



AI GUIDE SERIES · BOOK FOUR

AI Safety and Scam Prevention

Patrick Thomas

AI GUIDE SERIES · BOOK FOUR

AI Safety and Scam Prevention

Protect Yourself and Your Family in the Age of AI Fraud

AI GUIDE SERIES · BOOK FOUR

AI Safety and Scam Prevention

*Protect Yourself and Your Family in the Age of AI
Fraud*

Patrick Thomas

AI GUIDE SERIES

COPYRIGHT NOTICE

Copyright © 2026 Patrick Thomas. All rights reserved.

No part of this publication may be reproduced, distributed, transmitted, stored in a retrieval system, or published in any form or by any means without prior written permission from the author, except for brief quotations used in reviews, commentary, or educational references where permitted by law.

First Edition, 2026

The information contained in this guide is provided for general educational and informational purposes only. The scam tactics and AI tools described in this book evolve rapidly; information reflects general guidance and may not reflect the current state of any specific threat or platform. For legal, financial, or other professional advice, consult a qualified professional.

The author shall not be liable for any loss, damage, or expense arising from the use of this guide or reliance upon any information contained herein.

For Carol — and everyone who turned a hard lesson into protection for the people they love.

AI Guide Series • Book Four · [trustedstranger.app](#)

Contents

INTRODUCTION	The New Wild West
PART ONE	Understanding AI in the Hands of Bad Actors
CHAPTER ONE	What AI Can (and Can't) Do
CHAPTER TWO	How Bad Actors Use AI
PART TWO	Spotting Scams Before They Find You
CHAPTER THREE	Phishing, Smishing, and Social Engineering
CHAPTER FOUR	Voice and Video: When You Can't Trust Your Ears or Eyes
CHAPTER FIVE	Fake Profiles, Fake Reviews, and AI- Generated Content
CHAPTER SIX	Financial Scams Powered by AI

Contents *(continued)*

PART THREE	Fighting Misinformation
CHAPTER SEVEN	What Is Misinformation (and Why Does It Spread)?
CHAPTER EIGHT	AI-Generated Fake News and Manipulated Media
CHAPTER NINE	Fact-Checking Like a Pro: The SIFT Method
PART FOUR	Protecting Yourself and Your Family
CHAPTER TEN	Locking Down Your Digital Life
CHAPTER ELEVEN	Talking to Your Family About AI Safety
CHAPTER TWELVE	What to Do When You've Been Targeted
PART FIVE	What Comes Next
CHAPTER THIRTEEN	The Future of AI Scams (and the Tools Fighting Back)
CHAPTER FOURTEEN	Your AI Safety Action Plan
REFERENCE	Glossary of Key Terms

Introduction

The New Wild West

The phone rang at 11:47 on a Tuesday night.

Carol, a 67-year-old retired nurse in Columbus, Ohio, grabbed it on the second ring. She didn't recognize the number, but she answered anyway — because the voice on the other end was her grandson Tyler's.

"Grandma." His voice was shaky. Scared. "I need your help. I'm in trouble and I can't call Mom and Dad right now."

Tyler — or the voice that sounded exactly like Tyler — explained that he'd been in a minor car accident in another city. He needed \$2,400 in gift cards for bail. He begged her not to call his parents.

Carol drove to the nearest Walmart at midnight. She bought the gift cards. She read the numbers over the phone to a stranger.

Tyler, of course, was asleep at home the entire time. The voice wasn't his. It was a near-perfect copy, built by an AI tool in minutes using clips pulled from Tyler's public social media videos. Carol lost \$2,400 that night — and more than the money, she lost the quiet confidence that she could trust her own senses.

That story is real. It happened last year. It's happening right now, in cities and suburbs and small towns across the country, to people who are smart, careful, and capable.

It's happening because the rules of the game just changed.

What's Changed — And Why It Matters Now

For most of human history, there were things you could trust. You could trust your own ears. You could trust that professional language meant a legitimate source. AI has quietly dismantled all of that.

The tools that allow a scammer to clone your grandson's voice cost almost nothing. The software that writes a convincing phishing email in flawless, personalized prose runs on the same platforms millions of people use every day. The technology that creates a fake video of a trusted public figure runs on a standard laptop.

This Book Is Not Here to Scare You

Fear without knowledge is useless. Knowing that AI-powered scams are getting more sophisticated doesn't protect you. Knowing how to recognize them does.

Our goal is simple: by the time you finish this book, you will know more about AI safety and scam prevention than roughly 95% of the people around you. You will be harder to fool. You will know what to do if something does slip through. And you'll be equipped to protect the people you care about — without becoming paranoid or distrustful of every digital interaction.

You're going to finish this book feeling smarter. More capable. More confident. Not more afraid.

How This Book Is Organized

Part One gives you a clear understanding of how AI actually works and how it's being misused.

Part Two walks through the most common and dangerous AI-powered scams in detail — phishing, voice cloning, fake profiles, financial fraud — with real examples, red flags, and immediate action steps.

Part Three tackles misinformation: fake news, manipulated images, AI-generated content designed to deceive. You'll learn the SIFT method for evaluating information before you share it.

Part Four covers locking down your personal digital security, having the AI safety conversation with your family, and what to do if you've already been targeted.

Part Five looks ahead at what's coming, what tools are being built to fight back, and how to stay informed without becoming overwhelmed.

Features in This Book

Each chapter includes practical features designed to make the informa-

PART ONE

Understanding AI in the Hands of Bad Actors

What the technology can actually do — and why that matters right now

Chapter One

What AI Can (and Can't) Do

Marcus was a middle school science teacher with a knack for explaining complicated things simply. When students asked what AI actually was, he'd pause, then say: "You know that really smart kid in class who reads every book in the library but still sometimes doesn't understand the joke?"

The kids always laughed. And Marcus would nod. "That's AI. Incredible knowledge. Sometimes no common sense."

At its core, AI — specifically the kind powering most of what we discuss in this book — is a system trained on enormous amounts of data to recognize patterns and generate useful outputs. Think of it this way: imagine you read every email ever sent, every book ever published, every website ever created. You'd develop an incredibly detailed sense of how language works — which words follow which other words, what a professional email sounds like, what a concerned grandchild says when they're in trouble.

That's essentially what large language models have done. They don't "think" the way you do. But they've become extraordinarily good at producing text, images, voice, and video that looks and sounds like it came from a human. That's the capability that matters most for our purposes.

