

Contents

- About the editors, ix
- List of Contributors, x
- Series Preface, xi
- Preface Book, xii
- Acknowledgements, xiii
- 1** Introduction, 1
Zeno Geraadts and Katrin Franke
- 2** AI-based Forensic Evaluation in Court: The Desirability of Explanation and the Necessity of Validation, 3
Rolf J.F. Ypma, Daniel Ramos, and Didier Meuwly
 - 2.1 Introduction, 3
 - 2.1.1 AI for Forensic Evaluation, 6
 - 2.2 The Desirability for Explanation and the Necessity of Validation, 7
 - 2.3 Explainability (and its Validity), 8
 - 2.3.1 Reasons to Pursue Explanations, 9
 - 2.3.2 Types of Explanations, 9
 - 2.3.3 Limitations of Explanations, 11
 - 2.4 Validation (and its Explanation), 11
 - 2.4.1 Measure the Method's Performance, 12
 - 2.4.2 Approach in Four Steps, 12
 - 2.4.3 Accountability, 16
 - 2.5 Conclusion, 17
- 3** Machine Learning for Evidence in Criminal Proceedings: Techno-legal Challenges for Reliability Assurance, 21
Radina Stoykova, Jeanne Mifsud Bonnici, and Katrin Franke
 - 3.1 Introduction: AI in the Intersection of Criminal Procedure and Forensics, 21
 - 3.1.1 Technical Fragmentation in Digital Investigations, 21
 - 3.1.2 Legal and Methodological Fragmentation in Digital Investigations, 22
 - 3.1.3 Specifics of ML-based Investigative Approach, 23
 - 3.1.4 Scope and Definitions, 25
 - 3.2 Legal Framework, 27
 - 3.2.1 The Fair Trial Principle, 28
 - 3.2.2 Necessity and Proportionality of Investigative Measures, 32
 - 3.2.3 The AIA Proposal, 33

- 3.2.4 AI System Development and Legislative Contradictions, 35
- 3.3 Machine Learning Pipelines: Techno-legal Challenges, 44
 - 3.3.1 Task + Purpose Limitation and Data Minimization, 44
 - 3.3.2 Dataset Engineering and Data Governance, 48
 - 3.3.3 Pre-processing for Input: Trade-offs between Accuracy and Computational Costs, 53
 - 3.3.4 Modelling, 56
- 3.4 AI Use in Investigations: AI System Design + Data Protection = Fair Trial?, 63
- 3.5 Conclusion, 66
- 4 Formalising Representation and Interpretation of Digital Evidence to Reinforce Reasoning and Automated Analysis, 74
Eoghan Casey and Timothy Bollé
 - 4.1 Introduction, 74
 - 4.2 Background and Related Work, 76
 - 4.3 Method, 77
 - 4.4 Representing Digital Traces, 79
 - 4.5 Representing Computed Similarity, 86
 - 4.6 Representing ML Classification, 89
 - 4.7 Representing Hypothesis Test Results (a.k.a. Inferences), 91
 - 4.7.1 Location Example, 93
 - 4.7.2 Identification Example, 95
 - 4.8 Effective/Reliable/Responsible Automated Analysis, 99
 - 4.9 Conclusion, 101
- 5 Servicing Digital Investigations with Artificial Intelligence, 103
Harm van Beek and Hans Henseler
 - 5.1 Introduction, 103
 - 5.2 Introduction To Hansken, 104
 - 5.2.1 Normalized Trace Model, 105
 - 5.2.2 Forensic Tool Application, 106
 - 5.2.3 Hansken's Application Programming Interfaces, 108
 - 5.3 Large Scale Application of AI Techniques, 109
 - 5.3.1 Rule-based AI Techniques Implemented in Hansken, 109
 - 5.3.2 Deep-learning AI Techniques Currently Implemented in Hansken, 111
 - 5.3.3 Deep-learning AI Techniques to be Implemented in Hansken, 115
 - 5.3.4 The application of large language models in digital forensics, 118
 - 5.4 Conclusions and Further Reading, 120
- 6 On the Feasibility of Social Network Analysis Methods for Investigating Large-scale Criminal Networks, 123
Jan William Johnsen and Katrin Franke
 - 6.1 Introduction, 123
 - 6.2 Previous Work, 125

