Staying Safe with AI

What to trust, what to check, and what to do differently

Dominic A. Wilson Ph.D.
Assistant Professor of Computer Science
University of Findlay

Way Public Library, Perrysburg, Ohio | September 15, 2026

Why I Am Here

Some background

1 What I do

I teach computer science at the University of Findlay. My research is about the security of AI language models; finding the ways they fail, and the ways they can be fixed.

2 Why this room

The failures I study in a lab are the ones landing in your voicemail, inboxes, and Facebook messages. A room like this is where that gets fixed.

3 Why now

AI tools have grown very useful and pervasive recently. These tools also produce deceptive voices, faces, and writing. Everyday habits for handling them have not caught up.

No technical background needed tonight, and nothing here costs money.

Where We Are Going Tonight

1. **What AI Is**
   Quick overview in plain English.

2. **When AI Is Turned Against Us**
   Voice clones, deepfakes, phishing, and the habit to stops most of it.

3. **Using AI Without Giving Yourself Away**
   Where typed words go, and what should never be entered into a chatbot.

4. **Using It Well, and Getting Help**
   Good uses, better questions, and who to call if something goes wrong.

01

What AI Is

Strip away the marketing and the science fiction, and what is left is a fast pattern prediction machine.

AI Has Been in Your Life for Years

It is not new. What is new is that it can write, talk, and make pictures.

1 Smartphone keyboard suggestions

Predicting the next word as you type.

2 Voice prompts

The system that hears "billing" and routes your phone call.

3 Maps and traffic

Arrival times predicted from millions of past trips.

4 Photos on a phone

Grouping photos with similar faces, finding objects, and improving lighting.

5 Streaming suggestions

Netflix and YouTube choosing what to show next.

6 Captions on TV

Live subtitles typed by a machine, not a person.

Here is what changed: AI can now produce things, not just sort, categorize, or predict information.

The New AI, in One Sentence

It predicts what comes next, again and again, whether it's words, sounds, or images.

What that means

- It has read an enormous amount of writing
- It learned which words tend to follow which
- It assembles what a good answer looks like

Two useful comparisons

1. **Autocomplete that went to college**
   - It finishes a sentence, just far better
2. **bright intern who has read everything**
   - One who will never admit to not knowing

AI Makes Things Up

The technical word is hallucination. Put simply, it is making things up.

It happens constantly

- Court filings citing cases that never existed
- Book and article citations that look real
- Confident wrong math and wrong dates
- Invented details about local people and places

Why it will not say it does not know

- It was rewarded for plausible-sounding answers
- A smooth guess and a solid fact look identical
- It has no internal check against the real world
- It will apologize, then invent a second answer

Fluent is not the same as true.

Three Things AI Cannot Do

1 It cannot verify

Some tools now search the web and cite sources, which helps. But nothing checks whether the summary matches the page. Asking 'are you sure?' produces more confident text, not more truth.

2 It cannot keep a secret

Typed text leaves your phone or computer. It can be stored, reviewed by staff, used to train future versions, or exposed in a breach.

3 It cannot care

The warmth in the reply is a writing style copied from us. It knows no one and holds no one's interests in mind.

02

When AI Is Turned Against Us

Criminals use the same AI tools we do. Here is what that looks like from the receiving end.

