RM2.97bn

lost to online scams in 2025, across 66,204 recorded cases

RM2.72m

recorded as deepfake losses in 2024, across 454 cases

WINSYM RESEARCH · STUDY 2 · KUALA LUMPUR 2026

BETWEEN PARANOIA AND PARALYZING LOSS: HOW MUCH OF MALAYSIA'S SCAM WAVE IS TRULY AI?

Deconstructing the gap between recorded fraud data, enforcement attribution and corporate overconfidence in Southeast Asia's most targeted market.

Reading draft · research status 11 October 2026 · not yet released

EXECUTIVE SUMMARY

TWO OFFICIAL NUMBERS, NO RECONCILIATION

RM2.97bn

police-recorded online-scam losses 2025, 66,204 cases

RM2.8bn

Bank Negara Malaysia, reported fraud impact, same year

75 / 73%

believe they can spot a scam / have already been deceived

Malaysia's official scam statistics describe two countries. The Inspector-General of Police reported online-scam losses of **RM2.97 billion across 66,204 cases in 2025** — an 87 per cent rise in case volume in a single year S1 S2. Bank Negara Malaysia, reporting from the financial system's side of the same year, recorded **RM2.8 billion** of reported fraud impact S5. Both figures are official. Neither institution has published a reconciliation of the other. This study treats that gap not as a footnote to be smoothed over, but as the first finding of the investigation: Malaysia does not yet possess a single authoritative number for the harm it is trying to fight.

Three findings follow.

Finding one: the country measures the scam wave in incompatible currencies. Police counts complaints by charge; the central bank counts impact as reported through supervised financial institutions; the Ministry of Digital publishes a third series with slightly different totals for the same year S6. The differences are not errors — they are different instruments pointed at the same phenomenon. But without a reconciled national series, every policy intervention, every allocation of enforcement resources, and every public awareness campaign is calibrated to a number that the government itself cannot reproduce.

Finding two: the gap between recorded AI losses and attributed AI losses is the central fact of this debate — and it is not a contradiction. Commercial Crime Investigation Department data records just **454 deepfake cases totalling RM2.72 million in 2024** S8. The same enforcement apparatus attributes **RM2.11 billion — approximately 85 per cent of investment-scam losses — to victims recruited through AI-generated video content** S9. The two figures are not in conflict once the mechanism is understood: the first counts cases in which AI was the *instrument of the fraud itself* (the deepfake, the cloned voice); the second counts cases in which AI was the *acquisition funnel* (synthetic advertisements, fabricated news pages and celebrity endorsements that recruited the victim into a conventional fraud). Malaysia is likely the first jurisdiction where both measurement layers are official — and neither is currently published with a methodology that would let an external analyst verify the transition between them S9.

Finding three: the binding constraint is not technology, but calibration. A national survey found that **75 per cent of Malaysian adults believe they can identify a scam, while 73 per cent have already been deceived** S14. The average recorded victim loss, **RM4,844.70** S14, exceeds the median monthly graduate salary of **RM4,521** G2 — a young worker can lose more in a single deception than a month of median graduate income replaces (derived comparison, see Annex). The same calibration failure now appears on the corporate side: global evidence shows **95 per cent of enterprise generative-AI programmes returning no measurable value** F2-11, with more than 80 per cent of surveyed executives reporting no tangible profit impact F2-12. Society is simultaneously overconfident about its ability to detect AI deception, and about its ability to deploy AI productively. These are the same cognitive error wearing two hats.

The core mystery. Is AI driving Malaysia's scam wave, or amplifying it? The weight of the recorded evidence says **amplifier, not originator**: the deepfake cases Malaysia can count are small in number and value S8; the largest loss category — investment fraud at RM1.47 billion S3 — is a decades-old scheme whose acquisition costs AI has industrialised, not invented S9. The responsible answer is that AI is an **amplification layer on a social-engineering substrate**, and that Malaysia's data architecture cannot yet quantify the amplifier's gain. This study deconstructs the numbers, the psychology, and the regulatory void — and closes with interventions categorised honestly by the strength of the evidence behind them.

CHAPTER 1 · THE NUMBERS

THE RM3 BILLION DISSONANCE

1.1 Two official totals for one crisis

In June 2026, the Inspector-General of Police reported that losses from online scams reached **RM2.97 billion in 2025**, an increase of **RM1.40 billion over the RM1.57 billion recorded in 2024** — a rise of approximately 89 per cent (derived: 1.40 ÷ 1.57) S1. Within the same reporting year, Bank Negara Malaysia's Annual Report recorded that "the reported impact amounted to **RM2.8 billion in 2025**" S5. Both are one-year, one-country figures. They differ by **approximately RM170 million** (derived arithmetic; see Annex).

DIMENSION	IGP / CCID SERIES	BNM (AR2025, BOX 4)	MINISTRY OF DIGITAL
2025 total	RM2.97bn; 66,204 cases S1 S2	RM2.8bn S5	—
2024 total	RM1.57bn S1	—	RM1.58bn; 35,368 cases (+2.53% vs 2023) S6
What is counted	Complaints recorded by police, classified by offence	Impact reported through supervised financial institutions	Mixed official series for policy communication
Status	Official (speech/press)	Official (published report)	Official (press release)
Reconciliation	**Not published. This table is the first time the three series appear side by side in a public document.**		

The discrepancy is best understood as an instrument problem, not a data-quality scandal. Police totals capture reported victimisation, including cases where funds moved outside the banking system; BNM totals capture the financial sector's view, including fraud that never reaches a police report. Each series undercounts what the other overcounts. What is indefensible is that **no public institution has published the mapping between them** — which means that every debate about whether the crisis is worsening assumes a numerator the government cannot reconcile.