6.3	Material and Methods, 127
6.3.1	Real-world Underground Forum Database Dumps, 127
6.3.2	Network Centrality Measures, 129
6.3.3	Measuring Association Using Bi-variate Analysis, 129
6.3.4	Topic Modelling Algorithms, 130
6.4	Experimental Setup, 130
6.4.1	Evaluating Network Centrality Measures for Forensics, 130
6.4.2	Our Novel Approach for Analysing Cybercriminal's Technical Skills, 133
6.5	Experimental Results and Discussion, 137
6.5.1	Correlation Testing, 137
6.5.2	Our Newly Proposed Method, 142
6.6	Conclusion, 145
7	Mapping NLP Techniques to Investigations and Investigative Interviews, 149 <i>Kyle Porter and Bente Skattør</i>
7.1	Introduction, 149
7.2	Criminal Investigation, 150
7.2.1	Investigative Interviews, 150
7.3	Assessing the Needs of Investigators in an NLP Context, 151
7.3.1	Mapping Interviewer Needs to Existing NLP Tasks, 151
7.4	Automatic Speech Recognition, 152
7.4.1	ASR Basics, 152
7.4.2	ASR, Digital Investigation, and the State of the Art, 153
7.5	NLP Basics, 154
7.5.1	Common Terminology, 154
7.5.2	Vector Space Models and Embeddings, 156
7.5.3	Modern NLP Models, 157
7.6	Text Extraction, 157
7.6.1	Entity Identification and Named Entity Recognition, 157
7.6.2	Named Entity Recognition Metrics, 158
7.6.3	NER Applied to Investigations, 159
7.6.4	Entity Linking, 159
7.6.5	Limitations of Using NER, 160
7.6.6	Extraction Methods outside NER, 161
7.7	Text Classification, 161
7.7.1	Classification Evaluation Metrics, 162
7.7.2	Text Classification and Digital Investigation, 162
7.7.3	Classification Limitations, 163
7.8	Text Reduction, 164
7.8.1	Thematic Extraction and Topic Modelling, 164
7.8.2	Topic Modelling and Digital Investigations, 165
7.8.3	Limitations of Topic Modelling, 166

- 7.8.4 Text Summarization, 166
- 7.8.5 Text Summarization and Digital Investigations, 167
- 7.8.6 Summarization Limitations, 167
- 7.9 Discussion and Conclusion, 167
 - 7.9.1 Future Work, 169
- 8 The Influence of Compression on the Detection of Deepfake Videos, 174**
Meike Kombrink and Zeno Geraads
 - 8.1 Introduction, 174
 - 8.2 Method, 178
 - 8.2.1 Dataset, 178
 - 8.2.2 Deepfake Detection, 180
 - 8.3 Results, 183
 - 8.3.1 Compressed Dataset, 183
 - 8.3.2 Algorithms, 184
 - 8.4 Discussion, 190
 - 8.4.1 Deepfake Detection, 190
 - 8.4.2 Compression, 191
 - 8.4.3 Future Work, 193
 - 8.5 Conclusion, 193
- 9 Event Log Analysis and Correlation: A Digital Forensic Perspective, 195**
Neminath Hubballi and Pratibha Khandait
 - 9.1 Introduction, 195
 - 9.2 Sources of Logs, 197
 - 9.2.1 End Host System Logs, 198
 - 9.2.2 Networking Devices and Security Applications, 203
 - 9.2.3 Application Logs, 207
 - 9.3 Need for Correlation, 208
 - 9.4 Correlation Techniques, 210
 - 9.5 Conclusions, 214
- 10 (Hyper-)graph Analysis and its Application in Forensics, 216**
Marcel Worring
 - 10.1 Introduction, 216
 - 10.2 Survey of Methods, 218
 - 10.2.1 Preliminaries, 218
 - 10.2.2 Tasks, 219
 - 10.2.3 Graph Neural Networks, 220
 - 10.3 Explainability and Visualization, 224
 - 10.4 Conclusion, 227
- 11 Conclusion, 230**
Zeno Geraads and Katrin Franke
- Index, 232