The Permanent Exposure

Why the quantum threat to blockchains is a public-key problem that already happened — and what AI does to the clock

A position paper on quantum, AI, and the limits of “be your own bank”

Draft 4.0 • Core edition • 14 August 2026

AI-assisted (Claude Fable 5); the human author sets the stance and owns every claim.

Abstract. Public-key cryptography secures every major blockchain, and a sufficiently powerful quantum computer would break the elliptic-curve signatures (ECDSA over secp256k1) that stand between a wallet and its coins. The usual telling of this story — ‘quantum will hack the blockchain’ — is both too dramatic and too vague. This paper makes a narrower, sharper claim. On a public ledger, the secret a quantum attacker needs is the *public key*, and every coin that has ever been spent has already published it, permanently, in the clear. ‘Harvest now, decrypt later’ — the practice of storing ciphertext today to break it once the hardware exists — normally requires an interception step and is bounded by how long data stays secret. For a blockchain there is no interception step and no time bound: the harvest is already complete. Roughly a quarter to a third of all Bitcoin — between roughly $256 billion and $446 billion at today’s price — sits on quantum-exposed keys right now, including Satoshi’s ~$70 billion (of some $109 billion in lost coins) that can never be moved to safety. **Ethereum, where the NFT world lives, is different — more exposed, but more defensible.** Its account model reuses one address, and any on-chain action — a purchase, a transfer, an approval, paying gas — reveals the key permanently, so an estimated 55–60% of all ETH (~$134 billion) is already exposed, and essentially every *active* collector’s wallet with it. (One exposed key takes the whole vault — every NFT, token, and coin at that address.) The offset is upgradeability: Ethereum’s permanently-lost ‘dormant’ core is about 0.1% of supply against Bitcoin’s ~8.5%, so most of the exposed ETH is defensible by a future migration — if owners act. A deliberately conservative, floor-conditional order-of-magnitude bound puts on the order of **a hundred** exposed NFT vaults worth over $1M and **several hundred to a thousand** worth over $200k; and because in fine art the token carries the authenticated original and its provenance, a key-break severs that value even though a copy of the image survives. Expert surveys put a quantum computer capable of breaking RSA-2048 at 28–49% within ten years, the highest on record; a March 2026 Google Quantum AI research paper (an arXiv preprint) estimates breaking the 256-bit curve needs fewer than 500,000 physical qubits, and fault-tolerant machines are targeted for 2029 — putting the late 2020s in the credible window. A deliberately simple migration-race model — meant to structure intuition, not to forecast — finds that even a well-coordinated upgrade would leave on the order of 12% of all Bitcoin still exposed when the machine arrives, most of it the unmovable core of lost coins. Meanwhile AI is doing to the classical attack surface what quantum threatens to do to the cryptographic one: autonomous agents already exploit around half of vulnerable smart contracts in tests, at roughly a tenth the cost of defending them. We separate what is demonstrated from what is speculative throughout, price everything at Bitcoin and Ether’s actual levels rather than the inflated figures in secondary coverage, and end with what an owner, a builder, and a policymaker can do before the hardware is ready.

This is a position / working paper, not peer-reviewed research. It argues a thesis and declares that it does so. Figures and claims are sourced; where a number is disputed, the range is given; where a claim is speculative, it is labelled.

Contents

Part I — The claim, stated precisely

1. The one secret, published forever
2. Why a blockchain is the worst case for ‘harvest now, decrypt later’

Part II — How much, and how soon

3. How much Bitcoin is actually exposed (and what it’s worth)
4. The clock: expert timelines and Mosca’s inequality
5. The migration race — and why even winning it leaves a remainder

Part III — Ethereum, NFTs, and the vaults at risk

6. ‘Be your own bank’ and the custody illusion
7. Ethereum is the worse case — and NFT collectors are the most exposed of all
8. How many vaults are actually at risk
9. The fine-art vault: what a key-break really takes
10. Three ways custody fails — and which one is new

Part IV — The accelerant

11. Patchageddon: AI on immutable, money-holding code
12. The state adversary who already harvests

Part V — What to do before the hardware is ready

13. For owners, builders, and policymakers
14. What would change my mind

Part I — The claim, stated precisely

1. The one secret, published forever

Start with what a Bitcoin or Ethereum wallet actually is. It is a pair of numbers: a private key you keep, and a public key derived from it that you can safely show the world. The mathematics that makes this safe — the elliptic-curve discrete logarithm problem — is easy to run forward (private key to public key) and, for a classical computer, effectively impossible to run backward. Every signature you make proves you hold the private key without revealing it. This is the machinery behind 'your keys, your coins.' It is also, in 1994, exactly what Peter Shor showed a quantum computer can run backward. A large enough quantum computer takes a public key and returns the private key that generated it.

The load-bearing observation of this paper is about *when* your public key becomes visible. On Bitcoin, funds usually sit behind a *hash* of the public key, not the key itself — and a hash is quantum-safe. But the moment you *spend* from an address, the transaction publishes the full public key so the network can verify your signature. From that block onward, the key is on the ledger forever, visible to anyone, in every copy of the chain on Earth. Address reuse, the Taproot format (which publishes the key by design), and the earliest 'pay-to-public-key' coins from 2009 all leave the key exposed permanently. Aggarwal and colleagues made the core point as early as 2017: it is the *signature scheme*, not the proof-of-work that secures mining, that quantum breaks.

So the vulnerability is not abstract and it is not future. If your wallet has ever sent a transaction, your public key is already published. A quantum computer does not need to break into anything, intercept anything, or catch you at the wrong moment. It needs only to read the public ledger — which is the one thing a public ledger is designed to let everyone do — and then do the math. The secret is already out. Only the machine that reads it is missing.

2. Why a blockchain is the worst case for 'harvest now, decrypt later'

'Harvest now, decrypt later' (HNDL) is the intelligence practice of vacuuming up encrypted data today and storing it until a quantum computer can decrypt it. The U.S. security agencies have warned about it since 2023; a 2025 Federal Reserve working paper describes it plainly: adversaries 'can download or intercept encrypted information today, store it, and later use a sufficiently powerful quantum computer to reveal its contents.' For ordinary traffic — an encrypted email, a TLS session — HNDL has two natural brakes. The attacker must *capture* the data (tap a cable, sit on a network), and the value decays: most secrets stop mattering after a few years.

A public blockchain removes both brakes. There is nothing to capture, because the data is already public — the whole design goal is that anyone can download the entire history. And nothing decays, because coins do not expire: a public key exposed in 2013 still guards spendable money in 2026 and will in 2036. The same Federal Reserve paper notes that even after a post-quantum upgrade, 'transactions that have already occurred would remain vulnerable.' For a blockchain, in other words, the harvest is not a future risk to be managed. It is a completed fact. Billions of dollars of public keys have already been collected, by everyone, and cannot be un-published.

It is worth being precise about the analogy, because it is imperfect in a way that cuts *against* comfort, not for it. Strictly, this is not HNDL: there is no ciphertext to intercept and nothing to decrypt — only a public key and

a future key-derivation via Shor’s algorithm. But that is exactly why the blockchain case is worse than the one the agencies warn about. Classic HNDL still needs the attacker to do the harvesting and still races the clock of data that eventually stops mattering. Here the harvesting is done for the attacker, by the protocol, and the data never stops mattering. The premises that make HNDL frightening are not merely met; they are met maximally.