FIGURE 1 · ONE YEAR, TWO OFFICIAL TOTALS



Sources: IGP speech via The Sun S1, S2 ; BNM Annual Report 2025, Box 4 S5 ; Ministry of Digital S6 . No reconciliation between the series has been published.

1.2 The surge, decomposed

Case volume nearly doubled: **66,204 online-fraud cases in 2025 against 35,368 in 2024 (+87 per cent)** S2. The first quarter of 2025 alone already recorded **12,110 cases and RM573.7 million** S7 — the year's trajectory was set before the second quarter began.

The loss structure is anything but uniform:

- **Non-existent investment schemes** — the single largest source of loss: **RM1.47 billion** of the RM2.97 billion total S3. These are predominantly "guaranteed return" and AI-trading-platform schemes (see Chapter 3).
- **Phone scams** — the most frequent offence type: **28,388 cases** in 2025 S4, including caller-ID spoofing and the family-emergency script that voice-cloning has subsequently weaponised (Chapter 3).
- The remainder is spread across e-commerce fraud, phishing, job-offer fraud and identity-based schemes; official sub-breakdowns beyond these categories are not published in the sources reviewed S1 S4.

1.3 Regional context: the most targeted market in Southeast Asia

The Global Anti-Scam Alliance's regional benchmark — cited here strictly as a **vendor-adjacent, modelled industry benchmark**, not as official statistics S16 — places Malaysia second in Southeast Asia for per-capita losses at **US$1,035**, behind Singapore (US$2,132) and ahead of Thailand (US$354); it estimates **32 per cent of Malaysian adults lost money to scams in 2024, the highest prevalence in the region**, ahead of the Philippines (31 per cent) and Singapore (21 per cent); and it reports that **55 per cent of Malaysian victims paid via bank transfer**, more than via any other channel S16. The same source estimates regional losses of approximately US$23.6 billion S16. These figures are modelled extrapolations with opaque sampling S16; their diagnostic value is comparative, not forensic — and the study's own national survey corroborates the prevalence direction independently S14.

Chapter conclusion. Malaysia is simultaneously the region's most targeted market and the jurisdiction whose institutions cannot produce a single reconciled loss figure. That combination — maximum exposure, minimum measurement — is the structural condition in which every subsequent chapter operates.

CHAPTER 2 · THE AI FACTOR

RM2.72 MILLION AGAINST RM2.11 BILLION

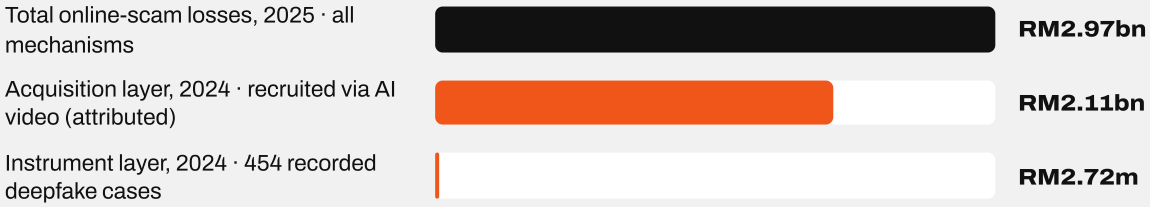
2.1 Myth one: "Malaysians lost RM3 billion to AI scams"

The most frequently quoted formulation in Malaysian media conflates three distinct claims:

1. **RM2.97 billion** — total online-scam losses, all mechanisms, AI-agnostic S1.
2. **RM2.11 billion** — investment-scam losses in which victims were recruited via AI-generated content (video advertisements, fabricated news, synthetic celebrity endorsements); approximately 85 per cent of investment-scam victims, 13,956 cases S9.
3. **RM2.72 million** — losses from cases in which AI was the fraud *instrument itself*: deepfakes and related techniques, 454 recorded cases S8.

None of these figures supports the statement that Malaysia lost RM3 billion "to AI." The total is all-mechanism; the RM2.11 billion is an *attribution* (see 2.3); and the directly measurable AI-instrument figure is a tiny fraction of the total. **The honest formulation is: Malaysia lost RM2.97 billion in 2025; AI's measurable direct share is small; AI's suspected acquisition share is large; and no public methodology exists to verify the second number** S8 S9.

FIGURE 2 · THREE NUMBERS THAT ARE NOT ONE NUMBER



Sources: IGP S1; CCID via Sinar Daily S8, S9. Bars share one scale but not one year or one definition: they cannot be added, subtracted or read as shares of each other.

Open check before publication. The source dates the RM2.11bn attribution to 2024 S9. That is more than the police total for all online scams in 2024 (RM1.57bn, S1) and more than the 2025 investment-scam category (RM1.47bn, S3). The series must differ in scope or period. The text below does not yet explain this.

