KAIA-OR · WHITEPAPER · SOCIAL ENGINEERING

A second look

Defending families and organizations against scams, business email compromise and impersonation, with a second look before anyone replies, pays or opens, and the decision left with people.

DOCUMENT ID	WP-3
VERSION	1.0
DATE	3 November 2026
AUDIENCE	Families and individuals; small-business owners; finance, security-awareness and risk leaders; enterprise and government security teams

Contents

1. Executive summary 3

2. The problem and how lures work 4
   1. The scale of the problem 4
   2. How modern lures work, in the abstract 4

3. The second-look model 6
   1. Why one-shot verdicts fall short 6
   2. How a second look works 6
   3. Context matters 7

4. Families and organizations 8
   1. Families and older relatives 8
   2. Organizations: invoice fraud, executive impersonation and staff abroad 8

5. Protecting the assistant itself, and honest limits 10

6. A personal and organizational playbook 11

7. What this means for you 12
   1. Glossary 12

© Kaia Technologies, Inc.

Executive summary

Many losses to online crime do not begin with a technical break-in. They begin with a person being persuaded: to reply, to approve, to open, to move money. Social engineering works because it targets judgement rather than software, and because it arrives in the same channels people use for work and family. It reaches the relative who lives alone, the clerk who processes supplier invoices, the executive on a trip and the official posted abroad.

Many tools respond with a one-shot verdict on a single item, or with nothing at all until after the loss. KAIA-OR offers something different: **a second look before anyone replies, pays or opens**.

- **Paste a message, or ask Hey KAIA out loud,** and see or hear the warning signs KAIA-OR recognizes and a suggested next step.
- **Check a link while it stays unopened.**
- **See what looks wrong, and when the check can't tell.** The suggested next step points to confirming through a channel you already trust. The decision stays with you.
- **In your own language.** Message checks recognize warning signs in English, French, Turkish and Hebrew.
- **For each person.** Each adult in a family, and each member of an organization, has their own second look on their own device. Membership never grants access to another member's messages.

This paper sets out the scale of the problem, how modern lures work in the abstract, why one-shot verdicts fall short, how the second-look model works for families and organizations, how the assistant itself is defended, the honest limits of the approach, and a short playbook for people and organizations.

> ABOUT THIS PAPER
>
> This paper describes the design and the committed product scope of KAIA-OR. Capability availability depends on the edition and deployment. To avoid teaching or imitating real lures, it never reproduces scam text, real brands or example web addresses.

© Kaia Technologies, Inc.

1. The problem and how lures work

The scale of the problem

Public reporting in more than one country describes the same trend: losses are large, they are rising, and persuasion rather than technical intrusion is at their centre. The figures below are market context. They describe reported losses and incidents, not the effect of any product.

PUBLIC FIGURES ON FRAUD AND SOCIAL ENGINEERING

Topic	What was reported
Reported internet crime losses, United States	$20.877 billion in reported losses in 2025, up 26% on 2024.[1]
Most-reported complaint type, United States	Phishing and spoofing, with 191,561 complaints in 2025.[1]
Reported losses of people aged 60 and over, United States	$7.748 billion in losses in 2025, up 59% on 2024, across 201,266 complaints, which were up 37%.[1]
Reported business email compromise losses, United States	About $3.05 billion in reported losses in 2025.[1]
Reported consumer fraud losses, United States	Consumers filed 3 million fraud reports in 2025 and reported $15.9 billion in losses.[2]
Breaches and attacks reported by UK businesses	43% of UK businesses reported a breach or attack in the previous 12 months, and 85% of those businesses experienced phishing.[3]

Two patterns stand out. Older people carry a disproportionate share of reported losses, often from far away from the relatives who would have helped. And organizations are often reached first through a message that persuades someone to act.

How modern lures work, in the abstract

Lures change their wording constantly. Their structure changes much less. The patterns below are described in the abstract, without reproducing any real message.

- **Haste.** The message leaves no time to think, with a deadline or a consequence that seems to arrive at once.
- **Authority.** It appears to come from someone with power over you: a manager, a public authority, a financial institution, a service provider.
- **Familiarity.** It appears to come from someone you know: a relative in trouble, a colleague, a long-standing supplier.
- **Money moved somewhere new.** A request to send money to a new destination, or to change where a regular sum goes.
- **Requests for sign-in details.** A request to share sign-in details, or to approve a sign-in you did not start.

© Kaia Technologies, Inc.

- **A link or a file.** A page that imitates a real one, or an attachment that asks to be opened.
- **A change of channel.** A move to a new number, a private chat or a phone call, away from the channels where the request could be checked.
- **Synthetic voices and faces.** Generated audio or video that imitates someone the target knows.

Most lures combine two or three of these. Their common goal is to separate a person from the moment of judgement. A useful defence therefore does not depend on recognizing each message. It needs to give people a reliable way to step back into that moment, before they reply, pay or open.

