

CPE

CURRICULUM:

Accounting and Auditing, Tax

LEVEL:

Intermediate

DESIGNED FOR:

Tax practitioners, CPAs in public practice and business and industry

OBJECTIVES:

Explore how machine learning (ML) works in day-to-day operations and show how natural language processing over SKU and contract text, combined with supervised ML, improves classification accuracy

KEY TOPICS:

Taxability classification, return preparation and filing, benefits of machine learning for sales and use tax compliance, right-sizing and cost-benefit, reducing false negatives in tax ML systems

PREREQUISITES:

None

ADVANCED PREPARATION:

Read [“AI-Powered Tax Compliance, Part 1: How Machine Learning is Revolutionizing Sales and Use Tax”](#) in the January/February 2026 issue of *Today’s CPA*.

TAKE THE ONLINE CPE QUIZ



Today’s CPA offers the self-study exam for readers to earn **one hour of continuing professional education credit**. The questions are based on technical information from the following article. If you score 70 or better, you will receive a certificate verifying you have earned one hour of CPE credit in accordance with the rules of the Texas State Board of Public Accountancy (TSBPA).

Take the CPE quiz online. Go to the News & Publications section of TXCPA’s website and select [Today’s CPA Magazine](#). Then scroll down on the page and click on “AI-Powered Tax Compliance, Part 2.”

The State Board stipulates that the quiz is valid for one year from its publication in *Today’s CPA*. Quizzes submitted after this one-year period will not be accepted.



AI-Powered Tax Compliance, Part 2

BY KARINA KASZTELNIK, PH.D.



In the first article of a two-part series, we explored how machine learning (ML) works alongside rules to deliver greater consistency,

speed and auditability in sales and use tax operations. We also described the data and governance prerequisites for adopting ML and shared a case study (Case Study 1) on automated nexus detection, demonstrating how a hybrid rules-plus-ML approach accelerates registrations, reduces manual review effort and provides stronger, audit-ready evidence. [Read the article online](#) in the January/February 2026 issue of *Today’s CPA*.

Building on Part 1’s foundation, this installment turns to day-to-day operations: taxability classification and return preparation and filing. We show how natural language processing over SKU and contract text, combined with supervised ML, improves classification accuracy while keeping false negatives (missed taxable items) within agreed risk tolerances. A confusion matrix and precision-recall focus translate model quality into operational decisions, with dual thresholds and human-in-the-loop review for boundary cases.

We then outline filing workflows that use anomaly detection, checklist automation and auditable decision logs to shorten cycle time and strengthen defense. The article closes with deployment options (SaaS, private instance, on-prem) and a right-sizing checklist to balance cost, control and time-to-value.

With exposure detection in place, the next bottlenecks are classifying what’s taxable and filing consistently across states. Here, ambiguity in product descriptions and jurisdictional differences can turn “95 percent accuracy” into operational risk if misses aren’t measured

and managed. This article operates ML for tax teams. We use confusion matrices and recall-centric evaluation to set thresholds aligned to regulatory risk, keep humans in the loop where confidence is low and log every decision for auditability. A practical filing pipeline – data normalization, anomaly checks, workpapers, and approvals – translates model outputs into timely, defensible returns. We conclude with deployment and cost guidance so teams can adopt ML at the right size without overbuilding.

CASE STUDY 2: TAXABILITY CLASSIFICATION

A technology company selling SaaS subscriptions and support services faced challenges accurately classifying transactions given states’ varied treatment of cloud services (Ezeife, 2025). The company implemented an ML-powered classification engine trained on historical invoices labeled by tax determinations. Features included SKU and contract descriptors (NLP over descriptions), customer location and select contract language cues (Galvix, 2023).

Validation and metrics. Using a time-based train/validation split (past periods → train; most recent periods → validate) to avoid leakage, the team optimized for recall on taxable items and monitored precision-recall (PR) curves. After iterative training with class-weighted (cost-sensitive) loss and probability calibration, classification accuracy rose to 95 percent while maintaining high recall on taxable classes. (See Table 1 for error balance and FN tolerance; Table 2 for per-state breakdown.)

Operational controls. Thresholds were tuned to regulatory risk and implemented as dual thresholds: transactions with high confi-



dence were auto coded; borderline cases were routed to human review via a work queue. Guard-rail rules (e.g., explicit nexus thresholds and always-taxable SKUs) remained in place to catch obvious cases even if the model under-fires. Manual coding hours decreased by ~50 percent

and decision/audit logs strengthened audit defense documentation. In a two-cycle shadow audit, observed false negatives in variance samples decreased by ~30–40 percent after threshold calibration and reviewer feedback were incorporated.

