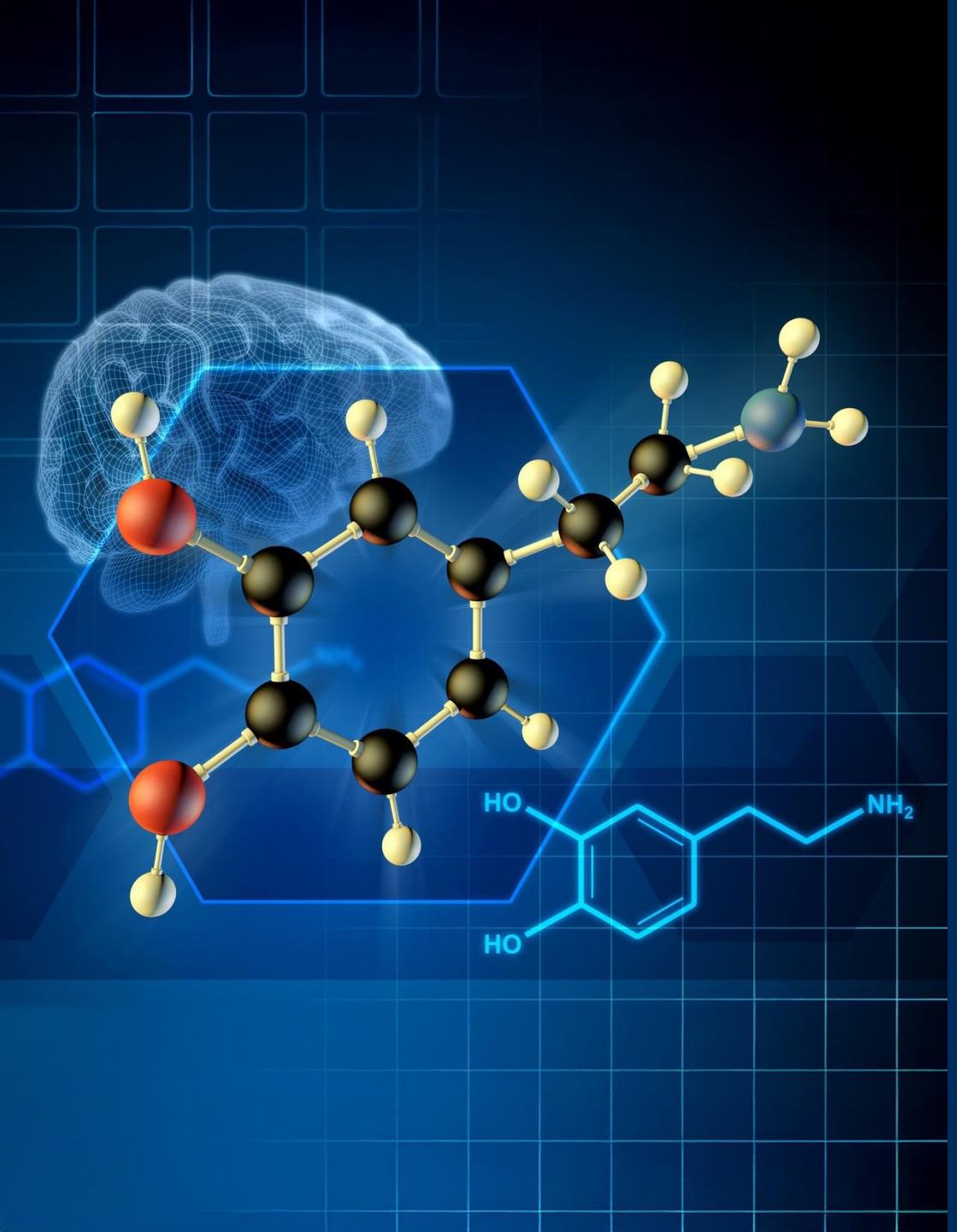


Kepner-Tregoe

Challenges
Opportunities
Results

What every graduate should know about LLM's and Cybersecurity