Figure 1. 'Harvest now, decrypt later' is already complete on a public blockchain

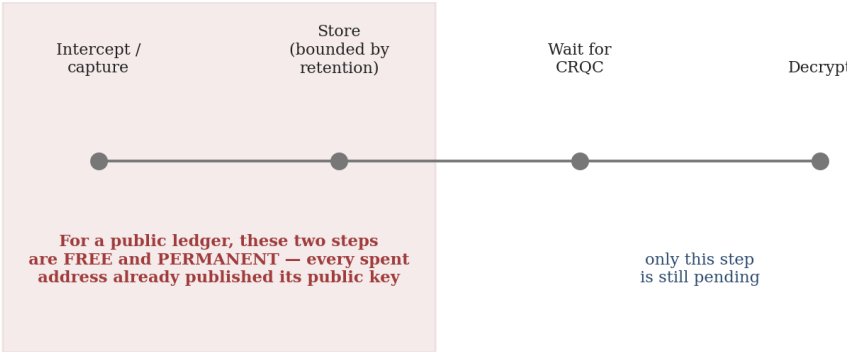

Figure 1. For a public ledger, the harvest is already complete.

Ordinary HNDL requires an interception step and is bounded by how long the data stays secret. A public blockchain has neither brake: every spent address already published its public key, permanently and to everyone. Only the decryption hardware is still pending.

Source: authors’ schematic; CISA/NSA/NIST (2023); Federal Reserve, Mascelli & Rodden (2025).

Part II — How much, and how soon

3. How much Bitcoin is actually exposed (and what it's worth)

Estimates of the exposed pool cluster in a striking band — between a quarter and a third of all Bitcoin — though it is a band, not a consensus: the four studies below count somewhat different things (see Appendix A) and span 4.0 to 7.0 million coins. Deloitte's well-known analysis puts it around a quarter of supply (roughly two million in the oldest pay-to-public-key coins and 2.5 million more in reused addresses whose keys are already public, measured against the smaller supply of its day). A 2026 Glassnode study puts it at 6.04 million BTC — 30.2% of today's supply — splitting it into a 'structural' core (old formats and Taproot) and a larger 'operational' layer created by address reuse. Chaincode Labs and Ark/Unchained land in the same band, at 31–35%.

One correction matters before we attach a dollar sign. Much of the recent quantum coverage quietly prices these coins at bull-market levels ($78,000–$100,000+), producing headline numbers near half a trillion or higher. This paper prices everything at Bitcoin's actual level on 12 August 2026 — $64,085 — and applies it consistently. At that price the exposed pool is worth between **$256 billion** (Deloitte's ~4M coins) and **$446 billion** (Ark's 35%). The number is smaller than the scare headlines and still enormous — larger than the market capitalisation of most companies on Earth.

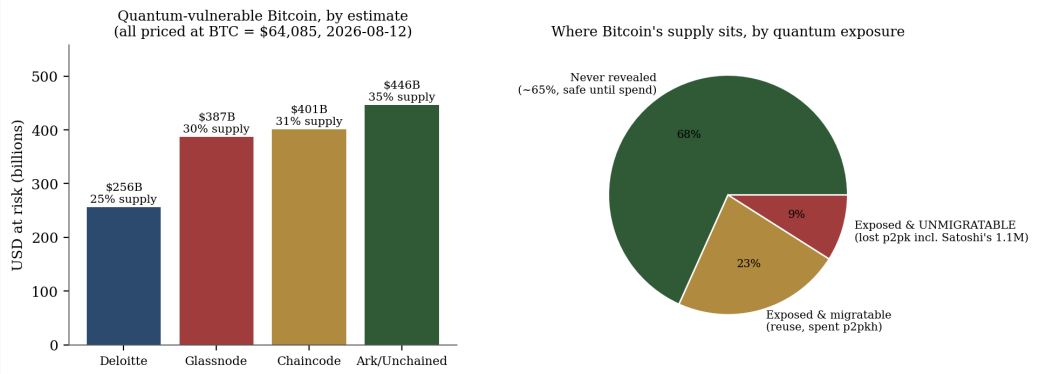

Figure 2. Quantum-vulnerable Bitcoin, honestly priced.

Left: four independent estimates of the exposed pool, each converted at the same $64,085 price rather than the inflated figures in secondary coverage. Right: how the supply splits — about two-thirds has never revealed a public key and is safe until it spends; the rest is exposed, and a hard core of that (lost pay-to-public-key coins, ~1.7M in all, including Satoshi's ~1.1M) can never be migrated at all.

Sources: Deloitte; Glassnode (2026-05-21); Chaincode Labs (2025); Ark/Unchained (2026); price Fortune (2026-08-12). Model: exposure_result_v4.json.

The right-hand panel carries the part people miss. About 65% of Bitcoin sits in addresses that have never revealed a public key — they are safe until the day they first spend. The exposed remainder is the near-term battlefield, and within it is a hard core that can *never* be defended: coins in the oldest format whose owners are gone. The total unmovable core is on the order of 1.7 million coins (about **$109 billion**); within it, the roughly 1.1 million coins attributed to Satoshi — a figure that is itself disputed — are worth about **$70 billion** and are the largest single block. No living person can move these to a safe address, because moving them requires the very signature that would expose them, and no one holds the key. They will sit in the open until either a quantum computer takes them or the community votes to freeze them. Both options are, in their own way, the

end of a promise.

4. The clock: expert timelines and Mosca’s inequality

How soon is this? Nobody knows, and anyone who claims a date is selling something. What we have is the best available proxy: the Global Risk Institute’s annual survey of quantum experts, which asks how likely it is that a quantum computer will break RSA-2048 within various horizons. Two caveats belong up front. First, the survey asks specifically about breaking *RSA-2048* in 24 hours, which is a proxy, not the Bitcoin question; breaking Bitcoin’s elliptic-curve signatures is a distinct resource problem, and by most recent estimates a somewhat *easier* one, so an RSA clock likely under-states the Bitcoin risk rather than over-stating it. Second, these are expert opinions, not measurements. With that stated: the 2025 survey, published in March 2026 and covering 26 specialists, put the ten-year probability at **28–49%** — the highest ten-year reading in the report’s seven-year history. Half the respondents put it at roughly even odds or better within a decade; more than two-thirds see even odds or better at fifteen years, and over 90% at twenty. The central ten-year reading rose from the low-20s to the high-30s in a single year. The clock is not merely ticking; the experts moved it forward.

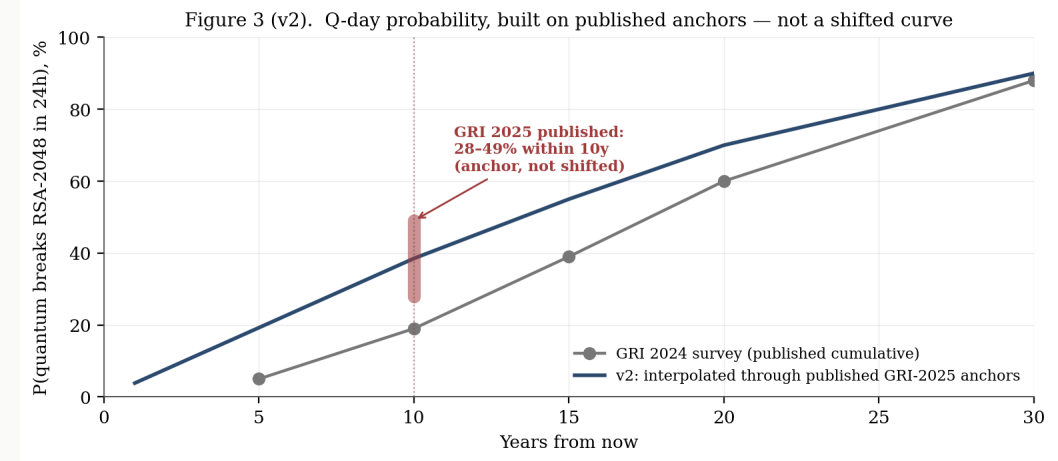

Figure 3. Q-day probability, built on published anchors — not a shifted curve.