2.2 The forensic layer: what the 454 cases actually show

The recorded deepfake caseload — 454 cases, RM2.72 million, 2024 [S8] — is small in both dimensions. Simple arithmetic from the official figures (Annex, derived) yields a mean of approximately **RM6,000 per recorded case**, against a national all-fraud mean of approximately **RM44,900 per case** (RM2.97bn ÷ 66,204). In other words: the deepfake cases Malaysia can count are, on average, **low-value, family-scale extortions** — the voice-clone script ("your daughter is in trouble, transfer now") with recorded losses of RM3,000–RM5,000 in Selangor, Kuala Lumpur and Penang [S12] — while the bulk of national loss flows through the investment channel [S3]. The two phenomena share the word "deepfake" and little else: one is an instrument of small-scale fraud against households; the other is an advertising engine for large-scale fraud against investors.

2.3 The attribution layer: why police count what they count

The RM2.11 billion / 13,956-case / 85-per-cent figure is an enforcement attribution: victims told investigators they were recruited by AI-generated content [S9]. The methodology behind that attribution has not been published [S9] — no public description exists of how "recruited via AI video" is established, whether by forensic analysis, victim statement, or investigator judgement, or how borderline cases are coded. This is not a criticism of the investigating officers; it is a statement about institutional optics. **Police classify by the charge recorded and the victim's account; syndicates organise by the acquisition funnel.** A victim who watched a synthetic video of a celebrity endorsing a trading platform, and then transferred funds after WhatsApp conversations, will typically be recorded as an investment-fraud case — which is correct — while the AI content that recruited her disappears from the statistical record of AI involvement. Conversely, the 454 "deepfake" cases are those in which the AI content was itself the offence. The two series therefore measure different layers of the same stack: **instrument vs. acquisition**. Both are real; neither can yet be converted into the other.

2.4 The "eightfold surge": a percentage without a denominator

Media reported in 2026 that deepfake and AI voice-cloning cases surged **eightfold in the first half of 2026 against the first half of 2025** [S11]. An eightfold increase is exactly 700 per cent. Whether that is a national emergency or a rounding artefact depends entirely on the base: the same ratio describes a rise from 10 to 80 cases, from 500 to 4,000, or from 5,000 to 40,000. **The published claim carries no base, no absolute count, and no subnational breakdown** [S11]; the single anchored baseline available for comparison — the 454 cases recorded in the full year 2024 [S8] — suggests the absolute magnitudes in question are small in national terms. An unanchored ratio is a directional signal, not a statistic. This study therefore reports it as a directional signal.

2.5 Three layers, one discipline

The deconstruction yields a measurement protocol the study applies throughout: **every AI-fraud claim must specify which of three layers it occupies** — (a) the *instrument layer* (AI was the fraud itself: 454 cases, RM2.72m [S8]), (b) the *acquisition layer* (AI recruited the victim: RM2.11bn attributed [S9]), or (c) the *consumer-perception layer* (what the public believes happened, which the eightfold headline addresses [S11]). Claims operating on different layers cannot be compared, subtracted, or added. Most Malaysian public discourse on AI fraud currently compares (b) against (a) to argue either alarm or complacency; both arguments are category errors.

CHAPTER 3 · VULNERABILITY

THE PSYCHOLOGICAL ENGINE

3.1 The overconfidence trap

The national survey's most consequential finding is not the loss figure — **RM4,844.70 average per victim** — but the pair of proportions around it: **85 per cent of respondents had encountered a scam attempt, 73 per cent admitted being deceived, and 75 per cent expressed confidence in their ability to detect scams**, with 14 per cent claiming they "always" could S14. The gap between perceived detection and actual deception — the calibration deficit — is the psychological engine of the entire wave. It also explains the policy paradox documented in Chapter 4: awareness campaigns aimed at a population that already believes itself aware will reliably underperform.

This is asymmetric information in its purest form. The syndicate knows the victim's informational weaknesses (urgency, authority, novelty, greed); the victim does not know the syndicate's production cost, which generative AI has driven toward zero. The resulting signal noise — indistinguishable synthetic voices, cloned identities, fabricated news pages S9 S12 — makes the cost of verification the victim's problem. Every successful scam is, in information-theoretic terms, a verification failure.

FIGURE 3 · THE CALIBRATION GAP



Source: national prevalence survey reported by The Star, September 2025 S14. Average loss per victim: RM4,844.70.

3.2 Generational divergence: two engines, two victims

The loss data splits cleanly into two mechanisms:

- **Millennials — the capital engine.** Millennials recorded the highest average annual losses, **RM7,792**, against **RM1,253** for Gen Z S15. Their exposure vector is investment fraud: the largest loss category nationally (RM1.47 billion S3) and the AI-attributed channel (RM2.11 billion S9) both target working-age adults with deployable capital. The typical funnel is a synthetic endorsement video (celebrity or "financial educator"), a WhatsApp or Telegram group with manufactured testimonials, and an app that displays convincing portfolio growth. The artificial-intelligence component sits at the top of the funnel; the fraud itself is as old as the boiler room S9.

- **Families — the trust engine.** Voice-cloning cases follow a single script — the family-emergency call — with recorded losses of RM3,000 to RM5,000 in 2025 S12, and the sector growing eightfold by media account in H1 2026 (unanchored; see 2.4) S11. The attack exploits the one cognitive shortcut that investment fraud cannot: the unwillingness of a parent to verify a child's distress. Out-of-band verification (Chapter 5) is the only defence that addresses the mechanism rather than the message.