Dr Clive King

clking@kepner-tregoe.com

clk5@aber.ac.uk

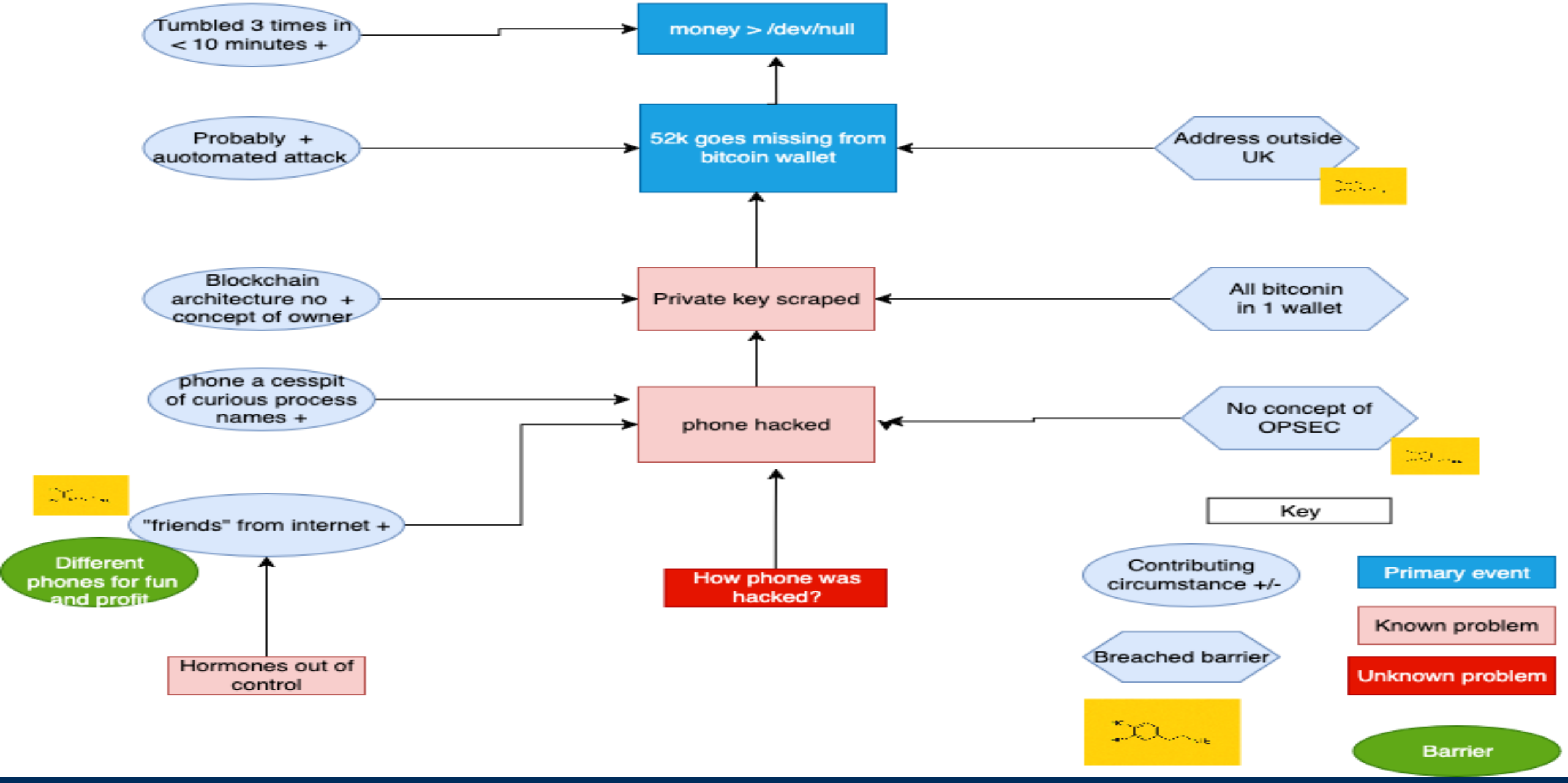
December 2025

Ruth

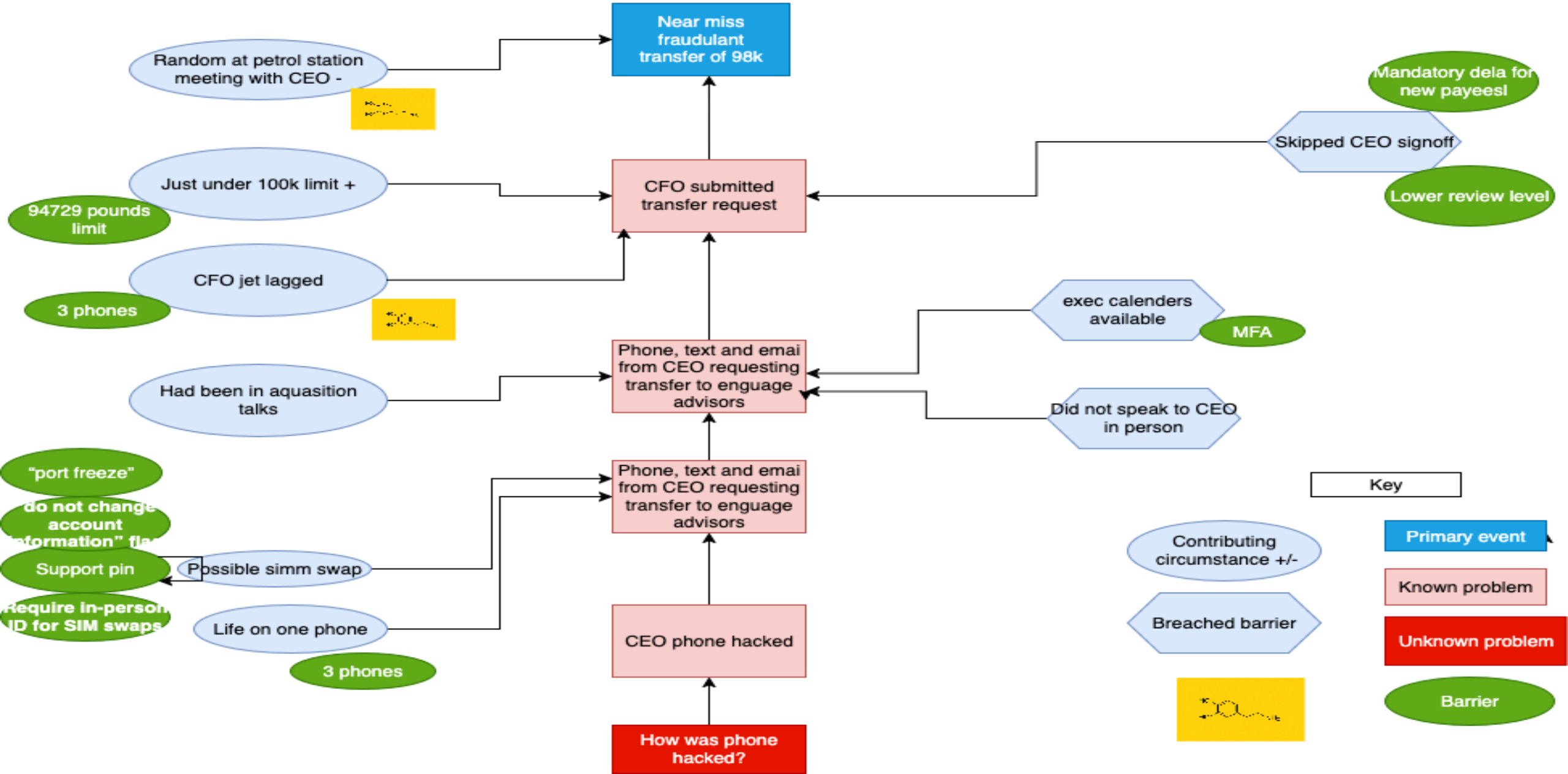
Dream scam target

Trust model rooted in 1960’s

Porsche fund drained Incident Map



98k CFO almost scam Incident Map



Share your last criminal act?