The grey curve is the 2024 expert survey (published cumulative). The blue curve interpolates the migration model’s hazard through the *published* GRI-2025 anchors; the red bar marks the 2025 published range of 28–49% within ten years. Under this construction the near-term probability stays modest (about 4% within a year, 19% within five) — no fabricated front-loading. These are opinions about RSA-2048, used as a proxy; treat them as a trend, not a measurement.

Sources: GRI/evolutionQ Quantum Threat Timeline 2024 & 2025 (Mosca & Piani). Model: exposure_result_v4.json.

How firm are the underlying numbers? Firmer than a year ago, and moving the wrong way for holders. The resource estimates for breaking Bitcoin’s and Ethereum’s curve (both use 256-bit elliptic-curve signatures) have collapsed. A peer-reviewed 2022 study (Webber et al.) put it at 13–300 million *physical* qubits; in March 2026 a team from Google Quantum AI, the Ethereum Foundation, and Stanford estimated that breaking the 256-bit curve needs **no more than about 1,200 logical qubits and fewer than 500,000 physical qubits**, running in roughly eighteen minutes — a roughly twenty-fold reduction against the previous best resource estimate (Litinski 2023, ~9 million physical qubits) in three years. This is a primary research estimate (an arXiv preprint, not yet peer-reviewed), not secondary coverage. Two caveats keep it honest: it is a *resource estimate*, not a machine — today’s hardware has on the order of hundreds of physical qubits, not hundreds of thousands — and the gap between ‘we know how many qubits it would take’ and ‘someone has built them’ is exactly what the

roadmaps and the expert surveys are trying to time. Set it beside the hardware roadmaps — IBM has publicly targeted a fault-tolerant machine (‘Starling’, 200 logical qubits) for 2029 — and the late 2020s stop looking distant. To be careful about dates: no credible source names 2028 specifically, and this paper does not; the honest near-term anchors are IBM’s 2029 target and the Global Risk Institute’s roughly one-in-four by about 2030. But 2028 sits on the early edge of that window rather than outside it, and the resource estimates keep pulling the edge closer.

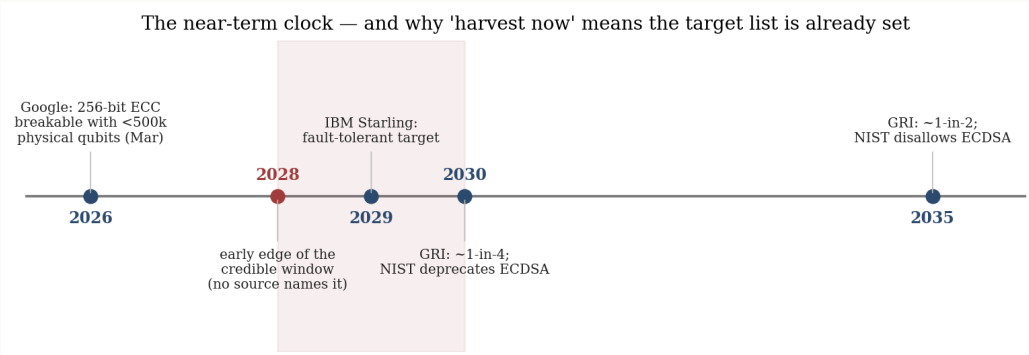

Figure 4. The near-term clock — and why waiting does not help.

The credible near-term markers: Google’s 2026 sub-500k-qubit estimate, IBM’s 2029 fault-tolerant target, and the Global Risk Institute’s ~1-in-4-by-2030 / ~1-in-2-by-2035 readings, against NIST’s deprecate-2030 / disallow-2035 schedule. 2028 is marked as the early edge of the window, not a sourced date. Because every exposed key is already harvestable, the keys do not get safer as the window narrows — the wait simply ends.

Sources: Google Quantum AI/EF/Stanford (2026-03); IBM (2025-06); GRI (2025); NIST IR 8547.

The reason a probability, not a date, is still enough to act on is Mosca’s inequality. Let X be how long your data must stay secret, Y how long it takes to migrate your systems to quantum-safe cryptography, and Z the time until a capable machine exists. If X + Y > Z, you are already too late — the machine arrives before you finish moving, and anything that had to stay secret past that point is lost. Here the inequality is trivially satisfied, and that triviality *is* the finding: for a blockchain X is effectively infinite (coins never expire), so X + Y exceeds any finite Z automatically. There is no value of the quantum timeline at which a permanently-exposed key becomes safe to leave exposed. A modest Z — a roughly one-in-three chance in a decade — is a live emergency not because the number is large but because the other two terms make the comparison a foregone conclusion.

5. The migration race — and why even winning it leaves a remainder

Suppose the community does the right thing and builds a quantum-safe address format — the work is already under way (Section 13). Migrating is still not a software update you push overnight. Each owner must actively move coins from an exposed address to a safe one, which means every dormant wallet, every lost key, every holder who isn't paying attention stays exposed. To see what that implies, we model the migration as an S-curve — slow, then fast, then a long tail — under three scenarios, and race it against the published Q-day anchors from Section 4. This is a toy model, built to structure intuition rather than to forecast; its value is in decomposing the residual, not in any single percentage.

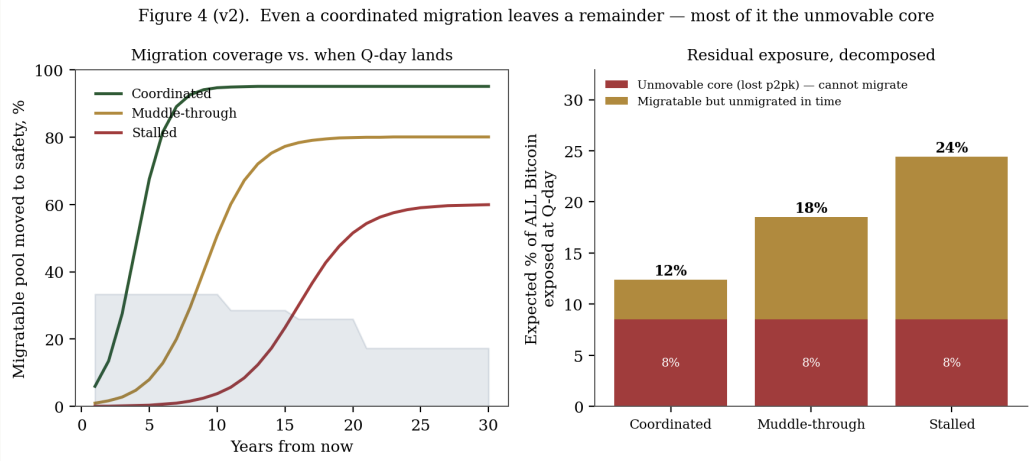

Figure 5. Even a coordinated migration leaves a remainder — most of it the unmovable core.

Left: three migration scenarios (how fast the movable exposed coins reach safety) against the year-by-year Q-day probability (shaded). Right: the expected share of *all* Bitcoin exposed at Q-day, split into the unmovable core of lost coins (red, which no migration can rescue) and the migratable-but-unmigrated increment (gold, the only part migration speed actually changes).

Model: exposure_result_v4.json (Q-day hazard on published GRI-2025 anchors; unmovable core ~1.7M BTC from Ark/Unchained). Scenarios are illustrative, not forecasts.

The decomposition is the honest part. Roughly **8% of all Bitcoin** is exposed at Q-day in *every* scenario, because that is the unmovable core of lost coins and no migration can touch it. On top of that floor sits the part migration speed actually governs: a coordinated push — half the movable pool secured within four years — adds only a few points, for a total near **12%**; a realistic muddle-through leaves about **18%** (roughly 17–20% across the low and high Q-day anchors); a stalled migration leaves around **24%**. The lesson is not that migration is pointless — it clearly shrinks the gold band — but that migration *alone* cannot get the number near zero, because the red floor is fixed by the ledger's own permanence. Getting past that floor requires freezing coins, a political act the likes of which Bitcoin has never attempted.

