# Contents

**Acknowledgments —— V**

**List of Abbreviations —— IX**

**1 Introduction —— 1**

**2 Opacity in Machine Learning —— 10**
2.1 Machine Learning —— **10**
2.1.1 General Remarks —— **10**
2.1.2 Statistical Learning Theory —— **13**
2.1.3 Deep Neural Networks —— **17**
2.2 The Problem of Opacity —— **20**
2.3 Explainable Artificial Intelligence —— **26**

**3 Overcoming Opacity with Explanations —— 33**
3.1 Scientific Explanation —— **33**
3.1.1 The Hempelian View —— **34**
3.1.2 The Statistical-Relevance Account —— **41**
3.1.3 Applications to Machine Learning —— **43**
3.1.4 The Pragmatic Perspective —— **47**
3.2 Epistemic Explanation —— **53**
3.2.1 The First-Person Perspective on Explanation —— **54**
3.2.2 Explanation as a Reduction of Surprise —— **58**
3.3 Means-End Epistemology and Explanation —— **67**

**4 A Means-End Account of Explainable Artificial Intelligence —— 73**
4.1 The Account —— **73**
4.2 Descriptive Component —— **78**
4.3 Normative Component —— **81**

**5 Towards Effective Guidelines for Machine Learning —— 87**
5.1 The Means-End Account in Practice —— **87**
5.2 Appropriate Explanations —— **88**
5.3 Implications for Effective Guidelines —— **93**

**6 Conclusion —— 101**

**Bibliography —— 105**

**Index —— 113**