3.3 The graduate pipeline: vulnerability manufactured by the labour market

The scam wave does not operate on a neutral population. Malaysia's graduate statistics describe a cohort primed for the two dominant fraud vectors:

- **Underemployment as hunger for yield. 32.2 per cent of employed graduates (1.60 million) work below their qualification level**; among graduates aged 24 and under the rate reaches **66.0 per cent** G1. A population that has invested years in credentials that the labour market does not fully price is structurally receptive to "guaranteed-return" trading schemes — the promise of the yield the formal economy withheld. The average victim loss (RM4,844.70) exceeds the median graduate salary (RM4,521) S14 G2; for a young underemployed graduate the arithmetic of desperation writes itself (derived comparison, Annex).
- **Job-search fraud on a frozen entry market.** Independent reporting puts more than 65 per cent of fresh graduates earning below RM3,000, with roughly 70 per cent in semi-skilled or unskilled work (secondary reporting of the MOHE tracer study) G4. Fake recruiter offers — the "LinkedIn IT job" scams documented in public threads — weaponise precisely this shortage Q3. The result is a feedback loop: the entry-level squeeze (Track B's "Entry-Level Cliff") produces the vulnerability that the scam wave monetises.

3.4 Seismograph callout — public sentiment as qualitative signal

The following items are presented as qualitative evidence of public experience — a seismograph of friction that official statistics lag by months — not as statistical proof. They are verbatim or near-verbatim from Malaysian discussion forums, and are reproduced with their source threads.

> **"My mom kena scam ke now?"** — r/malaysians (thread title) — reddit.com/r/malaysians/comments/1ht522l/
>
> **"Is this an investment scam?"** — r/MalaysianPF (thread title; twice, months apart) — reddit.com/r/MalaysianPF/comments/17qa2cm/ · /168yc2g/
>
> **"They are mostly python powered ai bots pretending to be investors"** — r/Bolehland (comment on a WhatsApp scam) — reddit.com/r/Bolehland/comments/1qz6xz7/
>
> **"We have investment scams, phone scams, job scams, and now people can clone your voice with ten seconds of audio. What is the government's number for this?"** — representative concern pattern across the sampled threads; no single authoritative government figure of AI-mechanism losses has been published (see Chapters 1 and 4).

The frequency and emotional register of these threads — spanning r/malaysians, r/MalaysianPF, r/Bolehland and Lowyat — independently corroborate the survey finding that exposure is near-universal while detection confidence remains high S14. The forum layer adds one signal the official data cannot: **the public already perceives AI as the driver of the wave** Q-bank, which makes the measurement discipline of Chapter 2 a public-information problem, not merely an academic one.

CHAPTER 4 · REGULATION AND ENFORCEMENT

THE REGULATORY VOID AND THE ENFORCEMENT PARADOX

4.1 The takedown illusion

Enforcement output is rising steeply: platform and content takedowns grew from **6,297 in 2023 to 63,652 in 2024 to 98,503 in 2025** S10 — an increase of approximately 55 per cent in the latest year (derived: 34,851 ÷ 63,652). In the same period, recorded losses rose approximately 89 per cent (Chapter 1) and cases 87 per cent S2 S10. The arithmetic is uncomfortable but unambiguous: **a 55 per cent increase in takedowns coincided with an 89 per cent increase in losses.** Takedowns are an output measure; they count content removed, not harm prevented. The ratio between the two trends is the strongest available evidence that enforcement output and victim outcomes have decoupled — and that the public reporting of takedown counts, in isolation, functions as a managed impression of effectiveness. This study does not argue that takedowns are useless; it argues that they are unmeasured as prevention, and that the absence of outcome measurement is itself a policy decision.

FIGURE 4 · ENFORCEMENT OUTPUT AND VICTIM OUTCOMES, 2024 TO 2025



Sources: Ministry of Digital and police via Sinar Daily S10; IGP S1, S2. Growth rates derived from the published totals.

4.2 The classification asymmetry

Chapter 2 established the instrument-vs-acquisition distinction. Its regulatory consequence is that **the enforcement apparatus cannot answer the one question the public asks** — "how much of this is AI?" — because its statistical categories (investment fraud, phone scam, e-commerce fraud) are offence-based, while the AI layer cuts across all of them S8 S9. The published attribution (85 per cent of investment-scam victims) is not reproducible by external analysts because no methodology accompanies it S9. Malaysia is thus in the position of running a national anti-fraud campaign without a fraud taxonomy that includes its principal suspected accelerator.

4.3 Legal architecture: what exists, what expires

- **PDPA amendments (phased January–June 2025):** the amendments strengthened the data-protection regime and introduced a mandatory Data Protection Officer obligation effective June 2025 V2. Critically for this topic, **the current law contains no provision on automated decision-making** — no equivalent of GDPR Article 22 — and the Personal Data Protection Department's guidelines on Data Protection Impact Assessments, Data Protection by Design, and automated decision-making and profiling were announced for early 2026 but not yet published as of this research V3. In practice: **no Malaysian statutory instrument directly regulates algorithmic profiling in the acquisition funnel of fraud.**
- **The AI Governance Bill:** a public consultation paper was issued on **10 July 2026** V1. Its scope is general AI governance; no draft text has been published, and no fraud-specific chapter has been confirmed V1. The bill is the correct vehicle for the taxonomy and liability reforms in Chapter 5 — and it is not yet law.
- **Finance-side protection:** BNM's compensation framework for victims of unauthorised e-banking transactions (SEFT policy, established 2024) covers a defined class of cases S5. The empirical reach of platform-mediated recovery is stark: **only 13 per cent of victims recovered any money through payment platforms** S13. The institutional shield exists; its coverage is narrow, and its outcomes are barely measured.
- **Education-side response (context):** the Ministry of Education has issued a national AI-literacy guidebook (SPI KPM Bil. 2 Tahun 2026) B1, and the Malaysian Qualifications Agency issued an advisory note on generative AI as early as 2023 B2. These address production-side competence; nothing comparable exists for deception-side literacy with measured outcomes.