- Must have arguably criminal intent and exclude motoring
- August 1985 Hathersage, Peak District
- Drank bottle of milk at 5am from milkmans crate drop
- Left 50p on crate [chatgpt says 20p per pint in 1985]
- I struggle to think like a criminal

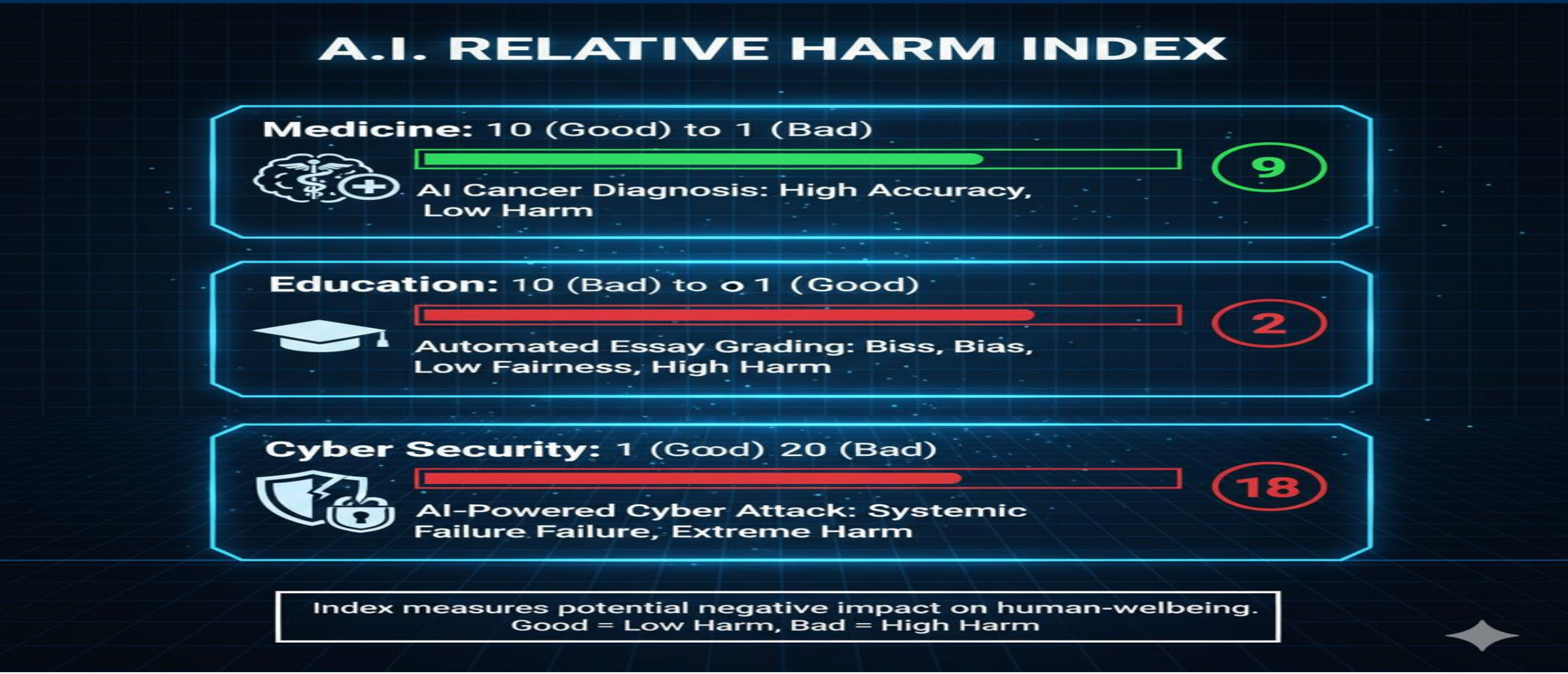
- Thinking like a criminal is a foundation of successful cyber defense
- Pen testing mindset – test the defenses. [white hat]
  - Robust Safety net

- Open question : how to give graduate enough of the criminal mindset to defend themselves and their organization, but not to equip them for a life of crime ?

