence of a required record or evidence item shall not be interpreted as confirmation that the corresponding activity has been completed.

The evidence package shall be maintained in accordance with Chapter 23, Documentation and Evidence Management, and the relevant traceability requirements of Chapter 22.

26.10 Safety Review Independence

The level of independence applied to the Safety Review shall be appropriate to the safety relevance and organizational requirements of the AI system.

Where independent review or confirmation measures are required, the responsible reviewer shall have appropriate competence and sufficient independence from the activities being reviewed.

The person performing the final release approval shall have the authority and competence defined in the applicable Authority and Responsibility Matrix.

26.11 Release Decision

Based on the results of the Safety Review, the authorized release authority shall make a formal release decision.

The decision shall identify, as applicable, the approved product or AI system, configuration or version, intended use, applicable restrictions, relevant residual risks, assumptions, deviations or waivers, and the effective release status.

The release decision shall be explicitly documented as, for example, **approved for release**, **approved with defined restrictions**, or **not approved**, according to the organization's defined release process.

A release decision shall apply only to the configuration and conditions for which the Safety Review was performed.

26.12 Release Configuration and Status

Only the configuration that has been subject to the applicable Safety Review and release approval shall be released.
The released configuration shall be uniquely identifiable and shall be traceable to the relevant model, data, software, hardware, configuration, V&V results, safety assessment and release decision.
Any subsequent safety-relevant change shall be subject to the Change Management process defined in Chapter 18 and shall not be considered covered by the existing release approval without an appropriate impact assessment.

26.13 Release Records

The formal release record shall provide objective evidence of the decision and its basis.
The record shall identify the review date, reviewed configuration, responsible reviewers, release authority, review outcome, applicable restrictions or conditions, significant open issues, deviations or waivers and references to the supporting evidence package.
Release records shall be maintained as controlled records and retained in accordance with Chapter 23.

26.14 Post-Release Feedback

The release decision shall not terminate the AI safety lifecycle.
Operational monitoring, incidents, field feedback, changes in the operating environment, newly identified scenarios, performance degradation or cybersecurity events may provide information that affects the validity of the release decision.
Where such information indicates that safety assumptions or release conditions may no longer be valid, the organization shall initiate the applicable processes for investigation, change, re-verification, re-validation, safety re-assessment and, where necessary, re-approval or withdrawal.
The objective of Safety Review and Release is to ensure that an AI system is released only on the basis of a controlled, configuration-specific and evidence-based safety decision, with applicable risks, limitations, deviations, interfaces and residual risks explicitly understood and formally addressed.

27 AI Safety Case

27.1 General

For safety-relevant AI systems, the organization shall establish and maintain an appropriate AI Safety Case.

The AI Safety Case shall provide a structured and reasoned argument demonstrating that the applicable safety objectives and requirements have been addressed and that the AI system is acceptably safe for its defined intended use and operating conditions, subject to identified assumptions, limitations and residual risks.

The Safety Case shall be supported by objective evidence generated throughout the AI lifecycle. It shall therefore provide a logical relationship between safety claims, supporting arguments, assumptions, requirements, analyses, verification and validation results, assessments and the final safety decision.

The level of detail and rigor of the Safety Case shall be appropriate to the safety relevance, complexity, autonomy, intended use and potential consequences of the AI system.

27.2 Safety Case Structure

The Safety Case shall normally be structured using a hierarchy of **claims, arguments and evidence**.

A top-level claim may be expressed as:

Claim:

The AI system is acceptably safe for its defined intended use and operating conditions, subject to the stated assumptions, limitations and residual risks.

The top-level claim shall be decomposed into sufficiently specific supporting claims and arguments to demonstrate why the conclusion is justified.

The generic argument structure defined in this chapter shall be adapted to the characteristics of the individual AI system. Not every project will require identical claims or the same level of decomposition.

27.3 Intended Use and Operating Conditions

The Safety Case shall demonstrate that the intended use and relevant operating conditions of the AI system are adequately defined.

The argument shall address, as applicable, the intended functionality, operating environment, users, interfaces, Operational Design Domain, applicable system boundaries, assumptions of use and relevant foreseeable misuse.

The Safety Case shall clearly identify conditions outside the validated or approved operating envelope where these may affect the safety conclusion.

27.4 Hazards and Safety Risks

The Safety Case shall demonstrate that relevant hazards, hazardous behaviors and safety risks have been systematically identified, analyzed and evaluated.

The argument shall be supported by the applicable risk analyses, hazard analyses, SOTIF analyses, Functional Safety analyses and other relevant safety activities.

Where AI-specific characteristics contribute to risk, such as uncertainty, insufficient scenario coverage, performance limitations, unexpected behavior or distribution changes, the corresponding risks and controls shall be addressed.

27.5 Safety Requirements and Safety Objectives

The Safety Case shall demonstrate that appropriate safety objectives and requirements have been established and addressed.

Requirements shall be traceable to their originating safety objectives, hazards, risks, assumptions or applicable external requirements and to the corresponding implementation and verification or validation evidence.

Where requirements are allocated between AI and conventional system elements, the Safety Case shall demonstrate that the relevant interfaces and responsibilities are adequately addressed.

27.6 Controlled AI Development

The Safety Case shall provide evidence that the AI development lifecycle has been appropriately controlled.

The argument shall consider, as applicable, AI requirements, architecture and model development, training processes, configuration management, toolchain management, verification and validation, change management and release control.
The purpose is not merely to demonstrate that defined processes exist, but that the processes were effectively applied to the configuration on which the safety conclusion is based.

27.7 Data as a Safety Argument Element

Where data contributes to the safety behavior of the AI system, the Safety Case shall demonstrate that the relevant data lifecycle has been appropriately controlled.
The argument shall address, as applicable, data requirements, data sources, selection and preparation, labeling, quality, representativeness, coverage, limitations, preprocessing and the relationship between datasets and the corresponding model versions.
The Safety Case shall demonstrate the relationship between the data used for development and the evidence supporting the safety conclusion.

27.8 AI Model and Behavior

The Safety Case shall demonstrate that the AI model and its relevant behavior have been sufficiently verified and validated for the intended use and operating conditions.
The argument shall be supported by appropriate evidence concerning model requirements, model configuration, performance, robustness, scenario coverage, boundary conditions, uncertainty, limitations and relevant unexpected behaviors.
Statistical performance results shall not automatically be considered sufficient evidence of safety where additional safety characteristics, constraints or system-level considerations are applicable.

27.9 Functional Safety, SOTIF and Cybersecurity Interfaces

Where applicable, the Safety Case shall demonstrate that the interfaces to Functional Safety, SOTIF and Cybersecurity have been adequately addressed.
For Functional Safety, the argument shall establish the relationship between the AI contribution and applicable safety goals, functional and technical safety requirements, assumptions, safety mechanisms and integration evidence.

For SOTIF, the argument shall address intended functionality, functional insufficiencies, performance limitations, foreseeable misuse, relevant scenarios and unknown or insufficiently understood unsafe behavior as applicable.
For Cybersecurity, the argument shall address cybersecurity threats that may affect AI safety, including the integrity of relevant data, models, software, configurations and deployment or update mechanisms.
Evidence generated by another discipline shall only be used within the Safety Case where its applicability to the respective safety claim has been established.

27.10 Operational Limitations and Continuous Assurance

The Safety Case shall identify relevant operational limitations, assumptions and conditions necessary for maintaining the safety conclusion.
Where safety depends on monitoring, fallback mechanisms, human supervision, restricted operating conditions or other operational controls, these shall form part of the Safety Case.
The Safety Case shall also address how operational monitoring and continuous assurance provide feedback on the continued validity of the safety argument.
Where operational evidence indicates that an assumption, safety claim or risk control may no longer be valid, the Safety Case shall be reviewed and updated as appropriate.

27.11 Change and Configuration

The Safety Case shall be associated with a defined and controlled system configuration.
Changes to the AI model, data, software, hardware, operating conditions, ODD, safety requirements, cybersecurity controls or other safety-relevant elements shall be evaluated for their potential impact on the Safety Case.
Where a change affects the validity of an existing claim, argument or evidence item, the applicable verification, validation, safety assessment and release processes shall be initiated.
The Safety Case shall therefore evolve with the AI system and shall not be treated as a static document created only at the end of development.