© Kaia Technologies, Inc.

2. The second-look model

Why one-shot verdicts fall short

The usual response to a suspicious message is a tool that returns a single verdict on a single item. That approach has three weaknesses.

- **False comfort.** No checker recognizes all lures. A clean verdict on a new lure reads as permission, and it is exactly what a well-crafted lure is designed to obtain.
- **No context.** A verdict on one pasted item knows nothing about the person's trips, the events around them, or the language they think in.
- **No decision support.** A verdict says "bad" or "fine" but not what to do next, and the next step is where the outcome is decided.

How a second look works

KAIA-OR replaces the verdict with a second look. The person stays in charge; the product helps them see what they might have missed.

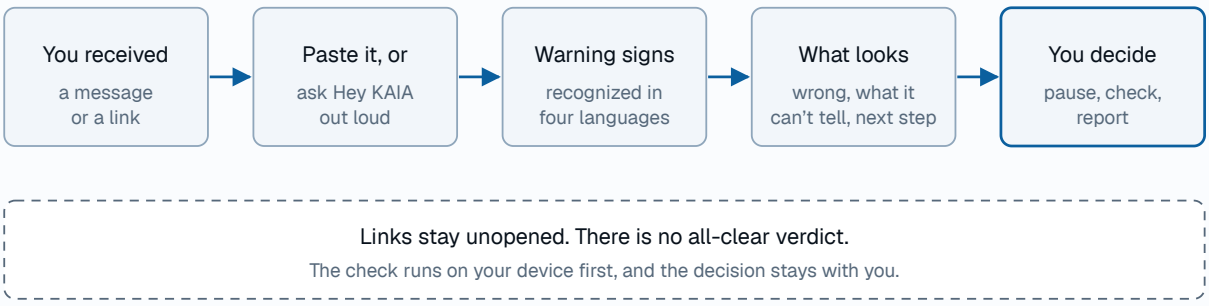


Figure 1. The second-look model. The figure is conceptual and does not describe an implementation.

- **Paste it, or say it.** A person pastes a message into KAIA-OR, or asks Hey KAIA out loud, which helps people who prefer to speak rather than type.
- **See the warning signs.** KAIA-OR shows the warning signs it recognizes, such as requests for sign-in details or demands for money, a change of channel, or a sender who does not match the claimed identity.
- **Links stay unopened.** A person can check a link without opening it.
- **See what looks wrong, and when the check can't tell.** Where the evidence is not enough, KAIA-OR says so plainly instead of guessing.
- **A suggested next step.** Typically, confirming through a channel you already trust: a number you already have, an address you know, a colleague you can ask in person. The decision stays with you.

Checks run on the device first. The messages and links a person checks stay on their device, and nothing is sent to a cloud AI service unless they agree first.

© Kaia Technologies, Inc.

Context matters

A lure is most convincing when it fits the moment. A message about a disrupted journey reads differently while you are travelling. A message about a local emergency reads differently when one is reported near you. Because the second look sits in the same product as each person's own trip and event picture, a person can compare what a message claims with what their event picture shows, including how well each event is corroborated. The context does not decide for them. It gives them one more fact to weigh before they act.

© Kaia Technologies, Inc.

3. Families and organizations

Families and older relatives

The person most at risk is often the one least likely to ask for help in time: an older parent who lives in another city, a relative who is embarrassed to say they are unsure, an adult who reads English as a second language. KAIA-OR is designed around their situation rather than the buyer's.

- **Each adult has their own check.** Each adult you invite, even when they live apart, gets their own KAIA-OR and their own second look, on their own device.
- **In their own language, by voice.** When a message impersonates a relative, a financial institution or a delivery service, each member can paste it in, or ask Hey KAIA out loud, and hear the warning signs and a suggested next step in English, French, Turkish or Hebrew.
- **No one reads anyone else's messages.** Membership never grants access to another member's messages, devices, location or saved places. Help comes from each person having their own second look, not from someone else looking in.

Families can strengthen this with a simple agreement made in advance: an unexpected request for money from a relative is checked by calling that relative back on a number already saved, never on a number given in the request.

Organizations: invoice fraud, executive impersonation and staff abroad

For organizations, the most damaging lures target the moment when someone has the authority to move money or data.

- **Invoice and supplier fraud.** A message that appears to come from a known supplier announces new payment instructions, or an invoice arrives that is almost, but not quite, like the usual one.
- **Executive impersonation.** A message that appears to come from a senior person asks for an unusual transfer, a confidential document or a favour that skips the usual process. This is the core of business email compromise.
- **Staff abroad.** Travelling staff receive messages that exploit their trip, the local situation or the distance from colleagues who could confirm a request.

KAIA-OR gives each member of an organization their own second look before anyone pays. A changed supplier instruction or an unusual transfer request can be pasted in, to see the warning signs and a suggested next step. Links stay unopened. Each traveller gets their own second look, in English, French, Turkish or Hebrew, alongside what changed in the official travel advisory for their own trip.