How cybercriminals think

- They don’t care about you or the consequences for you … only money
  - You are business, not a person. Just a target and a potential payout
- Specialized
  - Zero-day discovery
  - Target identification
  - Intrusion
  - Ransomware as a service
  - Social engineers
  - “Accountants”
- Flow of money is planned in advance
- Know their effort vs reward curve well : Move on to find a easier target
- Different mindset for
  - Nation states with political agenda
  - Script/prompt kiddies

A.I. Relative Harm Index [AIRHI]



Dopamine vs Defence: The Risks We Allow

1 Allowed Risks
2 LLM Danger
3 Training Fails
4 Usability v Security
5 Harder Solutions

Dopamine =	shortcuts e.g. sharing docs and opening docs from colleagues
Dopamine =	click the link
Dopamine =	weak or repetitive passwords – quick access (less thinking)
Dopamine =	Dopamine = more system 1 behaviours – auto pilot mode when emails arrive
***Dopamine =**	*'yes remember my password' - saved passwords onto a Credential Manager (without MFA)*

*contentious example

Living proof email link training does not work

Troy Hunt runs ';--Pwned

Few humans alive are more aware about Phishing attacks

"You know when you're really jet lagged and really tired and the cogs in your head are just moving that little bit too slow? That's me right now, and the penny has just dropped that a Mailchimp phish has grabbed my credentials, logged into my account and exported the mailing list for this blog. I'm deliberately keeping this post very succinct to ensure the message goes out to my impacted subscribers ASAP, then I'll update the post with more details. But as a quick summary, I woke up in London this morning to the following:"

Even the **most** aware can have System 1 failure moments

Training alone cannot defend us against ourselves

01 Email cues

Cues are the properties of an email that either compel a user to click on a fraudulent link or attachment or alert the user that the email may be a phish

02 Premise alignment

a measure of how closely an email matches the work roles or responsibilities of an email's recipient or organization.

NBIST Phish scale
https://theconversation.com/how-letting-your-mind-wander-can-reset-your-brain-259854
https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.2276.pdf

Training tells us to spot these 2 things.

BUT

- These require System 2 Thinking
- Reading email is largely System 1
- Attention fatigue
- Dopamine drives our attention to be occupied

LLM's Switch off system 2: Are we more vulnerable to phishing?

Phishing attacks exploit System 1 shortcuts

- Clicking without scrutiny
- Trusting familiar looking messages
- Emotional/urgency triggers

Dopamine-driven behaviours and environments drive more use of System 1

- Scrolling, multi tasking, chatting with LLMs

Study shows reduced cognitive load & neural engagement drop when using LLMs:

- Poor decision making
- Lower threat detection - reduced situational awareness of potential phishing cues

Over-reliance and use of LLMs can **erode our cognitive resilience**

Kepner-Tregoe

Legacy madness : keeping 55 year old email

Email reality

- Globally accessible email address
- Attachments which execute on local machine (SVG executes JavaScript)
- Links to globally accessible resources
- Low probability of message send failure (spam folder)
- Low barrier to email address and provider creation
- No central authority

Bless Raymond Tomlinson in 1971 when security was not really a thing

What would make email secure (ish)

- Restricted visibility email address
- No attachments or sandbox execution only
- No access to global resources, only local
- Low bar for spam folder
- Identity validation on email address creation
- No anonymous email addresses
- Central authority to administer accounts and provider creation

A fraction more secure (DKIM, DMARK, etc.)

LLM’s makes hacking at scale practical

- Prompt injection
  - Direct
  - Indirect
- Data poisoning
- Social engineering at scale
- Deepfakes and voice cloning for fraud
- AI-written malware that adapts
- Proof of concept exploits for the curious
- Autonomous attack agents
- OSINT agents
- Steal the model
- Compressed attack timelines and overwhelmed defenders

