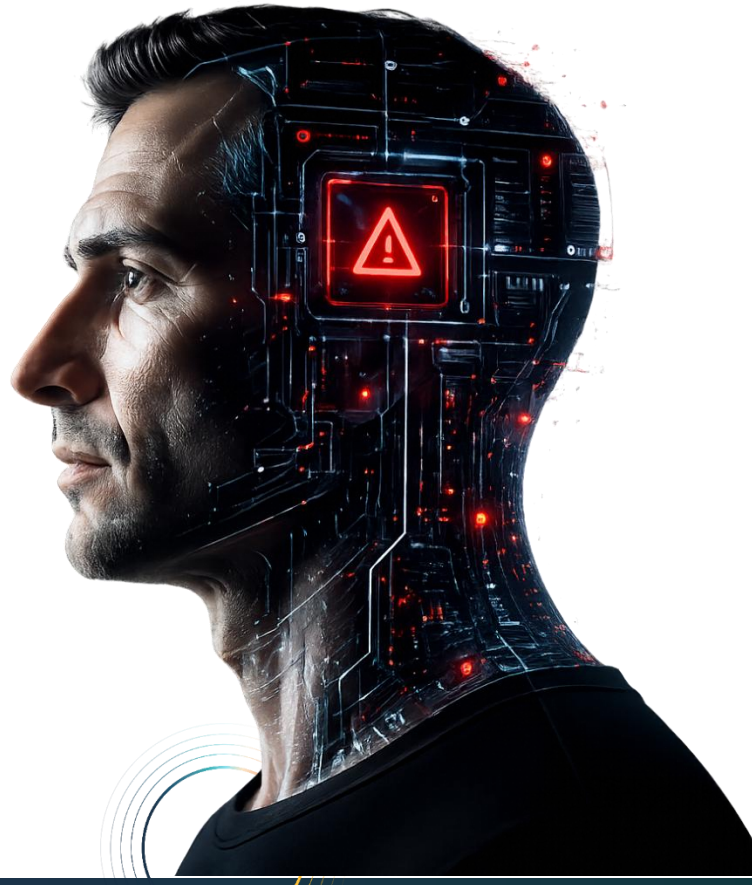
ramsac
the secure choice

HUMAN FIREWALL 2.0

The Ultimate Cybersecurity Upgrade

A practical toolkit to help you and your team build stronger judgement in the age of AI: what to watch for, how to respond, and the habits that hold it all together.

The Toolkit



Technology protects systems. People protect decisions.

Rob May | Executive Chairman, ramsac | rob.may@ramsac.ai

WHAT'S INSIDE

PART 1

The idea in brief

For many years, cybersecurity has been treated as either a technology problem or a training problem. Buy better tools or teach people better habits. Both matter, but neither gets to the heart of what's changed.

Artificial intelligence has removed the old warning signs. Poor spelling, clumsy phrasing, an unfamiliar tone, all the things we trained people to look for, are no longer reliable. AI can now write, speak and design more convincingly than most of us.

That means cybersecurity has become a judgement problem. Every technical control, eventually, hands a decision back to a person: whether to trust an email, approve a payment, or act on an urgent request. Technology can support that decision. It can't make it for you.

> **The organisations that thrive won't have the best technology. They will have the best judgement.**

Human OS is the name for that upgrade, not a new piece of software, but a better default response to uncertainty. This toolkit sets out three parts of it: the threats worth knowing about, the habits worth building, and the tools worth keeping close to hand.

PART 2

The AI threat guide

None of this is designed to make you anxious about AI. It's designed to make sure you recognise the shape of these tactics when they turn up in your inbox, your phone, or your calendar, because increasingly, they will.

1. AI-written phishing and business email compromise

AI can produce a fluent, error-free email in someone else's tone of voice in seconds, using nothing more than a few examples of how they normally write. The old advice, look for typos and clumsy grammar, no longer applies.

Watch for:

- A request that's slightly out of character in timing, channel, or tone, even when the wording is perfect.
- Urgency paired with secrecy: "don't mention this to anyone else yet."

2. Voice cloning and vishing

A convincing clone of someone's voice can now be built from a small amount of public audio, a conference talk, a voicemail greeting, a video. Attackers use it over the phone to impersonate an executive, a supplier, or a family member.

Watch for:

- An urgent call asking for immediate payment, access, or information.
- Pressure to act without calling back: "don't ring me, I'm about to go into a meeting."

3. Deepfake video and video-call fraud

Real-time video deepfakes have already been used to authorise large payments, with an attacker impersonating a senior leader convincingly enough to pass a live video call. This is no longer a future risk; it's an active one.

Watch for:

- Any high-value or unusual request made on video, especially one that discourages a follow-up check.
- Slightly stilted movement, delayed reactions, or odd lighting, though these cues are getting harder to spot.

4. Personalised social engineering

Everything posted online, your LinkedIn updates, press coverage, conference appearances, event photos, can be pulled together by AI into a detailed picture of how someone thinks, works, and communicates. That picture lets an attacker write something that feels like it was made specifically for you, because increasingly, it was.

Watch for:

- A message that references real projects, colleagues, or events, but arrives through an unexpected or unverified channel.
- Familiarity is not the same as authenticity. Knowing something about you doesn't make a message genuine.

5. Cloned websites and fake login pages

AI has made it fast and cheap to clone a legitimate-looking website or spin up a login page that's nearly identical to the real one, often to harvest passwords or payment details.

