Mitzpe Institute
בהירות למי שפועל

Forensic intelligence for those who act.

READER EDITION

DOSSIER

Borrowed Fluency

◇ URIEL ZEHAVI

CLASSIFICATION Reader Edition
DISTRIBUTION Open · Free to share with attribution

IN THIS EDITION

Borrowed Fluency

Every safeguard between Boko Haram and a working answer from AI failed.

A *munzir* in the Lake Chad basin had a bomb that would not go off. The wires ran in a way that seemed to be stopping detonation, and he could not put the question himself, because his rank did not permit him to work the AI model on his own. So he went to his *qaid*, the last rank in ISWAP authorised to use a large language model, and asked him to put the question to ChatGPT. The model answered. Its answers, the *munzir* said, "were not very clear." Then the *qaid* "contacted some people about how to put the question, and then it gave us useful information on how exactly to connect the wires, and it worked." Somebody outside Nigeria knew how to ask. "They call people in the network for this kind of help every day."

The capability had offered itself. Islamic State operatives — "the white guys," the commander said, before clarifying that he meant men from Libya, France and Arab countries with a lighter complexion than their own — arrived at the Lake Chad stronghold with laptops assigned to this purpose and no other, with VPNs and encryption software, and with a projector. Each battalion of five hundred fighters sent its top people, thirty to fifty leaders drawn from the whole of ISWAP's territory, "from Timbuktu to Tumbuma," into one room to watch a screen. Below them access thins by rank and stops at *qaid*. Internet is a separate gate that generally reaches only mid-level and senior leadership, lower ranks use phones without it, and foot soldiers carry devices with no SIM card at all. Answers come back in English, get rendered into Arabic, and travel downward as written and oral briefings to men who will never see the interface.

The models produced usable output the moment somebody phrased the question correctly, and the skill to phrase the hard ones was unevenly held (and partly imported — flown in, paid from abroad, or telephoned). Every safeguard between this group and a working answer failed. What stopped them going further was not a guardrail. It was simply that the people who knew how to ask were somewhere else, and had to be paid.

The dependency did not thin as the skill spread through the group. It acquired a staff — permanent units of bomb-makers, gun specialists and engineers who, in a commander's phrase, "don't go to war," whose job is to hold the accounts, work the models and pass the answers down. Accounts held in the names of others — supporters or, in cases, via the identities of the dead. Subscriptions paid by "leaders in Sudan and all over."

In June 2025, OpenAI's systems flagged a Canadian user's account, reportedly for conversations about gun violence, and banned it. The account went to human review, and roughly a dozen employees reportedly weighed whether it warranted referral to law enforcement. The company concluded it did not meet the criteria for reporting. Some eight months later Jesse Van Rootselaar applied some of that research and killed eight people in Tumbler Ridge,

British Columbia, and then herself. The provider noticed, documented, and applied a threshold it had set itself and published nowhere so it could ignore it.

Executive Summary

Terrorist use of frontier AI has been documented in the field exactly once — inside Boko Haram's two factions, in a working paper built on defector interviews that mostly stop in 2024. Read it and the limit on what the group could do is never the model. Islamic State operatives brought the laptops, taught the phrasing, opened the accounts and paid for them from outside Nigeria. The teaching spread through the ranks. The accounts and the cover stayed with the men who had brought them.

That paper's author sees the arrangement growing self-sufficient, and my own judgment — a judgment, not a finding — is that she is right, and that these groups are getting better at this. The interviews stop in 2024. Whatever the constraint was then, it is looser now. That is an argument for moving on it while it is still there. A bottleneck you can still see is one you can still act on.

Which puts the question at identity and payment, where screening of exactly that kind already runs at consumer scale. It is aimed at the wrong man. It checks whoever the account is registered to — a supporter, or a dead man, as far as the paper can establish — while the money arrives from Sudan. Paying a designated group's subscription, knowing who they are, is already a federal crime. Nobody is looking.

None of this contains the threat. It buys friction, and it buys a record, and both are worth more now than they will be later.

THERE WAS NO SHORTAGE OF MODEL. THERE WAS A SHORTAGE OF PEOPLE.

The binding constraint on terrorist use of frontier AI is human expertise and brokered access, not model capability and not refusals. Antonia Juelich's field study of Boko Haram for the Cambridge Programme on AI Science & Policy, a peer-reviewed working paper. Fifty-seven interviews with 27 former ISWAP and JAS members across two rounds in Borno and Adamawa, twelve of whom knew nothing of their faction's AI use. Fifteen of them, however, revealed some actionable findings.

Islamic State operatives reached the Lake Chad stronghold and ran a projector session for the top people of every five-hundred-man battalion — an estimated thirty to fifty leaders pulled from the whole of ISWAP's territory.

The trainers kept coming back through 2023 and into 2025. Underneath, *qaid* (roughly equivalent to "Commander") is the last rank permitted to touch a model, and the men below it push problems upward and receive written and oral briefings in return. "We are not allowed to access the computers," an ISWAP *naqib* said (more or less a squad leader or "Captain"). "They are the masters."

The obvious reply is that this says nothing special about AI. ISWAP takes drones, IED technique and strategy through the same Islamic State channel, so of course the models came that way too. Which, of course, holds for how the capability arrived, but it falls short for the what of what it left behind. A delivered drone is theirs once it lands. An account is somebody else's for as long as it works, and it can be switched off by a person who has never met the downstream user.

Both factions built non-combat units out of their senior technical talent, men who in the paper's summary "don't go to war" — five to twenty people, twenty-three in the JAS unit at Sambisa, sitting directly under the leadership. Those units use the premium subscriptions (though the accounts belong to other people). Juelich writes that "the specifics of whose accounts are used remains somewhat unclear," and states one thing flatly: "account setup and management is closely coordinated with ISIS-linked intermediaries."

The models produced usable output the moment somebody phrased the question correctly.

"Trial and error can kill you. AI gives you accuracy," one commander said.

Juelich calls prompt skill "unevenly distributed and partially dependent on foreign assistance." Refusals ran on the same circuit — "the 'white guys' taught us how to bypass restrictions" — and the trained men's pretext is that they need it for a movie.

Fighters had watched motorcycles jump bridges in a movie and asked the model how to do it on their own machines across the distance they needed. The humans brought the idea and the model fitted it to their equipment.

By the account of the commander who described it, eighteen men died practising and eight got across, and the next assault breached the trenches.

Juelich did not ask about AI upfront, tracing conflict experience first to reduce "the risk of respondents calibrating their answers to what they expected I wanted to hear," and she tested some participants against a plain search-engine page to check they could tell it from a chatbot. Her respondents are defectors who left months earlier, and she acknowledges the study "does not capture the most recent developments." We extend her limitation. Retrospective self-report degrades for details, dates, frequencies and causal attributions, and "AI enabled X" is a causal attribution about a detail. However, what these interviews carry well is the shape of the arrangement.

Two months before Juelich's paper appeared, three counterterrorism scholars writing in *CTC Sentinel* observed that terrorists "do experiment, and thereby develop novel approaches, but they rarely do so at the frontiers of knowledge," and posit that knowledge and skill are what limits them. Her fifteen respondents are the field evidence for an argument the practitioners had already made. The tools sat on the open web the entire time. What had to travel was people.

THE BROKER SUPPLIES THREE THINGS, AND ONLY ONE OF THEM DIFFUSES

The original unit trained about ten people across twelve camps, taken "rank-by-rank," who then instructed others. The cascade reaches down to a single keyboard, where a *munzir* (local commander) sat beside his *qaid* (area/regional commander) and was "trained bit by bit." That is real skill transfer, and it makes the diffusion objection serious.

Account access, however, did not travel with it. The AI units held the accounts, and one trained user did not know who had opened his: "I don't know who installed it and signed up. I was just told what platforms to use." Subscriptions came from outside the theatre, paid by "leaders in Sudan and all over."

Deniability did not travel either, and the surveillance anxiety that makes it valuable runs through the CASP data. A former *naqib* explained that platform knowledge stays inside the AI unit because "they are worried that the military will know that they are using this technology." ISWAP generally gates internet access to mid- and senior-level leadership and keeps accounts that "can't be linked to us" behind VPNs and encryption, an effort Juelich reads as designed to "prevent information leakage and evade state surveillance." Of ISWAP she writes: "AI usage has not been diffused throughout the entire organization."

The arrangement acquired staff, at an opportunity cost Juelich reads as institutional commitment. She reads that institutionalization forward. AI fluency "does not lie with any single individual but is embedded in organizational processes, making it more resilient to personnel attrition and harder to disrupt." More sharply: "LLMs supply the technical and tactical information that once required an external expert, reducing reliance on periodic visits that could be delayed, denied, or interdicted," so that "a single training intervention can have compounding effects."

The model replaces the visit — and that visit could be delayed, denied or interdicted. But the substitution runs both ways. A visit needs someone who can be stopped at a border. An account needs a subscription that can be stopped by a company, from anywhere, without notice (and that leaves a record of who paid for it).

Sure, a group that reached comparable use unaided would defeat the control surface we are proposing. And, yet, would also be a case nobody has observed to date. Across the period the accounts describe, the concentration held.

For the moment, there is a real bottleneck — and its eventual closing is a reason to act while it is still an interdiction point. A reading Juelich's own paper supplies: "the concentration of knowledge within a small team can also make it a high-value counterterrorism target."

