



How-To Guide:

Evaluating AI K-1 Extraction Tools

Accounting firms and K-1 recipients in general face mounting complexity in handling Schedules K-1, K-3 and related footnotes. Manual processes are costly, error-prone and slow, especially at scale. As firms adopt AI-driven extraction tools, the challenge shifts from if to how—how to evaluate vendors, how to measure quality and how to ensure compliance, scalability, and trust.



This guide provides a practical framework for evaluating vendors — ensuring you ask the right questions across six key categories, understand why they matter and can compare answers confidently.

Introduction.

Evaluating AI K-1 extraction solutions is not about flashy demos but about trust, compliance and real-world performance. By asking the right questions — and understanding why they matter — you can identify vendors that deliver not only automation, but also accuracy, scalability and long-term confidence in compliance outcomes.

HERE ARE THE CATEGORIES, THE QUESTIONS & WHY EVERY COMPANY SHOULD CARE:

Compliance, Security, Technology Stack & Governance

QUESTION	WHY IT MATTERS
Confirm SOC 2 Certification	Ensures data security with industry-standard controls. Without SOC 2, firms face exposure to breaches and disqualification in RFPs.
What is the AI model training process? Is training conducted in a closed, private environment (e.g., on-premises or private endpoints, with no OpenAI/public use)? Is any client data sent through the model accessible, stored, or reviewed by third parties, including vendors?	This question is important to confirm that AI training occurs in a secure, closed environment and that client data is not exposed to public models or third parties. It provides assurance around regulatory compliance (e.g., IRC §7216, privacy laws) and helps safeguard against unauthorized storage, access, or use of sensitive information.
Please describe the AI models used to address the complexities and variations specific to Schedule K-1. How many models are employed?	This helps assess whether the solution can handle the wide variations in Schedules K-1 with accuracy and reliability. Red flag might be if there is a single generic model.
What is the frequency of A.I. model trainings? Is there an automated, auditable process for model training? Are there controls in place to validate model outputs before the model is released?	The frequency and governance of AI model training directly affect the accuracy, reliability, and compliance of the solution. An automated, auditable training process ensures transparency and accountability, while validation controls safeguard against inaccurate outputs or unintended data use before deployment. Together, these practices build trust that the AI is both effective and compliant with regulatory and client standards.
What types of AI models power your solution? Specify whether they are proprietary, rely on public models (e.g., OpenAI, Claude, etc.) or follow a hybrid approach.?	Clarifying whether a solution is built on Proprietary, Public, or Hybrid models is essential to assess security, reliability and regulatory compliance. Public models often raise concerns around data exposure and may return less reliable answers when handling industry-specific content. Proprietary models, in contrast, are trained for the organization's needs and provide greater accuracy, explainability and control over innovation, customization and long-term support. Hybrid models can balance the scale and agility of public models with the precision and safeguards of proprietary approaches. There are also accuracy considerations: Public models can hallucinate, lack contextual accuracy and aren't optimized for regulated environments. Proprietary models are purpose-built, producing domain-relevant and consistent results, especially when compliance and reporting are at stake. Hybrid models offer flexibility, but success depends on strong guardrails and governance to ensure outputs remain reliable.

K-1 Extraction Capabilities & Intelligence

QUESTION	WHY IT MATTERS
Can the solution extract from overflow footnotes, Schedule K-3, Section 199A, UBI, and Form 926?	Most critical K-1 details live in supplemental footnotes. Missing these creates compliance gaps.
Does the system handle activity-based reporting and passive vs. active distinctions?	Ensures accurate classification of income, critical for compliance and tax planning.
Does the system cover all 42 state jurisdictions and residency nuances?	State complexity is where many tools fail; coverage reduces costly manual fixes.
Does the system support K-1s received for ALL legal entity types? (1040, 1120, 1065, 1041, 990, 1120S)	Many technologies focus only on the Form 1040. This is important to make sure a single solution can support all K-1 extraction needs.

User Interface, Workflow & Aggregation

QUESTION	WHY IT MATTERS
Can the platform review single K-1s and aggregate unstructured footnote data?	Improves transparency and reduces manual review burden.
Can it automate fund-level aggregation for multi-tier partnerships and fund-of-funds?	If the AI tool can't handle fund-level aggregation for multi-tier partnerships and fund-of-funds, it may solve extraction at the surface level but fail where the real value lies. This is the difference between "document processing" and true tax intelligence.

Data Ingestion, Efficiency & Connectivity

QUESTION	WHY IT MATTERS
Does the solution support API ingestion and bulk uploads?	Ensures scalability and efficiency in large tax engagements.
Can it integrate with CCH, Thomson Reuters, and other compliance platforms?	Reduces manual entry and speeds workflows.
What is the average extraction speed and cost per K-1?	Determines ROI and scalability in peak season.

Advanced Analytics & Insights

QUESTION	WHY IT MATTERS
Does the solution provide anomaly detection and year-over-year comparisons?	Identifies issues early, enabling consulting opportunities.
Does the platform provide summary reports to support filing determinations and risk identification for federal, state and international tax reporting purposes?	Unlocks value beyond compliance by highlighting planning opportunities.

Onboarding & Support

QUESTION	WHY IT MATTERS
How quickly can new clients be onboarded and live in the system?	Indicates the usability of the technology if onboarding is guided.
What training and self-service resources are available?	Reduces support tickets and increases user adoption.
What ongoing support (SLAs, dedicated reps, hours) is included?	Strong support ensures long-term satisfaction and retention.

Scoring Methodology

To ensure a fair and meaningful comparison across vendors, responses are scored using a weighted average across the six evaluation categories. Each category is assigned a weight based on its relative importance to overall success. Vendors are then rated on a 1–5 scale within each category, multiplied by the category weight, and combined into a total score. This approach balances quantitative scoring with strategic priorities, allowing firms to identify the best-fit solution rather than the lowest-cost or fastest-demo option.

DOWNLOAD OUR EXCEL VERSION FOR THE QUESTIONNAIRE AND SCORECARD.

EXCEL VERSION