REAL TALK

Think about the last time you got a spam email obviously full of typos. Now imagine that same email rewritten by a system that has studied millions of legitimate bank communications. Same message, different execution. Suddenly it reads like something your actual bank would send. That's not a hypothetical. That's Tuesday.

What AI Does Well — and Where It Struggles

What AI does very well: Writing convincing text in any style, generating realistic images of people and places that don't exist, cloning a human voice from a short audio sample, creating video of people saying things¹³ they never said, personalizing messages using real data about a specific person

The Intern Analogy

Here's another way to think about AI that may be even more useful: imagine AI as the most well-read intern you've ever met.

This intern has read everything. Their breadth of knowledge is genuinely astonishing. Ask them to draft a professional email and they'll produce something flawless. Ask them to summarize a 50-page document and they'll do it in two minutes. Ask them to write a phishing email pretending to be your bank — and they'll do that too, without hesitation, because they have no judgment, only capability.

This intern also has a serious problem with confidence. Ask a question they can't answer and instead of saying "I don't know," they'll often just make something up — confidently, with perfect grammar and a calm tone. This is called *hallucination*, and it's one of the most important AI behaviors to understand.

Finally, this intern has no loyalty. They'll work for you, or for someone trying to defraud you. They don't care either way.

MYTH VS. REALITY

Myth: AI is so advanced it can do anything — there's no way to tell if something is AI-generated.

Reality: AI is powerful, but it has real limitations and tells. There are specific patterns, inconsistencies, and behavioral quirks that trained observers can spot — and this book will teach you many of them.

AI Being Wrong vs. AI Being Weaponized

Sometimes AI causes harm by accident — wrong medical advice, a cited source that doesn't exist. This kind of harm is a flaw in the tool, not malicious.

¹⁴ Then there's AI being deliberately pointed at a target by someone who wants to cause harm: a scammer using voice cloning to impersonate your family member; a fraud ring generating thousands of personalized

Chapter Two

How Bad Actors Use AI

To understand why AI-powered fraud has exploded, it helps to look at exactly what changed — not just in capability, but in cost and accessibility.

Five Tools That Changed the Game

1. AI writing tools killed the grammar red flag. Scammers can now generate polished, personalized emails, texts, and messages in any style. The old reliable tell — broken English and awkward phrasing — is gone.

2. Voice cloning exploits the most trusted channel: a familiar voice. AI tools can now generate a convincing voice clone from as little as three to ten seconds of audio — a voicemail, a social media clip, a podcast appearance.

3. AI-generated images create fake people who have never existed. Photorealistic portraits — ideal for fake dating profiles, fake "testimonial" photos, and fake identity documents — cannot be reverse-image-searched back to a real person.

4. Deepfake video puts real people in fabricated scenes. A trusted public figure endorses a fraud. A CEO announces a fake merger. A family member appears to make an urgent request.

5. AI chatbots run 24/7 fraudulent "customer service" — fake bank support lines, fake tech support chats, fake romantic relationships — without a human on the other end.

The Economics of AI Fraud

A traditional phone scam operation might need twenty people to run 200 calls a day. An AI-assisted operation can run 2,000 personalized attempts simultaneously with one person managing it. The cost-per-attempt has dropped by orders of magnitude.

Think of it like spam email: even a 0.1% response rate on a million attempts is profitable. AI has brought that economy-of-scale logic to far more sophisticated attacks. This isn't a trend that's going away. It's a new baseline.

Who Gets Targeted — And It's Not Who You Think

There's a common assumption that AI scams primarily target older adults. And older adults are disproportionately targeted for certain scams, particularly family emergency fraud and romance scams.

But AI-powered fraud targets everyone. Young adults are heavily targeted by crypto investment scams and fake job opportunities. Professionals are targeted by executive impersonation fraud. Parents are targeted through fake "your child is in trouble" calls. Tech-savvy users are targeted by convincing fake security alerts.

The common factor is not age or education level. It's catching someone at a moment of urgency, distraction, or emotional vulnerability. Stress overrides skepticism. AI scammers know this — and knowing that helps you reset your guard in exactly those high-pressure moments.

MYTH VS. REALITY

Myth: Scammers only target older, less tech-savvy people.

Reality: AI-powered scams target people of all ages and technical backgrounds. The targeting is based on emotional vulnerability in a moment of urgency, not age or sophistication. Young adults lose significant amounts to crypto and investment scams every year.

How Data Breaches Fuel AI Scams

AI-powered scams become dramatically more convincing when the scammer has real information about you. Billions of records have been exposed in data breaches over the past decade — names, addresses, phone numbers, email addresses, account numbers. That data is bought and sold on criminal marketplaces.

When a scammer has your name, phone number, employer, and last four digits of your credit card, they can build a phishing message that feels personal and legitimate. When they have your voice from a public

PART TWO

Spotting Scams Before They Find You

How the most dangerous AI-powered fraud works — and the specific tells that give it away

Chapter Three

Phishing, Smishing, and AI-Powered Social Engineering

The email looked perfect.

It had the Chase Bank logo in the upper left corner — the right shade of blue, the right font. The sender's name said "Chase Customer Security Team." The subject line: "Action Required: Unusual Sign-In Detected on Your Account." The body was professional, calm, and urgent — explaining that a sign-in attempt had been detected from an unrecognized device two states away, and that Maria needed to verify her identity within 24 hours or her account would be temporarily locked.

Maria worked in marketing. She'd deleted dozens of obvious phishing emails. But this one felt real. The grammar was flawless. The logo looked right. She'd actually been traveling the previous week.

She clicked the button. She entered her username and password on a page that looked exactly like Chase's real website. By the next morning, \$4,200 had been transferred out.

What Is Phishing?

Phishing is any attempt to trick you into giving up sensitive information — passwords, account numbers, credit card details — by pretending to be someone you trust.

Email phishing remains the most common — fraudulent emails impersonating banks, government agencies, tech companies, or retailers.

Spear phishing is a targeted version, customized with real information about a specific individual.

Smishing is phishing via text message: "Your package could not be delivered. Click here to reschedule."

Vishing is voice phishing — phone calls from fake customer service agents, government officials, or tech support claiming you need to take immediate action.

AI has turbocharged all four. Every element of a phishing attack — the personalized details, the professional tone, the fake deadline — can now be generated in seconds.