Asked what weapons were prohibited, many respondents named poison, most often for the indiscriminate harm it causes, including among "the innocent." One gave the enforcement: "It is a general rule of engagement. You would get killed right away if you did this." A senior ISWAP respondent qualified that ban: "Chemical or biological weapons are allowed. Traditionally, they are prohibited. But they have been legitimized."

Juelich finds these interpretations are not uniform and may vary by rank and ideological position. The qualification runs through a redefinition, poison reclassified out of the category of weapons, and through reasoning that tracks the logic of the 2003 al-Fahd *fatwa*: an enemy's use of the forbidden licenses your own. Another respondent set its expiry: "Today, this is the rule; tomorrow it may change." The expertise constraint narrows hardest to biology, where the physical handling no text carries is the barrier. Juelich asked about CBRN directly and found no programmes, and she reads chemical weapons as the likelier near-term pursuit, on Islamic State precedent and low technical barriers. One respondent claimed a foreign operative, guided by AI, instructed members on making ammunition laced with an agent causing "bleeding from nose and eyes," which only senior commanders could use.

THIRTEEN PEOPLE FLEW IN WITH THE CAPABILITY

The evidentiary base so far is a single working paper, built on defector interviews inside one group across one window. The UN Security Council's Analytical Support and Sanctions Monitoring Team, as at late 2024, Member States reported that "13 ISIL (Da'esh) trainers had arrived in the Lake Chad basin from the Middle East, facilitating the acquisition, assembly and deployment of drones that were believed to have been used in these attacks." Vincent Foucher, who also peer-reviewed the Juelich paper, reports the same arrival through defector testimony. Malik Samuel's field reporting on seven instructors, which is where the paper's own syllabus comes from, records it — drone operations, vehicle-borne IED assembly, electronic device hacking, and advanced combat tactics. ISWAP's first rudimentary armed drone attack came in late December 2024, grenade-equipped, against installations in Borno and Yobe.

Seven more followed across 2025, against one the year before. S/2026/44 records the components acquired "through open commercial channels, which were later reassembled." That channel was open before the trainers arrived.

Islamic State provinces owe the General Directorate of Provinces "monthly reports on the ongoing military situation… and, in return, receive advice on future strategy and tactics," and the al-Furqan office covering the Lake Chad Basin, headed until his killing in the northeast in May 2026 by Abu Bakr ibn Muhammad ibn Ali al-Mainuki, is where ISWAP's reports go.

Thirteen men flying in is that structure operating normally, which is what "The Jihadist Continuum" argued about the shared doctrinal bloodstream running under movements Western analysis keeps filing separately.

The Monitoring Team's AI findings are propaganda-scoped, and S/2026/44 [apologies for the UN's naming conventions—though I have bigger gripes about them than just that] deflates them without leaving it there: groups were "increasingly adept at seamlessly integrating artificial intelligence tools… Although this did not represent a step change in capability, it underlined the increasing challenge that these tools pose to the international counter-terrorism effort." Nothing in the drone finding bears on AI transfer. The

syllogism is sitting right there — al-Furqan supervises other provinces, ISWAP obtained model access under al-Furqan, therefore the provinces have it. The evidence does not establish it. Juelich draws that inference herself and marks it speculation, adding that "it remains unknown if similar knowledge- and resource-sharing obligations extend to AI." We decline our own principal source's inference.

The drone case is the compounding-effects reading in its strongest form: thirteen men arrived once, and the attack rate went from one to seven. It compounds because a drone is not an account. Once the parts are assembled the capability is theirs, and nobody can switch it off from a distance.

The drone finding is evidence about drones.

THE REGIONAL EVIDENCE THINS AT EVERY STEP BACK

Follow the citations behind the claim that Hezbollah, Hamas and the Houthis already use AI operationally, and it thins at every step back toward its sources. That is a finding about the literature, not about the world. Bad sourcing for a claim is not evidence against it. *Generating Terror*, in *CTC Sentinel*, cited as evidence that groups are doing something, is a red-team jailbreak experiment documenting zero instances of any group doing anything. Hezbollah coverage leans on a conditional hypothetical, "if you're running a campaign for Hamas, Hezbollah…" — except Gabriel Weimann closed that thought with "This isn't hypothetical—it's already happening."

The Media Line's own pull-quote block drops the closing sentence, so the detachable conditional is manufactured at publication. And Weimann wrote *Generating Terror*, so the chain's first two nodes are one man. GNET's *AI Jihad* describes the propagation of the "IDF wears diapers" imagery and is careful about where it came from [Daniel Siegel, "AI Jihad: Deciphering Hamas, Al-Qaeda and Islamic State's Generative AI Digital Arsenal," GNET, 19 Feb 2024 — Siegel attributes the *verbal* claim to Abu Obeida and describes the imagery in the passive, though two other sentences in the same piece loosen toward "amplified by Hamas"], and ICCT restates it as content that "appeared to be generated by a Hamas-affiliated group using generative AI," on a single citation to that same GNET piece, which names no creator for the imagery itself.

Vincent Foucher's defector testimony corroborates the CASP account of the trainers' arrival, and Foucher peer-reviewed the CASP paper. A chain that reads as independent corroboration is partly one man agreeing with himself, and the discipline we are applying to ICCT and The Media Line applies to us too. Martini finds UN counterterrorism documents acknowledging that no concrete evidence of terrorist interest in AI has been found and, in the same breath, calling it "prudent to assume" that these groups "are aware of this technology" — caveats she reads as "immediately counterbalanced by statements that discard uncertainty."

Thirteen AI-vendor threat-intelligence reports, February 2024 to February 2026 — Microsoft and OpenAI jointly, OpenAI six more times, Google's GTIG three times, Anthropic three — name no designated terrorist organisation as a user of their models. Hamas appears there as a content theme, and APT42

is described as having researched the Israel–Hamas conflict as a topic. The same reports name Iranian state clusters repeatedly, Crimson Sandstorm, CyberAv3ngers, Storm-2035 and Storm-0817, doing phishing, debugging, reconnaissance and influence content.

That record carries one sentence and no stronger: no vendor has ever publicly named a designated organization as so designated.

The IRGC has been a designated Foreign Terrorist Organization since April 2019, and the clusters above are named as state actors every time. These are curated disclosures shaped by legal and commercial considerations, and a designated-FTO hit carries disclosure risk a state-actor hit does not.

Google's GTIG scopes its reporting to government-backed actors by its own statement. OpenAI assessed that the particular CyberAv3ngers interactions it disrupted "did not provide… any novel capability" beyond "publicly available, non-AI powered tools," a judgment about those interactions that carries no further.

The Houthi drone-AI claim fails technical sourcing. Sanaa Center's study documents operator-controlled FPV, defined there as drones giving their operators "real-time video-feed access to the drone's flight trajectory and surrounding environment," and it makes no navigation-, autonomy- or seeker-architecture claim of any kind.

None of it, however, establishes absence. The training documented in the Lake Chad basin happened in person, orally and behind a rank gate. And open-source monitoring never saw a minute of it — which is exactly as true of Hezbollah, Hamas, Palestinian Islamic Jihad and the Houthis as it is of ISWAP.

A monitoring record that cannot see something should not be construed as evidence the thing is missing. What collapses under inspection is the published record, in both directions at once: it cannot establish that these groups are using these tools, and it cannot establish that they are not. That is the state of the evidence, and it is worse than either camp admits.

Tehran supplies the one documented regional positive. A small Iranian production outfit's representative told the BBC in April 2026 that the Iranian government is a "customer" placing "direct commissions for several projects," which corroborates an IRGC command relationship without establishing one, and the *Christian Science Monitor* has it as production houses with "some IRGC connection." Tasnim amplified the "Lego-style" wave, and Israel's National Cyber Directorate attributed the campaign to "the cyber war waged by the IRGC."

TWENTY CENTS STRIPS THE REFUSALS

Concede all of it. Since January 2026, open-weight models have lagged frontier closed models by an average of four months on the Epoch Capabilities Index, slightly larger than the three-month average Epoch measured for January 2023 to October 2025. Epoch carries its own caveat: "This gap may be

understated: open-weight models tend to perform worse on private benchmarks compared to closed models, plausibly because they more aggressively optimize for public benchmarks." UK AISI's cyber-specific gap moved the other way, narrowing from six-to-ten months to four-to-seven. They measure different quantities, and both land in months. Months is what denying the top tier buys.

Qi (ICLR 2024) broke GPT-3.5 Turbo's alignment with ten examples for under twenty cents through the vendor's own fine-tuning API, and found even benign fine-tuning degrades it. Zhan (NAACL 2024) removed GPT-4's RLHF protections with 340 examples at 95 percent success.

If a closed, hosted, refusal-bearing model strips as cheaply as a downloaded one, the refusal layer is not yet doing the work it is credited with. That is an indictment of how it is built, not an argument for doing without it. To be sure, in the field, it never became the constraint.

CASP's report admits that because they kept accounts across multiple providers, a single refusal or suspension rarely mattered. Its evidence section says something narrower and more useful — respondents were not aware of account suspension being an issue at all, the *munzir* saying "they are very careful" that accounts do not get suspended, and a JAS *qaid* describing the AI units as responsible for keeping alternatives reachable if one is.