A caution on the model, stated plainly. The migration curves are invented scenarios, the Q-day anchors are opinion surveys about a proxy problem, and the exact percentages should not be read as forecasts. What is robust is the *shape* of the result: a fixed floor set by lost coins, plus a modest increment set by migration speed, giving a double-digit residual across every reasonable assumption. Appendix A gives the arithmetic so a reader can substitute their own beliefs.

Part III — Ethereum, NFTs, and the vaults at risk

6. ‘Be your own bank’ and the custody illusion

‘Be your own bank’ is the moral centre of self-custody: no counterparty, no permission, no one who can freeze you. It is a real freedom and this paper does not sneer at it. But it comes with a property people rarely price in — the same one from Section 1. A bank can rotate your account number after a breach; a blockchain cannot un-publish your public key. Every transaction you sign to *use* your money is also the act that exposes you. Self-sovereignty and permanent exposure are the same coin.

The intuition that a big custodian must be safer turns out to be backwards in at least one important way. Glassnode’s 2026 analysis found that some large exchanges hold the vast majority of their coins on already-exposed keys — on the order of 85% at one, essentially all at another — because heavy transaction volume means constant key exposure, while one well-run custodian kept exposure near 5% through disciplined address hygiene. A careful self-custodian who never reuses an address can be *less* quantum-exposed than a household-name exchange. Custody quality, not custody model, is what determines exposure. The uncomfortable corollary: when Q-day approaches, the coins on the most-used, most-liquid, most-institutional addresses may be among the most exposed.

7. Ethereum is more exposed — and NFT collectors are among the most exposed of all

Everything so far has used Bitcoin for its numbers, because Bitcoin is where the exposure data is cleanest. But the NFT world does not live on Bitcoin; it lives on Ethereum, and Ethereum reveals far more keys. The reason is the account model: where Bitcoin can hand out a fresh address per payment and let a careful holder keep funds behind never-spent, pubkey-hidden addresses, an Ethereum account is one long-lived address you reuse for everything. The Ethereum Foundation’s own documentation states the consequence plainly — ‘for any account that has sent a transaction, the public key is exposed onchain’ — and once it is exposed, it stays exposed. Any on-chain action does it: buying, selling, or transferring an NFT, approving a marketplace, or simply paying gas.

One clarification matters, because it is easy to overstate. Signing an *off-chain* message — an OpenSea or Blur offer, a token permit, a ‘sign in with Ethereum’ challenge — also mathematically reveals your public key, as Project Eleven notes, but only to the *counterparty* you signed for, not to the permanent public ledger. That is a real leak, not the same permanent, everyone-can-see-it exposure that an on-chain transaction creates. The reason it barely softens the conclusion is that there is essentially no way to be an *active* collector without also having done the on-chain actions — you cannot buy, sell, or move a piece without one — so an active wallet’s key is almost always already permanently public. On the aggregate numbers a 2026 analysis (‘Quantum Horizon’) puts the exposed share at 55–60% of all Ether on the most defensible reading — worth on the order of $134 billion — with an earlier Deloitte study putting it above 65%; either way, roughly double Bitcoin’s ~30%. And the account model means the damage per break is total: one derived key takes *everything* at that address — the ETH, every token, every NFT — in a single signed transaction.

But ‘more exposed’ is not the whole story, and it would be dishonest to stop there. Ethereum is also more *defensible* than Bitcoin, for the same reason it is more exposed: it is upgradeable. The slice of ETH that is permanently unrecoverable — the honest analog to Satoshi’s lost coins, funds whose owners are gone and can never act — is, by the Ethereum Foundation’s own estimate, about 0.1% of supply, against the roughly 8.5%

unmovable core on Bitcoin. The other ~55% of exposed ETH is held by owners who *can* still act, on a chain that *can* still change its signature rules. So the accurate framing is not ‘Ethereum is simply worse’ but the sharper pair: Ethereum exposes far more keys, and it retains far more ability to save them — if the owners and the protocol move in time.

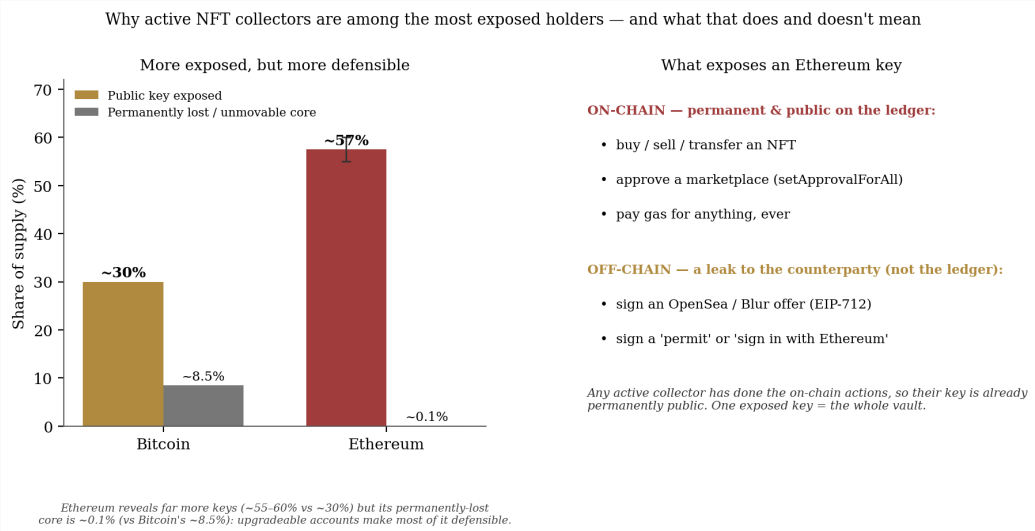

Figure 6. Ethereum: more exposed, but more defensible.

Left: Ethereum’s exposed share (~55–60%) is roughly double Bitcoin’s, but its permanently-lost ‘dormant’ core is ~0.1% against Bitcoin’s ~8.5% — most exposed ETH is still defensible by migration. Right: on-chain actions expose your key permanently and publicly; off-chain signatures are a lesser, counterparty-only leak. Any active collector has done the on-chain actions.

Sources: ethereum.org (2026, incl. the ~0.1% dormant figure); Project Eleven (2025-08); Gershteyn & Alber, arXiv:2606.14484 (2026); Deloitte (~2021, earlier corroboration).

There is a real mitigation on the Ethereum side, and it is worth knowing. Because Ethereum is upgradeable and supports smart-contract accounts, a wallet can in principle switch to a quantum-safe signature scheme in a way a plain Bitcoin address cannot. The 2025 Pectra upgrade (EIP-7702) lets an ordinary account delegate to smart-contract code, and a proposed native account-abstraction step would let each account choose its own signature verification — so users could move to post-quantum signatures without waiting for a single chain-wide migration. Vitalik Buterin has sketched a ‘recovery fork’ that would revert the blocks in which mass theft becomes visible and let honest users prove ownership with a STARK proof *without revealing the old key*. This is genuinely better than Bitcoin’s position. But be clear about what it is and is not. It is an *emergency* response to visible *mass* theft — a contentious, uncertain hard fork that the community would have to agree to execute in a crisis, not a personal undo button. It would do nothing for a quiet, single-vault theft that never rises to a network-wide emergency; the underlying exposed key stays valid until the owner migrates; most collectors have taken no such step; and a migration transaction is itself a key-exposing broadcast a fast-enough attacker could watch and front-run. Better odds are not safety, and an emergency backstop is not a lock on your own door.