How Big AI Fraud Is

Federal Trade Commission figures for 2025, released June 2026

$15.9B

reported lost to fraud in the US in 2025, the highest on record

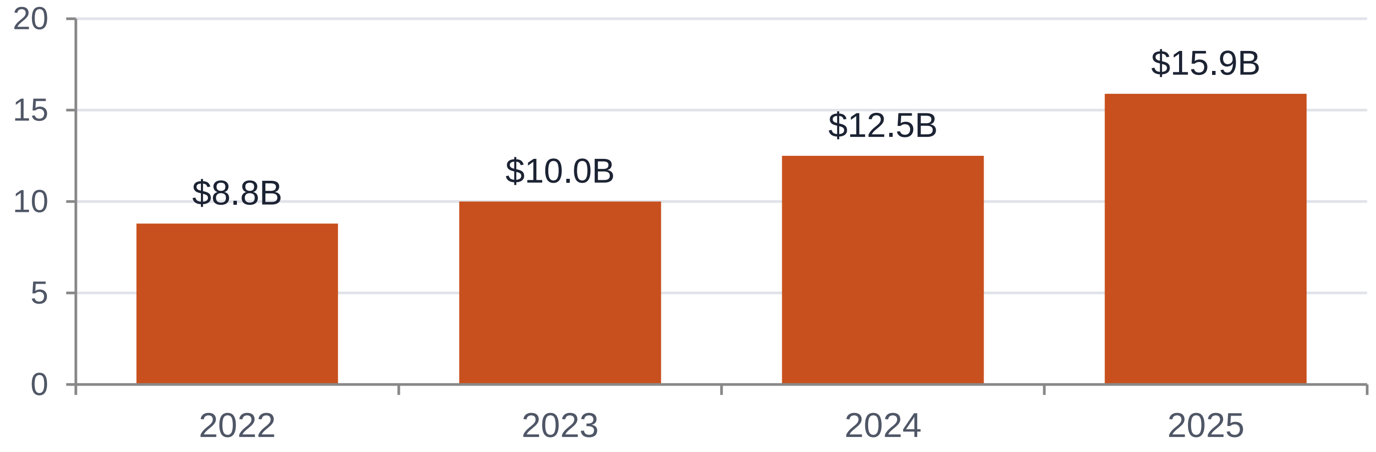
~ 5 x

growth in reported losses since 2020

$3.5B

lost to imposter scams, the most reported category

US reported fraud losses, billions of dollars



Worth knowing

- Typical reported loss: a few hundred dollars
- But very large losses are rising
- When older adults lose, they lose more
- Most victims never report at all

Data source: FTC Consumer Sentinel Network data for calendar year 2025

What Changed for the Criminals

The advice we were given ten years ago is dated

The old scam

- Broken English and obvious typos
- Generic wording sent to everyone
- One criminal, one victim at a time
- A stranger's voice on the phone
- A photo anyone could tell was doctored
- **Advice: look for spelling mistakes**

The AI-assisted scam

- Flawless grammar
- Personalized from public posts
- Thousands of conversations at once
- A known voice, cloned from a clip
- Realistic photos, video, and documents
- **Advice: judge the request, not the writing**

Scam 1: The Voice on the Phone

The family emergency call, now in a familiar voice

How someone would get a voice

- A few seconds of audio is all it takes
- A Facebook video, a livestream, a voicemail greeting
- The free tools do the cloning in minutes

How the call goes

- The reported ones come late at night, crying
- An accident or arrest, then a 'lawyer' takes the phone
- Do not tell Mom. There is no time.

What people are told to do

- Hang up, whatever they say
- Call back on the number already saved in the phone
- Call one other family member before money transfers
- Ask something the internet does not know
- If they want gift cards, wire, crypto or cash, it is fraud
- Lock down who can see personal videos online

Scam 2: The Face on the Video Call

Seeing is no longer believing, and it happens live

1 Fake video meetings

A case used in class: staff wired millions after a call where every colleague was generated.

2 The fake grandchild

A short call, poor lighting, a bad connection. Just enough to confirm the fear already there.

3 The fake investigator

A 'bank fraud officer' on camera, with a badge and a screen share. One could be built this afternoon.

Tells worth relying on

- They resist a callback and press to stay on this call
- They move the conversation to a new app: WhatsApp, Telegram, Signal
- The run into camera trouble the moment a specific question comes up

What to do

- End the call first
- Reconnect on a number already on hand
- Never on a link they sent

Scam 3: The Friend Who Is a Machine

The most expensive ones start with kindness, not a threat

The pattern

- A warm stranger from an app, a game, or a wrong-number text
- Daily messages, good listening, no pressure at first
- AI writes them, which is how one operation runs hundreds at once
- Always a reason they cannot meet or video chat
- Then a crisis, or an investment app with lovely returns
- Small withdrawals work. Big ones need a fee, then a tax

Checks worth making

- Run their photo through a reverse image search
- Be wary of new accounts with few real connections
- Money plus an online-only relationship is fraud, every time
- Never move savings into an app a new friend suggests
- Tell one trusted person about any new online friend
- If someone close is caught in one, stay kind. Shame keeps them in

Scam 4: Messages That Look Perfect

Email, text, and voicemail that pass every test

1 The bank

An unusual sign-in or a frozen account, needing action now.