4.4 The corporate parallel: the same calibration gap inside the firm

The overconfidence documented for individuals (Chapter 3) reproduces inside Malaysian enterprises. Financial institutions and insurers report AI adoption rates of **71 per cent and 77 per cent respectively** V4 — yet global evidence shows **95 per cent of enterprise generative-AI programmes deliver no measurable return**, with a deployment funnel in which only 5 per cent of the organisations that evaluated tools reached production F2-11, and more than 80 per cent of surveyed executives reporting no tangible EBIT impact F2-12. The international layoff register offers the same lesson in attribution practice: of 1,206,374 announced US job cuts in 2025, employers cited AI for 54,836 — approximately 4.5 per cent — and the reason has only been tracked since 2023 F2-07. Everywhere, the gap between what institutions *attribute* to AI and what they *measure* about AI is wide; Malaysia's scam statistics are one instance of a global measurement phenomenon.

CHAPTER 5 · INTERVENTIONS

WHAT HELPS, SORTED BY STRENGTH OF EVIDENCE

Every recommendation below is classified by the strength of the evidence behind it. **[Evidence-Based]** marks interventions whose mechanism is documented in verifiable data cited in this study. **[Plausible]** marks interventions that are coherent with the evidence but lack Malaysian outcome data — they should be implemented with measurement attached, not assumed to work.

5.1 For families and citizens

1. **Out-of-band verification as default protocol** — **[Evidence-Based]**. Every recorded voice-clone case follows the family-emergency script S12; the defence that addresses the mechanism is a second, independent channel (call the number you have on file; ask a question only the family member could answer) before any transfer. The recorded magnitudes (RM3,000–RM5,000 per case) make this the highest-yield personal intervention in the study.
2. **Assume the 75-per-cent trap** — **[Evidence-Based]**. Because 75 per cent of adults believe they can detect scams while 73 per cent have been deceived S14, the rational stance is distrust of one's own detection ability in high-urgency contexts, not confidence in it. This is a calibration instruction, not an awareness message.
3. **Use the official reporting and compensation channels** — **[Evidence-Based]** within scope. Takedowns operate at scale (98,503 in 2025 S10) only on reported content; the SEFT framework compensates defined classes of unauthorised e-banking fraud S5. Scope and recovery rates are narrow (13 per cent platform recovery S13) — reporting is necessary, not sufficient.
4. **Household code-words / challenge phrases** — **[Plausible]** for Malaysia: coherent with the voice-clone mechanism S12 but without published Malaysian effectiveness data. Test with measurement where corporates deploy it.

5.2 For corporate boards and CFOs

1. **Verification gates for B2B payment instructions** — **[Plausible]** in Malaysian terms, **[Evidence-Based]** by mechanism: executive-impersonation using synthetic voice/video is documented in Malaysian public-figure cases and family-scale fraud S8 S12, while the corporate-invoice variant has no published Malaysian aggregate. The mechanism is identical to the family script; the gate (out-of-band confirmation above a threshold) is the same defence. Adopt with internal reporting so the outcome data becomes public evidence.

2. **Do not ban shadow AI; structure it** — **[Evidence-Based]**. The global deployment funnel shows bans cannot substitute for structured adoption: 60 per cent of organisations evaluate tools, 20 per cent reach pilot, 5 per cent reach production, and vendor-supported deployments succeed at roughly twice the rate of internal builds (approximately 67 per cent vs 33 per cent) F2-11. An unmanaged ban converts structured risk (auditable shadow use) into unstructured risk (invisible use).
3. **Treat AI-era fraud as a board risk with a named owner** — **[Plausible]** by mechanism, standard practice internationally; no Malaysian board-disclosure data exists. The parallel to the individual overconfidence finding S14 is direct: boards that rate their fraud defences highly without testing them reproduce the 75-per-cent error at firm scale.
4. **Demand ROI measurement before scaling AI spend** — **[Evidence-Based]**. With 95 per cent of enterprise programmes returning no measurable value F2-11, the prudent position is that an unmeasured programme is a failed programme until demonstrated otherwise — the same measurement discipline this study applies to scam statistics.

