

Table of Contents

Preface	V
I. On the Scope	1
1. The Scope of the AI Act	1
a. On AI Systems	1
(1) Introduction	1
(2) A Deeper Dive: Qualitative and Quantitative Aspects of AI Systems	4
(3) On Ends and Beginnings: AI Infrastructure and AI Systems	13
(4) Solutions From the Field of AI Safety, Computation and UML	20
(5) Conclusion	24
b. Risk-based Approach	25
c. Personal Scope	27
(1) Providers	27
(2) Product Manufacturers	29
(3) Deployers	29
(4) Importers and Distributors	29
d. Territorial Scope	30
e. On AI Literacy	33
(1) Operationalizing AI Literacy	34
i. Objectives	35
ii. Training Needs, Including Target Groups and Levels of Expertise	36
iii. Training Content	36
iv. Training Methods	38
v. Training Frequency	38
vi. Evaluation	38
(2) Final Remarks on AI Literacy	39
2. The Scope of the Book	39
II. Design	41
1. Idea	41
2. Legal Requirements Engineering	42
3. Intended Purpose	43
a. Usability Engineering	47
b. Human-factor Engineering	49
4. Initial Risk Categorization and Assessment	53
III. Development	57
1. AI System Requirements	57
a. Quality Management System	57

Table of Contents

- (1) Types of Quality Management 58
- (2) Structure and Main Components..... 59
- (3) Software and AI System Quality Models 62
- (4) Operationalizing the Legal Requirements Under Article 17..... 65
 - i. First Line..... 66
 - ii. Second Line 67
 - iii. Third Line..... 69
- b. Risk Management 71
 - (1) Risk Identification..... 72
 - (2) Risk Assessment 74
 - i. Factors for Estimating Likelihood 75
 - ii. Factors for Estimating Impact..... 76
 - (3) Risk Mitigation 81
 - (4) Risk Acceptance 82
 - (5) Risk Tolerance..... 83
 - (6) Establishing Controlling, Monitoring and Incident Response Processes..... 83
 - i. Controlling 83
 - ii. Monitoring 83
 - iii. Incident Response Process 84
- c. Human Oversight..... 86
- d. Accuracy, Robustness, and Cybersecurity of AI Systems (Article 15 AI Act) 94
 - (1) Overview 94
 - i. Guiding Principles for the AI System Life Cycle (Article 15(1))..... 94
 - ii. Benchmarking and Standardization, Article 15(2) ... 101
 - iii. Documentation Obligations, Article 15(3) 102
 - iv. Resilience, Article 15(4)..... 103
 - v. Cybersecurity, Article 15(5) 105
 - (2) Detailed Explanations 112
 - i. Significance of Harmonized Norms and Standards in the AI Act 112
 - ii. Consistent Performance Throughout the AI System Life Cycle: Universal Lessons in Machine Learning 117
 - iii. What Are Feedback Loops and How Should They Be Addressed?..... 122
 - iv. “Engineering Safety in Machine Learning” 127
 - (a) Standardization: Accumulated Experience..... 127
 - (b) AI Engineering Best Practices 129

- (c) Case Study: Safety and Robustness of AI Systems in the Automotive Sector. 132
 - (d) Inherently Safe Design in Machine Learning. . . . 134
 - (e) A Framework for Testing AI Systems. 136
 - (f) AI Red Teaming 138
 - (g) NIST Test Platform “Dioptra” 139
- 2. Overview of Data Governance & Data Management (Article 10) 141
 - a. Machine Learning and Training Data in a Nutshell. 141
 - b. Mandatory Quality Criteria for Training, Validation, and Testing of AI Systems, Article 10(1) 142
 - c. Data Governance & Data Management, Article 10(2) 148
 - d. Standardization of Data (Quality) Management 148
 - e. Definitions and Metrics 150
 - (1) Training Data, Article 3(29) 150
 - (2) Validation Data, Article 3(30) 151
 - (3) Validation Dataset, Article 3(31). 151
 - (4) Testing Data, Article 3(32) 152
 - f. The Individual Elements 154
 - (1) Relevant Design Choices, Article 10(2)(a) 155
 - (2) Data Collection Processes and the Origin of Data, and in the Case of Personal Data, the Original Purpose of Data Collection, Article 10(2)(b). 156
 - i. Data Sources in Machine Learning Practice. 156
 - ii. Datasheets for Datasets. 157
 - iii. Data Protection Law 162
 - (3) Relevant Data-Preparation Processing Operations, such as Annotation, Labeling, Cleaning, Updating, Enrichment, and Aggregation, Article 10(2)(c) 170
 - i. Annotation and Labeling 170
 - ii. Data Cleaning 175
 - iii. Updating 176
 - iv. Enrichment 176
 - v. Aggregation 177
 - (4) Formulation of Assumptions, Particularly Regarding the Information that the Data are Intended to Measure and Represent, Article 10(2)(d). 178
 - (5) Assessment of the Availability, Quantity, and Suitability of the Necessary Data Sets, Article 10(2)(e). 178