8. How many vaults are actually at risk

This is the question a collector actually asks: not ‘what fraction of a chain’ but ‘how many wallets like mine, holding real value, are sitting on exposed keys.’ It deserves an honest answer, and honesty starts with a caveat that is itself part of the finding: *no source publishes a count of wallets by dollar value*. So what follows is a deliberately rough, order-of-magnitude bound, and we show its seams. Two things are sourced. Analysts at CoinGecko identified at least twenty NFT wallets each holding $4.3–19.2 million in 2023; and CryptoPunks has on the order of three to four thousand distinct holders — a convenient lower anchor, though a rough one, because it uses one collection’s floor as a stand-in and so *undercounts* every wallet whose value is in other collections. Draw the simplest heavy-tailed curve between those two anchors and read off the tiers. Because the CryptoPunks floor is volatile — about $60,000 at today’s Ether price, but $100,000 or more when Ether was higher in 2025 — we run it under both floors and report the span, rather than pretending to a single number.

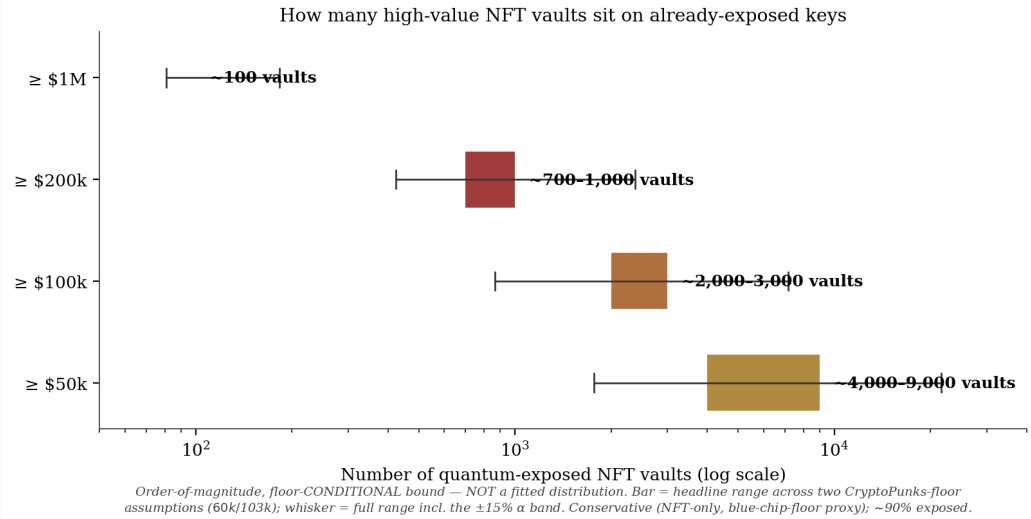

Figure 7. How many high-value NFT vaults sit on already-exposed keys.

An order-of-magnitude, floor-conditional bound — not a fitted distribution. The bar is the headline range across two CryptoPunks-floor assumptions ($60k / $103k); the whisker adds the ±15% tail-exponent band. It is conservative (NFT-only, using a blue-chip floor as a proxy that undercounts non-Punk value). The ~90% exposure multiplier — that active Ethereum wallets are almost all already exposed — is the well-sourced part.

Model: exposure_result_v4.json. Anchors: CoinGecko Research (2023-09); CryptoPunks holder data / floor (Jul 2025 & 2026).

Read as a range, the bound says something like this: on the order of **a hundred NFT vaults hold over $1 million** (roughly 81–183 across the assumptions), **several hundred to a thousand hold over $200,000** (about 424–2379), a few thousand hold over $100,000, and something in the mid-thousands to five figures hold over $50,000 — and, applying the ~90% exposure rate that follows from Section 7, nearly all of them sit on keys that are already public. The honesties: this is a bound, not a count, because the data to do better does not exist; it is *conservative*, because it uses NFT value only and a single-collection floor, so it ignores the ETH and tokens in those same vaults and every collection beyond the anchors; and it is *uncertain to a factor of several*, which is why it is stated as a range and rounded to one figure. What is not uncertain is the shape of the answer: this is not a handful of billionaires’ wallets, it is on the order of thousands of serious collectors, each holding a five-, six-, or seven-figure vault on an exposed key — and the fact that nobody has measured it properly is itself why the risk goes unmanaged.

It also helps to see why even a ‘safe’ vault holds a real, seizable number rather than a paper one. A single piece’s worth is bracketed by three simultaneous market signals: the highest standing bid or collection offer, which is what a buyer will pay right now (say $25,000); the last comparable sale, the ‘like-market’ value (say $100,000); and the owner’s asking price (say $200,000). A vault is therefore not a guess — it is a live, liquid range with a floor a buyer will execute against today. Blur’s bidding incentives went so far as to push standing bids *above* buy-now prices on some collections, which tells you how deep and real the bid stack under a blue-chip vault can be. A quantum thief does not steal a notional valuation; they steal into a market that is standing ready to pay.

9. The fine-art vault: what a key-break really takes

The NFT market created, and then mostly un-created, an extraordinary amount of paper wealth. Whole-market volume peaked near $25 billion a year in 2021 and 2022 and has fallen to under a billion; the fine-art segment specifically — the one-of-one and generative work on SuperRare, Foundation, Art Blocks and Nifty Gateway, as opposed to large-supply profile-picture collections — collapsed from $2.9 billion in 2021 to $197 million in 2024, a roughly 93% fall on an annual basis (thinning to $23.8 million in the first quarter of 2025), with active art traders down 96%. A widely-cited 2023 study found 95% of surveyed collections effectively worthless, and on-chain analysis found 58% of 2022 Ethereum NFT trading was wash trading. So the first honest answer to ‘how much NFT wealth is at risk’ is that most of the notional wealth already evaporated on its own. What remains — the canon — is real, concentrated, and precisely the thing a quantum key-break would take.

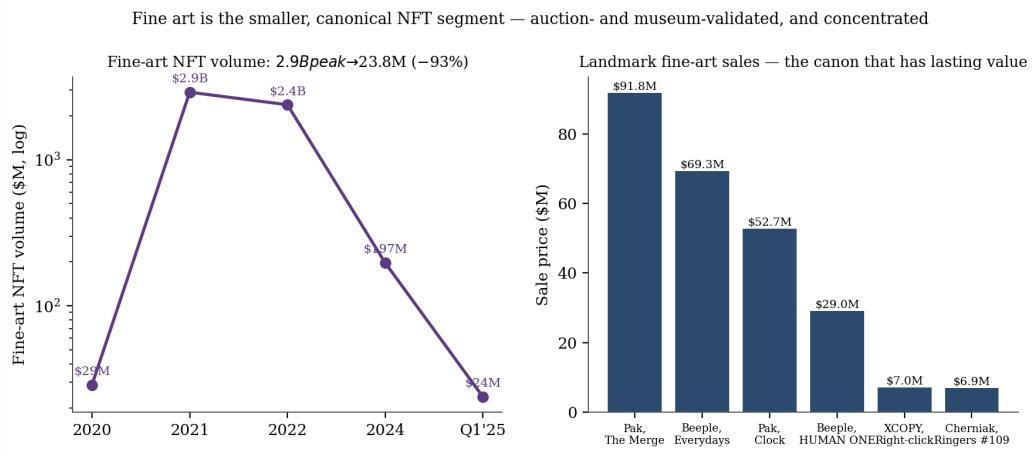

Figure 8. Fine art is the smaller, canonical segment — validated and concentrated.

Left (log scale): the fine-art NFT segment fell ~93% from its $2.9B peak, so what is left is the durable, canonical value rather than the froth. Right: landmark sales — Pak’s *The Merge* ($91.8M, the largest sale by value, though a mass edition of 312,686 units rather than a single work), Beeple’s *Everydays* ($69.3M), XCOPY, Cherniak’s Ringers — the works now entering auction houses and museums.