Tech Against Terrorism's CT-AI benchmark measured compliance and not usefulness. We take their policy argument anyway, because it does not rest on the measurements. TAT suggests, short of establishing, that the pipeline is governable where "open weights" in the abstract is not. Google's threat intelligence documents the automating registration of premium LLM accounts, and reads the method as procuring "high-tier AI capabilities at scale while insulating their malicious activity from account bans." That is one observation.

"Interdiction" does not mean prevented from obtaining AI capability. Everything above is a route around a block placed on the holder — accounts across several providers, automated registration, alternatives kept reachable. A block placed on the payer is the one none of those routes replaces, because each of them still has to be bought. It buys friction and visibility, and it contains nothing.

The Van Rootselaar ban and the Florida State messages are both artifacts of a hosted account. Local inference produces neither, which is what our own recommendation costs: driving actors onto local models denies us the signal, because the whole value of the logged channel is that someone is on it.

THE LABS MODEL NOVICES AND STATES. THE KILLING HAPPENED IN BETWEEN

Anthropic's two published thresholds imagine two actors. In the Responsible Scaling Policy v2.2 — superseded on 8 July 2026 by a v3.4 that reframes the threshold as "Non-novel chemical/biological weapons production" — CBRN-3 turns on capability that would "significantly help individuals or groups

with basic technical backgrounds (e.g., undergraduate STEM degrees) create/obtain and deploy CBRN weapons," and CBRN-4 on capability that would "substantially uplift CBRN development capabilities of moderately resourced state programs." ISWAP is a group with basic technical backgrounds, so the first of those describes it. Neither measures what the group was actually doing, because neither is about conventional weapons.

Terrorism is named at design time and missing at run time. OpenAI's Preparedness Framework v2 names "terrorists consulting a model to debug the development of a biological weapon" as an example of misuse its safeguards must minimise, the word's only appearance in the document. Anthropic's Responsible Scaling Policy and Usage Policy name terrorism, and the Claude Opus 4.6 system card documents a mustard-gas elicitation failure. Across the same thirteen vendor threat-intelligence reports, none names a designated terrorist organisation. A state-actor frame does not explain that: OpenAI's and Anthropic's own taxonomies reach non-state actors down to ransomware and extortion.

The safety establishment is calibrated to CBRN and cyber, and what Lake Chad has already produced — unjamming a rifle, identifying a captured weapon, retreat drills, force allocation, which base to hit — sits beneath every published threshold and kills anyway. AI told ISWAP fighters to wash a jammed rifle with diesel. It taught them a 200-man assault could sometimes be 20, after 60 died in the larger version.

OpenAI's standard defence is that this is information "that could be found broadly across public sources." GovAI's March 2026 technical report holds that "the extent to which current AI systems have enabled explosives misuse by terrorists beyond existing tools remains unclear, but is likely limited." Fine, the information the *qaid* needed was public, and yet he still telephoned the network to learn how to ask.

OpenAI made the argument itself in January 2024, against its own interest: biorisk information sits "just a quick internet search away," and "other factors, such as the difficulty of acquiring wet lab access or expertise in relevant disciplines… are more likely to be the bottleneck." The 2026 measurements split on that line.

None of that settles whether the danger is real. Casper, Krueger and Hadfield-Menell show that evidentiary bars set high enough systematically neglect risk. RAND's Brent and McKelvey hold that contemporary foundation models increase biological weapons risk, and the Virology Capabilities Test cuts the same way. Our claim is about where the constraint sits. The men who learned to move jammed rifles to the back of a formation cleared none of those thresholds along the way.

IDENTITY CHECKS POINTED AT THE WRONG PERSON

Screening of exactly this kind may already run on frontier AI access at consumer scale, and the evidence for it is two interested parties. Forbes reported in April 2025 that OpenAI screens users across 225 countries and territories

through the identity vendor Persona, on ChatGPT and on the API — a staff byline in a Persona funding profile built on an interview with Persona's chief executive. The 225 counts the countries and territories that screening reaches, not an OpenAI list. Persona's customer page quotes OpenAI's integrity product lead, Jake Brill, on the company's wish to "make sure people or entities that are sanctioned are not using our services." That page is undated vendor marketing quoting a customer employee.

OpenAI's own published material confirms none of it. Its Privacy Policy returns nothing for sanction, screen or OFAC, and where OpenAI names Persona in its own documentation it names it for age. What OpenAI does publish is the allowlist: 208 entries for consumers and 188 for the API as of 20 July 2026. Russia, China, Iran, Syria, Cuba, North Korea, Belarus, Venezuela and Hong Kong are absent from both. Sudan, South Sudan, Libya, Somalia, Iraq, Afghanistan, Yemen, and "Palestine" are present on both.

The screen runs against the nominal account holder, and the field evidence says that holder is a real person with no link to the group. ISWAP's former war strategist told CASP that the people who open the accounts cannot be linked to the group, and that "we have leaders in Sudan and all over who do the subscriptions for us." Asked about fake email addresses he denied them, saying the accounts belong to "real people elsewhere," and two respondents pointed instead to supporters or to deceased individuals. The paper records that "the specifics of whose accounts are used remains somewhat unclear," and the *munzir*, not allowed to set up an account himself, supplied the reason: "I don't know who installed it and signed up. I was just told what platforms to use." Payment originates outside the operational theatre, and the rank-gating that produces the evidentiary limit produces the screening blindness.

Identity proofing on AI access is supposed to be infeasible. Yet, Google built it years ago for other reasons. Its Cloud signup FAQ says a card is required because "we ask for your name, address, and payment method to verify your identity. This reduces redundant accounts and fraudulent use on production infrastructure," and Google's own HTML bolds *to verify your identity*. Five proofing methods sit behind the verification gate, among them an email-history check by a third-party data aggregator, queried on where an address has been used. Google Cloud Billing's India section already mandates beneficial-ownership-grade KYC, down to an individual holding executive management control, through Aadhaar and DigiLocker. A regulator demanded, and it got built. The third ground on AI Studio's block page is age, so what exists is identity-grade infrastructure standing up for a different purpose.

Anthropic has no coherent account of what its own check is for. It runs Persona's government-ID-and-selfie check discretionarily, on unpublished triggers, with no way for a user to enrol, where OpenAI's Verified Organization is a standing documented gate. As of 20 July 2026 Anthropic publishes three supported-region lists on three hosts, at 184, 175 and 177 entries. The two that disagree carry no date. The one dated 16 March 2026 is the third. Nine countries separate the first two: Central African Republic, Eritrea, Ethiopia,

Libya, Mali, Nicaragua, Somalia, South Sudan, and Sudan. That is an inconsistency in the paperwork and says nothing about who can actually reach the service.

Anthropic's Consumer Terms, effective 8 October 2025, bar credential sharing and making an account available to others, resale, and access from embargoed countries or by anyone on the OFAC SDN list or the Commerce Entity List. Anthropic separately lists account creation from an unsupported location as an express ban ground. Every element CASP documents already breaches an instrument the provider wrote itself. [You really can't trust an industry to properly regulate itself.]

Tech Against Terrorism observed in 2023 that users "likely exploited free tools to mitigate the risk of identification via payment for paid services," and built nothing on it.

GNET documented extremists cost-sharing subscriptions and stopped there.

Jason Blazakis reached designated-FTO plus LLM plus anti-money-laundering in February 2026, prompt-tested a Hamas fundraising request, found refusals, read them as "modest reassurance" — then said himself that the testing was narrow and the real problem lay elsewhere. He never reached the payer either.

OpenAI's February 2025 report, its IUVM ban and its CyberAv3ngers assessment catalogue bans, recidivism, multi-account operation and cross-provider hopping with no substantive mention of KYC, sanctions or payment.

Two of them saw the payment layer. None of them followed it to whoever was paying.

BUYING ISWAP A SUBSCRIPTION IS ALREADY A FEDERAL CRIME

Anthropic's Consumer Terms bar anyone appearing on the OFAC Specially Designated Nationals list, and its Supported Regions allowlist is a sanctions-derived jurisdictional control carried as its own distinct violation category. In threat-intelligence enforcement actions, providers cite conduct and never designation. OpenAI banned IUVM, which the SDN list annotates "Linked To: Islamic Revolutionary Guard Corps (IRGC)-Qods Force," and its account of the ban reaches the listing nowhere.

Federal Code defines material support to include "any property, tangible or intangible, or service," expressly enumerating "training" and "expert advice or assistance" — further defined as "advice or assistance derived from scientific, technical or other specialized knowledge." ISWAP has been a designated Foreign Terrorist Organization and a Specially Designated Global Terrorist for at least the past eight years.

Whoever buys that subscription from abroad walks right into the pit of § 2339B. Its knowledge element is disjunctive on the statute's own terms — designation, or terrorist activity, or terrorism, any one of the three.

Across OFAC's published civil-penalty lists from 2020, not one respondent is an AI or machine-learning company, and the closest analogues are both software. SAP SE paid $2,132,174 in April 2021, penalised in part for "the sale of cloud-based software subscription services accessed remotely" reaching users in Iran. Exodus Movement paid $3,103,360 in December 2025, where staff "recommended that these users obscure their location in Iran using Virtual Private Networks (VPNs) to avoid the sanctions compliance controls." Remotely accessed subscriptions and brokered evasion already sit inside the regime.