Kepner-Tregoe

A.I. enables novel attack surfaces

Any University I.T. Security policy

Exists to protect the institutions I.T. infrastructure

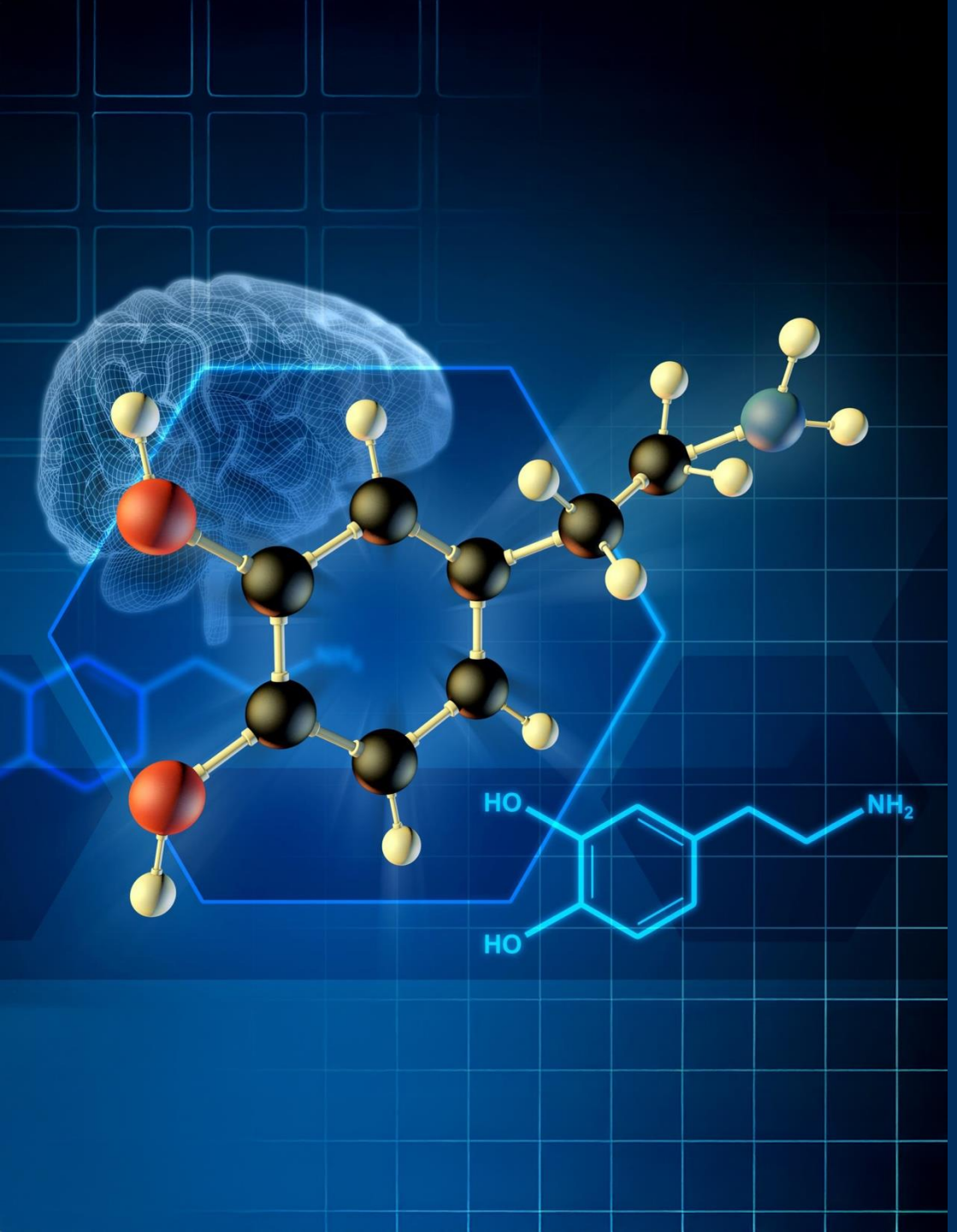
Anti-pattern for educating (any) students around cyber security

- Only what
  - Not how
  - Not why
  - Not what if

- An institutions students in relative terms not the risky actor they once were

What do *all* graduates need to know about Cyber Security & A.I. exploits

- Awareness
  - Scale
- Perspective
  - Criminal mindset
- Mechanism
  - Dangers
- Preventative
  - Stop
- Contingent
  - Response



Kepner-Tregoe

Challenges
Opportunities
Results

Don’t start from here

Fix the foundations

Dr Clive King

clking@kepner-tregoe.com

clk5@aber.ac.uk