27.12 Residual Risk and Safety Conclusion

The Safety Case shall explicitly address residual risks remaining after implementation of the identified safety measures.
The argument shall demonstrate that residual risks have been evaluated against the applicable risk acceptance criteria and that any assumptions, restrictions, monitoring measures or compensating controls necessary for their acceptance are defined.
The final safety conclusion shall be consistent with the approved release decision and shall not claim a level of safety beyond that supported by the available evidence.

27.13 Evidence Supporting the Safety Case

The Safety Case shall be supported by objective evidence appropriate to the claims and arguments. Such evidence may include requirements, safety objectives, hazard and risk analyses, assumptions, datasets, model documentation, training records, verification and validation results, test and simulation reports, reviews, assessments, audit results, configuration records, change and deviation records, operational evidence and release decisions.
Each significant evidence item shall be identifiable, traceable to the applicable configuration and maintained in accordance with Chapter 23, Documentation and Evidence Management.

27.14 Traceability and Consistency

The Safety Case shall be consistent with the traceability framework defined in Chapter 22.
Claims shall be traceable to their supporting arguments and assumptions, and arguments shall be supported by appropriate objective evidence.
The Safety Case shall also remain consistent with the applicable requirements, risk assessments, V&V results, safety assessments, deviations and release configuration.
Contradictions between the Safety Case and other controlled safety records shall be identified and resolved.

27.15 Review and Approval of the Safety Case

The Safety Case shall be subject to formal review before it is used to support a significant safety decision or release.

The review shall determine whether the claims are adequately supported, the arguments are logically consistent, the assumptions are explicit, the evidence is sufficient and traceable, and the conclusion is justified for the defined configuration and operating conditions.
Where the applicable safety classification requires independent review or confirmation measures, these shall be performed by appropriately competent and sufficiently independent personnel.
The Safety Case shall be approved by the authorized responsible persons in accordance with the Authority and Responsibility Matrix.

27.16 Maintenance of the Safety Case

The Safety Case shall be maintained throughout the AI lifecycle.
It shall be reviewed following significant changes, safety incidents, newly identified risks, changes in operating conditions, changes to applicable requirements, significant V&V results or other events that may affect the validity of its claims or arguments.
Superseded versions shall be retained where required for traceability and historical reconstruction.

27.17 Safety Case Records

The organization shall maintain controlled records demonstrating the development, review, approval and revision history of the Safety Case.
The records shall identify the applicable AI system and configuration, relevant claims and arguments, assumptions, evidence references, reviewers, approval status and significant changes.
The Safety Case shall be retrievable as part of the controlled safety evidence package.

The objective of the AI Safety Case is to provide a structured, traceable and evidence-based argument that the AI system is acceptably safe for its defined intended use and operating conditions, and to make the basis, assumptions, limitations and residual risks of that conclusion transparent and reviewable.

28 Training and Awareness

28.1 General

The organization shall establish and maintain an AI Safety Training and Awareness process to ensure that personnel whose activities may affect AI safety have the knowledge and competence appropriate to their responsibilities.

Training shall be considered part of the overall competence management process defined in Chapter 7 and shall support the effective implementation of the AI Safety Management System throughout the AI lifecycle.

The extent and depth of training shall be proportionate to the individual's role, responsibilities, authority, level of involvement and potential impact on AI safety.

28.2 Training Needs

The organization shall identify training needs based on the responsibilities and activities of personnel involved in AI safety-related work.

Training needs shall consider the applicable AI technologies, lifecycle activities, safety classification, applicable standards and regulations, organizational processes, tools and methods, and the interfaces with other safety disciplines.

Training needs shall also be reviewed when significant changes occur, including changes to applicable standards, AI technologies, processes, responsibilities, tools or safety-related findings.

28.3 AI Safety Training Programme

The organization shall establish an appropriate AI safety training programme covering the knowledge and skills required to perform assigned responsibilities.

Depending on the role and applicability, training may address AI safety principles and the AI lifecycle, applicable standards and regulatory requirements, AI safety risk management, data and model management, verification and validation, Functional Safety, SOTIF, Cybersecurity, configuration and change management, incident management and the organization's AI Safety Management System.

Training shall not be limited to theoretical knowledge. Where personnel perform safety-relevant engineering or assessment activities, practical knowledge and application of the relevant methods and processes shall also be addressed.

28.4 Role-Specific Training

Training shall be tailored to the responsibilities of the individual role.
Personnel involved in AI or ML development shall receive training appropriate to AI development, model behavior, data management, model limitations and relevant safety requirements.
Personnel responsible for data activities shall understand the safety relevance of data quality, data selection, representativeness, traceability, data changes and associated limitations.
Personnel performing verification and validation shall have appropriate knowledge of the applicable V&V methods, acceptance criteria, scenario coverage, robustness, uncertainty and interpretation of AI-specific results.
Personnel responsible for Functional Safety, SOTIF or Cybersecurity interfaces shall understand the relevant interfaces and the potential interaction between AI safety and their respective disciplines.
Personnel involved in configuration, change, release or incident management shall understand the safety implications of uncontrolled changes, incorrect configurations, incidents and incomplete safety evidence.
Management and release authorities shall receive sufficient awareness training to understand their responsibilities, decision-making authority, escalation obligations and the importance of evidence-based safety decisions.

28.5 Awareness

In addition to role-specific technical training, the organization shall establish appropriate AI safety awareness among personnel whose activities may influence AI safety.
Awareness shall include the principle that AI safety is a lifecycle responsibility and is not limited to testing or final product approval.
Personnel shall understand the importance of reporting safety concerns, unexpected AI behavior, data issues, process deviations, configuration discrepancies and other conditions that may affect safety.

Personnel shall also understand that commercial, schedule or delivery pressure does not override established safety requirements or release criteria.

28.6 Training Methods

Training may be provided through formal courses, workshops, practical exercises, supervised activities, mentoring, internal knowledge-sharing, external training or other suitable methods.
Where practical competence is required, training shall be supplemented by appropriate practical application, supervised work or demonstration of competence.
The selected training method shall be appropriate to the complexity and safety relevance of the subject matter.

28.7 Training Effectiveness

The organization shall evaluate the effectiveness of training where appropriate.
Effectiveness may be demonstrated through examinations, practical exercises, observed performance, technical reviews, supervised activities, assessment results or other suitable methods.
Completion of a training course shall not automatically be considered sufficient evidence of competence where the assigned role requires demonstrated practical capability.
Where training is found to be ineffective or competence remains insufficient, additional training, supervision or other corrective measures shall be established.

28.8 Initial and Continuing Training

Personnel shall receive appropriate training before performing safety-relevant activities independently.
Training shall be maintained throughout the lifecycle of the AI Safety Management System. Refresher or additional training shall be provided where necessary due to changes in technology, standards, processes, tools, responsibilities or identified competence gaps.
Personnel returning to safety-relevant activities after significant periods of absence shall receive additional training where necessary to ensure that their competence remains current.

28.9 Training Records

Training and competence-related records shall be maintained as controlled records.
Records shall identify, as applicable, the person, training subject, training date, training provider, applicable role, training status and evidence of successful completion or competence evaluation.
Where training contributes to formal qualification or authorization for a safety-relevant activity, the relationship between the training record and the applicable qualification or authorization shall be traceable.
Training records shall be managed in accordance with the Documentation and Evidence Management requirements of Chapter 23.

28.10 External Personnel and Suppliers

External personnel and suppliers performing safety-relevant activities on behalf of the organization shall have appropriate competence for their assigned responsibilities.
The organization shall determine the applicable competence and training requirements and shall obtain appropriate evidence where necessary.
Where external personnel are required to follow the organization's AI safety processes, they shall receive sufficient awareness or process-specific training to perform their activities correctly.

28.11 Training and Organizational Change

Significant changes to the AI Safety Management System, applicable standards, technology, tools, organizational responsibilities or safety processes shall be evaluated for their training implications.
Where a change introduces new responsibilities, methods, technologies or safety risks, appropriate training shall be completed before personnel are expected to perform the affected activities independently.