5.3 For regulators (BNM, KESUMA, Ministry of Digital, police)

1. **Mandatory reconciliation of national fraud databases** — **[Evidence-Based] and the first priority.** The RM2.8 billion-vs-RM2.97 billion gap S5 S1 and the third series from the Ministry of Digital S6 must be reconciled in a single published national series with defined scope per component. This is not a research recommendation; it is a bookkeeping obligation whose absence makes every downstream decision unmeasurable. The study's Table 1 is offered as the starting schema.
2. **Publish the AI-attribution methodology** — **[Evidence-Based]** as a precondition for public trust. The 85-per-cent / RM2.11 billion attribution S9 must be accompanied by its coding rules (forensic vs. statement-based, borderline handling) or withdrawn from official communication. An unreproducible official number is worse than no number: it converts statistics into rhetoric and gives the alarmist and the sceptic equal license.
3. **Standardise the AI-fraud taxonomy (instrument vs. acquisition)** — **[Evidence-Based]** by the contradiction documented in Chapter 2 S8 S9. A three-layer classification — instrument / acquisition / perception — should become the national reporting standard, and the eightfold-type claims S11 should be published with absolute bases or not at all.
4. **Liability signals toward ad networks and platforms** — **[Plausible]**: the RM2.11 billion acquisition layer flows through advertising and messaging infrastructure S9; shifting defined liability (or a verification duty) toward the distribution layer is the coherent policy response. No Malaysian precedent exists; international comparison is required before design.

5. **Close the automated-decision gap in the PDPA and give the AI Governance Bill a fraud chapter** — **[Evidence-Based]** on the legal facts V3 V1: the law currently has no automated-decision-making provision, and the bill's consultation is the moment to add profiling-verification duties and a fraud-specific obligations chapter.
6. **Require outcome reporting for takedowns** — **[Evidence-Based]**: takedown counts (98,503 S10) must be paired with outcome measures (reported-loss trajectories by category) so that enforcement output and victim outcomes are reported as one system. The 55-per-cent-output/89-per-cent-loss divergence (Chapter 4) is the baseline this reporting would make visible.

CONCLUSION

THE AMPLIFIER, MEASURED

Malaysia's scam wave is not primarily an artificial-intelligence phenomenon — it is a social-engineering phenomenon whose marginal costs artificial intelligence has collapsed, whose reach it has industrialised, and whose measurement the state has failed to organise. The record supports alarm about the amplifier and precision about the numbers: RM2.97 billion lost S1, 66,204 victims counted S2, 454 deepfake instruments recorded S8, RM2.11 billion of recruitment attributed S9, and a population that is 75 per cent confident and 73 per cent deceived S14. Between the paranoia that reads every figure as AI-driven and the paralysis that reads none of them as actionable lies the only defensible position: measure the layers separately, reconcile the totals publicly, and defend the human being at the point of verification.

Scams and deepfakes are the externalised symptoms of a deeper structural friction: an economy adopting advanced AI tools faster than its workforce and governance structures build verified execution capability. The same calibration gap that lets a parent trust a cloned voice, and lets a board trust an unmeasured AI programme, is the defining vulnerability of Malaysia's AI transition.

The forthcoming winsym Research flagship study — *The Entry-Level Cliff & The Competency Trap* — examines that structural friction directly: why near-full employment (2.9 per cent A1-DS1) coexists with 66 per cent underemployment among young graduates G1, why 95 per cent of enterprise AI programmes return no measurable value F2-11, and why Malaysia's RM182.96 billion in approved data-centre investment carries a declared employment yield of 7,561 jobs F3-09. This study's final sentence is that study's first question: what, exactly, is Malaysia building the capability to do?

ANNEX

METHOD AND DERIVED ARITHMETIC

Evidence conventions. Every factual claim carries the source ID of the dossier row from which it is drawn (register below). Figures marked derived are simple arithmetic from published official figures (listed in this annex) and are labelled as such in the text. Claims marked secondary were not verified against a primary page in the underlying research and are flagged in the register. The qualitative forum register is presented as a seismograph of public discourse, not as statistical evidence Q-bank. Vendor-adjacent figures (GASA S16) are labelled at each point of use. All totals are as published; nothing has been converted between currencies. Research window: dossiers compiled and double-verified 11 October 2026 (see underlying dossiers for per-figure verification logs).

Derived arithmetic register (all from figures cited above):

1. Loss growth 2024→2025: RM1.40bn ÷ RM1.57bn ≈ +89% S1.
2. BNM-vs-IGP 2025 gap: RM2.97bn – RM2.8bn ≈ RM170m S1 S5.
3. National mean per recorded case 2025: RM2.97bn ÷ 66,204 ≈ RM44,900 S1 S2.
4. Mean per recorded deepfake case 2024: RM2.72m ÷ 454 ≈ RM6,000 S8.
5. Average victim loss vs median graduate salary: RM4,844.70 – RM4,521 ≈ RM324 S14 G2.
6. Takedown growth 2024→2025: 34,851 ÷ 63,652 ≈ +55% S10.
7. AI-cited share of announced US cuts 2025: 54,836 ÷ 1,206,374 ≈ 4.5% F2-07.

Explicitly excluded: the "140 scam attempts per person per year" survey claim could not be re-verified against its primary source in the underlying research (flagged as a weakness of S14 in the study-2 dossier) and is not used in this manuscript. The eightfold claim is used strictly as a directional signal with its missing denominator identified S11.

Primary/secondary flags in the register: all S-series financial totals verified on page; G4 (fresh-graduate salary reporting) secondary; S15/S11/S13 survey-and-news derived, verified on page at the source named; F2-11/F2-12 adopted from the Track B evidence base with the caveat that they describe non-Malaysian samples.