Red Flags That Survive the AI Era

The old red flags — bad grammar, weird formatting — are no longer reliable. Here are the ones that still work:

WATCH OUT FOR

The sender's domain doesn't match the company. "Chase Customer Service" sent from "chase-alerts.net" is not Chase. Always read the full email address, not just the display name.

Artificial urgency. "Act within 24 hours or your account will be locked." Real companies rarely threaten immediate closure via unsolicited email.

Asked to click a link to verify sensitive information.

Legitimate companies virtually never ask for your password or full account number via an email link.

The link URL doesn't match. Hover over any link (without clicking) to preview the actual destination. If it doesn't go to the official company website, don't click.

The message arrived unsolicited. You didn't initiate a transaction, yet you're being asked to "confirm" something.

Payment in unusual forms. Gift cards, wire transfers, cryptocurrency, or Zelle to a new contact are all major red flags.

Asked to keep it confidential. Any message that says "don't discuss this with others" is designed to prevent you from getting a reality check.

Social Engineering: The Six Psychological Triggers

Phishing works because it exploits human psychology. Understanding the levers scammers pull helps you recognize when you're being manipulated.

24

Authority — the message appears from someone you're conditioned to comply with (your bank, the IRS, your boss)

Chapter Four

Voice and Video: When You Can't Trust Your Ears or Eyes

For almost all of human history, hearing someone's voice was proof enough.

If the voice on the phone sounded like your daughter — the way she pauses when she's nervous, the slight upturn at the end of her sentences — then it was your daughter. That certainty is gone.

How Voice Cloning Works

Voice cloning uses AI to capture the acoustic qualities of an individual's voice, then generate new speech that sounds like that person saying anything you type. The amount of source audio required has dropped dramatically — today's tools can generate a convincing clone from as little as three to ten seconds of audio. A voicemail. A video posted to social media. A clip from a podcast.

The clone isn't perfect — it can struggle with unusual emotional range and sometimes has a slight synthetic quality in longer samples. But for a short, urgent phone call from a "scared teenager"? It's convincing enough to work.

The Grandparent Scam, Supercharged

The grandparent scam is one of the oldest telephone frauds in existence. In its classic form, a caller impersonated a grandchild in trouble — in jail, in a hospital — and asked for money wired immediately.

With voice cloning, the scam has become dramatically more convincing. The caller doesn't just sound like a scared young person — they sound like *your specific grandchild*. After the initial call, a second person often gets on the line playing a "lawyer" or "bail bondsman" — providing instructions for buying gift cards and urging secrecy from the parents.

REAL TALK

Robert, 71, received a call that sounded unmistakably like his grandson Derek — crying, barely coherent, asking for \$3,500 after an accident. "Don't tell Mom and Dad," Derek begged. Robert started to get in his car. Then he stopped. A few months earlier, his

Setting Up a Family Safe Word

This is the single most important action step in this chapter.

A family safe word is a pre-arranged word or phrase that any family member can be asked to provide in order to verify their identity during an unexpected or urgent call. It should be:

Random and memorable — not your pet's name or last name. A random noun works well: "pineapple," "lighthouse," "accordion."

Known only to the family — never shared with others, never posted online.

Used consistently — every family member should know that anyone can ask for the word at any time, and that asking is not an insult.

The rule: if someone calls claiming to be a family member and the situation is urgent or unusual, ask for the code word before doing anything else. A real family member will know it. A voice clone won't.

TRY THIS NOW

This week, have the code word conversation with your immediate family. It takes five minutes.

1. Pick a random word together. Write it down somewhere safe.
2. Tell every member: "If anyone calls me claiming to be you in an emergency, I'm going to ask for this word. Please ask for it too if I ever call you in a panic."
3. That's it. You're now protected against the most convincing voice cloning attacks.

Deepfakes: When Your Eyes Can't Be Trusted Either

A deepfake is an AI-manipulated video in which a real person appears to say or do something they never actually said or did. Deepfakes are used in

Chapter Five

Fake Profiles, Fake Reviews, and AI-Generated Content

She found him on a Facebook group for hiking enthusiasts.

His name was David. He was 58, recently widowed, retired from civil engineering. His profile photo showed a distinguished-looking man with kind eyes and a weathered trail hat. He posted thoughtful comments. He remembered details from her posts. They messaged for three weeks, then started talking on the phone.

There was no David. The photo was generated by an AI image tool — a face that had never existed. The conversations were guided by AI tools that keep romance scammers on-script. By the time Linda discovered the truth, she had sent \$22,000 in three separate "emergency loans" to a man she had never met in person.

Spotting AI-Generated Profile Photos in 10 Seconds

AI-generated faces are ideal for scammers because they cannot be reverse-image-searched to a real person. Traditional fake profiles used stolen photos — a reverse image search could reveal the source. AI-generated faces have no source. They exist nowhere else.

The tell, when there is one, is in the details. AI-generated faces have characteristic imperfections: earrings that don't quite match, hair that merges oddly with the background, eyes that are unnaturally symmetrical, or teeth that are too perfect in a slightly uniform way.

Look at the background. AI images often blur or warp unnaturally near the subject's hair.

Check the ears and earrings. Asymmetrical or oddly shaped jewelry is a common tell.

Look at the teeth. Too-uniform, too-perfect teeth with an unnatural sheen.

Check the eyes. Unusual symmetry, slightly glowing irises, or an uncanny "too clear" quality.

Run a reverse image search. Right-click the photo and choose "Search image with Google." If it doesn't appear elsewhere, combined

AI-Generated Reviews: The Trust Problem

AI-generated fake reviews are now a significant problem on every major platform. A product with 4.7 stars and 847 reviews might have hundreds of those reviews written by AI, submitted through fake accounts.

WATCH OUT FOR

Signs a review may be AI-generated: overly generic praise ("Great product! Highly recommend!") with no specific details; unusual specificity without personal context; perfect grammar with a slightly formal tone that doesn't match how people actually talk; very similar phrasing across multiple reviews; reviewer profile that is new or has reviewed dozens of very different products; reviews all posted within a short window of time; suspicious ratio — thousands of reviews for a brand-new product.

Romance Fraud in the AI Era

Romance scams have been around as long as the internet. AI has made them faster, cheaper, and more scalable.