Deployment note. The solution launched as tenant-isolated SaaS; guidance for private-instance/VPC or on-prem deployment (with key management, data-residency and audit logging) is provided in Considerations and Challenges in Part 1 of this series.

Main point. ML improves consistency and speed of taxability determinations when trained on well-labeled history and paired with governance, thresholds and human-in-the-loop review. See Figure 1.

Whenever the model isn't sure, it flags the transaction for a tax specialist to double-check. This step ensures that a human expert makes the final call on tricky cases. As more transactions are reviewed and confirmed, the system keeps learning and gets better over time.

The goal is to make tax coding faster and more accurate. Instead of spending hours manually reading every invoice, tax teams can focus on the exceptions that truly need attention. This combination of automation and human review helps businesses keep up with complicated tax rules.

CASE STUDY 3: RETURN PREPARATION AND FILING

A global manufacturer faced challenges in reconciling transactional data prior to preparing its monthly sales tax returns (EY, 2023). Discrepancies between the ERP records and tax engines resulted in delays in filing. To address this issue, the company

implemented a machine learning-based anomaly detection system (Aladebumoye, T. 2025). This model compared expected tax amounts with calculated figures, identifying any outliers. As a result, the time required for monthly reconciliation was reduced from 12 days to just four days and filing delays were decreased by 70 percent.

Main points from the case study: Machine learning-powered anomaly detection enhances data integrity and accelerates the preparation of tax returns.

ANOMALY DETECTION IN RETURN PREPARATION

Figure 2 shows how machine learning helps catch mistakes before tax returns are filed. It begins by collecting transaction data from different sources, such as sales records and the company's tax calculation system. All of this data is brought together and reconciled, which means any inconsistencies are identified and corrected so everything matches up.

When the model compares current data to expected results, it looks for anything unusual or suspicious. These unusual items are called

Figure 1. Taxability Classification with NLP and ML



Natural language processing interprets product descriptions; the ML model predicts taxability codes. Tax specialists review flagged transactions for accuracy.

Source: Compiled by Author

Table 1. Confusion Matrix for Taxability Classifier

Rows = Actual, Columns = Predicted

| Actual \ Predicted | Taxable | Not Taxable | Row Total |
|--------------------|----------------------|--------------------------|-----------|
| Taxable | TP (correct taxable) | FN (missed taxable) | TP + FN |
| Not taxable | FP (over-taxed) | TN (correct non-taxable) | FP + TN |
| Column total | TP + FP | FN + TN | N |

Operational goal: Keep FN at or below the agreed tolerance; threshold tuned to regulatory risk

Key Metrics:

Recall (Taxable) = $TP / (TP + FN)$ → prioritize high recall to reduce miss risk

Precision (Taxable) = $TP / (TP + FP)$

False-Negative Rate (FNR) = $FN / (TP + FN)$

Accuracy = $(TP + TN) / N$

anomalies. For example, if a transaction shows a much higher or lower tax amount than normal, the model will flag it.

Flagged anomalies are automatically sent to tax professionals on the team. The tax specialists then review each flagged item to determine whether it's an error or an acceptable exception. This process helps ensure all information on the tax return is correct and ready to be submitted.

By catching problems early, the company avoids filing inaccurate returns that could lead to penalties or audits. This system also reduces the time employees spend manually combing through records. In short, anomaly detection works like a safety net that makes tax filing more accurate, efficient and reliable.

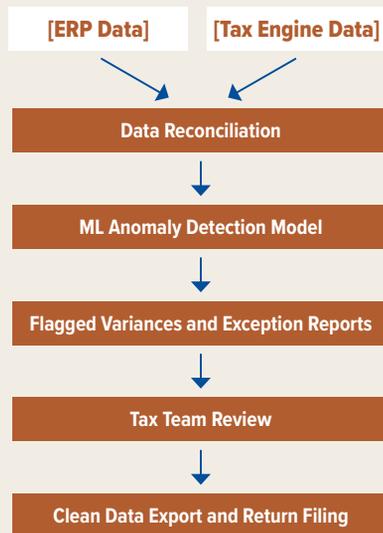
After launch, the team instituted a monthly shadow audit on jurisdictions with recent rule changes. Combining cost-sensitive thresholding with a human review queue reduced observed false negatives in flagged variances by ~35 percent over two filing cycles, while maintaining manageable reviewer workload.