Sources: DappRadar (2025); Christie’s / Sotheby’s / Nifty Gateway; Art Basel & UBS Survey of Global Collecting 2025.

That this value is durable, not a fad, is now underwritten by the institutions whose job is to say what art lasts. Christie’s and Sotheby’s have run dedicated digital-art sales (Christie’s first AI-art auction in 2025 sold 28 of 34 lots, 37% of bidders new to the house); the Art Basel & UBS Survey of Global Collecting 2025 found 51% of high-net-worth collectors had bought digital art and their average allocation to it rose from 3% to 13% in a year; and in December 2025 the Museum of Modern Art accepted sixteen on-chain works — eight CryptoPunks and eight Chromie Squiggles — into its collection, called ‘one of the most significant institutional

endorsements of onchain art to date.’ The canon is being written now, and it is being written on Ethereum.

So: could an artist like XCOPY — the pseudonymous British creator whose *Right-click and Save As Guy* sold for $7 million and who released all his work under CC0 in 2022 — be hacked, and the art itself changed? Here precision matters, because the honest answer is ‘partly, and not the part you fear most.’ An NFT is a token on-chain that *points to* the artwork. If a wallet is compromised — by a stolen key, a phishing signature, or in future a quantum attack — the attacker can transfer the token and steal *ownership*, and for a quiet single-vault theft that transfer is effectively irreversible (Ethereum’s emergency recovery fork, Section 7, would not be triggered for one wallet). What they generally cannot do is *repaint the picture*. When the art is stored content-addressed on IPFS or Arweave, or fully on-chain as with Autoglyphs, the image is locked to its cryptographic hash: change one pixel and it is a different file with a different address. ‘Who owns it’ and ‘what it is’ are two different questions, secured differently. This is not hypothetical for XCOPY specifically: his primary market, SuperRare, stores the media content-addressed on IPFS, so his canonical pieces are hash-locked — their *ownership* can be stolen by any wallet compromise, quantum or otherwise, but their *image* cannot be repainted by taking a key.

Figure 6. Could XCOPY be 'hacked and the art changed'? What a compromised key actually does

Storage of the artwork	Stolen key lets attacker…	Can the ART change?	Why
Plain web URL	TRANSFER token	**CAN change / vanish**	hosting problem — art mutable
IPFS / Arweave (content-addressed)	TRANSFER token	**CANNOT repaint**	hash-locked to the bytes
Fully on-chain (SVG / bytecode)	TRANSFER token	**CANNOT repaint**	e.g. Autoglyphs

Ownership can always be stolen irreversibly by transferring the token; the pixels of a content-addressed or on-chain piece cannot be altered by stealing a key. 'Who owns it' ≠ 'what it is.'

Figure 9. What a compromised key actually does to NFT art.

A stolen key always lets an attacker transfer the token — stealing ownership irreversibly. Whether the *art* can change depends entirely on where it is stored: a plain web link can rot or be swapped (a hosting problem, not a key-break), while a content-addressed or on-chain piece is fixed to its hash and cannot be repainted by stealing a key.

Source: authors’ analysis; Moxie Marlinspike (2022).

The genuine vulnerability on the ‘art’ side is different and older: many NFTs point to a plain web address, not a hash. In 2022 the cryptographer Moxie Marlinspike minted a demonstration NFT that displayed a different image depending on who looked at it — one thing on OpenSea, another on Rarible, a poop emoji in a personal wallet — because, in his words, ‘what you bid on isn’t what you get.’ That is a hosting problem: whoever controls the server or buys the lapsed domain controls the picture. It has nothing to do with quantum computers or broken keys. So the layered answer for a collector is: a well-constructed, hash-locked XCOPY cannot be repainted by any key attack, quantum or otherwise — but its *ownership* can be stolen the instant a wallet is compromised, and a carelessly-stored piece that points at a dead link can quietly become nothing at all.

And here is why ‘only the ownership’ is not the reassurance it sounds like — stated carefully, because it is easy to overclaim. In fine art the market has long priced *authenticity* over inimitability: it pays for the provenanced original, not the absence of copies, which is why a perfect reproduction of a masterwork is worth a

tiny fraction of the original. An NFT is a machine for asserting that authenticity on-chain, so when a key-break transfers your XCOPY away, the pixel-identical JPEG still on IPFS is little comfort: what has been severed is the authenticated chain of ownership on which the value rides. But two honest qualifications keep this from proving too much. First, an NFT's 'original' is a *social and registry convention*, not a physically fixed object like the panel the Mona Lisa is painted on — it can be reassigned. Marketplaces routinely freeze and delist stolen tokens (which cripples the thief's ability to sell), and an artist or community can choose to re-canonise a replacement. Second, for CC0 work such as XCOPY's, the copyright is already given away, so what the thief takes is precisely the ledger entry and its provenance, nothing about the image. The right conclusion is therefore narrower than 'the art is gone' and still serious: a key-break severs a real, market-relevant, but socially-contingent claim — your standing as owner of the authenticated original — and whether any of it can be restored depends on collective action you do not control.

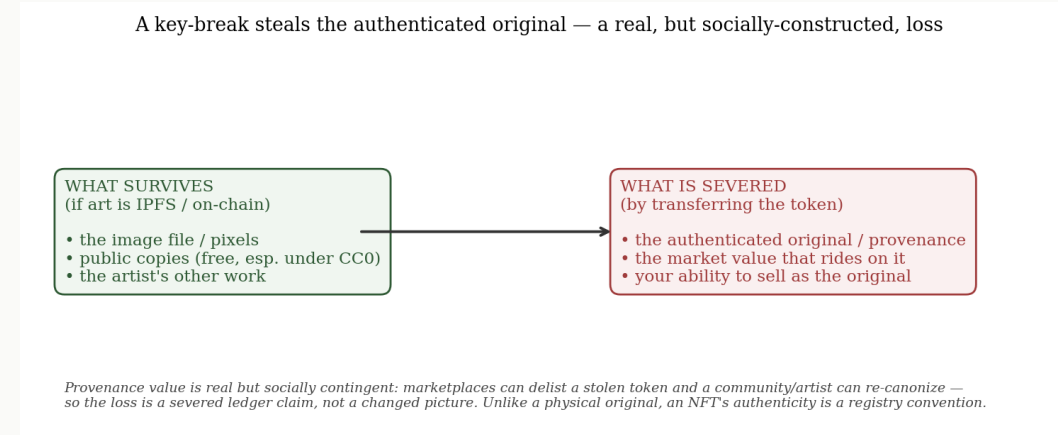

Figure 10. A key-break severs the authenticated original — a real but socially-constructed loss.

For a content-addressed or on-chain piece the image cannot be altered, so 'the art is safe' is half true. Transferring the token severs the provenance and the market value that rides on it. Unlike a physical original, though, an NFT's authenticity is a registry convention: marketplaces can delist a stolen token and a community can re-canonise, so the loss is a severed ledger claim, not a changed picture.

Source: authors' analysis; Unit London on authenticity-as-value (2022).

10. Three ways custody fails — and which one is new

It is essential not to blur three different failures, because the crypto community's fear of quantum often borrows credibility from thefts that had nothing to do with cryptography. To date, *no* major NFT or wallet theft has come from breaking the underlying math. Every one has been either (a) a stolen seed phrase, usually via malware — as when the collector 'NFT God' lost everything in 2023 to a malicious search-ad download, a hardware wallet defeated by seed exposure, not a key-break — or (b) social engineering: a phishing signature the victim was tricked into approving. Seth Green's Bored Ape, the $1.7 million OpenSea phishing wave, the Bored Ape Instagram hack, and North Korea's NFT-phishing campaigns are all category (b). The victim authorised the theft; the cryptography held.