The pattern is consistent: a fake profile (often with an AI-generated photo) makes contact through a dating site, social media, or online community. The scammer builds rapport over days or weeks, often using AI tools to help craft messages. Eventually, a crisis emerges — medical emergency, business trouble — and the victim is asked for money.

Key warning signs: the relationship moves very fast emotionally; the person is always too busy or too far away to meet; they never do a live, spontaneous video call; their story has inconsistencies; and eventually, there's a request for money.

TRY THIS NOW

³² If you or someone you know is developing an online relationship with someone you haven't met in person, request a live video call where you ask them to do something spontaneous — wave, hold up

Chapter Six

Financial Scams Powered by AI

James was 47, financially sophisticated, and had invested in the stock market for twenty years. He knew about scams. He'd even given his elderly mother advice about avoiding them.

Then a contact he'd met through a professional networking app introduced him, gradually and convincingly, to a cryptocurrency investment platform with remarkable returns — 12%, 14%, then 18% monthly. The platform's interface was sleek and professional. His "returns" appeared in real time. He could even make small withdrawals at first, which convinced him the gains were real.

He invested \$140,000 before the platform disappeared overnight. Along with all his money.

This is called "pig butchering" — an investment fraud technique where scammers "fatten the pig" (build trust and encourage ever-larger investments) before the "slaughter" (disappearing with everything). AI has made every step of this fraud faster, cheaper, and more convincing.

The Most Common AI-Powered Financial Scams

Crypto investment fraud ("pig butchering") builds fake relationships that lead to fake investment platforms with fabricated returns. AI generates the platform interface, the "trading data," and even maintains the fake relationship through scripted conversation tools.

Fake celebrity investment endorsements use deepfake video to show a trusted financial figure, celebrity, or public personality endorsing a fraudulent scheme. These spread through social media ads and targeted video content.

Business email compromise (BEC) uses AI-generated emails that impersonate a vendor, partner, or executive to redirect a payment to a scammer's account. Often no one realizes until a legitimate invoice goes unpaid.

IRS and government impersonation combines AI-generated voicemails, emails, and even official-looking documents to create the im-

The Gut Check

Before any financial decision involving an unsolicited opportunity or urgent request, run through these questions:

WATCH OUT FOR

The Financial Gut Check — Pause and Ask:

Did I seek this out, or did it come to me unsolicited?

Is there time pressure — am I being told I must act now?

Is the return unusually high? (Anything reliably above 8–10% annually is a significant red flag.)

Is payment being requested in an unusual form — gift cards, cryptocurrency, wire transfer to a new account?

Can I independently verify this person, company, or platform through sources I already trust?

Am I being asked to keep this confidential or move quickly before consulting family or a financial advisor?

Any "yes" answer to questions 2–6 is a reason to stop, verify, and wait before acting.

Verifying Financial Professionals and Platforms

Every legitimate financial advisor registered with FINRA or the SEC is searchable through free government databases. A legitimate investment opportunity can withstand scrutiny. One that can't is telling you something important.

FINRA BrokerCheck: finra.org/brokercheck — verifies registered brokers and firms

SEC Investment Adviser Search: adviserinfo.sec.gov — verifies registered investment advisors

36 **ScamAdviser:** scamadviser.com — checks websites for fraud indicators

BBB Scam Tracker: bbb.org/scamtracker — reports of fraud by

SKILL UP · PART TWO — SPOTTING SCAMS

Your 5 most important takeaways from Part Two:

Artificial urgency is always a red flag. Real institutions allow time for verification. Any pressure to act immediately is manufactured.

The family code word protects against voice cloning. Set it up this week — it costs nothing and may save everything.

AI-generated faces have tells — check earrings, teeth, eyes, and background consistency before trusting a profile.

Unsolicited investment opportunities should be verified through FINRA and the SEC before you consider them. No exceptions.

Verify before you act. Hang up and call back through a number you already have. Don't click — go directly to the official site or app.

PART THREE

Fighting Misinformation

*How fake news and manipulated media work — and the simple method
that cuts through both*

Chapter Seven

What Is Misinformation (and Why Does It Spread)?

In the spring of 2020, a video began circulating on social media claiming that 5G cell towers were connected to a new respiratory illness. Within days, dozens of cell towers had been vandalized in the UK. Engineers received death threats. The claim was false. The people who believed it weren't stupid — they were scared, uncertain, and living through a frightening moment. The misinformation found the crack in their defenses and pried it open.

Misinformation isn't new. What's new is the machinery. AI has given anyone the ability to produce believable misinformation at scale — and social media platforms have built the most efficient distribution networks ever created for emotional content.

Three Terms You Need to Know

Misinformation: False or inaccurate information spread without deliberate intent to deceive. The person sharing it thinks it's true.

Disinformation: False information deliberately created and spread to deceive or manipulate. The person spreading it knows it's false.

Propaganda: Information — true, false, or selectively presented — used to promote a particular agenda or viewpoint. It may contain facts, but they're selected and framed to influence rather than inform.

Why it matters: understanding the intent behind false information helps you think about who benefits from it spreading — which is often a useful question to ask.

Why Misinformation Spreads: The Science

Your brain is not designed to be a neutral fact-processor. It's designed to keep you alive, connected to your social group, and oriented toward things that matter to you emotionally. That makes you vulnerable to misinformation in specific, predictable ways.

Confirmation bias is the tendency to accept information that confirms what you already believe and scrutinize — or reject — information that challenges it. Misinformation is often designed to confirm existing

Tribalism and in-group identity: We trust information from our "tribe" and are skeptical of information from outgroups. When a piece of content reinforces your group's identity or paints an outgroup negatively, it gets shared at high speed regardless of whether it's true.

The illusory truth effect: The more times you've seen a piece of information, the more likely you are to believe it's true — even if you only saw it shared, not verified. Repetition creates the feeling of familiarity, and familiarity creates the feeling of truth. This is why misinformation campaigns repeat the same false claims over and over.

REAL TALK

A false claim spread on social media during a local election: a well-known community leader had been arrested for fraud, with a "leaked document" that looked official. By the time the truth emerged — no arrest, no document, a completely fabricated story — the false version had been shared thousands of times. The correction was shared a few hundred times.

This pattern holds almost universally: corrections travel much less far than the original misinformation. Being slow to share, and fast to verify, is far more valuable than the reverse.