28.12 Training Records and Continual Improvement

Training results, competence gaps, audit findings, incidents, assessment findings and operational experience shall be used as inputs to the continual improvement of the training programme.
Recurring errors or findings may indicate that existing training is insufficient or that training content, methods or competence requirements need to be revised.

The effectiveness of the overall training programme shall be periodically reviewed by the responsible management function.

The objective of Training and Awareness is to ensure that personnel performing or influencing safety-related AI activities possess and maintain the knowledge, skills and awareness necessary to perform their responsibilities effectively and to recognize and appropriately respond to conditions that may affect AI safety.

29 Incident and Problem Management

29.1 General

The organization shall establish and maintain a controlled process for the identification, reporting, investigation, containment and resolution of AI-related safety incidents and problems.

AI-related safety incidents shall be recorded and investigated when they indicate actual or potential impact on the safety of an AI system, its intended functionality, its operating assumptions or its associated safety controls.

Incident and Problem Management shall cover the complete lifecycle, including development, verification, validation, integration, release, deployment and operation.

The process shall ensure that safety-relevant information is captured promptly and that appropriate actions are taken to prevent recurrence and protect affected products, users and systems.

29.2 Identification and Reporting

Personnel shall be provided with appropriate means to report suspected AI safety incidents, anomalies and problems.

An incident may originate from development activities, testing, simulation, production, field operation, customer feedback, operational monitoring, cybersecurity monitoring, audits or other sources.

Personnel shall report suspected safety-relevant conditions without waiting for complete confirmation of the underlying cause.

The initial report shall contain sufficient information to identify the affected system or product, the observed behavior, relevant operating conditions and the available configuration or version information.

29.3 Examples of AI Safety Incidents

AI-related safety incidents may include unexpected or unsafe AI behavior, significant model degradation, unacceptable false-negative or false-positive behavior, unsafe interactions with other system elements, data integrity issues or model integrity issues.

They may also include cybersecurity events affecting AI behavior or the integrity of AI-related assets, operation outside defined safety assumptions or validated operating conditions, unexpected changes in model performance, or other conditions indicating that an approved safety assumption or control may no longer be valid.

The examples are not exhaustive. Any event that may affect an established safety claim, safety requirement, risk control or release condition shall be evaluated for potential safety impact.

29.4 Initial Assessment and Classification

Reported incidents shall undergo an initial assessment to determine their significance and required response.

The assessment shall consider the observed or potential safety impact, affected functionality, affected products or configurations, operating conditions, frequency or recurrence, potential exposure and whether the incident may indicate a systemic problem.

Where appropriate, incidents shall be classified according to defined organizational criteria so that response, escalation and investigation activities are proportionate to their significance.

Where the incident may affect an already released product, the organization shall promptly determine whether containment, operational restriction, customer communication, field action or other protective measures are necessary.

29.5 Containment

Where an incident presents an immediate or potentially unacceptable safety risk, appropriate containment measures shall be initiated without waiting for completion of the full investigation.

Containment may include restricting operation, disabling an affected function, preventing further deployment, suspending a release, applying additional monitoring, isolating affected configurations, reverting to a previously approved configuration or other appropriate protective measures.

Containment decisions shall be documented and shall identify the affected products, configurations and operating conditions.
Containment shall not be considered a permanent resolution unless the applicable safety assessment determines that continued operation under the defined conditions is acceptable.

29.6 Incident Investigation

Safety-relevant incidents shall be investigated by personnel with appropriate competence and, where necessary, sufficient independence from the original development or operational activity.
The investigation shall seek to determine the technical and organizational causes of the incident and shall consider contributing factors rather than only the immediate failure mechanism.
Depending on the nature of the incident, the investigation may examine the AI model, training data, data processing, software, hardware, tools, configuration, operating environment, ODD, human interaction, cybersecurity conditions, monitoring mechanisms and applicable assumptions.
The investigation shall preserve relevant evidence and ensure that the configuration associated with the incident can be identified.

29.7 Safety Impact Assessment

The investigation shall determine the actual or potential safety impact of the incident.
The assessment shall consider affected safety objectives, requirements, hazards, risks, assumptions, safety mechanisms, V&V evidence and release conditions.
Where applicable, the organization shall determine whether the incident affects Functional Safety, SOTIF, Cybersecurity or other safety-related disciplines.
The investigation shall also determine whether the incident indicates that existing safety assumptions, risk controls or Safety Case arguments are no longer valid.

29.8 Root Cause and Problem Management

Where an incident indicates an underlying systemic or recurring problem, the organization shall initiate appropriate Problem Management activities.
The objective shall be to identify and eliminate or adequately control the underlying cause rather than repeatedly treating individual occurrences.

Problem analysis may identify weaknesses in requirements, data, model development, V&V, configuration management, tools, processes, competence, monitoring, cybersecurity controls or organizational responsibilities.
Recurring incidents, similar anomalies or repeated deviations shall be evaluated for common causes and systemic corrective action.

29.9 Corrective and Preventive Action

Corrective actions shall be established based on the results of the investigation and safety impact assessment.
Actions may include changes to requirements, data, models, software, hardware, safety mechanisms, monitoring, operational restrictions, cybersecurity controls, development processes or organizational controls.
Where appropriate, preventive actions shall also be established to reduce the likelihood of similar incidents occurring elsewhere.
Corrective actions affecting an approved configuration shall be managed in accordance with Chapter 18, Change Management.

29.10 Re-Verification, Re-Validation and Re-Assessment

The organization shall determine whether the incident or resulting corrective action requires re-verification, re-validation or safety re-assessment.
This determination shall consider the affected functionality, configuration, requirements, safety assumptions, risk controls and existing evidence.
Where a change is introduced, the applicable sequence shall be applied:
Incident → Investigation → Safety Impact Assessment → Corrective Action → Change Management → Re-Verification / Re-Validation → Safety Re-Assessment → Re-Approval, where required.
Completion of corrective action or successful testing shall not automatically restore the previous release status.

29.11 Cybersecurity-Related Incidents

Where a cybersecurity event may affect AI behavior, data, models, software, configuration or deployment integrity, the incident shall be coordinated with the applicable Cybersecurity Management process.

The investigation shall determine whether the cybersecurity event has a potential or actual safety impact.

Safety and cybersecurity investigations shall exchange relevant information while maintaining clear responsibilities for the respective technical and management decisions.

29.12 Data and Model Integrity Incidents

Incidents involving potentially corrupted, manipulated or otherwise compromised datasets, model parameters, model weights or related configuration shall receive particular attention.

The organization shall determine the affected versions and the extent to which the integrity issue may have propagated into development, V&V, released configurations or operational systems.

Where the integrity of previously generated safety evidence may be affected, the validity of that evidence shall be reassessed rather than assumed to remain valid.

29.13 Communication and Escalation

The organization shall define communication and escalation requirements for significant AI safety incidents.

Depending on the nature and impact of the incident, communication may be required between engineering functions, management, customers, suppliers, certification or assessment bodies, regulatory authorities or other affected parties.

Communication shall be accurate, controlled and consistent with applicable contractual, regulatory, safety and cybersecurity requirements.

Significant unresolved incidents shall be escalated to the defined safety and release authorities.

29.14 Closure

An incident shall remain open until the required investigation, safety impact assessment, containment and corrective actions have been completed or an appropriately authorized decision has been made regarding its disposition.

Closure shall be supported by objective evidence demonstrating that the defined actions have been completed and, where required, their effectiveness has been verified.

Where an incident results in a permanent change to the AI system or its safety controls, the corresponding change, configuration, V&V, assessment and release records shall be linked to the incident record.

29.15 Incident Records and Traceability

Incident records shall be maintained as controlled safety records.

Records shall provide traceability, as applicable, between the incident, affected product and configuration, observed behavior, operating conditions, investigation results, root cause, safety impact, containment, corrective action, V&V, safety assessment, communication and final disposition.

Incident records shall be managed in accordance with Chapter 23, Documentation and Evidence Management, and shall remain available for future safety analysis and continual improvement.

29.16 Interface with Continuous Assurance

Incident and Problem Management shall form an integral part of the Continuous Assurance process defined in Chapter 17.

Operational monitoring, customer feedback, field data, anomaly detection and safety-event monitoring shall provide inputs to the incident process.

