

BRIEF CONTENTS

Foreword	xvii
Acknowledgments	xix
Introduction	xxi
PART I: AI AND SECURITY FUNDAMENTALS.....	1
Chapter 1: What Is AI?	3
Chapter 2: Working with Models	29
Chapter 3: AI Threats	75
PART II: ATTACKING AND DEFENDING AI	109
Chapter 4: Attacks and Weaknesses	111
Chapter 5: Defenses, Controls, and Mitigations	155
PART III: THE AI SECURITY ECOSYSTEM	189
Chapter 6: Red Teaming AI	191
Chapter 7: Attacking and Defending with AI	221
Chapter 8: AI Safety	243
Chapter 9: AI Governance	281
Chapter 10: What's Next for AI Security?	309
Conclusion: A New Kind of Hacker	329
Figure Credits	331
Index	333