How AI Changed the Misinformation Landscape

Producing convincing misinformation used to require effort. You needed to write well, find images, build a narrative, and distribute it. AI has removed most of that barrier. Now, a single person can:

- Generate a fully written fake news article in seconds, complete with a plausible publication name, author byline, and professional formatting
- Create a photorealistic image of an event that never happened
- 42 Produce a deepfake video of a real public figure saying something they never said
- Build an entire ecosystem of fake corroborating content — multiple

Chapter Eight

AI-Generated Fake News and Manipulated Media

The photograph appeared to show the aftermath of a massive explosion at a government building. The rubble was convincing. The smoke was convincing. The emergency vehicles in the background were convincing. It was shared hundreds of thousands of times before anyone looked closely enough to notice that the shadows fell in two different directions — and that one of the firefighters had six fingers on his left hand.

AI image generation has characteristic tells, if you know to look. But most people don't look. They see a shocking image, feel an emotional response, and share it before their analytical mind has a chance to engage.

How Fake Articles Are Built

A fake news operation now follows a predictable, cheap production process:

First, an emotionally resonant topic is chosen — political figures, health crises, disasters, scandals. Second, AI writing tools generate a full article — headline, byline, body copy, pull quotes — in under a minute, written in the style of legitimate journalism. Third, a fake publication is built with a plausible name ("National Daily Report"), a professional template, and a fake editorial team. Fourth, fake social media accounts share the article across platforms. Fifth, other fake sites "cover" the same story, creating the illusion of independent corroboration.

The whole operation can be run by one person and take a few hours. Ten years ago, it would have required a team and significant resources.

WATCH OUT FOR

Before accepting or sharing a news article, check:

Source: Is this a publication I've heard of with a verifiable history?

URL: Watch for subtle misspellings (ABCnews.com.co, not ABCnews.com).

Date: Old articles are frequently recirculated as if they're new.

Byline: Can you find the author anywhere else? Real journalists have a traceable history.

AI-Generated and Manipulated Images

There are two main categories: AI-generated images (created entirely from scratch by tools like DALL-E or Midjourney) and manipulated images (real photos that have been altered). Both are now accessible to anyone.

To investigate a suspicious image:

Step 1 — Look carefully at the details. Zoom in on hands, text, backgrounds, and lighting. AI images often have subtle errors in these areas. Manipulated images often show seams or inconsistent lighting at the edit point.

Step 2 — Run a reverse image search. Right-click and select "Search image with Google," or upload to images.google.com. This shows where the image has appeared before — revealing its original context.

Step 3 — Use a detection tool. FotoForensics (fotoforensics.com) uses error level analysis to highlight areas that may have been digitally altered.

Three Case Studies in AI Misinformation

The health hoax: AI-generated articles claimed an over-the-counter medication "had been proven" to treat a serious illness. The citations linked to real journals with slightly altered URLs. The "experts" were AI-generated identities. People bought out pharmacy shelves. *Lesson: Always verify medical claims through established health organizations.*

The political deepfake: A 90-second video appeared three weeks before a local election, appearing to show a candidate confessing to a crime. By the time video forensics experts confirmed it was a deepfake, the election was over. *Lesson: Deepfakes are most effective when they confirm existing suspicions — which is exactly when they deserve the most scrutiny.*

The fake celebrity statement: A fabricated post attributed to a well-known celebrity appeared to endorse a controversial political position. Shared millions of times before the celebrity issued a denial. *Lesson:*

Always check the official verified account of someone claiming to have made a statement.

Chapter Nine

Fact-Checking Like a Pro: The SIFT Method

There's a common mistake people make when they first learn how bad the misinformation problem really is: they go too far in the other direction.

They become the person who says "Well, how do we really know that?" about everything. They mistake cynicism for critical thinking. Those are not the same thing.

A cynic rejects everything and trusts nothing. A skilled fact-checker starts with genuine curiosity, follows a reliable process, and ends with a well-grounded conclusion — which is sometimes "this is true," sometimes "this is false," and sometimes "the evidence isn't clear yet."

This chapter gives you that process.

The SIFT Method

SIFT was developed by digital literacy educator Mike Caulfield. It stands for four steps that are fast enough to use in real life — not just when you have an hour and a laptop, but when you're scrolling on your phone and something catches your eye.

S — Stop. Before reading, sharing, or reacting: stop. Notice if you're feeling a strong emotional response. That reaction is the signal to slow down, not speed up. Ask: Do I know this source? Is this worth my time and attention?

I — Investigate the Source. Before reading the content, investigate who is behind it. A quick search on the publication or author takes 30 seconds. You're not looking for perfection — you're looking for a track record. A real outlet has a Wikipedia entry, media coverage, and mentions in reliable sources. A fake byline leads nowhere.

F — Find Better Coverage. Look for other coverage of the same claim from sources you already trust. If a story is real and significant, multiple independent reliable outlets will be reporting on it. This step also surfaces fact-checkers: Snopes, PolitiFact, FactCheck.org, and AFP. Fact Check often appear with their verdicts when you search a viral claim.

SIFT in Action: A Guided Walkthrough

The claim: A post circulates with the headline "Government Scientists Confirm: New Study Shows Common Breakfast Food Causes Memory Loss in Adults Over 50." It links to a site called "HealthTruthDaily.net." It has 47,000 shares. Your aunt sent it to you.

S — Stop. You feel mild alarm. You want to share it. Pause. This is a health claim. Health claims deserve careful scrutiny.

I — Investigate the Source. Search "HealthTruthDaily.net" in Google. Result: No Wikipedia entry. No mainstream media mentions. The "About" page says it's "dedicated to health truths the mainstream medical system ignores." Red flag. Not a credible publication.

F — Find Better Coverage. Search "breakfast food memory loss study over 50" in Google News. Result: No coverage from CDC, NIH, or major outlets. Snopes hasn't covered it. If a major legitimate study found this, medical reporters would be all over it. No coverage = almost certainly not a real study.

T — Trace the Claim. The article claims to cite a study "published in the Journal of Neurological Sciences." Search that journal and the topic. No such study exists. The citation is fabricated.

Verdict: False. Don't share. Consider letting your aunt know — privately, kindly.