FATF's *Crowdfunding for Terrorism Financing*, October 2023, mentions artificial intelligence exactly once in the whole document, concerning algorithms that "identify prospective donors," and nowhere treats AI providers as obliged entities.

The one adjacent case where enforcement met facts resembling these closed without a penalty. Cloudflare voluntarily self-disclosed to OFAC that it "may have inadvertently allowed its network and associated products to be accessed or used by some customers in apparent violation of U.S. economic sanctions laws." Per its Form 10-Q filed 8 May 2026: "On April 9, 2026, we received a No Action Letter from OFAC, stating that OFAC was closing its review without penalties or further action." No Action Letters state no rationale, and closure can reflect remediation, self-disclosure credit or where an agency puts its people. Cloudflare is a content delivery network that says it inadvertently served embargoed-country users and SDN-listed entities "identified in OFAC's counter-terrorism and counter-narcotics trafficking sanctions programs," and the scenario here is knowing supply to a designated organization through a third-party payer. The nearest analogues are settlements over remotely accessed subscriptions, and OFAC gave no reasons in either.

BRUSSELS BUILT IT, WASHINGTON WITHDREW IT

Enforced since November 2025, the EU forbids providing AI services (access, training, fine-tuning) and GPU/quantum computing to the Russian government and Russian entities. (The only exception allows strict open-source contributions). A proposed US rule that would have forced cloud providers to verify foreign customers renting raw GPU hardware—ignoring everyday AI subscriptions—was officially scrapped in December 2025.

The same drafting reached two state legislatures and left both. California's attempt would have made computing-cluster operators collect "the means and source of payment… or virtual currency wallet or wallet address identifier" and retain IP addresses. Gavin Newsom vetoed it on 29 September 2024. Its successor bill signed a year later contains "customer," "identifying information" and "Internet Protocol" zero times each, and its one surviving computing-cluster reference is CalCompute. Massachusetts's attempt carried the same language at first, but the redraft that replaced it in October 2025 contains neither "operates a computing cluster" nor "basic identifying information." MassCompute survives.

The AI Diffusion Rule reasoned that "restricting the export of model weights, while allowing access to the most advanced AI models through other methods, such as application programming interfaces, can unlock the beneficial uses of AI for users across the world while mitigating the national security and public safety risks posed by these models." That holds about hardware and says nothing about who holds the account.

Brussels is the nearest thing to a precedent and it is not close. Article 5n bars supplying the service to Russian counterparties and imposes no duty to identify a customer anywhere in the regulation. Nobody has built the thing this Dossier is asking for.

THE WARNING WAS GENERATED, LOGGED, AND SENT NOWHERE

Kenneth Cooper of ATF's San Francisco Field Division, at the joint LVMPD, FBI and ATF press conferences of 2 and 7 January 2025, assessed Livelsberger's Las Vegas device on 2 January, prefacing it with "I want to be careful with my language here, but": "the level of sophistication is not what we would expect from an individual with this type of military experience." He added that it was "too early" to say whether the components had been wired for a coordinated detonation at all. The coverage carried the assessment and dropped both hedges. The consumer fireworks he used are "designed specifically to not mass detonate." The lead birdshot packed around them never became shrapnel, and Cooper read it as "an attempt to weaponize the device over and above the explosive and incendiary effects" — deliberate, which cuts against reading the device as ineptitude. The AI allegation against Gann survives only in the criminal complaint and never reached his indictment. Nashville and Minneapolis carry no AI in any official finding, the only AI in Nashville's coverage being the weapons-detection system that failed.

The cases carrying the most AI evidence are disproportionately the cases with no defendant. Livelsberger, Van Rootselaar, Soelberg, Rupnow, Westman and Henderson all died, and Palm Springs charges Park while Bartkus, who did the prompting, died in the blast. That pushes the AI record out of criminal court, where it would be adversarially tested, and into civil pleadings, coroners' findings and police briefings.

GovAI's own read of current systems is deflationary, and the same report holds the strongest case against us. The 7/7 and 7/21 London cells built near-identical hydrogen-peroxide devices from designs that, in the UK Home Office account GovAI quotes, required "no great expertise" and could be learned from open sources. The 7/7 cell took calls from Pakistan that increased in frequency while the devices were being built, and killed 52. The 7/21 cell had little ongoing contact, prepared its peroxide improperly, and all four devices failed. GovAI names the improperly prepared peroxide as what failed the 7/21 devices, and offers mentorship as one factor separating the two cells. Interactive, responsive troubleshooting is the part a language model could take over, and GovAI puts even that in the conditional, about future systems. It adds that a significant number of terrorist bombs do not detonate, which "indicates there is room for technical uplift." "The

information is publicly available" is the wrong test, because in both London cells it was.

What the Western investigative record documents AI contributing is target selection and outcome modelling. At Florida State, Ikner asked ChatGPT about the student union and was told it "experiences its busiest periods during weekday lunchtimes, typically between 11:30 a.m. and 1:30 p.m." The complaint records that exchange at ¶109, and police place his attack inside that window. The same complaint has the model disclaiming any "official threshold" and then offering that "3 or more people killed (excluding the shooter) is often the unofficial bar for widespread national media attention." Records released by the Florida State Attorney's Office show he exchanged more than 13,000 messages with ChatGPT over more than a year. Ikner did not get chemistry from the model, and the targeting knowledge he did get was added to what he already brought. Red-team evaluations do not test that category, and the Nigerian case used the models for it too — twenty fighters where two hundred had been the habit, and a rally point to retreat to together.

GovAI's account of how these plots get caught is worth adopting outright: counter-terrorism against bomb plots "relies significantly on informants and surveillance of extremist networks — methods that assume plotters will reach out to other terrorists," and mentorship-seeking is one of the avenues that gets them foiled. A model removes that exposure and creates a logged one in its place. In our own case a *qaid* still needed a person to phrase the question, which is 2024-era practice, and practice changes. If models close that gap the bottleneck moves to the account layer alone. That is a thinner place to stand than it sounds: the measure that secures the layer is the one that pushes people off it.

That OpenAI considered alerting law enforcement about Van Rootselaar's account is on the record. That she reportedly "described scenarios involving gun violence over the course of several days," and the roughly dozen employees who weighed it, rest on people familiar with the matter, the employee figure reaching us only through outlets crediting the Wall Street Journal. Per OpenAI's own letter of 26 February 2026 to Canada's AI Minister, the account went to human review to determine whether usage policies were violated and "whether the account warranted referral to law enforcement."

A pathway existed. OpenAI applied it. And answered themselves that it was best to look away. No statute compels referral. The only mandatory provider-reporting duty in US law is scoped to child sexual exploitation and expressly disclaims any duty to monitor. No transparency report anywhere discloses how often the discretionary criterion fires. Altman apologized on 23 April 2026, and OpenAI conceded, that "we would refer the account banned in June 2025 to law enforcement if it were discovered today." After her name became public the company found a second ChatGPT account its repeat-violator detection had missed. The ban bound an account. She opened another one.

The constraint is human expertise and brokered access. Not model capability, and not refusals. An account layer built to notice who is paying would

generate the attribution that local inference will never produce, and a referral pathway already waits on the other end of it. Attribution nobody is generating, feeding discretion nobody is auditing. The visibility of an adversary's use of a monitored system is itself a security asset, and a measure that collapses that visibility can subtract more than it adds. Call that legibility as deterrent. In practice it would run through the payer rather than the holder and through the account rather than the model. Juelich's own key findings say that "similar training has likely reached other affiliates." Every route to the model ran through a person, every one of those people left a payment trail, and the enforcement record does not contain a single instance of anyone reading one.

References

Grouped by evidentiary tier, not alphabetically: in this Dossier what a claim rests on is as much the finding as the claim. Live figures carry the date they were observed, so a later provider fix dates the finding rather than falsifying it.

PRIMARY FIELD EVIDENCE

1. Juelich, Antonia. *"God has helped us, and so will AI": How the Terrorist Group Boko Haram Uses Frontier AI.* Cambridge Programme on AI Science & Policy (CASP), University of Cambridge, Frontier AI Working Paper Series No. 1/2026, July 2026. https://casp.ac/reports/ai-enabled-terrorism (PDF: https://casp.ac/__l5e/assets-v1/8e5796ad-c373-4dea-8470-466263cd2125/ai-enabled-terrorism-report.pdf). Accessed 17 July 2026. A peer-reviewed working paper, not published in a peer-reviewed journal, and carrying no DOI; its Acknowledgements, at p. 76, name Gary Ackerman, Vincent Foucher and Alan Z. Rozenshtein as peer reviewers. 57 interviews, 27 former ISWAP and JAS members, 15 carrying every AI finding. Page citations here are to the Full Report PDF, 93 pp., whose printed folios run 1:1 with the PDF sequence — not to the Extended Executive Summary or the Short Summary & Key Excerpts, which are separate documents. PDF archived alongside this edition.