Conversely, lessons learned from incidents shall be fed back into monitoring criteria, safety risks, requirements, validation scenarios, datasets, model development, safety mechanisms and other relevant lifecycle activities.

29.17 Continual Improvement

Incident and problem information shall be periodically reviewed to identify trends, recurring causes and weaknesses in the AI Safety Management System.

Lessons learned shall be used to improve processes, training, risk controls, V&V strategies, monitoring, change classification and other relevant controls.

Where an incident reveals a systemic weakness, the organization shall address the underlying management-system issue rather than limiting the response to the affected individual product or event.

The objective of Incident and Problem Management is to ensure that AI-related safety incidents and problems are promptly identified, contained, investigated and resolved, that their safety impact is properly assessed, and that appropriate corrective actions, re-verification, re-validation and re-assessment are initiated to maintain the validity of the AI Safety Case and approved release configuration

30 Continuous Improvement

30.1 General

The organization shall continually improve the suitability, adequacy and effectiveness of the AI Safety Management System.

Continuous Improvement shall ensure that experience gained throughout the AI lifecycle is systematically evaluated and used to strengthen safety processes, controls, competence and organizational effectiveness.

Improvement shall address both identified deficiencies and opportunities to increase the effectiveness, robustness and efficiency of the AI Safety Management System.

Continuous Improvement shall not be limited to corrective actions following failures. The organization shall also proactively identify potential improvements based on changes in technology, AI methods, operating experience, standards, regulations and emerging safety knowledge.

30.2 Inputs to Continuous Improvement

The organization shall establish mechanisms for collecting and evaluating information relevant to the effectiveness of the AI Safety Management System.

Inputs shall include, as applicable, internal audit results, external assessment findings, safety assessments, incidents and problems, field experience, customer feedback, operational monitoring, deviations and waivers, lessons learned, changes in AI technology and methods, updates to applicable standards and regulatory developments.

Information from one AI project may also provide relevant lessons for other projects where similar technologies, processes, risks or safety assumptions are involved.

30.3 Evaluation of Improvement Needs

Relevant information shall be evaluated to determine whether it indicates a need to modify existing processes, controls or organizational arrangements.

The evaluation shall consider whether an identified issue is isolated or systemic and whether similar risks or weaknesses may exist in other products, projects or lifecycle activities.

Where appropriate, the organization shall assess potential effects on existing safety cases, safety requirements, risk controls, competence requirements, monitoring arrangements and released configurations.

30.4 Lessons Learned

Lessons learned shall be systematically captured and evaluated.

Lessons may originate from successful activities as well as from incidents, audit findings, assessment findings, unexpected AI behavior, deviations, failed V&V activities, operational experience or changes in technology.

Relevant lessons learned shall be incorporated into appropriate processes, guidelines, training programmes, engineering methods, V&V strategies, risk controls and other applicable elements of the AI Safety Management System.

Lessons learned shall be communicated to relevant personnel where their application may prevent recurrence or improve safety performance.

30.5 Improvement of AI Safety Processes

Where improvement needs are identified, the organization shall establish appropriate actions and assign responsibility for their implementation.

Improvement actions may address AI safety planning, risk management, data management, model management, V&V, continuous assurance, change management, configuration management, toolchain management, incident management, documentation and evidence management, training or organizational responsibilities.

Changes to controlled processes shall be implemented through the applicable configuration and change-management arrangements.

30.6 Technology and AI Method Changes

The organization shall monitor developments in AI technologies and methods that may affect the effectiveness of the AI Safety Management System.

Changes such as new model architectures, training methods, data-generation techniques, AI development frameworks, deployment technologies or verification methods shall be evaluated for their potential impact on existing safety processes and controls.

Where new AI methods introduce safety characteristics or risks not adequately addressed by existing processes, the organization shall update the applicable management-system elements and competence requirements.

30.7 Standards and Regulatory Developments

Changes to applicable standards, regulations, certification schemes and customer requirements shall be monitored and evaluated.

Where such changes affect the applicable AI safety framework, the organization shall determine the impact on existing processes, projects, products, safety evidence and certification or conformity assessment activities.

Required changes shall be planned and implemented in a controlled manner.

30.8 Effectiveness of Improvement Actions

Improvement actions shall be reviewed to determine whether they achieved their intended result.

Where an action was intended to prevent recurrence of a problem or finding, its effectiveness shall be evaluated using appropriate objective evidence.
Ineffective actions shall be reassessed and supplemented or replaced as necessary.
The organization shall avoid treating repeated findings through repeated short-term corrections where the underlying systemic cause has not been addressed.

30.9 Management Review and Continuous Improvement

Information from the Continuous Improvement process shall provide input to management review.
Management shall periodically evaluate the overall effectiveness of the AI Safety Management System, including significant trends, recurring findings, incidents, assessment results, operational experience, competence and resource needs, technology developments and emerging risks.
Management shall ensure that appropriate resources and authority are available to implement necessary improvements.

30.10 Improvement Records and Traceability

Continuous Improvement activities shall be documented to the extent necessary to demonstrate the basis, responsibility, implementation and effectiveness of significant improvement actions.
Where an improvement affects a controlled process, requirement, responsibility, training programme or technical control, the resulting change shall be traceable through the applicable Change Management and Configuration Management processes.
Improvement records shall be maintained in accordance with Chapter 23, Documentation and Evidence Management.

30.11 Continual Improvement of the Safety Case

Where improvements, incidents, operational experience or new knowledge affect the assumptions, claims, arguments or evidence supporting an AI Safety Case, the Safety Case shall be reviewed and updated as appropriate.
The organization shall ensure that the Safety Case remains consistent with the current approved configuration, applicable safety requirements, risk controls and available objective evidence.

30.12 Organizational Learning

The organization shall promote an environment in which safety-related concerns, unexpected results, mistakes and lessons learned can be reported and used for improvement.
Personnel shall be encouraged to identify potential weaknesses without avoiding or suppressing information because of project, schedule or commercial considerations.
Organizational learning shall contribute to the development of stronger processes, better competence and more effective safety decision-making.
The objective of Continuous Improvement is to ensure that the AI Safety Management System remains effective, current and capable of addressing evolving AI technologies, risks, operating experience, standards and regulatory expectations, while systematically transforming findings and lessons learned into sustained improvements in AI safety.

31 Management Review

31.1 General

Top Management shall periodically review the suitability, adequacy and effectiveness of the AI Safety Management System.
Management Review shall provide a formal mechanism for Top Management to evaluate whether the AI Safety Management System remains appropriate for the organization's activities, technologies, products, safety responsibilities and external environment.
The review shall consider information from across the AI lifecycle and from the management-system processes established in this manual. It shall provide the basis for management decisions concerning resources, responsibilities, competence, processes, risks and required improvements.
Management Review shall be performed at planned intervals and additionally when significant changes, incidents or other circumstances indicate that an earlier review is necessary.

31.2 Management Review Inputs

Management Review shall consider relevant information concerning the performance and effectiveness of the AI Safety Management System.

This shall include, as applicable, safety performance, internal and external audit results, incidents and problems, operational and field experience, open safety risks, deviations and waivers, significant changes, safety assessment results, status of corrective and improvement actions, competence and training, resource requirements, customer requirements and feedback, applicable standards and regulatory developments.

Management shall also consider significant changes to AI technologies, development methods, organizational responsibilities, products, operating environments or business activities that may affect the AI Safety Management System.

31.3 Safety Performance

Management shall review whether the AI Safety Management System is achieving its intended safety objectives.

The review shall consider relevant safety indicators, significant risks, safety incidents, operational monitoring results, recurring findings, unresolved issues and trends that may indicate deterioration or improvement in safety performance.

Where performance indicators reveal emerging safety concerns, Top Management shall ensure that appropriate investigation, escalation and corrective action are initiated.

31.4 Audit and Assessment Results

Internal audit results and relevant external assessment, certification or regulatory findings shall be reviewed.

Management shall consider the significance and recurrence of findings and whether they indicate isolated deficiencies or systemic weaknesses in the AI Safety Management System.

The status and effectiveness of corrective actions associated with significant findings shall also be reviewed.