2 A package

A delivery held over a small fee, addressed by name.

3 Medicare or IRS

A benefit at risk, with a deadline on a weekend.

4 A utility

A shutoff this afternoon unless payment comes immediately.

5 A subscription

A renewal charge no one made, and a number to call.

6 A boss or pastor

A short note asking for gift cards, quietly.

Never use the link or number in the message. Open the app, or type the address yourself.

Scam 5: Things That Do Not Exist

Whole storefronts, charities, and news events built in an afternoon

What is being manufactured

- Stores with invented products and glowing reviews
- Charities that appear hours after a disaster
- Job listings that only collect personal data
- Rentals for homes the poster has never seen
- News photographs of events that never happened
- Fake obituaries built to farm clicks and donations

The checks to run

- Search the business name plus the word scam
- Look for an address and a phone a human answers
- Pay by credit card, the strongest way to claw money back
- Give only to already-known charities
- If one site alone has the amazing deal, walk away
- For a photo, ask where it came from before sharing it

Hang Up. Look It Up. Call Back.

If one thing sticks tonight, make it this one.

1. Every scam here needs a decision within minutes.
2. No real bank or family member is harmed by a ten minute delay.
3. Look the number up independently: the card, the bill, the contacts list.
4. Anyone who argues against a callback has already shown who they are.

Set Up a Family Code Word

One short conversation, and it holds for years

1 Choose a phrase

Pick something no algorithm could scrape. Not the dog's name, not a birthday, not anything that has appeared on Facebook.

2 Share it out loud

In person, or on an outbound call. Everyone in the household, plus adult children and grandchildren. Never in a text or an email.

3 Agree the rules

Any emergency money request needs the phrase and a callback. The family never asks for gift cards or wire transfers. Nobody minds being verified.

Then practice it once, so the first use is not during a real emergency.

If the phrase slips the mind, hang up and execute the callback rule.

Six Red Flags, in Any Message, from Anyone

Judge the request, not the writing

1 Unexpected contact

It arrived out of nowhere and you didn't start it.

2 A rush

A deadline measured in minutes or hours.

3 Secrecy

Do not tell family, staff, or the bank.

4 An odd payment

Gift cards, wire, crypto, an app, or a cash courier.

5 A change of channel

Move to telegram, WhatsApp, or a link they provide.

6 Strong emotion

Fear, love, shame, or excitement, turned up fast.

Two or more together means stop and verify, before anything else happens.

If It Has Already Happened

Speed matters, and so does not carrying it alone

1 Stop, call the bank

Send nothing more, whatever the promises about recovering what is gone. Call the bank or card issuer now. Fast reports get reversed far more often.

2 Report it

reportfraud.ftc.gov
ic3.gov (FBI)
Ohio Attorney General : 800-282-0515
Perrysburg City Police : 419-872-8001

3 Close the doors

Change passwords, turn on two-step verification, and freeze credit at Equifax, Experian, and TransUnion. All three are free.

Beware of recovery scams. Anyone promising to get the money back for a fee is the second wave of the same crime.

03

Using AI Without Giving Yourself Away

The risk is not only what criminals do with AI. It is also what we hand over ourselves.

Where Typed Words Actually Go

Typing into a chatbot is closer to posting than to whispering

What happens to typed text

- It travels to a company's servers
- It is usually stored, often for a long time
- Staff may review conversations for safety
- It may train the next version of their AI model
- It can surface in a breach or a lawsuit

What can be controlled

- Check settings for history and training controls
- Turn off training on personal data where offered
- Delete old conversations no longer needed
- Paid and work versions have stricter terms
- Free tools inside social apps are the loosest

A useful test: would this be fine on a screen at church?

What to Never Type Into a Chatbot

1 Identifiers

Social Security, driver's license, or passport numbers.

2 Money

Account and card numbers, logins, security answers.

3 Health records

Test results, diagnoses, anything tied to a name.

4 Other people

A family's private business, a friend's diagnosis.

5 Work material

Client files, student records, contracts, personnel matters.

6 Passwords

No exceptions, not even to ask whether one is any good.

For help with a real document, retype your question with the names and numbers removed.

Health, Money, and Legal Questions

Use it to understand. Do not use it to decide.