Source register

ID	PUBLISHER / AUTHOR	SOURCE
S1, S2, S3, S4	IGP Mohd Khalid Ismail (speech) / The Sun	thesun.my/news/malaysia-news/crime/online-scam-losses-hit-rm2-97-billion-in-2025-says-igp/
S5	Bank Negara Malaysia, Annual Report 2025, Box 4	bnm.gov.my/publications/ar2025/box4

ID	PUBLISHER / AUTHOR	SOURCE
S6, S7	Ministry of Digital	digital.gov.my (Garis Panduan AI release)
S8	CCID via Sinar Daily; OECD AI Incident Monitor	sinardaily.my/article/737264/... · oecd.ai/en/incidents/2024-08-28-d707
S9, S10, S12	CCID / Ministry of Digital via Sinar Daily	sinardaily.my/article/737264/...
S11	Newswav	newswav.com/article/deepfake-scams-targeting-malaysian-families-have-surged-eightfold-A2607_DBvld0
S13	The Vibes	thevibes.com/articles/news/113513/...
S14	National prevalence survey / The Star	thestar.com.my/news/nation/2025/09/03/most-malaysians-fall-victim-to-scammers-survey-shows
S15	Newswav	newswav.com/article/scams-now-a-national-crisis-...
S16	GASA "State of Scams in SEA 2025" (vendor-adjacent) / Asianews.network	asianews.network/singapore-scam-victims-top-southeast-asia-losses-at-us2132-per-person-malaysia-second
G1, G2	DOSM Graduates Statistics 2024	dosm.gov.my/uploads/release-content/file_20260401145140.pdf
G4	MOHE tracer study via CNA (secondary)	channelnewsasia.com (Malaysia underemployment coverage, May 2026)
V1	NAIO consultation via Baker McKenzie	bakermckenzie.com (Malaysia AI Governance Bill, Jul 2026)
V2, V3, V4	Pertama Partners; Ministry of Digital	pertamapartners.com (PDPA amendments & AI governance)
B1	Ministry of Education (KPM)	moe.gov.my (Buku Panduan Literasi AI, SPI Bil.2/2026)
B2	Malaysian Qualifications Agency via Bernama	bernama.com (MQA advisory note, 2023)
F2-07	Challenger, Gray & Christmas	challengergray.com/blog/2025-year-end-challenger-report-...
F2-11	MIT NANDA	report text via MIT Sloan Q (mlq.ai)
F2-12	McKinsey State of AI 2025 (via secondary commentary)	linkedin.com/pulse/ai-value-crisis-2025-2026-ben-carroll-c7w4e
A1-DS1	DOSM Labour Force Statistics Dec 2025	dosm.gov.my/portal-main/release-content/labour-force-statistics-december-2025
F3-09	Johor State / Bernama / TNB via The Sun & The Star	thesun.my (Johor data centre approvals)
Q3	r/malaysians [job-scam thread]	full thread register in the underlying study-2 dossier §8

RESEARCH APPENDIX

FROM THE DOSSIER

Four working sections of the underlying research file that the manuscript draws on: the myths, the timeline, the scenarios found in the literature and the list of what nobody measures.

Ten myths, checked

1. **"Malaysians lost RM2.97 billion to AI scams in 2025."** The RM2.97bn is *total* online-scam losses; AI-specific *recorded* losses are 454 cases / RM2.72m in 2024 S8. Even the RM2.11bn "AI-video" figure is attribution, not verified mechanism S9. The honest headline is: **AI's recorded share is tiny; AI's suspected share is huge; nobody measures the gap** — that gap is the study.
2. **"BNM and the police contradict each other (RM2.8bn vs RM2.97bn) — one must be wrong."** They measure different things: bank-reported fraud impact vs police-recorded complaints. Both are official; neither is reconciled in public S5 vs S1. The fact that Malaysia cannot reconcile its own numbers is a finding, not a typo.
3. **"Deepfake cases surged eightfold — AI fraud is exploding."** The eightfold claim has no published denominator S11. It may be a small-base effect (e.g. 10→80 cases). Pair it always with the 454-case baseline S8.
4. **"85% of scam victims were fooled by AI videos" (the most-quoted 2025 stat).** Police attribution from victim accounts; the method is not published; it sits awkwardly beside the tiny recorded deepfake count. It should be cited as an *estimate by enforcement*, not a measured fact S9 weakness.
5. **"Takedowns are winning the war (98,503 takedowns!)."** Takedowns rose ~55% (63,652→98,503) in the same year losses rose ~89% S1/S10. Enforcement output is not prevention outcome.
6. **"Malaysians are naive; education will fix this."** 75% of adults believe they can spot scams; 73% were deceived anyway S14. Overconfidence is the vulnerability — "awareness campaigns" without this insight repeat the error.
7. **"GASA data is official."** GASA is an industry alliance (report co-produced with the vendor Gogolook); per-capita losses are modelled, not reported S16. Useful, never official; always labelled VENDOR-ADJACENT in this dossier.
8. **"The AI Governance Bill will stop AI scams."** The bill is a general governance framework, only in public consultation since 10 Jul 2026 V1; it will not retrofit fraud detection, and the PDPA still lacks automated-decision provisions V3. Scams evolve faster than statute.
9. **"Voice cloning is brand-new."** Voice-mimicry fraud predates generative AI; AI only industrialised it S12 cases are small sums. The *cost curve* is the news, not the existence.

10. **"Only the elderly fall for AI scams."** Millennials have the highest average losses (RM7,792/yr — S15); Gen Z is deceived at scale [S14]; the biggest loss category (investment scams, RM1.47bn — S3) targets the working-age. "It won't happen to me" is the demographic myth the syndicates exploit.