Table of Contents

- (6) Examination of Possible Biases Affecting Health and Safety, Fundamental Rights, or Leading to Discrimination Prohibited Under Union Law Article 10(2)(f) and Appropriate Measures to Detect, Prevent, and Mitigate Possible Biases Identified According to Article 10(2)(g) 179
 - (7) Identification of Relevant Data Gaps or Shortcomings Preventing Compliance and Appropriate Mitigation Measures, Article 10(2)(h). 179
- g. Combating Bias and Discrimination, Articles 10(3), (4), and (5). 180
 - (1) Recitals and Ethics Guidelines for Trustworthy AI by the High-Level Expert Group on Artificial Intelligence (HLEG AI, 2019). 180
 - (2) Fundamental Rights Agency (FRA), LIBE, HLEG AI and the Toronto Declaration 184
 - (3) Research and Science: It’s Not Just the Data, Stupid!. . . . 184
 - (4) International Standardization. 185
 - (5) Relevant, Sufficiently Representative, Free of Errors and Complete in view of the Intended Purpose. 187
 - i. The Purpose of the System 187
 - ii. Relevance 188
 - iii. Representativeness 189
 - iv. Error-Free 190
 - v. Completeness 191
 - (6) Balanced Statistical Characteristics in Datasets 191
 - (7) Geographically, Contextually, Behaviorally, or Functionally Typical Datasets 193
 - i. Geographical Setting. 193
 - ii. Contextual Setting. 194
 - iii. Behavioural Setting. 195
 - iv. Functional Setting. 196
 - (8) Processing of Sensitive Data for the Analysis and Mitigation of Biases 197
 - (9) AI, Bias, and European Anti-Discrimination Law: An Overview 201
 - i. Legal Framework 201
 - ii. Scope of Application and Protected Characteristics. . 202
 - iii. Direct and Indirect Discrimination. 202
 - iv. Justification. 203
 - v. Positive Measures 204
 - vi. Reversal of Burden of Proof. 204

vii. Legal Consequences	205
viii. Respondents	205
3. Testing & Compliance	206
a. Sandboxes	207
(1) Sandboxes in the AI Act	208
(2) Establishment and Operation of AI Sandboxes.....	209
(3) A Case for Participating in AI Sandboxes?	210
(4) Sandbox Plan.....	212
(5) Exit Reports.....	213
(6) Consequences	213
(7) Processing of Data Within the Sandbox	214
b. Testing in Real World Conditions Outside AI Regulatory Sandboxes	214
c. Conformity Assessment	215
(1) What?.....	215
(2) When?	217
(3) Who and How?	219
(4) Exceptions Make The Rule	221
(5) Why?	222
d. Harmonised Standards, Common Specifications & Presumption of Conformity	225
(1) Harmonised Standards.....	225
(2) Presumption of Conformity.....	226
(3) Common Specifications.....	228
(4) Codes of Conduct and Guidelines.....	230
e. Placing on the Market	231
4. Technical Documentation	233
IV. Deployment	235
1. Providers	235
a. The Obvious.....	235
b. Documentation Keeping (Article 18) and Automatically Generated Logs (Article 19)	236
c. Risk Management	238
d. Human Oversight.....	238
e. Transparency and Provision of Information to Deployers and/or End Users	240
f. Post-market Monitoring (Article 72), Corrective Action & Duty of Information (Article 20), Reporting Serious Inci- dents (Article 73) and Cooperation with the Authorities.	244
(1) Goal and Purpose.....	244
(2) Key Features	245
(3) Post-market Monitoring Plan	246

Table of Contents

- (4) Exit Report. 251
 - (5) Interplay with Other Systems and Processes. 252
- 2. Deployers 254
 - a. The Obvious: Due Diligence, Use According to Instructions for Use (Logs, Human Oversight), Transparency, Monitoring and Duty of Information, Reporting 254
 - (1) Due Diligence: The Legal Deep-dive 254
 - (2) Use According to Instructions for Use (Logs, Human Oversight) 256
 - (3) Transparency 257
 - (4) Monitoring and Duty of Information. 258
 - b. Data Governance 260
 - c. Fundamental Rights Impact Assessment (FRIA) 261
 - (1) Who? 263
 - (2) What? 264
 - (3) Why? 266
 - i. Make Adjustments and Implement Additional Measures 267
 - ii. To Assess the Acceptability of Residual Risks 267
 - (4) End Result 267
 - d. Consulting the Works Council. 270
 - e. Data Protection Impact Assessment. 271
 - f. Article 25 Obligations Along the AI Value Chain 272
- V. Special Considerations 275**
 - 1. IP, TDM and GenAI in the AI Act 275
 - a. General Purpose AI in the AI Act 275
 - (1) AI Model Providers and AI System Providers 276
 - (2) General-purpose AI Model Provider’s Obligations. 279
 - (3) General-purpose AI System Provider’s Obligations 283
 - (4) Copyright and General-purpose AI Models. 283
 - b. Training and Training Data 284
 - (1) GenAI Needs Data (and It Needs a Lot of It) 284
 - (2) Memorization and Overfitting. 287
 - (3) Text and Data Mining 288
 - i. The Copyright in the Digital Singe Market Directive (CDSM Directive). 288
 - ii. TDM and Transformer Models. 289
 - iii. Three-Step Test 293
 - iv. Conclusion 294
 - c. Use, Inputs and Outputs. 295
 - (1) Use and Inputs 296

- (2) Outputs 297
 - (3) The Curious Case of Retrieval-Augmented Generation (RAG) 298
- 2. AI in Financial Institutions 300
 - a. Two Important Caveats 301
 - b. One Additional (Major) Burden. 302
 - c. Where It Gets Easier 302
- 3. Biometric and Emotion Recognition Systems 304
 - a. Biometric Identification. 305
 - b. Biometric Verification 309
 - c. Biometric Categorisation. 309
 - d. Emotion Recognition. 311