CAROLINE WONG

THE AI CYBER-SECURITY HANDBOOK

THE AI CYBER-SECURITY HANDBOOK

CAROLINE WONG

THE AI CYBER-SECURITY HANDBOOK

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

The manufacturer's authorized representative according to the EU General Product Safety Regulation is Wiley-VCH GmbH, Boschstr. 12, 69469 Weinheim, Germany, e-mail: Product_Safety@wiley.com.

Library of Congress Cataloging-in-Publication Data has been applied for:

Paperback ISBN: 9781394340866
ePDF ISBN: 9781394340880
epub ISBN: 9781394340873
oBook ISBN: 9781394421732

Cover Design: Wiley
Cover Image: © 300 librarians/Shutterstock

For everyone who shows up with heart in a world that rewards speed.
For the ones who build, protect, and imagine better systems—and do it with empathy.
And for my family, who teach me every day that true resilience isn't technical—it's human.

About the Author

Caroline Wong is a cybersecurity executive, author, and educator with more than 20 years of experience leading global information security, risk, and compliance programs. She has held leadership roles at eBay, Zynga, Symantec, Cigital, Cobalt, and Teradata.

Her first book, *Security Metrics: A Beginner's Guide* (McGraw-Hill, 2011), was inducted into the Cybersecurity Canon Hall of Fame in 2022. Caroline also teaches through LinkedIn Learning and hosts the podcasts "Humans of InfoSec," "The AI Security Edge," and "Wake Up."

Her work bridges technology, psychology, and resilience—exploring how artificial and human intelligence can work together to build a safer digital future.

Contents

Preface

My first cybersecurity book came out in 2011, when AI still felt like science fiction. Back then, the industry was obsessed with metrics, compliance, and control frameworks. Today, we're grappling with something far more dynamic—intelligent systems that learn, create, defend, and sometimes deceive.

This book builds on that foundation, connecting two decades of lessons in security, risk, and resilience to the new realities of AI-powered defense and AI-enabled attack.

It was born out of countless conversations with practitioners, researchers, and executives trying to make sense of this shift. Each asked some version of the same question: What does AI really mean for cybersecurity?

I wrote this book to answer that question as honestly and practically as possible. You won't find hype here. What you'll find instead are frameworks, stories, and patterns that reflect the reality of AI's dual nature—its power to help and its potential to harm.

As both an engineer and a humanist, I've always believed that security is fundamentally about people—how we think, what we value, and how we make decisions under uncertainty. The same is true for AI. Behind every model and algorithm are human choices about data, ethics, and trust.

My hope is that this book helps you see AI and cybersecurity not as competing forces, but as intertwined disciplines shaping the future of resilience. Whether you're a student, a researcher, a policymaker, or a security leader, I hope you'll find ideas here that challenge you to think deeper, act smarter, and lead with empathy, in a world where technology never stops changing.

Acknowledgments

No book is written alone.

This one, especially, was shaped by the insights, conversations, and challenges shared by so many people across the cybersecurity community.

My deepest thanks to Gisela Hinojosa, my best friend and the technical editor for this book. Gisela taught me nearly everything I know about offensive security—and about the curiosity, precision, and persistence that define great hackers. She also reminded me that security, at its best, is creative work.

To my Wiley editorial team, thank you for your patience, structure, and belief in this project. You guided it from a "crazy" idea to a finished book.

To my colleagues and friends across the industry—thank you for pushing me to ask better questions and never accept easy answers. To the IANS community, for the candid and thoughtful discussions that shaped so many of these ideas. To the guests and listeners of The AI Security Edge—thank you for bringing your

honesty, curiosity, and experience to every conversation. Your thoughts gave this book depth and perspective.

If this book changes how even one person thinks about the relationship between AI, security, and humanity, it's because of the people who made it possible.

Introduction

Artificial intelligence is changing cybersecurity faster than most people expected. It's accelerating attacks, transforming defenses, and redefining how we think about risk, resilience, and responsibility.

This book is for practitioners, leaders, and learners who want to understand—not just what's happening—but why it matters and how to respond.

When I started researching AI's role in cybersecurity, I noticed a widening gap between hype and reality. Executives were asking how AI could make their defenses smarter, but few understood the new dependencies and vulnerabilities introduced by AI systems themselves. Meanwhile, defenders were being hit with nonstop marketing noise and buzzwords, struggling to tell the difference between genuine innovation and empty claims.