Watch for:

- Always navigate directly, never through a link in an email or message.
- Check the web address carefully, small misspellings and unfamiliar domains are the main tell.

6. AI-generated invoices, contracts, and CVs

A highly convincing invoice, contract or CV, correct branding, plausible formatting, confident tone, can now be generated in moments. Documents alone are no longer proof of anything.

Watch for:

- Any change to bank details, however official the paperwork looks, verified independently before it's actioned.
- Unusual candidate claims checked through a reference call, not just a document.

7. QR code phishing ("qrishing")

A branded, professional-looking QR code is trivial to generate and place on a poster, a parking sign, or an email. Scanning it can lead straight to a fake login page.

Watch for:

- Any change to bank details, however official the paperwork looks, verified independently before it's actioned.

PART 3

The behaviour playbook

Three upgrades sit at the heart of Human OS. None of them require new technology. All of them require practice.

Upgrade 1

Trust behaviour, not appearance

Appearance stopped being a reliable test of authenticity the moment AI learned to write a perfect email. The better question isn't "Does this look genuine?" It's "Does this behave like the person I know?"

An attacker can copy a logo, a tone of voice, even a colleague's writing style in seconds. What's much harder to copy is a pattern of behaviour built up over years.

Before you act, ask:

- Is this how they normally communicate?
- Would they usually make this request by this channel?
- Is the urgency consistent with how they normally work?
- If I challenged this request, would they expect me to?

Upgrade 2

Create thinking space

Attackers rely on urgency to lower the quality of your thinking. “I need this today.” “I'm about to board a flight.” The goal isn't just speed, it's stopping you from noticing that something doesn't fit.

Permission to pause is one of the most valuable things a leader can give their team. Not permission to procrastinate, permission to stop for long enough to ask whether something makes sense.

Build the pause in:

- Take a breath before clicking approve on anything financial or sensitive.
- Pick up the phone rather than replying to the same channel.
- Ask a colleague to take a quick second look.
- Reward people for taking an extra minute, not just for moving fast.

Upgrade 3

Verify independently

Organisations that handle uncertainty well share one habit: they never rely on a single source of information for an important decision. Cybersecurity deserves the same discipline.

The principle is simple. When a decision matters, don't verify it through the same route that delivered it.

In practice:

- Bank detail change request? Call a number you already have, never one in the email.
- Login link? Don't click it, open your browser, and use your own bookmark.
- Unexpected Teams or WhatsApp message? Call the person, or walk over to them.
- Make verification normal. Nobody should feel accused for asking a confirming question.

What leaders can do

Habits spread from the top. A leader who visibly pauses, checks, and verifies makes it normal for everyone else to do the same.

- Model the pause yourself, visibly, especially on financial or sensitive requests.
- Praise people for verifying, even when the request turns out to be genuine. You're rewarding the habit, not the outcome.
- Build a no-blame reporting culture. If someone realises they've made a mistake, what matters most is how fast they tell you, not how they're punished for it.
- Give the three upgrades five minutes in a team meeting or induction, little and often beats a single annual session.

Scripts worth borrowing

Verification only works if people feel comfortable doing it out loud. These are useful lines for exactly that moment.

"I just want to double check this before I go ahead, could we hop on a quick call?"

"This is a bit unusual for us, would you mind if I verified it through our normal process first?"

"Nothing personal, I always make a quick call on payment changes. It's just how we do things here."

"Can you send that from your usual email, or shall I call you back on the number I already have?"

PART 4

Practical tools

Trust behaviour, not appearance

Run through this before you click, reply or approve anything that feels even slightly off.

- Am I being asked to act quickly, or to keep this quiet?
- Does this match how this person or organisation normally contacts me?
- Am I being asked to click a link, open an attachment or change a payment detail?
- Could I verify this through a different channel before I act?

If you've ticked any of the first three and can't confidently tick the fourth, stop and verify before you go any further.

The payment and bank detail verification protocol

Use this every time a bank detail, payment amount or payee changes, no exceptions, regardless of how senior the request appears to come from.

1. Do not action the request immediately, even if it's marked urgent.
2. Do not reply to the email or message that contained the request.
3. Find a phone number for the requester from an existing, trusted source, never one supplied in the message itself.
4. Call and confirm the change verbally before anything is actioned.
5. Log the change and who verified it, so there's a record if questions arise later.

If you think you've made a mistake

Speed matters far more than blame. The faster something is reported, the more can usually be done about it.

1. Tell your IT or security team immediately, don't wait to be sure.
2. Don't try to fix it yourself first, that can sometimes make recovery harder.
3. Write down what happened while it's fresh: what you received, what you clicked or sent, and when.
4. If money has been sent, contact your bank straight away, speed genuinely affects the chances of recovery.

Quick reference

Keep this table close to hand or share it with your team as a one-page reminder.

Upgrade	Old default	New default
Trust behaviour	"It looks right, so it must be right."	"Does this behave like the person I know?"
Create thinking space	Respond immediately, every time.	Pause when the stakes justify it.
Verify independently	Reply to confirm, using the same channel.	Check through a separate, trusted channel.

Keep the conversation going

This toolkit is the starting point, not the end of the discussion. Back at your desk, the question worth asking isn't "How do we become more secure?" It's "How do we help our people make better decisions?"

For more resources, or to talk through how these habits might work in your organisation, get in touch: ramsac.com/cybersecurity

Rob May
Executive Chairman, ramsac | rob.may@ramsac.ai