Reliable Fact-Checking Resources

General: Snopes.com · FactCheck.org · PolitiFact.com · AFP Fact Check (afpfactcheck.org)

Images and video: images.google.com · FotoForensics.com · InVID / WeVerify browser extension

Health claims: CDC.gov · WHO.int · MedlinePlus (medlineplus.gov)

Financial claims: SEC.gov · FINRA · BrokerCheck

PART FOUR

Protecting Yourself and Your Family

Practical steps to lock down your digital life and build lasting protection

Chapter Ten

Locking Down Your Digital Life

There's a house down the street from where Sandra lives that has every security system imaginable — a Ring doorbell, motion-sensor floodlights, a monitored alarm. And the side gate is always unlocked.

Digital security works the same way. Most people have some protections in place. But somewhere, usually in the places that feel less important or more inconvenient, there's an unlocked gate. This chapter is a practical, action-oriented walkthrough of the changes that will make the biggest difference for the most people.

You don't need to become a cybersecurity expert. You need to close the gate.

Your Three Most Important Security Upgrades

1. Use a password manager. The single most impactful security decision most people can make is to stop reusing passwords and start using a password manager. When a company gets hacked, stolen credentials are tested across other websites automatically. If you use the same password for your email, your bank, and your Amazon account, one breach can expose all three.

A password manager creates and stores a unique, strong, random password for every account. You only need to remember one master password. Reputable options: **Bitwarden** (free and open-source), 1Password, Dashlane, or the built-in managers in Chrome and Safari.

2. Enable two-factor authentication (2FA). 2FA adds a second step to your login — usually a code sent to your phone or generated by an app — so that even if someone has your password, they still can't get in. Enable it on your highest-value accounts first: email, banking, social media, and any account connected to payment information.

Authenticator apps (Google Authenticator, Authy, Microsoft Authenticator) are more secure than SMS codes. But even SMS-based 2FA is vastly better than nothing.

3. Check your breach exposure. Go to haveibeenpwned.com and enter your email addresses. If your email appears in known data breaches.

Privacy Audit Checklist

Work through this checklist at your own pace:

Passwords & Access: Set up a password manager with unique passwords for all major accounts · Enable 2FA on email, banking, and social media · Check haveibeenpwned.com for breach exposure · Change any passwords that appeared in breaches

Social Media Privacy: Set profile to "friends only" (not public) · Remove or limit birth date, hometown, employer, phone number from public view · Audit what old posts are visible · Review which apps have access to your social accounts (Settings → Apps)

Phone & Device: Enable screen lock with PIN, fingerprint, or face recognition · Keep operating system and apps updated · Review app permissions — does that flashlight app need your contacts?

Data Removal (optional but powerful): Search your name on data broker sites (Whitepages, Spokeo, BeenVerified) · Submit opt-out requests · Consider a service like DeleteMe or Canary to automate ongoing removal

Social Media: The Scammer's Research Tool

Social media profiles are one of the primary sources scammers use to personalize attacks. A public profile showing your full name, employer, hometown, family members, and recent travel gives a scammer everything they need to craft a convincing targeted message.

WATCH OUT FOR

Oversharing patterns that give scammers valuable targeting information: posting that you're traveling · sharing that you've made a large purchase · publicly announcing family milestones · tagging elderly family members in photos · sharing health conditions publicly · complaining publicly about a company (scammers impersonate the company to "resolve" your complaint).

Chapter Eleven

Talking to Your Family About AI Safety

Diane had been reading about AI scams for weeks. She knew exactly what she needed to tell her mother — and she'd been putting it off for two months.

"I just don't want her to feel like I think she's stupid," Diane said. "She worked in finance for thirty years. If I come in with this 'let me teach you about scams' energy, she's going to shut down."

Diane was right. *How* you have the AI safety conversation matters as much as *whether* you have it. A clumsy or condescending approach can actually backfire — leaving people defensive, embarrassed, or so overwhelmed that they disengage.

The Core Principle: Lead with Empowerment, Not Fear

The goal is to equip people with knowledge, not frighten them into anxiety. Fear makes people feel helpless. Knowledge makes them feel capable.

The most effective approach: frame the conversation as sharing something interesting and new — not correcting a vulnerability. Lead with the story. Let them react. Invite them into the solution.

Talking to Older Parents

What doesn't work: "You need to be careful — there are scams that target people your age." → Implies they're already a likely victim. Puts them on the defensive.

What works: "I've been reading about this new type of scam that's genuinely hard to spot — even for really savvy people. It uses AI to clone voices. Can I tell you about it?" → Frames it as new information, not a character assessment. Invites, doesn't lecture.

Follow up with collaborative action: set up the family code word together, add the FTC scam alert number to their contacts (1-877-382-4357), do a quick social media privacy check together, and agree on a "check with family first" rule before sending money anywhere.

The Family AI Safety Meeting

The most effective single step any family can take is a short, structured conversation where everyone learns the same information at the same time — and where protective protocols get established together. It doesn't need to be formal. It can happen at dinner. It should take about 20 minutes.

Minutes 1–5: Share a Story. Tell Carol's story from the introduction, or another relevant example. Keep it concrete and specific. Don't editorialize — just tell the story.

Minutes 6–10: Three Things to Know. (1) AI can clone voices — a call that sounds like me might not be me. (2) AI can create fake photos and fake profiles — a person who looks real online might not exist. (3) Urgency is a red flag — real emergencies allow time to verify.

Minutes 11–15: Set Up the Code Word. Pick a word together. Write it down. Each person commits to asking for it in any unexpected emergency call from a family member.

Minutes 16–20: Establish the Rule. Agree: before sending money anywhere based on an unexpected request — phone, email, or message — we call each other first to verify. No exceptions. No offense taken.

REAL TALK

After reading about AI safety online, Tom printed out a short one-page summary and brought it to Thanksgiving dinner. He didn't give a lecture. He just told his mother's story — she'd almost fallen for a voice clone scam the previous year — and pulled out the summary. "I put together something for all of us. It's just three things to remember."

His 78-year-old father, who had bristled at every technology conversation for years, listened carefully. His teenage niece asked follow-up questions. His mother found herself contributing to the conversation rather than receiving it. They set up a family code

Chapter Twelve

What to Do When You've Been Targeted

First: it's not your fault.

That's not a platitude. It's an important truth. The people who designed these attacks are professionals. They have studied human psychology, tested their approaches on thousands of targets, and refined their methods to specifically overcome the defenses that intelligent, careful people put up. The tools they use — AI voice cloning, deepfakes, personalized phishing — were not available even five years ago.

When someone is injured in a car accident because another driver ran a red light, we don't ask why they didn't swerve faster. AI fraud is the same.