Good uses

- Translate a diagnosis or a denial letter into plain English
- Build a list of questions for a doctor or attorney
- Explain a term on a bill or statement
- Summarize a long policy to know what to ask
- Practice a difficult conversation beforehand

Do not use it for

- Diagnosing a symptom or deciding whether to seek care
- Doses, interactions, or changing a prescription
- Whether to sell, buy, or move an investment
- Filing taxes or drafting a document to be signed
- Anything where being confidently wrong is costly

It is an excellent translator and a poor professional. Take its answers to a human, not instead of one.

Talking With Children and Grandchildren

The conversations worth having, without turning it into an inspection

1 Homework

A real line between help and substitution. Ask them to explain it back.

2 AI companions

Apps built to feel like a friend. Some teens spend hours a day there.

3 Fake images

Apps that generate explicit images of real classmates.

4 Not everything is real

Much of their feed is generated. Teach the habit of asking the source.

5 Scams reach them too

They are targeted through games, marketplaces, and job offers.

6 Keep the door open

The goal is that they speak up when something goes wrong.

04

Using It Well

None of this is a reason to avoid AI. Use it every day, with eyes open.

What It Is Genuinely Good At

Jobs where a fast, imperfect draft is worth a great deal

1 **Drafting**

A first version of a letter, a complaint, or a hard note.

2 **Shortening**

Turn a long document into something readable.

3 **Explaining**

Anything, at any level, and again more simply.

4 **Translating**

Between languages, and jargon to plain English.

5 **Organizing**

Packing lists, trip plans, agendas, project steps.

6 **Rehearsing**

Practice an interview or a hard conversation.

The pattern to go by: strong where the result can be checked, weak where it cannot.

How to Get Better Answers

Most disappointing results come from a one line question

1 Give it context

State who is asking, who it is for, and what is already known. 'I am helping my mother, who is 82, read a letter from Medicare' beats 'explain Medicare'.

2 State the goal

Length, format, and tone. A short letter, a bulleted list, six questions, plain language at an eighth grade level. It follows instructions closely.

3 Push back

Follow-up costs nothing. Too long, make it warmer, the deadline is wrong, try again. The second and third answers are usually much better.

Ask for sources, then open them. If a link does not exist, discard the whole answer.

The Two Source Rule

Anything touching health, money, safety, or a vote gets confirmed somewhere else before anyone acts on it.

1. This covers AI answers, social media, and forwarded news.
2. Two sources means two independent ones, not one copied twice.
3. For a photo, ask where it originally came from.
4. Sharing something false is how a scam finds its next person.

In the News This Week

Why AI leaders spent the weekend asking each other to slow down

What happened

- Anthropic's CEO called for slowing capability gains. OpenAI and xAI CEO agreed.
- The worry: AI now helps build the next AI
- It followed a researcher resigning in public with alarming warnings.

Where it is contested

- No pause, no deadline, nothing binding
- Critics ask whose lead a slowdown would lock in
- It concerns future systems, not today's apps

None of it changes tonight's advice. The scams are already here.

Five Things to Take Home

If nothing else, remember to slow down and verify.

1	**Fluent does not mean true**	Confidence is a writing style. Check what matters.
2	**Hang up and call back**	On a number looked up independently, never one they gave you.
3	**Pick a family code word**	This week. Ten minutes. Tell everyone, practice once.
4	**Keep secrets out of chatbots**	No ID numbers, no accounts, no passwords.
5	**Verify before acting**	Two independent sources for health, money, or safety.

Where to Get Help

Free, real, and staffed by people

Report a scam

- reportfraud.ftc.gov
- ic3.gov (FBI)
- Ohio Attorney General: 800-282-0515

Learn more

- consumer.ftc.gov, plain language and free
- aarp.org/fraud, with a scam tracking map
- staysafeonline.org, passwords and two-step login

Talk to someone

- AARP Fraud Watch Helpline: 877-908-3360
- The bank fraud line printed on the card
- Perrysburg City Police non-emergency line: 419-872-8001

Protect credit

- Freeze it at Equifax, Experian, TransUnion
- Free, and it can be lifted any time
- Do this before anything happens, not after

Questions

With thanks to Way Public Library and the League of Women Voters of the Perrysburg Area, Mindful Matters series

Dominic A. Wilson Ph.D. , Assistant Professor of Computer Science, University of Findlay