UN AND OFFICIAL DOCUMENTS

2. United Nations Security Council, Analytical Support and Sanctions Monitoring Team. S/2025/482, 24 July 2025, ¶¶23, 25. The trainers passage is at ¶25, where the Monitoring Team is relaying a member-state report rather than making a finding of its own — "Member States reported that 13 ISIL (Da'esh) trainers had arrived" — and where the drones are ones "that were believed to have been used in these attacks." The same paragraph dates ISWAP's first "rudimentary" armed drone attack, against military installations in Borno and Yobe States, to late December 2024: "The armed drones were equipped with grenades."
3. United Nations Security Council, Analytical Support and Sanctions Monitoring Team. S/2026/44, ¶¶24, 126, 132. ¶132 verbatim: ISWAP "increased its cache of unmanned aerial vehicles through the acquisition of parts for such vehicles through open commercial channels, which were later

reassembled." The "step change" passage is scoped by the report itself to propaganda — "greater proficiency in the use of artificial intelligence, primarily in propaganda" — and the sentence deflates and then re-inflates: "Although this did not represent a step change in capability, it underlined the increasing challenge that these tools pose to the international counter-terrorism effort."

4. 18 U.S.C. § 2339A(b)(1) and § 2339A(b)(3) (material support; "training"; "expert advice or assistance"); 18 U.S.C. § 2339B.
5. ISWAP designations. The FTO designation is made by the Secretary of State under INA § 219 (8 U.S.C. § 1189) — ISIS-West Africa, 83 FR 8730, FR Doc. 2018-03995. The SDGT designation is made under Executive Order 13224. Both determinations were signed 13 September 2017 and sat unpublished for five months, appearing in the Federal Register on 28 February 2018, which is the operative date.
6. Council Regulation (EU) 2025/2033 of 23 October 2025 amending Regulation (EU) No 833/2014, OJ L, 2025/2033, 23.10.2025. http://data.europa.eu/eli/reg/2025/2033/oj. Article 17 replaces Article 5n; Articles 5n(1)(g) and 5n(1)(h) bind from 25 November 2025; derogations reaching (g) and (h) at 5n(9b) (authorisation of supply strictly necessary for Russian nationals' contributions to international open-source projects) and 5n(10) (humanitarian, diplomatic, critical-infrastructure and civil-nuclear grounds), and, in the parent Regulation, at Article 12b(2a) (continued supply strictly necessary for divestment or wind-down, to 31 December 2026). Cited as the enacted Regulation.
7. Council Regulation (EU) 2026/506 of 23 April 2026 amending Regulation (EU) No 833/2014. Amends Article 5n by addition alone; 5n(1)(g) and (h) stand.
8. Financial Action Task Force. *Crowdfunding for Terrorism Financing.* FATF, October 2023. It mentions artificial intelligence once and does not treat AI providers as obliged entities.

LEGISLATIVE AND RULEMAKING RECORD

9. Bureau of Industry and Security. Proposed rule, RIN 0694-AJ35, 89 FR 5698, 29 January 2024 (IaaS Customer Identification Program). Comments closed 29 April 2024; withdrawn 16 December 2025. 9a. Bureau of Industry and Security. *Framework for Artificial Intelligence Diffusion.* Interim final rule, FR Doc. 2025-00636, 90 FR 4544, 15 January 2025. The sentence cited here is in the preamble at 90 FR 4555, § III ("Overview of New Controls for AI Model Weights"), in the closing "In sum" paragraph; the parenthetical gloss "(APIs)" is not part of it and comes from 90 FR 4547.
10. California SB 1047 (2024), § 22604(a). Vetoed 29 September 2024. § 22604(a)(1) requires "(A) The identity of the prospective customer. (B) The means and source of payment… (C) The email address and telephonic contact information used to verify the prospective customer's identity." Internet Protocol retention sits separately, at (a)(4). The statute says "telephonic contact information," not "telephone number."
11. California SB 53 (2025). Signed 29 September 2025; CalCompute.
12. Massachusetts S.37 (Sen. Finegold, filed 27 February 2025), § 3(a)(1).

13. Massachusetts S.2630, redraft of 16 October 2025; MassCompute.
14. North Carolina S735, "AI Innovation Trust Fund," First Edition (S735v1), 2025 session; § 143B-472.83C(a)(1) and § 143B-472.83C(c). Official text: https://www.ncleg.gov/Sessions/2025/Bills/Senate/PDF/S735v1.pdf. Introduced 25 March 2025; in Senate Appropriations since; never redrafted.
15. The Russia (Sanctions) (EU Exit) Regulations 2019, SI 2019/855 (consolidated text). It contains nothing on artificial intelligence, machine learning, cloud or high-performance computing.
16. Nearby instruments that do not reach model or compute access: California SB 243 (knowledge-standard trigger); Texas TRAIGA § 552.054; EU AI Act Article 55 and Article 2(3); Online Safety Act 2023, s.64.

COURT AND ENFORCEMENT DOCUMENTS

17. Criminal complaint, C.D. Cal. 5:25-mj-00400 (Palm Springs), charging 18 U.S.C. § 2339A, ¶¶31–33.
18. *United States v. Alvarez et al.*, No. 2:26-cr-00113-EAS (S.D. Ohio), Doc #33, filed 9 July 2026 (UFC Freedom 250). Neither the twelve-page indictment nor the underlying complaint against Proper makes any AI reference.
19. *United States v. Gann*, criminal complaint ¶5(q)(i), and indictment. The AI allegation survives only in the complaint.
20. Treason Act 1842 (UK) — charging statute in *R v Chail*.
21. Office of Foreign Assets Control, enforcement release — SAP SE, April 2021, $2,132,174.
22. Office of Foreign Assets Control, enforcement release — Exodus Movement, Inc., December 2025, $3,103,360.
23. Office of Foreign Assets Control, Specially Designated Nationals and Blocked Persons list (IUVM entry, "Linked To: Islamic Revolutionary Guard Corps (IRGC)-Qods Force"); and Civil Penalties and Enforcement Information, 2020–2026, every year read in full.
24. Cloudflare, Inc., Form 10-Q filed 8 May 2026 (SEC), disclosing OFAC's No Action Letter of 9 April 2026 closing its review "without penalties or further action."
25. OpenAI, letter to the Hon. Evan Solomon, Canada's Minister of Artificial Intelligence and Digital Innovation, 26 February 2026, signed Ann M. O'Leary, VP Global Policy. The primary document for the June 2025 account flag, ban, human review and non-referral, and the strongest cite available for it. It is also the source of the closing concession, verbatim: "under our enhanced law enforcement referral protocol, we would refer the account banned in June 2025 to law enforcement if it were discovered today." The letter is dated 26 February 2026, two months before Altman's 23 April 2026 apology.
26. Wells, Georgia. *The Wall Street Journal*, 20 February 2026. Broke the Tumbler Ridge / Van Rootselaar account flag and non-referral. Paywalled. The employee count and the "scenarios involving gun violence" detail rest on people familiar with the matter.

PEER-REVIEWED AND PREPRINT LITERATURE

Entries 41–46 are grey literature — research reports, think-tank commentary, self-citation. Each is labelled.

27. Gade, Pranav, Simon Lermen, Charlie Rogers-Smith and Jeffrey Ladish. *BadLlama: cheaply removing safety fine-tuning from Llama 2-Chat 13B.* arXiv:2311.00117. Submitted 31 October 2023. Disclosure this Dossier owes about its own sources: one team at Palisade Research, submitted three hours apart from arXiv:2310.20624; BadLlama withheld weights, data and methodology, so the result is reported and unreproduced.
28. Lermen, Simon, et al. *LoRA Fine-tuning Efficiently Undoes Safety Training in Llama 2-Chat 70B.* arXiv:2310.20624. ICLR 2024 Workshop on Secure and Trustworthy LLMs — a workshop paper. Submitted 31 October 2023. Same disclosure: one team at Palisade Research, submitted three hours apart from arXiv:2311.00117. Carries the ~$200 figure. 28a. Qi, Xiangyu, Yi Zeng, Tinghao Xie, Pin-Yu Chen, Ruoxi Jia, Prateek Mittal and Peter Henderson. *Fine-tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!* arXiv:2310.03693. ICLR 2024, accepted as an oral, per the OpenReview conference record. Source of the closed-model figures, verbatim from the abstract: "we jailbreak GPT-3.5 Turbo's safety guardrails by fine-tuning it on only 10 such examples at a cost of less than $0.20 via OpenAI's APIs." The benign-fine-tuning degradation is the paper's own title claim. 28b. Zhan, Qiusi, Richard Fang, Rohan Bindu, Akul Gupta, Tatsunori Hashimoto and Daniel Kang. *Removing RLHF Protections in GPT-4 via Fine-Tuning.* arXiv:2311.05553. NAACL 2024, per the paper's own arXiv comment ("Accepted to NAACL 2024"). Verbatim: "remove RLHF protections with as few as 340 examples and a 95% success rate."
29. Arditi, Andy, et al. *Refusal in Language Models Is Mediated by a Single Direction.* arXiv:2406.11717.
30. Joad, et al. *There Is More to Refusal in Large Language Models than a Single Direction.* arXiv:2602.02132.
31. Casper, Stephen, David Krueger and Dylan Hadfield-Menell. *Pitfalls of Evidence-Based AI Policy.* arXiv:2502.09618.
32. Egan, Janet, and Lennart Heim. arXiv:2310.13625. Includes "Box 3: Learning from KYC in the financial sector"; aimed at compute and export-control entities of concern, not designated armed groups, and silent on payment.
33. Zhang, et al. arXiv:2602.23329, February 2026. Novices with LLMs 4.16× more accurate (95% CI 2.63–6.87) on *in silico* biosecurity tasks.
34. Hong, et al. arXiv:2602.16703, February 2026. Pre-registered, investigator-blinded wet-lab trial, n=153; 5.2% workflow completion with an LLM against 6.6% with internet alone, P=0.759.
35. *Virology Capabilities Test.* arXiv:2504.16137.
36. Brent, and McKelvey. *Contemporary Foundation AI Models Increase Biological Weapons Risk.* RAND PE-A3853-1, December 2025. A RAND Perspective — not peer-reviewed.