My goal with this book is to bridge that gap—to provide clarity, precision, and practical insight grounded in research and real-world experience.

Each chapter explores a different dimension of AI and cybersecurity: from AI-accelerated attacks and model manipulation to governance, resilience, and the future of human–machine collaboration. You'll see how attackers are using AI to exploit systems, how defenders are applying machine learning to stay ahead, and how organizations can build trustworthy AI ecosystems aligned with emerging global standards like NIST, ISO, and the EU AI Act.

This is not a book about fear—it's a book about agency. As AI becomes embedded in every layer of technology, each of us has a role to play in shaping how it's used, secured, and governed.

By the time you reach the end, I hope you'll feel more informed, more empowered, and better equipped to navigate the shifts unfolding at the intersection of AI and cybersecurity—because understanding AI isn't just about keeping up. It's about staying human while everything around us gets smarter.

Chapter Summaries

Part I—Cyber's New Player: AI on Both Teams

Part I explores how AI changes the tempo of offense and defense.

Chapter 1: The Age of Adaptive Adversaries—AI's Impact on the Security Balance of Power

The big picture: AI changes who gets to play and how fast the game moves. The advantage isn't in having AI—it's in knowing how to use it. You'll see how AI empowers both attackers and defenders, and why true resilience comes from mastering both.

Chapter 2: Smarter, Faster, Cheaper Recon: What Hackers Know Before You Even Notice

Reconnaissance has gone from manual to machine-speed. This chapter shows how AI automates OSINT, maps infrastructure, links identities, and profiles targets—quietly and continuously. You'll see why "recon" is now a living model of your organization and learn how to blunt it.

Chapter 3: Polymorphic to Predictive: How AI Supercharges Malware at Machine Speed

Malware no longer just "runs," it learns. This chapter unpacks polymorphism, metamorphism, mimicry, and reinforcement learning, and explains why static defenses falter. The takeaway: shift detection to behavior, context, and intent.

Chapter 4: Breaking the Rules: Novel Attacks and the Limits of Signature Detection

Bots act like people. Signature-based detection can't keep pace. This chapter explores AI-powered botnets, device fingerprint spoofing, CAPTCHA bypass, and commoditized "as-a-service" attack kits—and explains how to pivot from rules to risk scoring and real-time behavioral defense.

Chapter 5: Hyper-real, Hyper-targeted: Deepfakes, Phishing, and AI-powered Deception

Social engineering is now industrialized. LLMs and synthetic media enable flawless language, voice, and video. This chapter focuses on detection by context and intent, layered verification, and communication hygiene that scales.

Part II—Building Resilience with AI

Part II examines how defenders use AI to see earlier, decide faster, and act with confidence.

Chapter 6: Patterns of Life: Learning What's Normal and Spotting What's Not

Defense that learns. This chapter shows how to baseline users, systems, and apps; detect subtle anomalies; and reduce noise while catching "low and slow" attacks. Core idea: understand the "good" deeply to spot the "bad" early.

Chapter 7: From Alert to Containment: AI-augmented Incident Response

Incident response is a race against time. This chapter covers AI for triage, correlation, natural-language investigations, and automated containment. You'll build playbooks that shrink dwell time without sacrificing judgment.

Chapter 8: The Illusion of Trust: Poisoned Data, Tampered Models, and AI Supply Chain Risk

Dependencies define risk. This chapter unpacks poisoned data, tampered models, third-party tools, and hidden backdoors—and the controls that surface provenance, monitor continuously, and keep trust measurable, not assumed.

Chapter 9: Building Secure Model Supply Chains with AI-augmented Security Teams

Strong programs depend on strong teams. This chapter outlines roles, workflows, and guardrails for human-AI collaboration—from detection and vendor risk to stakeholder comms—so automation scales with accountability.

Part III—Staying Grounded: Trust, Risk, and Accountability

Part III explains how to lead with clarity when technology moves faster than policy. The takeaway is to light a path through uncertainty—and invite others to follow.

Chapter 10: Human Values, Machine Decisions: The Ethics of AI in Security

The future of security depends on where we choose to draw the line between autonomy and oversight—between what we delegate to machines and what we preserve for ourselves. This chapter examines

bias, privacy, enforcement, and the governance moves that keep humans responsible for outcomes—no matter how fast systems decide.

