

Beyond the Security Stack

Most accounting firms are equating security to purchasing the right software. Real security means much more.

Chapters

1 Introduction

2 Security controls need the same diligence as financial controls

3 Six essential tips to protect your firm

4 New standard: what the IRS expects from AI in the tax profession

5 What causes AI data leakage?

6 What to demand from your AI vendors

7 How Truss approaches these challenges

8 The bottom line

Ken Kantzer,

CTO and Co-Founder of Truss



Truss has run modern AI in production since 2023, so the guidance contained herein comes from operational reality rather than theory. Co-founder and CTO Ken Kantzer has spent 15 years in cybersecurity and engineering, leading security teams and building software for firms that handle deeply sensitive data. He was previously Vice President of Engineering at FiscalNote, a technology company specializing in global policy and market-intelligence software. He also was a founding partner at PKC and led all security consulting efforts there, which included developing secure applications for companies of every shape and size. Earlier, he was a Senior Consultant at Booz Allen Hamilton, working in cybersecurity across the defense, government systems, industrial infrastructure, and oil and gas sectors.

Introduction

Most accounting firms are equating security to purchasing the right software. Real security means much more.

Operational discipline vastly outstrips technical sophistication when it comes to security in tax and accounting. A well-designed security system has continuity; every control reinforces the next and every tool earns its keep. In weak designs, a firm can use the best tools on the market and remain exposed.

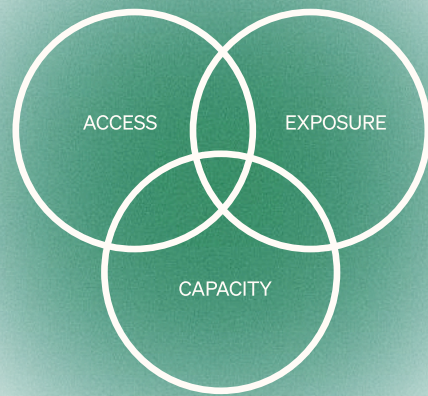
Tax and accounting firms own some of the richest data sets in the world economy: tax returns, SSNs, EINs, bank details, payment records, and ownership structures. A breach can be catastrophic, and is typically caused by a lack of verified controls, not a missing firewall.

There has been a wave of AI security research in the last three years, and it confirms that the controls are the key to preventing breaches. In some instances, AI assistants embedded within collaboration tools could be influenced by hidden instructions found within ordinary documents and emails. The assistant didn't break encryptions or bypass logins and firewalls. It followed instructions to retrieve data it already had permission to access.

One important lesson from those audits is that the worst vulnerabilities are hiding in plain sight. Across dozens of engagements, the truly dangerous findings were the ones that were easiest to discover and exploit.

Attackers are lazy and reach for the lowest-hanging fruit. Discoverability is what predicts exposure, and firms that stay safe are the ones who keep things simple to make sure the truest, most basic fundamentals are put in practice.

If access, exposure, and outbound capability coexist without safeguards, the system is destined to fail. Researchers call that the lethal trifecta.



Security Controls Need the Same Diligence as Financial Controls

Think of the way a smart firm protects its books: no individual handles everything, accounts get reconciled, and work gets double-checked. Security needs those same layers of protection.

Policies without monitoring are just strongly-worded suggestions. You can set up a perfectly encrypted client portal, but that still leaks if staff downloads returns onto their personal laptops. Multifactor authentication is great, but it only works until a staffer shares a public link that circumvents it.

Cross-tenant data is particularly vulnerable, as the dividing lines between multiple customers can become easily blurred. If one customer's data leaks into another's, it's rarely a break-in, but rather the system doing what it was told. The design didn't institute strict enough separation, leaving a path for one firm to view client records that belonged to another. Real security happens when controls back each other up.

When one customer's data leaks into another's, it's rarely a break-in — it's usually that the boundaries between accounts were drawn too loosely. The system did what it was told; the design just didn't demand strict separation. Real security comes from controls that back each other up.

Six Essential Tips to Protect Your Firm

- 1. Know your hardware visibility first.** If you can't see it, you can't secure it. Inventory everything that comes into contact with your network: home laptops, seasonal contractor devices, scanning stations, legacy tax-engine servers, NAS drives, cloud backups, outsourced IT, and payroll integrations. In 2013, an attack on Target exposed 40 million+ payment cards and data on roughly 70 million people, costing the company over \$200 million. The hackers entered through a third-party HVAC vendor's remote-access system.
- 2. Know your software, reduce the doors.** Every application you install could be an entry point, and most firms accumulate them without noticing. In 2017, the NotPetya attack spread through a compromised update to a Ukrainian accounting product called M.E.Doc, causing an estimated \$10 billion+ in global damage. Maersk alone reported roughly \$250-300 million in losses. That extra PDF tool or portal-login browser extension could open the gateway to disaster.
- 3. Security must be continuous.** A threat actor doesn't operate according to the tax calendar. In 2023, attackers exploited a vulnerability in the MOVEit file-transfer tool and affected hundreds of organizations and millions of individuals. That vulnerability existed long before it was exploited, and a critical weakness left open for three months is far more dangerous than one closed in three days.
- 4. Limit privilege and contain the blast radius.** It happens quietly: temporary access becomes permanent, administrative rights get handed out left and right, shared credentials become a matter of convenience. The 2017 Equifax breach began with an unpatched server, but broad internal access turned it into a large-scale catastrophe. Roughly 147 million people were affected and the company had to pay a settlement of roughly \$700 million, because one compromised account can create a huge blast radius.