37. Glazzard, Andrew, David McIlhatton and Paul Martin. “Will Generative AI Fundamentally Change Terrorist Threats?” *CTC Sentinel* 19:5, May 2026, pp. 25–31. Combating Terrorism Center, West Point. Issue PDF: https://ctc.westpoint.edu/wp-content/uploads/2026/05/CTC-SENTINEL-052026.pdf. The sentence cited here is at p. 27: “Terrorists do experiment, and thereby develop novel approaches, but they rarely do so at the frontiers of knowledge.” The constraint on experimentation the article names is knowledge and skill, reinforced on the same page by the note that CBR weapons still require “a great deal of skill and knowledge.”
38. Hamming, Tore R. “The General Directorate of Provinces: Managing the Islamic State’s Global Network.” *CTC Sentinel* 16:7, July 2023, pp. 20–27. https://ctc.westpoint.edu/wp-content/uploads/2023/07/CTC-SENTINEL-072023.pdf. Source of the “monthly reports on the ongoing military situation…” passage, at printed p. 23, under the section heading “Military Affairs”; the ellipsis elides only “in their respective region.” Note that the CASP bibliography at entry 1, p. 80, gives the page range as 20–29; the issue ends at printed p. 27.
39. Weimann, Gabriel, Alexander T. Pack, Rachel Sulciner, Joelle Scheinin, Gal Rapaport and David Diaz. “Generating Terror: The Risks of Generative AI Exploitation.” *CTC Sentinel* 17(1), January 2024, pp. 17–24. A jailbreak-testing study: the authors put 2,250 prompts to three platforms and report success rates by prompt type. Every reference to a named group in it sits inside a prompt the researchers wrote; it documents no instance of a group using the tools.
40. Martini. *Review of International Studies*, 24 March 2026. DOI 10.1017/S0260210526101843. https://doi.org/10.1017/S0260210526101843
41. Hunter, Connor A. Stewart, and Luca Righetti. *AI and Bomb Plots: Distinguishing Potential Effects from Language Models.* Centre for the Governance of AI, technical report, March 2026. https://govai.b-cdn.net/AI_and_Bomb_Plots_Distinguishing_Potential_Effects_from_Language_Models.pdf. Grey literature, and the report says so on p. 1: “GovAI technical reports have received extensive feedback but have not gone through formal peer review.” Its conclusion carries a global hedge this Dossier should not outrun: “The analysis in this report should be cautiously considered, as the data on terrorist bomb plots is necessarily limited and our findings remain exploratory.” Carries the “likely limited” assessment (p. 3), the 7/7–7/21 London comparison (Box 2, p. 13), and the informants-and-surveillance detection channel (p. 20). Printed folio = PDF page minus one.
42. Epoch AI. Epoch Capabilities Index — open-weight lag behind frontier closed models. Four-month average since January 2026, against a three-month average for January 2023 – October 2025. Figures as of 20 July 2026. Not peer-reviewed.
43. UK AI Safety Institute. Cyber-specific open-to-frontier capability gap, narrowed from six-to-ten months to four-to-seven. As of 20 July 2026. Government research body, not peer-reviewed.

44. Foucher, Vincent. ISPI, 16 July 2025. Defector testimony on the ISIL trainers' arrival in the Lake Chad basin. Foucher also peer-reviewed entry 1 — noted so the overlap is visible.
45. Samuel, Malik. "From the Levant to Lake Chad: ISIS Fighters Fuel ISWAP Resurgence." Good Governance Africa, 30 May 2025. https://gga.org/from-the-levant-to-lake-chad-isis-fighters-fuel-iswap-resurgence/. Cited as field reporting on the seven instructors. Not independent of entry 1: Samuel is upstream of it, cited there at p. 36 fn. 27 for the seven-instructor syllabus. 45a. Roggio, Bill. "US, Nigerian forces kill senior Islamic State leader." *FDD's Long War Journal*, 16 May 2026. https://www.longwarjournal.org/archives/2026/05/us-nigerian-forces-kill-senior-islamic-state-leader.php. Sole source for the killing of the head of al-Furqan, which it places in northeastern Nigeria on 16 May 2026. It renders the name "Abu Bilal al Minuki" — AFRICOM's form — and quotes the UN Monitoring Team's February 2026 report on that man verbatim ("assumed a more prominent role within the global [Islamic State] leadership, with some suggesting he may have become head of [the] General Directorate of Provinces"), which is S/2026/44 ¶7; that is the identity link between the two renderings. See also S/2026/44 ¶23: "the elevation of Abu Bakr ibn Muhammad ibn Ali al-Mainuki (not listed), head of the al-Furqan office of ISIL, to ISIL (Da'esh) core leadership." AFRICOM further billed him the Islamic State's "number two," and the piece records that analysts who track ISWAP doubt that billing. It sits awkwardly with entry 1, which at p. 37 has al-Furqan "apparently headed by ISWAP leader Habib Yusuf (Foucher 2024, 13–14), following the recent death of former co-leader Abubakar Mainok (Samuel 2026b)" — a compatible succession claim, except that whether "Abubakar Mainok" is the same man as al-Mainuki is unsettled.
46. Mitzpe Institute, prior publications — self-citation, not independent authority. "The Jihadist Continuum," *Mitzpe Dossiers*, 18 December 2025, https://mitzpe.org/dossiers/the-jihadist-continuum (cited in §3); "Two Middles," *Mitzpe Field Dossiers*, 24 April 2026, https://mitzpe.org/dossiers/two-middles (cited in §4 for the FARA-filed Clock Tower X LLC contract).

PROVIDER DOCUMENTATION AND VENDOR RESEARCH — THE INTERESTED PARTY'S OWN ACCOUNT, NOT INDEPENDENT AUTHORITY

Every item here is written by a party with a commercial or reputational stake in what it says — provider terms, vendor marketing, security-vendor research. Cited for what the party published, never as independent verification of it.

47. Persona. Customer case study on OpenAI quoting OpenAI integrity product lead Jake Brill on sanctions screening across 225 countries and territories. An undated vendor marketing page quoting a customer employee. Observed 20 July 2026; cited only for the Brill quote. No OpenAI-side confirmation exists.
48. Shrivastava, Rashi (staff). *Forbes*, 30 April 2025. Reports the 225-country screening claim for ChatGPT and the API. A Persona funding profile built

on an interview with Persona's chief executive — edited outlet, named staff reporter, but vendor-sourced. Figure as reported 30 April 2025; observed 20 July 2026.

49. Anthropic, access-control terms. Consumer Terms of Service, effective 8 October 2025, §2 (credential sharing; making an account available to others), §3 (resale), §12 (Export Controls; embargoed countries; OFAC SDN list; Commerce Entity List); support.claude.com article 8287232, "Verify your phone number," dateModified 19 May 2026 (no VoIP, no skip, one permanent unchangeable binding per account) and article 14328960, "Identity verification on Claude," dateModified 17 June 2026 (Persona; "a valid government-issued photo ID: the physical document, in hand"; live selfie; rollout framed as "routine platform integrity checks"), which are the sources for the phone binding and the discretionary ID check — the second, with trust.anthropic.com/subprocessors ("Persona: Claude Free/Pro/Max"), being the only Anthropic surface naming the vendor; and supported-regions documentation — three lists on three hosts, at 184, 175 and 177 entries as of 20 July 2026. The two that disagree carry no date; the dated one (16 March 2026) is the third. Nine countries separate the first two: Central African Republic, Eritrea, Ethiopia, Libya, Mali, Nicaragua, Somalia, South Sudan, Sudan.
50. OpenAI, access and identity documentation. Supported countries and regions — 208 entries for consumers and 188 for the API as of 20 July 2026; Verified Organization documentation (API identity verification); and the OpenAI Privacy Policy, which does not mention sanctions, screening or OFAC, and whose subprocessor URL 404s. Checked 20 July 2026.
51. Google Cloud. Signup FAQ, `https://cloud.google.com/signup-faqs` ("we ask for your name, address, and payment method to verify your identity…"); Google account verification methods, `https://support.google.com/accounts/answer/10071085` — the five proofing methods, including government-ID upload with manual review, credit-card authorisation, email-history inference and digital ID (Android-only), and naming no third-party vendor; the AI Studio block destination, `https://ai.google.dev/gemini-api/docs/available-regions` — three grounds for denial, the third being age ("you haven't yet verified your age"); and Google Cloud Billing documentation, "(India only) Verify your identity" (Aadhaar / DigiLocker, beneficial-ownership-grade KYC). Observed 20 July 2026.
52. Anthropic, safety documentation. Usage Policy (terrorism named). Responsible Scaling Policy, version 2.2, effective 14 May 2025, 23 pp.
    - The CBRN-3 quotation is from v2.2's capability-threshold table. v2.2 names the *threshold* CBRN-3; ASL-3 is the Deployment and Security Standard that crossing CBRN-3 requires. The two are not interchangeable.
    - v2.2 reads: "CBRN-4: The ability to substantially uplift CBRN development capabilities of moderately resourced state programs (with relevant expert teams)…" It does not itself define the ASL-4 standards: "We expect this threshold will require the ASL-4 Deployment and Security

Standards. We plan to add more information about what those entail in a future update."