Scenario 1: You Clicked a Suspicious Link

If you clicked a link in a phishing email or text but didn't enter any information:

Immediately close the browser tab. Do not click anything else on the page.

Run a malware scan using reputable software (Malwarebytes is free and widely respected).

Change the password for any account you were accessing when you clicked.

Enable 2FA on that account if you haven't already.

Monitor your email for unusual activity over the next several days.

Scenario 2: You Entered Information on a Fake Site

If you entered a password, account number, or personal information:

Change the compromised password immediately — and everywhere you used the same password.

Contact the real company the site was impersonating. They can flag your account for monitoring and lock it if necessary.

Enable 2FA on the account.

If you entered financial information, call your bank immediately. Ask

Scenario 3: You Sent Money

This is the hardest scenario — and the one where people most need to hear that recovery is possible. The sooner you act, the better the odds.

Bank or wire transfer: Call your bank immediately and ask them to recall or reverse the transfer. Act within hours — transfers can sometimes be recalled. Even if they can't reverse it, reporting starts a paper trail.

Credit or debit card: Call your card issuer and report it as fraudulent. Credit card charges have strong federal protections — you're often covered. Debit card fraud has fewer protections but is still reportable and often recoverable.

Gift cards: Call the gift card issuer immediately (number on the back). Report the card was used in a scam. Some issuers can freeze unredemed balances. Act within hours for any chance of success.

Cryptocurrency: Transactions are generally irreversible. Report it anyway — it creates a record. Be especially wary of "recovery services" that promise to get your crypto back for a fee. Most are scams.

Zelle / Venmo / Cash App: Report the transaction in the app immediately and contact the bank or service. Recovery varies by platform — act fast.

Where to Report

Reporting a scam doesn't guarantee recovery — but it matters. Reports build the data sets that law enforcement uses to identify patterns, track networks, and pursue prosecutions. Your report may directly contribute to shutting down an operation targeting thousands of others.

FTC: reportfraud.ftc.gov — the primary U.S. agency for consumer fraud reports

FBI IC3: ic3.gov — for internet-based fraud, especially if significant money was lost

Your state Attorney General: search "[your state] attorney general

SKILL UP • PART FOUR — PROTECTING YOURSELF AND YOUR FAMILY

A password manager + 2FA on your most important accounts is the highest-impact security upgrade you can make today.

The Family AI Safety Meeting — 20 minutes, one story, one code word, one rule — is more protective than any app or security tool.

Check your social media privacy settings now. Public profiles are a scammer's research tool.

If targeted: act fast. Call your bank, change passwords, and place a fraud alert or credit freeze within hours.

Report fraud to the FTC and FBI IC3. It costs nothing and may protect thousands of other people from the same attack.

PART FIVE

What Comes Next

The future of AI fraud — and the tools, habits, and principles that will keep you ahead of it

Chapter Thirteen

The Future of AI Scams (and the Tools Fighting Back)

In 2019, the idea that a scammer could clone your child's voice from a TikTok video and use it to extort you over the phone would have sounded like science fiction. In 2024, it was a documented pattern reported by the FBI.

The pace of change in AI is unlike anything in the history of consumer technology. What's experimental today becomes mainstream in eighteen months. But here's the other side: the same rapid development producing new threats is also producing sophisticated defenses. AI is not only a weapon — it's also a shield.

What's Coming: The Next Wave of AI Fraud

AI Agents and Automated Fraud. Current AI scams still require some human involvement. The next wave involves AI agents: autonomous systems that can conduct entire fraud operations from initial contact to final extraction with minimal human oversight — identifying targets through public data, initiating personalized contact, managing multi-day conversations, and guiding victims through a financial transfer, all automatically and simultaneously with thousands of targets.

Synthetic Identity Fraud. AI accelerates the creation of fictional identities — combining a real Social Security number (often from a child or someone with little credit history) with fake name, address, and birth date, plus AI-generated faces and documents to pass verification checks.

Hyper-Personalized Attacks. As AI tools become better at processing large amounts of personal data, attacks will incorporate your communication style, family relationships, recent life events, and documented vulnerabilities — all derived from publicly available data and data broker profiles. The defense is the same: limit your public data footprint and cultivate the pause-and-verify habit.

Real-Time Deepfake Video. Current deepfake technology is primarily pre-recorded. Real-time deepfake video — applying a face swap during a live call — is improving rapidly. Within a few years, it may be possible for a scammer to conduct a live video call appearing to be someone you know. The family code word system becomes even more critical

MYTH VS. REALITY

Myth: There's nothing regular people can do — AI fraud is just too advanced.

Reality: The fundamentals of fraud detection — verify identity through established channels, pause before acting on urgency, don't send money based on unsolicited contact — are resistant to technological advancement. Scammers need you to act before you think. Your best defense is the pause, and that doesn't require any technology.

The Shield: How AI Is Fighting Back

For every concerning development in AI-powered fraud, there's a corresponding defensive application in development.

AI-powered fraud detection at banks analyzes transactions in real time, flagging unusual patterns faster and more accurately than traditional rule-based systems. Banks are also beginning to deploy AI-powered voice analysis that can detect the acoustic signatures of cloned voices during phone calls.

Platform-level detection: Social media companies are deploying AI tools to detect deepfakes, identify coordinated fake account networks, and flag AI-generated content at scale. Some platforms are beginning to require disclosure labels on AI-generated content.

Consumer-facing detection tools: A growing ecosystem of free and low-cost tools helps regular people evaluate potentially fraudulent content — AI voice detection tools, AI image detectors, and browser extensions that analyze content in real time.

AI SAFETY HERO**The Global Anti-Scam Alliance (GASA)**

70 GASA is an international nonprofit that connects anti-fraud organizations, researchers, law enforcement, and industry partners to combat fraud at scale. They operate ScamAdviser (scamadviser.com), a free tool that provides real-time fraud risk assessments and links to resources for consumers.

Chapter Fourteen

Your AI Safety Action Plan

You made it.

And here's what that means: you now know more about AI safety and scam prevention than roughly 95% of the people around you. You understand how AI-powered fraud works, why it's effective, and — specifically — how to recognize it, respond to it, and protect the people you care about.

Carol, the retired nurse from our introduction, didn't get her \$2,400 back. But after she shared her story with her family, her daughter's neighbor recognized the same scam when it came for her — and hung up. One person's painful experience, shared with honesty and without shame, created a ripple of protection that touched people Carol had never met.