5. **Make the secure path the default.** Busy people get careless, so any control that depends on paying careful attention is more likely to fail. The safest configuration should happen automatically: MFA is mandatory, encryption is on, and files are private unless deliberately shared. Firm incidents usually come down to pressure mistakes, like overly broad folder permissions or portals that are so tricky to use that staff revert to using email attachments.
6. **Log everything and detect early.** Most breaches are detected weeks after they begin; some even take months. Logging allows you to keep track of who accessed what, when they did it, from where, and what changes over time. In 2023 and 2024, researchers demonstrated that Microsoft 365 Copilot could be influenced through hidden instructions embedded in emails or documents. Without logging and monitoring, that retrieval could blend into normal use.

New Standard: What IRS Expects from AI in tax profession

What was once experimentation without guidance is swiftly coming to a close. In June, the IRS Office of Professional Responsibility released its first-ever guidelines for tax professionals using generative AI. The message was one any reasonable person could agree upon in principle, but the specific technical warnings were crucial for firms to take to heart and follow up on with their vendors.

In particular, the IRS' warning on cross-client data leakage was the most noteworthy thing to come from the guidelines.

In the IRS' own words client privacy can be compromised *“when data generated for one client is repurposed by the program to respond to an inquiry concerning another client, or data compiled for a particular issue is spilled over into an algorithm and combined with a related tax issue involving a different client.”*

Imagine a preparer asking their AI assistant about a rental loss and the system bringing back figures from a different taxpayer’s return because it “learned” from that earlier case. That’s a direct consequence of a tool built on a pool of shared client information or memory, and its ramifications are potentially catastrophic.

A failure to address spillover can simultaneously implicate Circular 230 confidentiality duties, Section 7216 regulations on the use and disclosure of taxpayer information, and the Written Information Security Plan (WISP) that preparers now attest to at PTIN renewal. So if a vendor sold you a product with AI security shortcuts, it’s likely that product will quickly become your firm’s compliance headache.

What causes AI data leakage?

Successful systems reason in the moment rather than remembering past context. And there are three key design choices that can create risk.

	DURING ONE REQUEST	ACROSS SESSIONS	INDEFINITELY
Loose context handling Even without training or shared storage, systems can leak by loading the wrong records into a request.	Ends with the request		
Shared memory or storage Some products use “stateful” endpoints that retain context across sessions and users. Those may be convenient, but they also store client data in a place that’s too easy for other requests to reach and spill over into unwanted places.	Lives until someone purges it		
Training or fine-tuning on customer data When vendors feed client data into their model to “improve” it once, the information continues to influence the model indefinitely moving forward. This is the most permanent form of leakage and simultaneously the hardest to undo.	Baked into the model		

None of these principles are inherent to the use of AI in tax-related software, or any more general use you might encounter. They are architectural features that thoughtful designers avoid, and the most thoughtful of all designers worked around them from the start. The recent IRS guidance essentially invites firms to make AI software vendors show their work.

What to Demand from Your AI Vendors

The controls we discussed in this piece are internal and can apply to any business handling data worth protecting. But most accounting firms will meet AI through one of its vendors, and the IRS is putting those vendors under a microscope. So here's a checklist to evaluate tools that touch taxpayer data that you can demand your AI vendors match 100%.

- **No training on data.** Vendors and any of their underlying model providers should state in writing that client data is never used to train, fine-tune, or adjust model weights. This is the only way to defend against the ever-dangerous cross-client data leakage.
- **Stateless, non-retaining processing.** Endpoints that receive requests should not retain the data involved indefinitely. Retention is only acceptable if it's temporary and limited to abuse and safety monitoring.
- **Per-request context segmentation.** Requests should only receive the minimum data required to complete their corresponding tasks. That data can be drawn from only one client, and every client's context must be kept separate from all others.
- **Prompt injection defenses.** When the system reads documents from untrustworthy sources, it must guard against hidden instructions attempting to extract or cross-contaminate data.
- **Encryption and U.S. data residency.** All traffic in transit must be encrypted, and firms handling U.S. taxpayer data should verify where processing happens, with offshore processing subject to stricter scrutiny.
- **Human review and transparency.** Practitioners must review every AI-generated document before it goes to a client or the IRS, and they must be transparent with clients about how AI is used. The IRS was explicit about AI not being a replacement for sound judgment.
- **Independent attestation.** A SOC 2, Type II report audited by a third party verifies that controls are actively tested over time, rather than making claims.

How Truss approaches these challenges

AI was a staple component of Truss from the start in 2022, so we thought about these questions before shipping anything. Our goal was to be proactive about identifying the issues that could crop up in the future as a result of increased automation in the tax profession, because we knew that trend would be central to our livelihood as a business.



No training on customer data

Truss does not train, fine-tune, or adjust model weights on customer data, and neither do our model providers.



Stateless endpoints

Requests run through non-retaining endpoints, with only the temporary retention required for safety monitoring.



Per-request data separation

Max AI receives only the records needed for each request, with every client's data separated from all others', preventing the spillover the IRS is most concerned about.



Prompt injection controls

We control for prompt injection whenever a document arrives from an untrusted source.



Encrypted, U.S.-based processing

All traffic is encrypted in transit and processed on U.S.-based infrastructure.



SOC 2, Type II audited

Controls tested over time and attested by an independent third party, not just claimed in marketing copy.

Our system does not “learn” your firm. It reasons over the case in front of it during a request and forgets what it saw once that request is marked complete. We’re candid about our [Security Policy](#) and allow firms to review the specifics of how their data will be used during the procurement stage. We’ve always strived to exceed the level of transparency the IRS is now mandating from the entire industry.

The Bottom Line

Inventory both hardware and software, stay on top of patches and limit privilege, keep security the default, and log and monitor everything. These are the fundamentals, and they're only becoming more important. AI hasn't changed them, but it's raising the cost of messing them up. The top currency your firm has is trust, and that trust is earned by creating systems that won't crack under pressure.