The organization keeps control of membership without looking into people's devices. Roles manage membership, not people. Invitations are one-time and bound to an account, and access reviews are dated. The organization sees its sites, not its people's whereabouts, and it does not read the messages its members check.

© Kaia Technologies, Inc.

PROCESS BEATS PERSUASION

The strongest organizational defence against invoice fraud and executive impersonation is a process that no message can bypass: a change to where money goes is confirmed through a contact already on file, and large or unusual transfers need a second person. A second look supports that process; it does not replace it.

© Kaia Technologies, Inc.

4. Protecting the assistant itself, and honest limits

Social engineering aimed at the AI. As assistants read more of the content people receive, attackers try to persuade the assistant instead of the person. A message, a page or a document can contain hidden instructions meant to make an assistant ignore its rules, reveal information or take an action. Hey KAIA is designed to resist this. The content it is asked to check is treated as data to be examined, never as instructions to follow. Model output grants no authority: nothing an AI model produces can approve an action, change a permission or widen what an agent may do. Answers are checked against the evidence before you see them and cite it. No design can rule out manipulation, so the final safeguard is the same as elsewhere in KAIA-OR: the decision stays with people.

Honest limits. A second look is useful because it is honest about what it is.

- **Indicators, not verdicts.** KAIA-OR shows warning signs and suggests a next step. It never declares a message harmless, and the absence of warning signs is not an all-clear.
- **It checks what you give it.** It does not intercept messages or links before you see them, and it does not claim to recognize all scams.
- **Not antivirus.** File and spyware checks give risk indicators only, never a verdict.
- **When it can't tell, it says so.** Some messages cannot be judged from their content alone. KAIA-OR then says so, and the suggested next step is to confirm through a channel you already trust.
- **It cannot undo a loss.** The second look comes before a person replies, pays or opens. KAIA-OR cannot recover money already sent.

© Kaia Technologies, Inc.

5. A personal and organizational playbook

Three steps cover most situations: pause, check through a channel you already trust, and report. They are cyber next steps only. For natural hazards and public-safety events, follow the guidance of the publisher.

PAUSE, CHECK, REPORT

Step	For a person or a family	For an organization
Pause	Treat a request that leaves no time to think as a warning sign in itself. Do not reply, pay or open while unsure.	Make it normal, and welcome, for staff to stop and ask before acting on an unusual request, whoever it seems to come from.
Check	Paste the message into KAIA-OR for a second look, or ask Hey KAIA out loud. Check a link without opening it. Contact the sender through a number or address you already have, never one given in the message.	Confirm any change to where money goes through a contact already on file. Require a second person for large or unusual transfers. Give each member their own second look, including staff abroad.
Report	Tell the organization being impersonated through its official channels, and report fraud to the authorities in your country.	Report to your security contact, share the warning signs with colleagues, and review access regularly through dated access reviews.

A good playbook is short enough to remember under stress. The aim is not to make each person an expert in lures. It is to make the pause, the check through a trusted channel and the report a habit, with a second look close at hand.

© Kaia Technologies, Inc.

What this means for you

Individuals and families. Before you reply, pay or open, paste the message in or ask Hey KAIA out loud, and see the warning signs and a suggested next step in your language. Each adult in your family keeps their own KAIA-OR.

Teams and small businesses. A second look before anyone pays: a changed supplier instruction or an unusual transfer request can be checked first, with links left unopened.

Enterprise and government security teams. Each member of staff, including staff abroad, has their own second look in four languages. Model output grants no authority, and the organization sees its sites, not its people's whereabouts. To evaluate KAIA-OR, request early access or choose Contact us on the KAIA-OR website.

Glossary

Term	Meaning in this paper
Social engineering	Persuading a person to act against their own interest, for example by replying, approving, opening or moving money.
Business email compromise	Fraud in which a message impersonates an executive, a colleague or a supplier to redirect money or obtain data.
Second look	KAIA-OR's check of a message or link: the warning signs it recognizes, what it cannot tell, and a suggested next step.
Warning sign	A feature of a message that lures commonly share, such as requests for sign-in details or demands for money.
Hidden instructions	Text placed in content to manipulate an AI assistant. KAIA-OR treats such content as data, never as instructions.

Sources

1. FBI Internet Crime Complaint Center (IC3), 2025 Internet Crime Report, April 2026. Market context only.
2. Federal Trade Commission (FTC), FTC Testifies before the Joint Economic Committee on Agency's Efforts to Combat Fraud, March 2026. Market context only.
3. UK Department for Science, Innovation and Technology (DSIT) with the Home Office, Cyber Security Breaches Survey 2025, April 2025. Market context only.

© Kaia Technologies, Inc.

WP-3 · Version 1.0 · 3 November 2026

Capability availability depends on the edition and deployment. Capabilities are described as the committed KAIA-OR product scope.