BENEFITS OF MACHINE LEARNING FOR SALES AND USE TAX COMPLIANCE

Machine learning for sales and use tax compliance begins with the collection of large volumes of transaction data from various systems throughout the organization (McKinsey, 2025). Rather than hiring additional personnel to manage this data influx, machine learning can scale automatically to process millions of records without increasing headcount. By employing supervised learning, the system examines past examples to understand what correct tax decisions entail (Dennis, A, 2024). As it ingests more data over time, its predictions become progressively more accurate. Each transaction is evaluated according to consistent rules, ensuring that similar cases are handled uniformly.

This standardized approach minimizes the risk of human error and fosters consistency in decision-making. Furthermore, the model continuously monitors incoming data for anomalies or irregularities. When it detects an outlier, it promptly flags the item for tax professionals to investigate (Aladebumoye, T, 2025). This early detection enables teams to rectify issues before tax returns are submitted, helping to avoid penalties or audits. Every decision made by the machine learning

Figure 2. Anomaly Detection in Return Preparation



The ML model compares reconciled transaction data with expected tax calculations, flags anomalies and routes issues to tax professionals before filing.

Source: Compiled by Author

system is meticulously logged, creating a clear data trail that illustrates how each result was derived. This level of transparency is crucial should regulators or auditors require supporting documentation.

Tax professionals can swiftly review the data to clarify the rationale behind each classification or calculation. By automating repetitive tasks, staff members are free to focus on strategic initiatives that enhance value. This streamlined workflow empowers companies to navigate complex and evolving tax regulations more effectively (Ezeife, E., 2025). Ultimately, machine learning evolves into an indispensable asset that bolsters accuracy, consistency and efficiency in tax compliance.

RIGHT-SIZING AND COST-BENEFIT, ESPECIALLY FOR SMALL COMPANIES

Not every organization needs a full ML platform on day one. A pragmatic path is to begin with a thin NLP classifier plus human-in-the-loop or a rules-first + ML-assist hybrid for high-variance SKUs/states.

A simple decision frame is:

Expected savings = (manual hours avoided × fully loaded cost) + (penalties avoided + interest avoided)

Total cost = (subscription or infra) + (implementation + data prep) + (ongoing MLOps/governance).

If expected savings do not exceed total cost within 12–18 months, a lightweight or phased approach is preferred. Table 3 includes a short checklist to help teams estimate both sides of the equation.

REDUCING FALSE NEGATIVES IN TAX ML SYSTEMS

Metric focus. Report confusion matrices and recall/Precision-Recall curves per state, product family and filing period, optimizing thresholds to prioritize recall where the cost of a miss is high.

Cost-sensitive learning. Weight FN is more heavily than FP in loss functions; consider focal loss or class-weighting when taxable classes are rare.

Threshold and calibration. Calibrate probabilities (e.g., Platt/iso) and set dual thresholds: auto-approve when confidence is high; route to human review when near the boundary.

Hybrid rules + ML. Keep guardrail rules (e.g., explicit nexus thresholds, certain taxable SKUs) to catch obvious cases even if the model under-fires.

Active learning on misses. Continuously feed confirmed FN back into training; schedule error review sprints after each filing cycle.

Shadow audits and sampling. Before relying on automation, run parallel processing for 1–2 cycles, compare variances and oversample “at-risk” segments (new jurisdictions, new products).

Production monitoring. Track FN proxies (late registrations, amended returns, auditor adjustments) and set alerts on unexpected drops in recall. Maintain a live Model Card documenting limits, known risks and the current FN mitigation plan.

LOOKING AHEAD

The future landscape of sales tax compliance is set to be transformed by AI-driven solutions (Ezeife, E., 2025). While automation enhances precision, the role of CPAs remains crucial for the interpretation and validation of results.

Tax professionals will be essential in establishing the frameworks and ethical standards that govern the deployment of machine learning technologies. As regulatory bodies grow more adept with sophisticated technologies, there will be heightened expectations for processes that are transparent and easily interpretable.

However, reliance on technology alone will not suffice to address all compliance challenges. The foundational elements of human judgment, professional skepticism and a comprehensive understanding of tax legislation will continue to play a pivotal role in effective compliance strategies. Innovative firms will seek to synergize AI capabilities with the expertise of their tax teams, while also committing to ongoing education, ensuring that personnel acquire the skills necessary to oversee, critically assess and refine machine learning outputs.