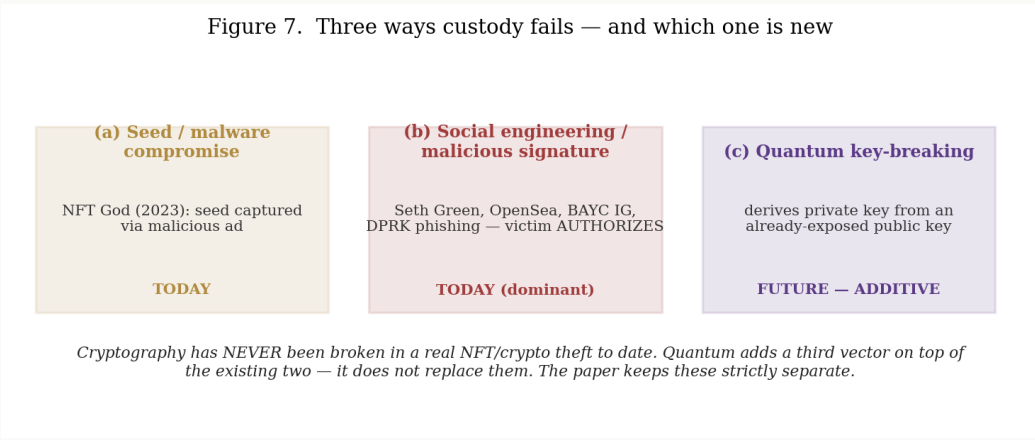

Figure 11. Three ways custody fails — and which one is new.

Today's thefts are seed/malware compromise or social-engineering signatures — the cryptography is never broken. Quantum key-breaking is a genuinely new, third vector that derives the private key from an already-exposed public key. It is *additive* to the existing two, not a replacement — and it is the only one that needs no human mistake at all.

Sources: Forbes (2022); Heimdal (2022); Cointelegraph (2023); SlowMist (2022).

Quantum is category (c), and its novelty is precisely that it needs no human error. The other two require you to click a bad link or run bad software. A quantum attacker needs only your public key, which is already on the ledger, and the machine. That is why it belongs in a different mental box: not 'another scam to avoid' but a background risk that accrues silently to coins whose owners did everything right. It is additive to the existing surface, and it is the one you cannot dodge with good hygiene once your key is exposed.

Part IV — The accelerant

11. Patchageddon: AI on immutable, money-holding code

Quantum is the threat to the cryptography. AI is already the threat to everything around it. Smart contracts are the most attractive target software has ever produced: the code is public and verified, it is immutable once deployed, and it holds money directly, so the path from ‘find a bug’ to ‘drain the funds’ is a single irreversible transaction. Into that environment arrives a technology that reads code tirelessly and cheaply. In late 2025 Anthropic’s red team ran ten AI models against 405 real contracts with known historical vulnerabilities; the models exploited 207 of them — 51% — for $550 million in simulated funds, and succeeded on contracts deployed after their training cutoffs, showing genuine capability rather than memorisation. Against thousands of fresh contracts with no known bugs, they found two previously-unknown exploitable flaws. A separate 2025 academic system, A1 (Gervais & Zhou), exploited 26 of 36 real incidents. All of this testing was done in simulators, never on live chains — but the capability is real and, by Anthropic’s measure, exploit revenue per unit of effort roughly doubled every 1.3 months across model generations.

The economics are the frightening part, and they are structural. The A1 authors found that an AI attacker becomes profitable at exploit values around $6,000, while a defender needs roughly $60,000 of value at stake to justify the cost of a professional audit — a ten-to-one asymmetry. That $6k/$60k gap is the empirical input; the population consequence is a matter of algebra, and we state it as such rather than dressing it up. For any heavy-tailed distribution of contract value with tail exponent α, the set worth attacking but not worth defending is $\mathbf{10^{\alpha} - 1}$ times the professionally-defended set — a pure function of the tail steepness. Across the plausible range (α from 1.0 to 1.5) that multiplier runs from about 9× to 31×. The multiplier is an assumption made visible, not a measurement; what is not an assumption is its sign. This is the ‘patchageddon’: not that any single contract is doomed, but that AI drops the cost of attack below the cost of defense across a vast long tail of smaller contracts that no auditor will ever be paid to look at.

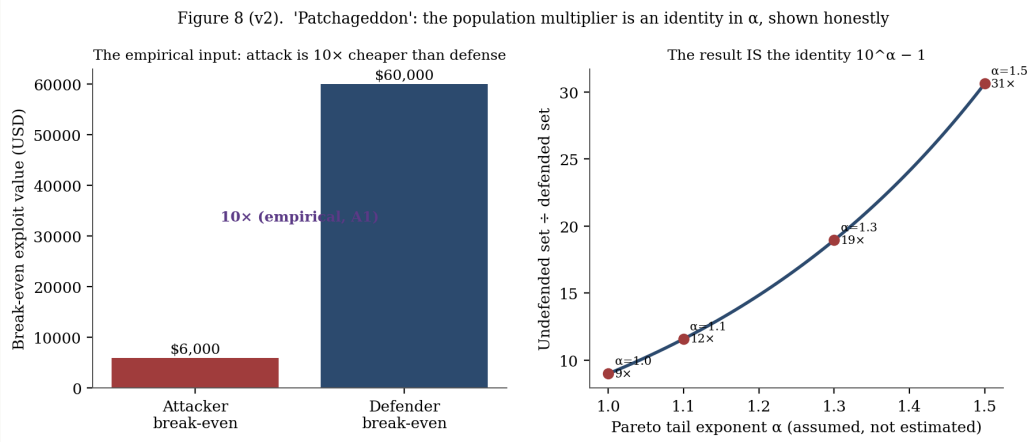

Figure 12. ‘Patchageddon’: the population multiplier is an identity in α, shown honestly.

Left: the empirical ten-to-one break-even asymmetry between AI attack and professional defense (A1). Right: the resulting population multiplier is exactly $10^{\alpha} - 1$ — a function of the assumed tail exponent, not a measured quantity. We plot the whole curve rather than pick a point, because the ‘result’ is the assumption. What survives regardless of α is that the undefended set is several times the defended one.

Model: exposure_result_v4.json. Break-even figures: Gervais & Zhou, A1 (2025). Capability: Anthropic Frontier Red Team (2025).

And the target cannot simply be patched. A deployed contract's code is immutable; fixing a bug requires either a pre-built upgrade mechanism (which is itself a single point of failure — one study found 61% of upgradeable systems controlled by a single key or small multisig), migrating every user to a new contract, or a chain-level hard fork. The canonical case is the 2016 DAO hack, whose only remedy was a contentious fork that permanently split Ethereum into two chains. 'Patching' an immutable ledger means fracturing the community and sometimes the asset itself. AI attackers move in seconds; consensus-based defense moves in weeks. The race is structurally unfair, and it is already being run: security firms report exploit patterns in the wild 'consistent with scripted or agent-driven reconnaissance,' hunting old, dusty, under-maintained contracts at a scale no human team could match.

One honest boundary. There is not yet a forensically-confirmed hack proven to have been generated end-to-end by an AI; the in-the-wild evidence is behavioural — speed, scale, simultaneity — not a smoking gun. And the further claim, that AI is shortening the quantum timeline itself by optimising error-correction, is real speculation voiced by credible people (the researchers behind Project Eleven and NEAR) but not a demonstrated result. We flag both as such. The demonstrated half — AI cheaply exploiting public, money-holding code — is alarming enough without the speculative half.