31.5 Incidents, Problems and Operational Experience

Management shall review significant AI safety incidents, problems, near misses and relevant operational experience.

The review shall consider their causes, safety impact, corrective actions, recurrence and implications for other projects or products.

Where incidents or operational experience indicate that existing safety assumptions, controls or processes may no longer be adequate, management shall ensure that appropriate actions are initiated.

31.6 Risks, Deviations and Changes

The review shall consider the status of significant open safety risks, deviations, waivers and safety-related changes.
Management shall ensure that unacceptable or insufficiently controlled risks are appropriately escalated and that temporary deviations or waivers do not remain open without adequate justification and control.
Significant changes to AI systems, technologies, intended use, operating conditions, organizational structure or applicable requirements shall be evaluated for their potential impact on the AI Safety Management System.

31.7 Resources and Competence

Top Management shall review whether sufficient resources and competent personnel are available to perform the required AI safety activities.
This shall include consideration of personnel capacity, technical competence, training needs, tools, infrastructure, independent assessment capability and other resources required to maintain the effectiveness of the AI Safety Management System.
Where competence gaps or resource constraints could affect safety, management shall establish appropriate actions and priorities.

31.8 Customer, Regulatory and External Requirements

Management shall review relevant changes in customer requirements, contractual obligations, applicable standards, regulations, certification requirements and market conditions.
The review shall determine whether these changes require modifications to the AI Safety Management System, applicable processes, competence requirements, product safety activities or conformity assessment arrangements.

31.9 Improvement Actions

Management shall review the status and effectiveness of significant improvement actions arising from audits, incidents, assessments, operational experience and previous Management Reviews.
Where improvement actions have not achieved their intended results, management shall ensure that additional or alternative measures are established.
Management Review shall therefore provide a link between organizational learning, Continuous Improvement and management decision-making.

31.10 Management Review Outputs

Management Review shall result in documented decisions and actions where necessary.
Outputs may include decisions concerning changes to the AI Safety Management System, safety objectives, processes, responsibilities, resources, competence and training, risk controls, monitoring activities, technical methods or improvement priorities.
Where management decisions affect controlled processes or safety-relevant configurations, the applicable Change Management and Configuration Management processes shall be followed.

31.11 Safety Escalation and Management Decisions

Top Management shall ensure that significant safety concerns receive appropriate priority and are not suppressed or deferred solely because of commercial, schedule or organizational considerations.
Where an issue cannot be adequately resolved within the responsible technical or management function, it shall be escalated through the defined authority structure.
Management decisions concerning significant residual risks, major deviations, release restrictions or continuation of operation shall be made by personnel having the appropriate authority and competence.

31.12 Management Review Records

Management Review results shall be documented as controlled records.
The records shall identify, as applicable, the date and participants, information reviewed, significant findings and trends, decisions taken, assigned actions, responsible persons and target dates.

Actions arising from Management Review shall be tracked to completion and their effectiveness evaluated where appropriate.
Management Review records shall be maintained in accordance with Chapter 23, Documentation and Evidence Management.

31.13 Relationship to Continuous Improvement

Management Review shall provide a formal management-level input to the Continuous Improvement process defined in Chapter 30.
Information and decisions resulting from Management Review shall be incorporated into the relevant improvement activities, while significant improvement results shall in turn provide input to subsequent Management Reviews.
This establishes a controlled management feedback cycle:
Safety Performance → Management Review → Management Decision → Improvement / Change → Implementation → Effectiveness Evaluation → Management Review

31.14 Frequency and Extraordinary Review

Management Review shall be performed at planned intervals defined by the organization.
An extraordinary Management Review shall be considered following significant safety incidents, major organizational or technological changes, substantial changes to applicable regulatory requirements, significant assessment findings or other circumstances that may materially affect the effectiveness of the AI Safety Management System.
The objective of Management Review is to ensure that Top Management periodically evaluates the effectiveness and continuing suitability of the AI Safety Management System and makes informed decisions on safety performance, risks, resources, competence, compliance and improvement needed to maintain effective control of AI safety.

32 Minimum Mandatory AI Safety Work Products

32.1 General

The organization shall establish and maintain the work products necessary to demonstrate the effective implementation of the AI Safety Management System and the safety of applicable AI systems.

The work products defined in this chapter represent the minimum set of controlled documentation and objective evidence expected within the AI safety lifecycle.

The applicability of individual work products shall be determined according to the AI system classification, safety relevance, intended use, lifecycle phase, applicable standards and regulatory requirements, architecture, risks and project-specific safety planning.

A work product identified as **Mandatory - Yes** shall be established for the applicable AI Safety Management System or AI project unless a formally documented and approved justification establishes that it is not applicable. Work products identified as **Where applicable** shall be established when the applicable project classification, safety analysis, standards, customer requirements or regulatory framework requires them.

The detailed content, format, review and approval criteria for these work products shall be defined in the applicable Level 2 procedures and Level 4 templates or checklists.

32.2 Minimum Work Product Set

ID	Work Product	Mandatory	Primary Purpose
AI-WP-01	**AI Safety Policy**	Yes	Defines management commitment, principles and overall AI safety objectives.
AI-WP-02	**AI Safety Management Plan**	Yes	Defines how the AI Safety Management System and applicable project activities are planned and controlled.
AI-WP-03	**AI System Classification**	Yes	Defines the safety relevance, classification and applicable lifecycle controls for the AI system.
AI-WP-04	**AI Safety Concept**	Where applicable	Defines the high-level safety objectives, strategy, assumptions and safety measures for the AI system.

ID	Work Product	Mandatory	Primary Purpose
AI-WP-05	**AI Safety Requirements**	Yes	Defines the safety requirements allocated to the AI system, AI component or AI function.
AI-WP-06	**Data Management Plan**	Yes	Defines the controls for safety-relevant data throughout its lifecycle.
AI-WP-07	**Dataset Specification**	Yes	Defines the required characteristics, purpose, scope, quality and limitations of relevant datasets.
AI-WP-08	**Model Specification**	Yes	Defines the intended purpose, relevant characteristics, requirements and configuration of the AI model.
AI-WP-09	**Verification Plan**	Yes	Defines the planned verification activities, methods, criteria and evidence.
AI-WP-10	**Validation Plan**	Yes	Defines the planned validation activities, intended-use conditions, scenarios and acceptance criteria.
AI-WP-11	**Verification Report**	Yes	Documents verification activities, results, anomalies, conclusions and supporting evidence.
AI-WP-12	**Validation Report**	Yes	Documents validation activities, results, limitations, conclusions and supporting evidence.
AI-WP-13	**Safety Risk Analysis**	Yes	Identifies, analyzes and evaluates safety risks and associated control measures.
AI-WP-14	**Safety Case / Safety Argument**	Where applicable	Provides the structured argument and evidence supporting the safety conclusion.
AI-WP-15	**Configuration Baseline**	Yes	Defines the controlled configuration on which development, V&V, assessment and release decisions are based.
AI-WP-16	**Change Records**	Yes	Documents safety-relevant changes, their impact analysis, implementation and approval.
AI-WP-17	**Deviation / Waiver Records**	Where applicable	Documents approved deviations, waivers, justification, safety impact and conditions.
AI-WP-18	**Safety Assessment Report**	Where applicable	Documents the independent or formal assessment of the applicable safety evidence and conclusions.
AI-WP-19	**Release Approval**	Yes	Documents the formal decision authorizing the defined AI configuration for release.

ID	Work Product	Mandatory	Primary Purpose
AI-WP-20	**Operational Monitoring Report**	Where applicable	Documents operational safety monitoring, relevant observations, incidents, trends and continuing assurance results.

Table 4: Minimum Workproduct Set

32.3 Management-System and Project Work Products

The work products shall be understood as belonging to two related categories.

The first category consists of **management-system work products**, which establish the organizational framework for AI safety. These include, in particular, the AI Safety Policy and the AI Safety Management Plan, together with organizational, competence, audit, training and management-review records defined elsewhere in this manual.

The second category consists of **AI project and product work products**, which provide objective evidence that the defined AI system has been developed, verified, validated, assessed and released under controlled conditions.

The same work product may therefore exist at different organizational levels. For example, the AI Safety Policy is primarily an organizational management-system document, while an AI System Classification and Model Specification are normally project- or product-specific.