Chapter 11: The Human Advantage: Leadership, Learning, and the Future of AI Security

By 2030, security is neither human-only nor machine-only. This chapter looks ahead to operating models, metrics, and leadership habits that balance speed with stewardship—and design programs that improve under pressure.

Note to the Reader

This isn't a textbook—it's a field guide for a field that's still being defined.

When I started writing *The AI Cybersecurity Handbook*, I wanted to create something I wish I'd had years ago: a book that connects the dots between AI, cybersecurity, and the people doing the work. The ideas here come from two decades of conversations—with builders, breakers, defenders, and dreamers.

You can read it front to back or dip into the chapters that matter most to you. If you lead teams, start with the sections on governance and resilience. If you work closer to the code or the data, explore the chapters on adversarial AI, model tampering, and AI-driven detection.

What you'll find throughout are frameworks, stories, and questions meant to make you think—not just about how AI changes security, but about how it changes us.

This book doesn't try to predict the future of AI and cybersecurity. It tries to help you navigate it—by asking better questions, noticing new patterns, and connecting what's technical with what's human.

The only thing I ask is that you stay curious.

How to Use This Book

The AI Cybersecurity Handbook is both a reference and a reflection—a guide to understanding how artificial intelligence is transforming cybersecurity. You can read it start to finish or skip directly to the sections most relevant to your work.

- If you're a security practitioner, you'll find practical insights into AI-driven threats and defenses: data poisoning, model tampering, adversarial detection, and more.
- If you're a leader or policymaker, focus on the chapters about governance, risk management, and AI assurance frameworks.
- If you're a researcher, a student, or simply curious, the case studies, definitions, and frameworks will help you connect theory to practice.

You can use this book as a learning tool, a leadership guide, or a reference for building your own AI and security strategy. Each section is designed to stand on its own, but together they tell a story—of how technology evolves, how humans adapt, and how we can create systems that are not only smarter, but stronger.

However you read it, I hope it helps you connect what's technical with what's human—and navigate this next era of cybersecurity with clarity, curiosity, and care.

Part I

Cyber's New Player: AI on Both Teams

Chapter 1
The Age of Adaptive Adversaries: AI's Impact on the Security Balance of Power

A New Tempo of Conflict: AI on Both Sides of the Cyber Battlefield

Each wave of technology has drawn us closer—to each other and to our creations. The telephone carried our voices across distance. The Internet carried our words across time. The cloud carried our data across systems. Artificial intelligence now carries our judgment—across the boundary between human and machine.

Each shift has redrawn the boundaries of risk and trust. Each has forced defenders to rethink what it means to secure something that's constantly changing shape. But none have changed the tempo of conflict quite like AI. What once unfolded over days or months now happens in milliseconds. Decisions that once required analysts and operators are now made by algorithms operating at speeds and scales that challenge traditional oversight.

Cybersecurity has always been a strategic game of moves and countermoves. Attackers probe, defenders patch. Detection leads to evasion; prevention leads to adaptation. The difference today is speed—and scale. We're no longer watching two human players trade moves across a static board. The board itself is expanding, and automation is increasingly shaping both sides of the conflict.

AI acts as a force multiplier for everyone who touches it.

For attackers, it means automation and scale: reconnaissance bots that map targets in seconds, polymorphic code that can vary its structure to evade detection, and social engineering campaigns that can be personalized at scale.

For defenders, AI means foresight and orchestration. It filters noise, prioritizes what matters, and connects the dots across oceans of data. It helps security teams anticipate potential attack paths, identify patterns no human could see, and respond before damage spreads. Properly directed, AI gives defenders what they've always needed most—time.

Yet with every advantage comes complexity. The modern battlefield isn't confined to networks and endpoints anymore. It now includes data pipelines, training models, and decision systems—invisible layers of logic that make choices on our behalf. Attackers target the data and logic that shape model behavior rather than just the underlying infrastructure. They manipulate prompts to influence model outputs or bypass guardrails. They exploit the assumptions baked into algorithms rather than the code itself.

This is the new equilibrium of cybersecurity. Attackers innovate at scale; defenders operationalize trust at scale. AI is reshaping the boundaries between offense and defense, creation and control. The same techniques that harden a model can also be used to deceive one. The same automation that powers defense can amplify an attack.

And so we arrive at the central idea of this book: AI is both sword and shield. It magnifies human potential on both sides of the fight. True resilience doesn't come from choosing one side or the other—it comes from mastering both.