- v2.2 has since been superseded. The v3 series carries no CBRN-4 threshold — the label survives only inside the changelog describing what v2.1 once added — and v3.4, effective 8 July 2026, reframes the live threshold as "Non-novel chemical/biological weapons production." v2.2 is cited because it was the policy in force across the window this Dossier covers; it is not the policy in force at publication.

52a. Anthropic. *System Card: Claude Opus 4.6.* February 2026, 213 pp. (revision carrying a 6 February 2026 changelog). The mustard-gas passage is at §6.2.5.3, on pilot investigations in a sandboxed GUI computer-use environment: "In one case, the auditor was able to elicit Claude Opus 4.6 to provide detailed instructions in an Excel spreadsheet for producing mustard gas." The failure occurred in GUI computer-use testing rather than text-only interaction, and Anthropic draws the inference itself — "our standard alignment training measures are likely less effective in GUI settings" — adding that Opus 4.5 "yielded similar results, suggesting that this gap is not new."

53a. OpenAI. *Preparedness Framework, Version 2*, last updated 15 April 2025, 22 pp. The passage cited here appears once, in §C.1, "Safeguards against malicious users": "Several of the Tracked Categories pose risks via malicious users leveraging the frontier capability to enable severe harm, such as professional hackers automating and scaling cyberattacks or terrorists consulting a model to debug the development of a biological weapon." It is an illustrative example inside the safeguards section, not a capability threshold. The word "terrorist" appears once in the whole document. 53. OpenAI, safety documentation. Preparedness Framework v2 — the "terrorists consulting a model to debug the development of a biological weapon" passage, at §C.1, "Safeguards against malicious users." An illustrative example inside the safeguards section, not a capability threshold; "terrorist" appears once in the whole document. See entry 53a. Also: "Building an early warning system for LLM-aided biological threat creation," 31 January 2024 (the provider's own argument that wet-lab access and expertise "are more likely to be the bottleneck"). 54. Vendor threat-intelligence corpus, February 2024 – February 2026: Microsoft with OpenAI; OpenAI ×6; Google Threat Intelligence Group ×3; Anthropic ×3 — thirteen reports. Read for the negative that none names a designated terrorist organisation. Includes OpenAI's February 2025 report ("While we cannot determine the locations or nationalities of the actors…"), its IUVM ban and its CyberAv3ngers assessment. Curated disclosures shaped by legal and commercial considerations, cited as such. 55. Google Threat Intelligence Group. *Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access.* 12 May 2026. https://cloud.google.com/blog/topics/threat-intelligence/ai-vulnerability-exploitation-initial-access. A standalone GTIG report, not an "AI Threat Tracker" edition. Carries the UNC6201 premium-account-automation finding; UNC6201 does not appear in the November 2025 or February 2026 reports. GTIG's own reading, quoted here, is that the actors were procuring "high-tier AI capabilities at

scale while insulating their malicious activity from account bans." 56. Pillar Security. "Operation Bizarre Bazaar," 28 January 2026. The Unified LLM API Gateway; a reseller on Netherlands bulletproof hosting. 57. Amazon Web Services. `AWSCompromisedKeyQuarantineV2` managed policy, edited 2 October 2024 to deny `bedrock:InvokeModel`. 58. Blazakis, Jason. *Terrorist Financing in the Age of Large Language Models.* Project CRAAFT II Research Briefing No. 4, RUSI Centre for Financial Crime and Security Studies, 11 February 2026, 14 pp. https://www.projectcraaft.eu/s/craaft-II-bp4-terrorist-financing-llms-february-2026.pdf. His phrase "baseline compliance architectures work as intended" is governed by *suggesting*, its subject is "the findings from this limited prompt testing," and it is prefaced as "modest reassurance" — and the very next sentence reverses it: "But the findings also underscore how narrow such testing can be… The true challenge lies not in blocking overtly criminal requests but in detecting and disrupting the persuasive infrastructure." Watermarking is one clause inside the first of five recommendations, not the paper's turn. What the briefing does not reach is the account and payment layer: subscriptions, payments, customer due diligence, sanctions screening and the SDN list go untreated. He stood at the door and did not open it; he did not pronounce the door sound. Note also that his endnotes cite gemini.com — Gemini Trust Company, not Google. 59. Tech Against Terrorism. CT-AI Benchmark, launched at the UN, early July 2026; 27 models, ~2,500 prompts. The full report is not public, and TAT measured compliance, not usefulness. This Dossier cites no figure from it: the rankings and compliance rates rest on a methodology nobody outside TAT can check, so this Dossier takes the policy argument and declines the counts. Also TAT, 2023: users "likely exploited free tools to mitigate the risk of identification via payment for paid services." 60. Global Network on Extremism & Technology (GNET). Reporting on extremists cost-sharing subscriptions. No author, title, date or URL is held for this item — it is the only entry in this apparatus naming no document. 60a. Siegel, Daniel. "AI Jihad: Deciphering Hamas, Al-Qaeda and Islamic State's Generative AI Digital Arsenal." GNET, 19 February 2024. https://gnet-research.org/2024/02/19/ai-jihad-deciphering-hamas-al-qaeda-and-islamic-states-generative-ai-digital-arsenal/. Source for the "Israeli Diaper Force" propagation. Siegel attributes the *verbal* claim to Abu Obeida of the Al Qassam Brigades and describes the imagery in the passive, by platform, naming no creator. Two other sentences in the same piece loosen toward attribution ("one by Hamas"; "amplified by Hamas through AI-generated content"), so he is careful about creation but not uniformly careful — which is where ICCT's reading found something to grip. Not to be confused with Criezis, "AI Caliphate" (GNET, 5 February 2024), a different piece. 60b. Nelu, Clarisa. "Exploitation of Generative AI by Terrorist Groups." International Centre for Counter-Terrorism (ICCT), 10 June 2024, tagged "Analysis / Short Read." https://icct.nl/publication/exploitation-generative-ai-terrorist-groups. The restatement node in the chain. Verbatim: "The content appeared to be generated by a Hamas-affiliated group using generative AI, and it aimed to undermine the Israeli army." ICCT never writes "Hamas created." The chain is documentary rather than inferred: the only citation ICCT attaches to that sentence is entry 60a,

a piece that names no creator. 61. Open-weight distribution platforms, all figures as of 20 July 2026. Ollama, model page for Llama 3.1 — 117.4M downloads across 93 tags, a cumulative counter published with no de-duplication methodology and not a count of users. Hugging Face, abliterated-variant survey — roughly 7,000 uploads resolving to about 3,378 variant names across 501 base models; one uploader accounts for 24 percent of the set, the top twelve for 39 percent. This survey is our own, run against the Hugging Face model index on 20 July 2026, not a third party's: `search=abliterated` returned 7,049 and the `other=abliterated` tag filter 8,487, overlapping 3,093 for a union of 12,443; of the 7,049, some 75 percent carry a quantisation or format tag (GGUF 3,744, MLX 762), and name-normalised de-duplication gives the 3,378 distinct names.

PRESS AND PRIMARY TEXT

Independent journalism and a primary religious text. This group exists because the material belongs to neither tier above.

62. Shea, Matt. BBC, 12 April 2026. The "Lego-style" Iranian generative-AI wave; a production outfit's representative calling the Iranian government a "customer" placing "direct commissions for several projects." The only documented regional positive in this Dossier rests on it.
63. *The Christian Science Monitor.* Iranian production houses with "some IRGC connection."
64. Israel National Cyber Directorate (Yossi Karadi). Attribution of the propaganda campaign to "the cyber war waged by the IRGC."
65. Tasnim. Amplification of the "Lego-style" wave. A hostile state-aligned outlet, cited for what it amplified and never as authority.
66. Mukhtar, Khaled. "New Technologies in Houthi Drones." *The Yemen Review, January–March 2025*, Sana'a Center for Strategic Studies, 21 April 2025, 6 pp. `https://sanaacenter.org/the-yemen-review/jan-mar-2025/24604` (PDF: `https://sanaacenter.org/files/New_Technologies_in_Houthi_Drones_en.pdf`). The quoted definition is at p. 4. Author writes under a pseudonym for security reasons; series funded by the government of the Netherlands.
    What it supports is operator-controlled FPV only, via its own definition — FPV drones "provide their operators with real-time video-feed access to the drone's flight trajectory and surrounding environment." Its subject matter is a February 2025 seized shipment: JetCat P500/P550-PRO-S jet engines, DJI MAVIC3 airframes, TARANIS X9D control systems. It makes no navigation- or seeker-architecture claim at all; its single mention of navigation is "maritime navigation," in the Implications section.
67. Al-Fahd, Sheikh Nasser bin Hamad. "A Treatise on the Legal Status of Using Weapons of Mass Destruction Against Infidels." *Fatwa*, May 2003. CASP does not quote the *fatwa* — it names and characterises it, at p. 60, in Juelich's own voice ("The last part is in line with retribution as a justification for WMD on which al-Qaeda and ISIS operatives have drawn, especially the fatwa entitled..."). No independent copy is held, and no

text of the *fatwa* is held at any remove. The attribution is researcher-inferred and never respondent-named: the interviewee supplied only the retribution reasoning.