32.4 Applicability and Tailoring

The minimum work product set shall be tailored according to the characteristics and safety relevance of the AI system.

Tailoring shall consider the intended use, potential consequences of incorrect AI behavior, system architecture, degree of autonomy, interaction with safety-related functions, applicable Functional Safety and SOTIF requirements, cybersecurity dependencies, regulatory requirements and customer requirements.

Where a work product identified as mandatory is considered not applicable, the rationale shall be documented and approved by the responsible authority.

Tailoring shall not be used to remove evidence necessary to substantiate a safety decision.

32.5 Work Product Development and Approval

Each applicable work product shall have a defined responsible owner and shall be subject to appropriate review and approval.

Work products shall be configuration-controlled and shall have a defined status, such as draft, under review, approved, released or superseded.

Where a work product provides the basis for a safety decision, its approval shall be completed before that decision is made.

Changes to approved work products shall be managed in accordance with Chapter 18, Change Management, and Chapter 20, Configuration Management.

32.6 Relationship to the AI Safety Lifecycle

The work products shall be generated and maintained at the lifecycle stages at which they are needed.

For example, the AI System Classification shall be established sufficiently early to determine the applicable lifecycle controls. Safety requirements and the Safety Concept shall support development activities, while Verification and Validation Plans shall be available before the corresponding activities are performed.

Verification and Validation Reports, Safety Assessment results and the final Configuration Baseline shall provide the evidence basis for the Release Approval.

Operational Monitoring Reports and other operational evidence shall subsequently provide feedback into Change Management, Incident Management, Risk Management and Continuous Improvement.

32.7 Work Product Traceability

Each applicable work product shall be traceable to the relevant requirements, lifecycle activities, configuration and safety decisions.

The organization shall maintain appropriate relationships between work products, including:

Safety Objectives → Safety Requirements → AI Requirements → Data / Model → Verification → Validation → Safety Evidence → Safety Assessment → Release Approval

Where applicable, Functional Safety, SOTIF and Cybersecurity requirements and evidence shall be integrated into this traceability structure.

32.8 Work Product Completeness

Before a defined lifecycle gate or release decision, the organization shall determine whether all mandatory and applicable work products have been established, reviewed and approved to the required extent.

Missing, incomplete or inconsistent work products shall be treated as an open issue and shall be assessed in accordance with the applicable risk, deviation and release processes.

A work product shall not be considered complete merely because a document exists. Its required content, review status, traceability and supporting evidence shall also be considered.

32.9 Work Product Control and Retention

All safety-relevant work products shall be maintained in accordance with Chapter 23, Documentation and Evidence Management.

The organization shall ensure that work products are uniquely identifiable, protected against unauthorized modification, retrievable and associated with the applicable configuration.

Superseded versions shall be retained where necessary to support historical traceability, change impact analysis, audits, assessments, incident investigations or reconstruction of previous safety decisions.

32.10 Relationship to Assessment

The work products defined in this chapter shall form an important part of the objective evidence available for internal and external assessment.

The existence of a work product alone shall not constitute evidence of compliance. The content, implementation, consistency, traceability and supporting objective evidence shall be considered.

Where an external assessment or certification scheme defines additional mandatory work products, those requirements shall be incorporated into the applicable Applicable Requirements List and project-specific safety planning.

32.11 Responsibilities

The **AI Safety Manager** shall ensure that the applicable work product framework is established and maintained.

Responsible technical and organizational functions shall create, maintain and review the work products within their areas of responsibility.
The Configuration and Change Manager shall ensure that work products are appropriately configuration-controlled.
The Safety Assessment and Release functions shall verify that the work products required for the applicable safety decision are available and sufficiently complete.

32.12 Continual Improvement of Work Products

The work product structure shall be periodically reviewed based on audit results, assessment findings, incidents, changes in AI technology, changes in applicable standards and regulatory requirements, and experience from AI projects.
Where recurring gaps or unnecessary complexity are identified, the organization shall improve the applicable procedures, templates, checklists and guidance.
The objective of Minimum Mandatory AI Safety Work Products is to establish a consistent and controlled set of documents and objective evidence that supports planning, development, verification, validation, assessment, release and continued safety assurance throughout the AI lifecycle.

33 AI Safety Gate Model

33.1 General

The organization shall establish and maintain formal AI safety gates throughout the AI lifecycle to ensure that progression to the next lifecycle phase is based on defined criteria, appropriate evidence and an authorized decision.
Safety gates provide controlled decision points at which the organization determines whether the AI system is sufficiently defined, mature and supported by objective evidence to proceed to the next lifecycle stage.
The gate model shall be applied proportionately to the safety relevance, complexity and classification of the AI system. The required evidence, review depth and independence shall be determined according to the applicable AI System Classification and Safety Management Plan.

A safety gate shall not be considered satisfied merely because the planned activities have been performed. The responsible authority shall determine whether the available evidence is sufficient to support the gate decision.

33.2 Gate Decision Principles

Each formal safety gate shall have defined entry criteria, required inputs, review activities, decision criteria and an authorized decision authority.

A gate decision shall normally result in one of the following dispositions:

- **Go** – the defined criteria are fulfilled and the lifecycle may proceed.
- **Go with Conditions** – the lifecycle may proceed under explicitly documented restrictions, actions or conditions that are formally approved and controlled.
- **No-Go** – the criteria are not sufficiently fulfilled and progression shall not occur until the identified issues have been resolved or otherwise formally addressed.

A “Go with Conditions” decision shall not be used to circumvent mandatory safety requirements or to permit uncontrolled accumulation of unresolved safety issues. Conditions shall have defined responsibilities, due dates or closure criteria and shall be subject to appropriate follow-up.

33.3 Gate 0 – AI Use-Case Approval

Key question:

Is AI appropriate for the intended function?

Gate 0 establishes whether the proposed use of AI is justified and appropriate for the intended function before significant AI development activities are initiated.

The organization shall define the intended use, intended functionality, operating context, relevant stakeholders and preliminary system boundaries. The suitability of an AI-based implementation shall be considered in relation to the intended function, foreseeable operating conditions and potential safety consequences.

Where the use of AI introduces safety-relevant uncertainties, limitations or dependencies that cannot reasonably be controlled, the organization shall determine whether an alternative technical approach is required.

The gate shall establish the basis for the subsequent AI System Classification and AI Safety Planning.

Typical outputs: approved use case, preliminary intended-use definition, initial safety relevance, initial assumptions and decision on whether AI development may proceed.

33.4 Gate 1 – Safety Classification

Key question:

What is the safety relevance of the AI system?

Gate 1 establishes the formal safety classification of the AI system or AI component.

The classification shall consider the intended function, potential consequences of incorrect or unexpected AI behavior, system architecture, degree of autonomy, interaction with safety-related functions, operating conditions, assumptions, applicable standards and regulatory requirements, and interfaces with Functional Safety, SOTIF and Cybersecurity.

The classification shall determine the level of lifecycle control, V&V, independence, documentation, assessment and release evidence required.

The resulting **AI System Classification** shall be approved by the authorized authority before safety planning and development activities are progressed beyond the defined entry point.

33.5 Gate 2 – Safety Concept Approval

Key question:

Are safety objectives, assumptions and requirements adequately defined?

Gate 2 confirms that the safety basis for development has been sufficiently established.

The organization shall review the applicable safety objectives, safety requirements, assumptions of use, intended operating conditions, relevant risks and high-level safety measures. Where applicable, the relationship to ISO 26262 safety goals and requirements, ISO 21448 SOTIF considerations and ISO/SAE 21434 cybersecurity dependencies shall be established.

The organization shall also determine whether the defined safety requirements are sufficiently clear to support subsequent AI development, data definition and V&V planning.

Where the Safety Concept is applicable, its approval shall provide the controlled basis for the subsequent development activities.

33.6 Gate 3 – Development Readiness

Key question:

Are data, tools, competence and development infrastructure ready?

Gate 3 determines whether the organization is sufficiently prepared to perform controlled AI development.

The review shall consider, as applicable, the availability and suitability of required datasets and data-management processes, development tools and toolchain controls, competent personnel, development and training infrastructure, configuration-management arrangements, cybersecurity controls and required development procedures.