Timeline

DATE	EVENT	SOURCE ID
28 Aug 2024	OECD records a Malaysian deepfake fraud incident	S8
2024	CCID records 454 deepfake cases / RM2.72m; investment-scam AI-video attribution RM2.11bn / 13,956 cases / ~85%	S8/S9
2024	Official losses 35,368 cases / RM1.58bn (+2.53% vs 2023); BNM SEFT compensation framework introduced	S6; BNM box
2025	12,110 cases / RM573.7m in Q1 alone	S7
2025	PDPA amendments phase in; DPO obligation effective June 2025	V2
May 2025	Selangor voice-clone case (RM5,000)	S12
Jun 2025	JPDP guidelines (DPIA, PbD, ADM) announced for Jan–Mar 2026	V3
Sep 2025	National survey: 85% exposed / 73% deceived / RM4,844.70 average	S14
2025	Full-year: RM2.97bn / 66,204 cases (IGP); RM2.8bn (BNM); takedowns 98,503	S1/S5/S10
2026 (H1)	Deepfake/voice-clone cases "eightfold" vs H1 2025 (media-reported)	S11
Jun 2026	Deepfake video of former minister Rafidah Aziz circulates; privacy complaint filed (The Star, Jun 2026 — coverage listed in notes/childC-candidates.md)	timeline/news
10 Jul 2026	AI Governance Bill public consultation opens (NAIO)	V1
2026	Budget 2026 measures active: MSME AI/cyber training +50% tax deduction; HRD Corp RM3bn training fund	Study 1 A1-GV5
Announced milestones	AI Governance Act (post-consultation); JPDP ADM/DPIA guidelines (early 2026); KPM AI-literacy guidebook in schools (SPI Bil.2/2026 — B1)	V1/V3/B1

Three scenarios, as found in the literature and data. Not a forecast.

Scenario A — Slow / "AI is a rounding error in the scam wave" (assumptions: AI-specific cases stay a small recorded share; takedowns keep scaling; banks' fraud models + compensation contain losses). Evidence: recorded deepfake losses RM2.72m vs RM1.57bn total (0.09%) S8; 98,503 takedowns S10; SEFT compensation (BNM box). Under this reading the story is *the same old scams with new wrapping* and the public answer is: protect yourself exactly as before.

Scenario B — Medium / "AI multiplies the existing scam machine" (assumptions: AI lowers cost of personalisation; the 85%-of-investment-victims attribution S9 is roughly right; losses keep compounding as in 2024→2025 (+89%)). Evidence: RM1.57→2.97bn trajectory S1; RM2.11bn AI-video attribution S9; 32% prevalence and US$1,035/capita S16; eightfold voice-clone surge S11. Under this reading, AI is the *amplifier*, and the answer is systemic (platform liability, detection, victim compensation).

Scenario C — Fast / "deepfake trust collapse" (assumptions: video/voice verification becomes impossible for laypeople; attack cost → near zero via open-source tools; regulators move slowly (AI Act not before 2027-28)). Evidence: global USD442bn S5; GASA growth projections; the overconfidence gap (75% think they'd spot it — S14); PDPA's missing ADM provision and unpassed AI Act V1/V3. The counterweight is the same as Scenario A: even in the worst case, *recorded* AI-specific losses today are tiny S8 — the scenario is capability-driven, not yet data-driven.

Known, contested, not measured

Known: recorded losses and cases (RM2.97bn/66,204; RM1.57-1.58bn/35,368 baseline) S1/S2/S6; loss structure (investment scams largest — S3; phone scams most frequent — S4); takedown scale S10; prevalence data (85%/73%/RM4,844.70 — S14); regional ranking (US$1,035, 32% — S16); protection rules (SEFT; PDPA DPO; AI bill consultation) V1-V3; graduate-pipeline vulnerability (32.2%/66% — G1).

Contested: how much of the wave is AI — police say RM2.11bn/13,956/~85% of investment-scam losses via AI videos S9 while *recorded* AI-specific losses are RM2.72m S8; BNM's RM2.8bn vs IGP's RM2.97bn for the same year S5 vs S1; whether takedowns help (losses +89% while takedowns +55% — S1/S10); whether voice-clone cases "surged eightfold" (no baseline — S11); whether GASA per-capita figures overstate (modelled vendor data — S16). **Not measured (no Malaysian data exists):**

1. No official breakdown of scam losses by AI technique (deepfake vs voice clone vs AI-text vs none).
2. No published methodology for the "85% of investment-scam victims came via AI videos" claim (police attribution, not forensic estimate).
3. No average or median loss per AI-specific case (only individual case examples — S12).
4. No scam-victim demographics from official data (age/income tables don't exist; only GASA survey).
5. No evaluation of takedown effectiveness; no platform-level loss attribution (which apps/ad networks carry the RM2.11bn?).
6. No recovery-rate statistics from banks (only the 13% survey number — S13).
7. No pre-2024 AI-scam baseline to measure the "eightfold" surge against S11.
8. No measure of scam *attempt* volume (only reported cases; the "140 attempts/person" survey claim could not be re-verified — excluded from fact table).
9. No legal analysis of liability chains (platforms, telcos, banks) under Malaysian law — the AI Governance Bill's scope is unknown (consultation only, Jul 2026 — V1).