ADDITIONAL DOCUMENTS AND REPORTING

68. Israel Defense Forces, Spokesperson's response to the Lavender reporting, 3 April 2024. No IDF-hosted publication of this statement was found. Its substance: the IDF "does not use an artificial intelligence system that identifies terrorist operatives or tries to predict whether a person is a terrorist," and the "'system' your questions refer to is not a system, but simply a database whose purpose is to cross-reference intelligence sources... This is not a list of confirmed military operatives eligible to attack." Fullest published text: *The Guardian*, "Israel Defence Forces' response to claims about use of 'Lavender' AI database in Gaza," 3 Apr 2024, 14.53 BST. https://www.theguardian.com/world/2024/apr/03/israel-defence-forces-response-to-claims-about-use-of-lavender-ai-database-in-gaza — the Guardian's copy is addressed to the Guardian's own query set ("your questions," twice), so it answers one outlet, not the public. CNN (3 Apr 2024) reproduces the same English independently. +972's rendering is materially different English and matches neither: consistent with a Hebrew statement to +972/Local Call and a separate English one to the Guardian, or with tailored per-outlet responses. The Guardian's partner status is asymmetric, and only its own claim is carried here: it says the testimony was "shared exclusively with the Guardian in advance of publication," while +972 never mentions the Guardian and credits only Local Call, whose "In partnership with" logo links to mekomit.co.il. *Times of Israel* (6 Apr 2024) truncates inside the quotation marks and dates the statement to 5 April; the discrepancy is unresolved.
69. Abraham, Yuval. "'Lavender': The AI machine directing Israel's bombing spree in Gaza." *+972 Magazine* and *Local Call*, 3 April 2024. https://www.972mag.com/lavender-ai-israeli-army-gaza/. Source of the "20 seconds" figure, which the source scopes to confirming the target was male, described as "the only human supervision protocol in place before bombing the houses of suspected 'junior' militants." The Guardian's companion piece (Davies and McKernan, 3 Apr 2024) states the testimony was "shared exclusively with the Guardian in advance of publication"; +972 does not mention the Guardian anywhere, and credits only Local Call.
70. Dwoskin, Elizabeth. "Israel built an 'AI factory' for war. It unleashed it in Gaza." *The Washington Post*, 29 December 2024. Carries the "three minutes to five hours" vetting figure — the *Post*'s own separate investigation, describing analysts vetting output from both the Gospel and Lavender. Not April 2024, and not a restatement of +972. Cited here to date the figure correctly; not used in the body, because pairing it with "20 seconds" would compare two different processes measured by two outlets nine months apart.
71. Neifakh, Veronica. "Terrorists Exploit AI for Propaganda and Operations, Exposing Critical Gaps in Tech Safeguards." *The Media Line*, 21 November 2024. https://themedialine.org/top-stories/terrorists-exploit-

ai-for-propaganda-and-operations-exposing-critical-gaps-in-tech-safeguards/. Source of the conditional hypothetical, spoken by Gabriel Weimann (University of Haifa). Full quotation: "If you're running a campaign for Hamas, Hezbollah, al-Qaida, or ISIS, you can use AI tools to create distorted images, fake news, deepfakes, and other forms of disinformation. This isn't hypothetical—it's already happening." The speaker does not leave it conditional; the outlet's own pull-quote block drops the second sentence and the closing period. Weimann is also lead author of entry 39, so the first two "independent" nodes in that chain are one man.

72. Google Threat Intelligence Group, the degradation pair — both dates pinned. (a) *Adversarial Misuse of Generative AI*, 30 January 2025, https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-ai — "current LLMs on their own are unlikely to enable breakthrough capabilities for threat actors." (b) *GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools*, 6 November 2025, https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools — "adversaries are no longer leveraging artificial intelligence (AI) just for productivity gains, they are deploying novel AI-enabled malware in active operations." (b) names (a) as its predecessor. The pair is not a flat reversal: the same November report holds its information-operations assessment explicitly unchanged — "we did not see evidence of successful automation or any breakthrough capabilities… similar to our findings from January" — and repeats that a third time in the 13 February 2026 edition. GTIG upgraded on AI-enabled malware and held steady on information operations across all three editions.
73. Bureau of Alcohol, Tobacco, Firearms and Explosives — no ATF document carries this; the source is the recording. Kenneth Cooper, Assistant Special Agent in Charge, ATF San Francisco Field Division, spoken remarks at joint LVMPD/FBI/ATF press conferences, LVMPD headquarters, Las Vegas, 2 and 7 January 2025. The "level of sophistication" quotation is on the C-SPAN recording of the 2 January press conference (closed captions, 4:13–4:14 pm EST), and is corroborated by Task & Purpose (Nieberg, 2 Jan 2025) and AP (Copp, Durkin Richer and Long). The recording carries two hedges the coverage dropped: Cooper prefaced the line with "I want to be careful with my language here, but," and his "too early" caveat attaches specifically to whether the components were wired for a coordinated detonation, not to the assessment at large. That quotation is from 2 January; the other three Cooper strings used here are from 7 January.
74. *Joshi v. OpenAI Foundation*, No. 4:26-cv-00222-MW-MJF (N.D. Fla.), Complaint, Doc. 1, filed 10 May 2026, 76 pp. Vandana Joshi as personal representative of the Estate of Tiru Chabba; eleven OpenAI entities plus Phoenix Ikner as defendants. Carries the 11:30–1:30 exchange at ¶109, p. 23, and the media-threshold exchange at ¶108. Related, and not consulted here: *Mall v. OpenAI Foundation*, 4:26-cv-00334, and *Askins v. OpenAI Foundation*, 4:26-cv-00335, both filed 14 July 2026 (CourtListener holds the dockets; no free document text). 74a. Brownlee, Briana. "1 year after FSU shooting, records reveal suspect's ChatGPT messages as victims

are honored." News4JAX / WJXT, 17 April 2026, section "New details from records." https://www.news4jax.com/news/florida/2026/04/17/1-year-after-fsu-shooting-records-reveal-suspects-chatgpt-messages-as-victims-are-honored/. Source of the Florida State Attorney's Office records and the 17 April 2026 date; corroborated on the custodian by News4JAX, "Florida attorney general targets OpenAI over ChatGPT's role in FSU campus shooting," 21 April 2026. The figure it gives is "more than 13,000," and what it describes are prosecution records released at the one-year mark, not messages admitted at trial: trial is set for October 2026, after this Dossier publishes.

75. 18 U.S.C. § 2339B(a)(1), (g)(4), (h). The knowledge element is disjunctive. § 2339A(b)(1) defines material support to include "any property, tangible or intangible, or service," carving out only medicine and religious materials. § 2339B(h) limits the "personnel" prong alone and is not a defence to a service theory.
76. National Telecommunications and Information Administration, U.S. Department of Commerce. *Dual-Use Foundation Models with Widely Available Model Weights: NTIA Report.* July 2024; released 30 July 2024; 72 pp. https://www.ntia.gov/sites/default/files/publications/ntia-ai-open-model-report.pdf. Recommends the government "should not restrict the wide availability of model weights for dual-use foundation models at this time" and instead "actively monitor and maintain the capacity to quickly respond to specific risks" (p. 46), with a three-part collect-evaluate-act framework at p. 47 and a Monitoring Template appendix at p. 48. The report is the assignment, not the audit: nothing in it speaks to whether the monitoring it recommends was ever carried out. EO 14110, which commissioned it, was rescinded 20 January 2025 and replaced by EO 14179 on 23 January 2025.

Tip? Share it securely via **signal: (@Uri.30)** or **proton: (uri.zehavi@proton.me).**

[FREE_PIECE_CTA]

ABOUT THE AUTHOR

Uriel Zehavi

EXECUTIVE DIRECTOR · MITZPE INSTITUTE

Uriel Zehavi is the executive director of Mitzpe Institute, where he writes the daily Israel Brief and co-directs its Institutional Intelligence practice with Modi Zehavi. He is the author of four books from Mitzpe Press and advises Israeli and allied clients on strategic communications for Western audiences.

ABOUT MITZPE INSTITUTE

Mitzpe Institute is an intelligence and strategy organization focused on Israeli security, diaspora threat assessment, and the strategic challenges confronting the Jewish people and the West. We produce daily intelligence, forensic analyses, strategic assessments, and tactical briefings read by policymakers, military leaders, and institutional decision-makers across more than 90 countries and all 50 US states — from the Knesset and the US Congress to allied governments, major policy institutions, and Jewish organizational leadership worldwide.

PASSED ALONG TO YOU?

The daily Israel Brief is free at mitzpe.org. The operational layer — Closed Dossiers, Vantage, and Watchwords — is members-only. Become a member, gift a membership, or book a briefing at mitzpe.org.