The organization shall also confirm that the planned development environment is capable of producing controlled and traceable AI configurations.

Development shall not proceed into safety-relevant activities where essential resources, competence or infrastructure are absent or where their absence could invalidate subsequent safety evidence.

33.7 Gate 4 – Verification Readiness

Key question:

Is the AI system sufficiently mature for formal verification?

Gate 4 establishes whether the AI system has reached a sufficiently controlled and stable configuration for formal verification activities.

The applicable model, software, data, configuration and relevant development artifacts shall be identified and controlled. Requirements and acceptance criteria shall be sufficiently defined to permit objective verification.

The organization shall confirm that the Verification Plan is available and that the verification environment, methods, tools, test cases and required competence are ready.

Where significant uncontrolled development changes remain likely to invalidate the verification results, formal verification shall not be considered ready.

33.8 Gate 5 – Validation Approval

Key question:

Does objective evidence demonstrate acceptable behavior?

Gate 5 determines whether the results of verification and validation provide sufficient objective evidence regarding the AI system's behavior for its intended use and relevant operating conditions.

The review shall consider the applicable Validation Plan and Validation Report, scenario coverage, boundary conditions, robustness, limitations, uncertainty, foreseeable misuse, relevant SOTIF considerations and interaction with the surrounding system.
Validation shall consider not only nominal performance but also whether the AI system behaves acceptably under the conditions relevant to its intended use.
Open anomalies, deviations, limitations and residual risks shall be explicitly evaluated.
Successful execution of tests alone shall not constitute automatic validation approval. The evidence shall be reviewed against the defined acceptance criteria and safety objectives.

33.9 Gate 6 – Safety Release

Key question:
Is the residual risk acceptable and is the evidence package complete?
Gate 6 is the formal safety release decision point.
Before release, the organization shall review the complete safety evidence applicable to the defined release configuration. This shall include, as applicable, requirements, risk analysis, data and model information, verification and validation results, integration evidence, Functional Safety/SOTIF/Cybersecurity interfaces, configuration baseline, deviations and waivers, safety assessment results and the Safety Case or Safety Argument.
The organization shall determine whether the residual risks are acceptable against the applicable acceptance criteria and whether assumptions, limitations, restrictions and operational conditions are sufficiently understood and controlled.
The release decision shall identify the exact configuration, intended use and applicable restrictions.
The release authority shall be competent and formally authorized. Where required by the classification or applicable conformity-assessment scheme, the required independent assessment or confirmation measures shall have been completed.
A completed Verification or Validation Report shall not, by itself, constitute authorization for release.

33.10 Gate 7 – Operational Monitoring

Key question:
Does operational evidence continue to support the safety argument?
Gate 7 establishes the continuing assurance of the released AI system during operation.

The organization shall monitor relevant operational evidence and determine whether the assumptions, limitations, safety requirements, safety mechanisms and safety argument remain valid. Operational monitoring shall consider, as applicable, AI performance, safety-related events, anomalies, model degradation or drift, changes in data distribution, operation outside validated conditions, cybersecurity events, customer and field feedback, and newly identified unsafe scenarios. Where operational evidence indicates that the safety basis may no longer be valid, the organization shall initiate the applicable Incident Management, Change Management, Re-Verification, Re-Validation, Safety Assessment and Re-Approval activities.

Gate 7 therefore represents a **continuing lifecycle control rather than a single one-time approval event**.

33.11 Gate Evidence and Records

Each safety gate shall have controlled evidence demonstrating:

- the gate applicable to the AI system;
- the defined entry and exit criteria;
- the required work products;
- the configuration under review;
- the review participants and responsible functions;
- identified issues, deviations and conditions;
- the gate decision;
- the decision authority;
- any required follow-up actions and closure criteria.

Gate records shall be maintained as safety-related evidence in accordance with Chapter 23, Documentation and Evidence Management, and shall be traceable to the applicable configuration and lifecycle phase.

33.12 Gate Responsibilities and Independence

The responsibilities and authorities for each gate shall be defined in the AI Safety Management Plan and the applicable Authority and Responsibility Matrix.

The person or function performing development shall not automatically have authority to approve the corresponding safety gate where independence is required by the safety classification, applicable standard, certification scheme or organizational rules.
For safety-relevant decisions, the organization shall ensure that the decision authority has sufficient competence, authority and access to objective evidence.

33.13 Gate Repetition and Re-Opening

A safety gate may be repeated or re-opened when subsequent information invalidates the original decision.
This may occur following significant changes, incidents, newly identified risks, cybersecurity events, changed operating conditions, changes to the ODD or intended use, significant model or dataset changes, new unsafe scenarios, failed assumptions or findings from operational monitoring.
A previous gate approval shall therefore not be considered permanently valid when the conditions or configuration on which the decision was based have materially changed.

33.14 Relationship to the AI Safety Lifecycle

The gate model shall be integrated with the lifecycle defined in Chapter 9 and the lifecycle entry and exit criteria defined in Chapter 10.
The relationship can be summarized as:
AI Use Case → Gate 0 → Classification → Gate 1 → Safety Concept → Gate 2 → Development → Gate 3 → Verification → Gate 4 → Validation → Gate 5 → Safety Release → Gate 6 → Operation → Gate 7 → Continuous Assurance
Changes occurring after release shall follow the controlled sequence defined in Chapter 18:
Change → Impact Analysis → Re-Verification → Re-Validation → Re-Assessment → Re-Approval → Release
Where the change affects the assumptions or evidence supporting an earlier gate, the relevant gate shall be re-opened or repeated as appropriate.

33.15 Gate Model and Minimum Work Products

The gate model shall be supported by the minimum mandatory work products defined in Chapter 32, **Minimum Mandatory AI Safety Work Products**.

Note: Since this is now also Chapter 32, I would renumber the previous chapter you gave me as **Chapter 31** and this one as **Chapter 32**, or, preferably, make the previous one Chapter 31 and keep the current chapter as 32. The two chapters are complementary:

- **Chapter 31 – Minimum Mandatory AI Safety Work Products**
- **Chapter 32 – AI Safety Gate Model**

That avoids having two different chapters numbered 32.

33.16 Continual Improvement of the Gate Model

The effectiveness of the gate model shall be periodically reviewed using audit results, assessment findings, incidents, deviations, release decisions, operational experience and lessons learned.
The organization shall evaluate whether gate criteria are sufficiently clear, whether decisions are consistently applied, whether evidence is available at the appropriate lifecycle stage, and whether recurring issues indicate weaknesses in planning, competence, V&V, risk management or organizational controls.
The objective of the AI Safety Gate Model is to ensure that progression through the AI lifecycle is controlled by explicit, evidence-based and authorized safety decisions, and that an AI system proceeds toward development, validation, release and operation only when the applicable safety criteria and evidence are sufficiently established.

34 Records

34.1 General

The organization shall establish and maintain records necessary to demonstrate the implementation, operation, effectiveness and continual improvement of the AI Safety Management System and the safety of applicable AI systems.
Records shall provide objective evidence that required activities have been performed, decisions have been made by authorized personnel, applicable requirements have been addressed, and safety-related conclusions are supported by appropriate evidence.

Records shall be maintained throughout the applicable AI lifecycle and for the retention period defined by the organization's document-retention policy and applicable contractual, regulatory, legal and certification requirements.

34.2 Minimum Records

As applicable to the AI Safety Management System and individual AI projects, the following records shall be retained:

Record Category	Examples
AI Safety Management	AI Safety Policy, AI Safety Management Plans, safety planning and gate decisions
Requirements	AI safety requirements, system requirements, applicable standards and regulatory requirements
Risk Management	Safety risk analyses, assumptions, risk evaluations and residual-risk decisions
Data	Dataset specifications, dataset versions, data-management records and relevant data-quality evidence
AI Models	Model versions, model configurations, training configurations and released model baselines
Verification and Validation	Test plans, test results, verification reports, validation reports and supporting evidence
Safety Assessment	Safety assessment reports, assessment findings and conclusions
Audits	Internal audit reports, external assessment records and corrective-action records
Deviations and Waivers	Deviation records, waiver records, justifications, conditions and approvals
Changes	Change requests, impact analyses, implementation records, re-verification/re-validation and approval records
Competence and Training	Training records, competence evaluations and qualification/authorization records
Release	Release decisions, release approvals, released configurations and applicable restrictions
Operational Safety	Monitoring records, incidents, problems, field feedback and operational safety evidence
Management Review	Management review inputs, decisions, actions and follow-up records

Table 5: Minimum Records

The actual record set shall be determined according to the AI System Classification, applicable lifecycle activities and project-specific safety planning.