12. The state adversary who already harvests

Who would actually run a quantum attack on the exposed pool? The honest threat model does not require imagining a new villain, because the ideal one already exists and is already in the business of stealing crypto at national scale. North Korea's Lazarus Group stole an estimated $1.34 billion across 47 hacks in 2024 — 61% of all crypto stolen that year — and a record $2.02 billion in 2025, including the $1.5 billion Bybit heist in February 2025, the largest cryptocurrency theft in history, formally attributed to the DPRK by the FBI. Cumulatively the regime has taken at least $6.75 billion. It already targets NFTs specifically (a 2022 campaign used some 500 phishing domains to steal over a thousand of them), and it already launders at scale through mixers and thousands of addresses.

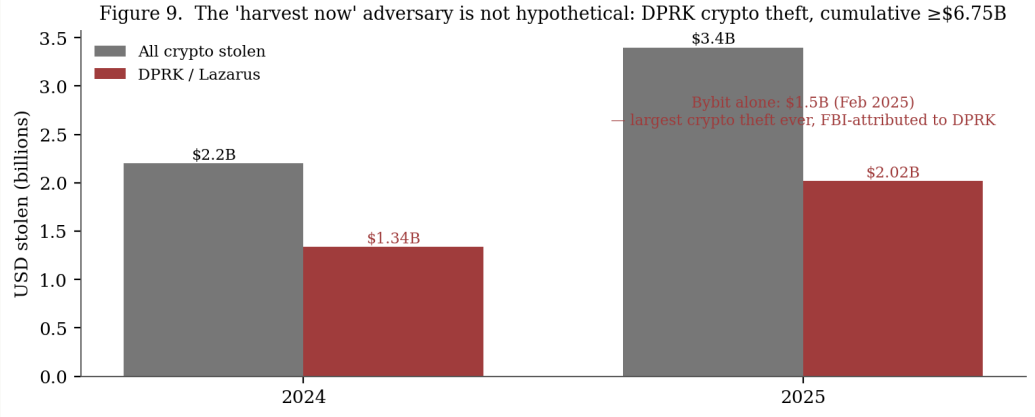

Figure 13. The 'harvest now' adversary is not hypothetical.

A state actor that already steals billions in crypto annually, already targets NFTs, and already launders at scale is precisely the entity with the motive and patience to run a quantum attack on the exposed-key pool the moment the hardware is within reach.

Sources: Chainalysis (2024, 2025); FBI PSA (2025).

This is what makes 'harvest now, decrypt later' more than a slogan for blockchains. A patient, well-funded state actor does not need the quantum computer to exist today. It needs only to note which enormous, unmovable balances are sitting on exposed keys — Satoshi's coins, dormant early-adopter wallets, heavily-reused exchange addresses — and wait. The target list is public, permanent, and already compiled. That is the whole uncomfortable point of this paper: the reconnaissance phase of the largest theft in history may already be complete, and its victims may already be chosen, years before the weapon is finished.

Part V — What to do before the hardware is ready

13. For owners, builders, and policymakers

The situation is serious but not hopeless, and the response divides cleanly by who you are.

For owners.

The single most protective habit is the oldest one in Bitcoin: never reuse an address, and keep meaningful balances in addresses that have never spent, so no public key is exposed until the moment you choose to move. Prefer wallets and formats that minimise key exposure; be aware that Taproot exposes the key by design. Understand that the dominant threat to you *today* is still category (a) and (b) from Section 10 — malware and phishing signatures — so hardware wallets, signature hygiene, and skepticism about ‘sign this’ prompts protect far more of your money right now than quantum worry does.

For NFT collectors specifically.

You are, per Section 7, among the most exposed holders on any chain, because being active means having signed — and every signature published your key. Take that seriously without panicking. Concrete steps: verify that each meaningful piece is stored content-addressed (IPFS/Arweave) or fully on-chain rather than behind a plain web link, so the artwork itself cannot rot or be swapped; treat the *token*, not the image, as the asset at risk, because provenance is the value. Consider segregating a genuine long-term cold vault — a wallet that *receives* pieces but from which you never sign offers, logins, or approvals — from a hot ‘trading’ wallet, since a receive-only address keeps its key hidden; move pieces in, not out and back. Follow Ethereum’s account-abstraction and post-quantum work (EIP-7702 and its successors) and be ready to migrate to a quantum-safe wallet scheme when it ships, understanding that the migration itself is a key-exposing act best done early rather than in a panic. None of this is a complete defence — there is no complete defence for an already-exposed key — but the receive-only cold vault is the closest thing a collector has to the never-spent Bitcoin address, and almost nobody in the NFT world is using it.

For builders.

The migration tools are being built and deserve support. Bitcoin’s BIP-360 proposes a quantum-resistant address type; the more aggressive BIP-361, authored in 2026 by Jameson Lopp and others, would block deposits to vulnerable addresses after three years and freeze the remainder after five — a wrenching step its own author says he dislikes but considers the least-bad option. Project Eleven has prototyped a zero-knowledge proof that lets a holder prove ownership without revealing the key, though it is unaudited and cannot rescue the oldest coins. Ethereum’s founder has sketched a ‘recovery fork’ using hash-based signatures and STARK proofs so wallets can switch to post-quantum validation without exposing old keys — a contingency, not a schedule. The standards exist: NIST finalised its post-quantum signature and key-exchange schemes in 2024 (with Falcon following). The bottleneck is not cryptography; it is coordination on an immutable, leaderless system.

For policymakers.

The regulatory system is beginning to name the risk — BlackRock's Bitcoin ETF added a quantum risk factor to its SEC filing in 2025, noting that any remedy 'would require widespread coordination within the decentralized Bitcoin community' and could 'involve contentious network forks.' NIST's own timeline deprecates classical elliptic-curve cryptography after 2030 and disallows it after 2035. The useful policy posture is to treat the blockchain migration as harder than the general internet migration, not easier, precisely because there is no central authority to mandate the upgrade and no way to un-expose historical keys. Funding the migration tooling, clarifying the legal status of a coin-freezing fork before it is needed in a panic, and applying Mosca's inequality honestly — long secrecy horizon, long migration, uncertain but non-trivial timeline — are the substance of quantum-readiness for this asset class.

14. What would change my mind

This is a position paper, so it owes you its kill conditions. Three findings would meaningfully weaken it. First, if the exposed-supply estimates collapse under careful on-chain scrutiny to a trivial share — if the 'quarter to a third' turns out to double-count or to include coins that are safe for reasons we've missed — the central spine breaks. Second, if someone demonstrates a clean, non-forking, retroactive way to migrate *already-exposed* keys to safety without their owners signing (and thus re-exposing) them, the 'permanent and irreversible' claim softens considerably. Third, if the expert timelines are shown to be systematically biased — if the GRI surveys consistently run years early, or the resource estimates are off by orders of magnitude in the safe direction — then the urgency, though not the mechanism, recedes.

Absent those, the argument stands on facts that are not really in dispute: public keys are exposed the moment coins are spent; that exposure is permanent; a quarter to a third of Bitcoin is exposed today; the hardware to exploit it is on a probability curve that experts keep revising toward us; migration is slow, partial, and politically fraught; and AI is already collapsing the cost of attacking everything around the cryptography. The community has treated all of this as a distant, abstract 'someday.' The claim of this paper is narrower and more urgent than the usual quantum scare, and harder to wave away: the harvest is already done. We are simply waiting, in public, for the machine that reads it.

> ***On this paper.*** *A position / working paper, AI-assisted (Claude Fable 5), with the human author setting the stance and owning every claim. Because it is a paper about AI and quantum risk written with AI assistance, that conflict is disclosed. Numbers live in a single audited data file with sources; the formal model runs as code and writes its results to JSON; disputed figures are given as ranges and speculative claims are labelled. All dollar figures use BTC = $64,085 and ETH = $1,910 as of 12 August 2026. Superseded drafts and the error log are kept with the project.*