That's what this knowledge does when you share it.

10 Things to Do This Week

Do these 10 things this week. Each takes less than 15 minutes.

Day 1 — Secure Your Accounts:

- ☐ 1. Set up a password manager (Bitwarden is free). Start with your email and bank.
- ☐ 2. Enable two-factor authentication on your email account.

Day 2 — Check Your Exposure:

- ☐ 3. Go to haveibeenpwned.com and check your email addresses.
- ☐ 4. Change any passwords for accounts that appeared in known breaches.

Day 3 — Lock Down Social Media:

- ☐ 5. Review your primary social media profile privacy settings. Set posts to "Friends Only." Remove phone number and birth date from public view.

Day 4 — Protect Your Family:

- ☐ 6. Establish a family code word. Write it down. Share it with immediate family.
- ☐ 7. Schedule (or have) the Family AI Safety Meeting.

Quick Reference Guide

Red Flags — Pause and Verify When You See:

Artificial urgency ("act within 24 hours")

Requests for payment via gift cards, wire transfer, or cryptocurrency

A caller who sounds like a family member but can't provide the code word

Investment returns that sound unusually high (above ~8–10% annually)

Any request to keep something secret from family

Links in unexpected emails or texts — go directly to the official site instead

Profiles or photos that show AI-generation tells

Breaking news from a source you've never heard of

Trusted Resources:

Fact-checking: Snopes.com · FactCheck.org · PolitiFact.com

Image verification: images.google.com · FotoForensics.com

Breach check: haveibeenpwned.com

Broker verification: finra.org/brokercheck · adviserinfo.sec.gov

Scam alerts: consumer.ftc.gov · aarp.org/fraudwatch

If You've Been Targeted:

Call your bank immediately — ask to recall transfers if possible

Change compromised passwords and enable 2FA

Place a credit freeze: Equifax, Experian, TransUnion (all three)

Report to FTC: reportfraud.ftc.gov

For internet crimes: ic3.gov

For identity theft recovery: IdentityTheft.gov

For support: AARP Fraud Watch Helpline 1-877-908-3360

The Principles That Don't Go Out of Date

The specific tactics of AI fraud will keep changing. New tools will emerge. New scams will be invented. The landscape a year from now will look different from today's.

But these five principles are as durable as human psychology — because they're built on how fraud actually works, not on the specific tools used to execute it:

Urgency is a weapon. Real institutions allow time for verification. Any pressure to act immediately, without time to think or consult, is manufactured.

Verify through channels you control. Don't call back the number that called you. Don't click the link in the email. Go directly to the official website, official app, or a phone number you already have on file.

Familiar doesn't mean safe. A voice that sounds like your child, a logo that looks like your bank, a profile photo that looks like a real person — none of these are proof of identity. Verify.

If it sounds too good to be true, it virtually always is. Not sometimes. Virtually always.

The pause is your superpower. Fraud requires you to act before you think. Every second you take to slow down, check, and verify is a second working in your favor.

A Final Word

We started this book with Carol — a retired nurse who drove to Walmart at midnight because she heard her grandson's voice telling her he needed help. She was not foolish. She was not careless. She was a loving person who encountered a sophisticated attack at a vulnerable moment.

And then she did something remarkable. She told her story. She taught her family. She turned her experience into armor — not just for herself, but for everyone around her.

Reference

Glossary of Key Terms

AI (Artificial Intelligence) — Computer systems trained on large amounts of data to recognize patterns and generate useful outputs, including text, images, voice, and video.

Chatbot — An AI-powered program that conducts text or voice conversations, used in customer service — and increasingly in fraud.

Confirmation Bias — The tendency to accept information that confirms existing beliefs and to scrutinize information that challenges them.

Credit Freeze — A restriction placed with credit bureaus that prevents new credit accounts from being opened in your name without your authorization. Free and highly effective against identity theft.

Data Breach — An incident in which unauthorized parties gain access to sensitive data — often resulting in personal information being sold on criminal markets.

Data Broker — A company that collects personal information from public and private sources and sells it to businesses and, unfortunately, anyone willing to pay — including scammers.

Deepfake — AI-generated or AI-manipulated video in which a real person appears to say or do something they never actually did.

Disinformation — False information deliberately created and spread to deceive or manipulate — unlike misinformation, the intent is knowing and malicious.

Fraud Alert — A notice placed with credit bureaus that asks creditors to verify your identity before opening new accounts. Free, lasts one year, and less restrictive than a credit freeze.

Hallucination (AI) — The tendency of AI systems to generate false information confidently — producing plausible-sounding but factually incorrect output.

Illusory Truth Effect — The psychological phenomenon by which repeated exposure to a claim increases belief in its truth, regardless of

Also in the AI Guide Series

Book One

AI Won't Replace You

But Someone Using AI Might

Book Two

AI for Everyday Life

50 Practical Ways to Save Time, Solve Problems, and Think Better with AI

Book Three

AI at Work

The 5D Work System for Modern Professionals

Book Four

AI Safety and Scam Prevention

How to Protect Yourself in the Age of AI-Powered Fraud

Book Five

AI for Veterans

Better Understand Benefits, Healthcare, and Government Services

Special Booklet

The AI Gold Rush

Why So Many People Are Selling AI Wealth — and What They're Really Selling

Find the complete series at trustedstranger.app

About the Author

Patrick Thomas is a facilities management professional, veteran, and technology enthusiast who helps everyday people understand practical applications of artificial intelligence.

His goal is to make emerging technologies approachable, practical, and useful for ordinary people — without requiring a technical background. He writes from experience, not theory: the tools, prompts, and strategies in his books are the ones he uses every day.

Patrick served in the United States military and brings a veteran's appreciation for clear, actionable guidance to everything he writes.

AI Safety and Scam Prevention is part of the AI Guide Series — a collection of accessible guides designed to help readers of all backgrounds benefit from artificial intelligence in their daily lives.

Learn more and find the complete series at trustedstranger.app

AI GUIDE SERIES

Learn more. Do more.

The AI Guide Series is a growing collection of practical books that help everyday people — professionals, veterans, families, and retirees — use artificial intelligence to save time, solve problems, and navigate complex systems.

No technical background required. No jargon. Just clear, actionable guidance for the tools that are reshaping how we work, learn, and live.



trustedstranger.app

Complete series & updates