34.3 Record Identification and Traceability

Records shall be uniquely identifiable or otherwise unambiguously associated with the relevant project, AI system, lifecycle phase and configuration.

Where relevant, records shall include or provide traceability to:

- the responsible organization or person;
- date and status;
- applicable AI system or product;
- relevant model, dataset or software version;
- applicable configuration baseline;
- related requirements or safety objectives;
- associated changes, deviations or incidents; and
- resulting decisions and approvals.

Records supporting safety decisions shall be sufficiently traceable to demonstrate the basis of the decision.

34.4 Record Integrity and Protection

Records shall be protected against unauthorized modification, deletion, loss, corruption or substitution.

Electronic records shall be maintained using controlled repositories or systems appropriate to their safety relevance. Access shall be controlled according to organizational responsibilities and authorization.

Where records are generated or maintained by external organizations, suppliers or service providers, their integrity, accessibility and relationship to the applicable AI system shall be appropriately controlled.

34.5 Configuration of Records

Safety-relevant records shall be associated with the configuration to which they apply.

In particular, verification, validation, assessment and release records shall be traceable to the relevant model, dataset, software, hardware, tools and other configuration items where these can influence the validity of the evidence.

Records from a previous configuration shall not automatically be considered valid for a changed configuration. Their continued applicability shall be determined through the Change Management and impact-analysis process defined in Chapter 18.

34.6 Record Retention

Records shall be retained for periods defined by the organization's document-retention policy, taking into account applicable:

- legal and regulatory requirements;
- contractual and customer requirements;
- certification and conformity-assessment requirements;
- product and system lifecycle;
- safety-relevant obligations;
- incident and liability considerations; and
- organizational requirements.

Where different retention requirements apply, the organization shall apply the applicable longer retention period unless otherwise legally or contractually determined.

Records shall not be destroyed while they remain necessary for an open safety investigation, incident, assessment, corrective action, legal obligation or other controlled activity.

34.7 Record Retrieval and Availability

Records shall remain retrievable for the period for which they are required.

The organization shall ensure that records can be made available, as appropriate, for:

- safety reviews and release decisions;
- internal and external audits;
- conformity assessments;
- regulatory activities;
- incident investigations;
- change-impact analysis;

- customer or contractual obligations; and
- reconstruction of historical safety decisions.

Archived records shall remain identifiable and readable for the required retention period.

34.8 Record Disposal

At the end of the applicable retention period, records may be disposed of in accordance with the organization's-controlled document-retention and disposal process.
Before disposal, the organization shall confirm that no legal, contractual, regulatory, safety, incident, assessment or other controlled obligation requires continued retention.
The disposal of records shall itself be controlled where required to demonstrate compliance with the retention policy.

34.9 Responsibilities

Process owners and responsible technical functions shall ensure that records generated by their activities are complete, accurate and appropriately maintained.
The AI Safety Manager shall ensure that records necessary to demonstrate the effectiveness of the AI Safety Management System are defined and controlled.
Configuration and Document/Evidence Management functions, where established, shall support identification, storage, retrieval, protection and retention.
Management shall provide appropriate resources to ensure that safety-relevant records remain available and protected throughout the required retention period.

34.10 Review and Continual Improvement

The organization shall periodically review the adequacy of its record-management arrangements, taking into account audit findings, assessment experience, incidents, retrieval difficulties, configuration discrepancies and changes in legal, regulatory or contractual requirements.
Deficiencies affecting the integrity, availability or traceability of safety records shall be treated as controlled issues and corrected appropriately.
The objective of Records Management is to ensure that reliable, protected and traceable records are retained for the required period and that the organization can reconstruct and

demonstrate the basis of its AI safety activities, decisions and conclusions throughout the applicable lifecycle.

35. Compliance Statement

35.1 General

[Company Name] shall establish, implement, maintain and continually improve an AI Safety Management System appropriate to the safety relevance, complexity, technology, intended use and operating conditions of its AI systems.

The organization shall ensure that AI safety is addressed throughout the applicable AI lifecycle, from use-case definition and classification through development, verification, validation, release, deployment, operation, monitoring, change and retirement.

The organization shall systematically manage the interfaces between AI safety and other applicable disciplines, including **Functional Safety**, **SOTIF** and **Cybersecurity**. Where applicable, interfaces to software update engineering, system engineering, quality management, regulatory compliance and other relevant processes shall also be established and maintained.

35.2 Compliance with this Manual

Compliance with this manual shall be demonstrated through objective evidence.

Such evidence shall include, as applicable, documented processes, responsibilities, plans, requirements, risk analyses, data and model records, verification and validation results, configuration records, assessments, audit results, safety cases, release decisions, operational monitoring records and management reviews.

The organization shall ensure that applicable requirements defined by this manual are implemented within the relevant organizational and project processes.

Where requirements are identified as **“where applicable”**, their applicability shall be determined and documented based on the AI system classification, intended use, safety relevance, applicable standards and regulations, contractual requirements and project-specific circumstances.

35.3 Relationship to Applicable Standards

Compliance with this manual shall not be interpreted as automatic compliance with any external standard, regulation or certification scheme.

Where external requirements are applicable, the organization shall identify and manage them through the Applicable Requirements List and associated project processes.

Compliance with standards such as ISO 26262, ISO 21448, ISO/SAE 21434, ISO 24089, ISO/PAS 8800 or other applicable standards shall be demonstrated against the respective requirements and applicable conformity-assessment criteria.

The AI Safety Management System provides the organizational and lifecycle framework for managing these requirements; it does not replace the technical or conformity-assessment requirements of the applicable standards.

35.4 Product Safety and Compliance

Compliance with this manual does **not** by itself constitute evidence that a specific AI product, AI component or AI-enabled system is safe.

Product safety shall be demonstrated through the applicable engineering activities, risk analysis, requirements, verification, validation, assessments, safety arguments, configuration control and release decisions defined for the particular AI system.

Similarly, compliance with this manual does not by itself constitute regulatory approval, type approval, certification or conformity with a particular standard.

Where a formal certification, conformity assessment or regulatory approval is required, the applicable independent assessment and decision processes shall be performed in accordance with the relevant scheme and requirements.

35.5 Management Commitment to Compliance

Top Management shall ensure that the AI Safety Management System remains suitable and effective for the organization's AI activities.

Compliance shall be supported by appropriate resources, competent personnel, defined responsibilities and authorities, controlled processes and objective evidence.

Commercial, schedule or operational pressures shall not justify the omission or circumvention of mandatory safety activities.
Nonconformities, deviations and unresolved safety issues shall be managed through the controlled processes defined in this manual.

35.6 Continuing Validity

Compliance shall be considered a continuing organizational responsibility rather than a one-time activity.
The organization shall monitor changes to AI technology, products, intended use, operating conditions, applicable standards, regulations, organizational structure and safety experience and shall determine whether such changes affect the AI Safety Management System or the validity of existing safety evidence.
Where necessary, the organization shall initiate the applicable change management, re-verification, re-validation, safety assessment, corrective action or re-approval activities.

35.7 Compliance Declaration

Based on the implementation and maintenance of the AI Safety Management System and the availability of appropriate objective evidence, **[Company Name]** may formally declare conformity with this manual within the defined organizational and applicable AI system scope.
Such a declaration shall identify the applicable scope and shall not extend beyond the evidence and formal assessments supporting the declaration.
The objective of this Compliance Statement is to establish the organization's commitment to maintaining an effective AI Safety Management System, to define the basis on which compliance with this manual is demonstrated, and to ensure that compliance with the management system is clearly distinguished from the separate determination of AI product safety, regulatory approval and conformity with external standards.
Product-specific compliance shall be demonstrated through the applicable assessment, verification, validation and certification activities.

END OF DOCUMENT