Organizations that adeptly combine automation with robust governance frameworks are likely to achieve a competitive edge in terms of accuracy, operational efficiency and stakeholder confidence. The upcoming years are poised for further advancements, with features like natural language processing and real-time reporting becoming industry standards. As these tools progressively evolve, CPAs who maintain proactive engagement and continuous learning will find themselves in optimal positions to guide their firms through this evolution. By integrating machine learning, tax professionals can help cultivate a compliance landscape that is not only efficient but also transparent, accountable and resilient.

Sales and tax compliance are becoming more complex as rules change rapidly and transaction volumes rise, especially for multi-

jurisdiction businesses. Manual processes no longer scale, making machine learning—combined with strong governance and professional oversight—essential. When supported by high-quality data, disciplined validation, clear exception handling and appropriate deployment choices, ML improves accuracy, scalability and audit readiness while reducing strain on tax teams. Used thoughtfully, it does not replace expertise; it amplifies it, enabling tax professionals to shift from reconciliation work to higher-value strategic advisory as tax laws continue to evolve.

References

- Aladebumoye, T. (2025). The role of AI in enhancing tax transparency and reducing evasion. *World Journal of Advanced Research and Reviews*, 25(1), 206–212. <https://doi.org/10.30574/wjarr.2025.25.1.0023>
- Galvix. (2023, June 9). Embracing the future: The role of automation in sales tax compliance. *Galvix Insights*. <https://www.galvix.com/article/the-role-of-automation-in-sales-tax-compliance/>
- KPMG. (2024). Tax Reimagined 2024: Perspectives from the C-suite. KPMG International. <https://kpmg.com/us/en/articles/2024/tax-reimagined-report-2024.html>
- McKinsey. (2025, March 12). The state of AI: Global survey. McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- Dennis, A. (2024, February). What AI can do for auditors. *Journal of Accountancy*. <https://editions.journalofaccountancy.com/article/What+AI+can+do+for+auditors/4705695/812308/article.html>
- Ezeife, E. (2025). AI-driven tax technology in the United States: A business analytics framework for compliance and efficiency. *Multidisciplinary Global Education Journal*, 5(1), 426–435. https://www.allmultidisciplinaryjournal.com/uploads/archives/20250219175220_MGE-2025-1-426.1.pdf
- Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). AI-Ethics by Design and Ethics of Use in Tax Fraud Detection: Survey Evidence on the Importance of Ethical Principles. arXiv preprint arXiv:2106.00326. <https://arxiv.org/pdf/2106.00326.pdf>

Table 2. Error Management and Auditability (Per-State Panel)

| State | TP | FP | FN | TN | Recall (Taxable) | FNR | Notes |
|-------|-------|-----|-----|-------|------------------|------|-------------------------------|
| CA | 1,240 | 95 | 44 | 3,210 | 96.6% | 3.4% | Threshold ↑ after rule change |
| TX | 980 | 62 | 38 | 2,705 | 96.3% | 3.7% | New SKU family added |
| NY | ... | ... | ... | ... | ... | ... | ... |

Table 3. Right-Sizing and Cost-Benefit Checklist

| Component | How to Estimate | Typical Source |
|----------------------------|--|---------------------|
| Manual hours avoided | Minutes per record × monthly volume × hourly rate | Time study; payroll |
| Rework/exception reduction | (# reworks × handling cost) × expected % reduction | Ticketing/ERP |
| Penalties/interest avoided | Historical penalties/interest × expected % reduction | Finance/tax logs |
| Cycle-time value | Overtime avoided + late fees avoided | Payroll; A/P |
| Analyst redeployment | Hours freed × rate (if truly reduced) | Manager estimates |
| Subscription/infra | Vendor quote or cloud T-shirt size | Vendor/cloud |
| Implementation | Internal + vendor hours × blended rate | PMO; SOW |
| Data prep/labeling | Records to map × minutes/record × rate | Tax ops |
| ML Ops/monitoring | Tooling + engineer hours | Eng. leads |
| Governance | Audit/docs time × rate | Risk/compliance |
| Security | KMS/VPC setup + reviews | Sec/IT |
| Change management | Training hours × headcount × rate | L&D |
| Human review operations | Cases/month × minutes × rate | Ops queue |

KARINA KASZTELNIK, PH.D., is Professor at University of Maryland Global Campus University in Washington, DC. Contact her at karina.kasztelnik@faculty.umgc.edu.

