TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION

The IRS Could Leverage Examination Results in Artificial Intelligence Examination Case Selection Models and Improve Processes to Evaluate Performance

May 19, 2025

Report Number: 2025-308-022

TIGTACommunications@tigta.treas.gov | www.tigta.gov

HIGHLIGHTS: The IRS Could Leverage Examination Results in Artificial Intelligence Examination Case Selection Models and Improve Processes to Evaluate Performance

Final Report issued on May 19, 2025 **Report Number 2025-308-022**

Why TIGTA Did This Audit

Artificial intelligence (AI) is a transformative technology that holds substantial promise for improving the IRS's examination efforts. The IRS is using Inflation Reduction Act of 2022 funding to deliver cutting-edge technology, data, and analytics to operate more effectively.

The IRS plans to use new analytics models to reduce the burden on compliant taxpayers by improving case selection. In August 2022, the Department of the Treasury reported that the IRS is using AI for examination workload selection.

This audit was initiated to determine the effectiveness of the Large Business and International and Small Business/Self-Employed Divisions' AI models in selecting returns and issues for examination.

Impact on Tax Administration

The Tax Gap was most recently estimated for Tax Year 2021 to be $688 billion annually, of which $542 billion (79 percent) is from underreporting. The no-change rate is a key metric that the IRS uses to measure the effectiveness of returns selected for examination. Current return selection models have resulted in a high percentage of examinations completed with no change to the tax liability. This results in the IRS using resources on unproductive examinations and unnecessarily burdening compliant taxpayers. AI models are intended to improve the process the IRS uses to select cases for examination.

What TIGTA Found

The IRS was proactively using AI prior to the December 2020 Presidential Executive Order, *Promoting the Use of Trustworthy AI in the Federal Government*. The IRS integrated statistical and machine-learning techniques into return selection processes. The IRS revamped how it selects returns and identifies issues for examination by using AI models trained on current return data rather than relying on historical examination results.

However, historical examination results are informative and should be used by the IRS to monitor and improve AI models when available. For example, the IRS could use examination results to improve return classification and return selection AI models that could potentially identify new areas of noncompliance.

The IRS should also consider evaluating ensemble machine-learning for improving the accuracy of identifying noncompliant taxpayers and narrowing the Tax Gap. Ensemble learning is an approach that combines multiple machine-learning algorithms to potentially improve performance by making more accurate predictions of which tax returns and/or issues to examine.

Additionally, the IRS has not established processes to evaluate whether the performance of AI models is better than prior methods or is achieving the intended objectives. Therefore, the IRS cannot readily demonstrate in real-world context if AI models are improving the IRS's overall examination compliance efforts. Not evaluating performance results is contrary to federal AI key practices to ensure accountability and responsible AI use.

What TIGTA Recommended

We recommended that the Chief Tax Compliance Officer, in partnership with the Chief Data and Analytics Officer where appropriate, require division commissioners to: (1) use governance processes to ensure that examination performance results are part of the monitoring and continuing refinement of the return classification and selection AI models; (2) refine AI models by incorporating ensemble machine-learning when appropriate; and (3) establish a measurement plan with appropriate metrics to monitor AI models to ensure that they are achieving the expected benefits and to correct any model drifts.

The IRS agreed with all three of our recommendations, agreeing with recommendations 1 and 3, subject to staffing constraints and anticipated new Department of the Treasury guidance on AI governance. Regarding recommendation 2, the IRS stated that it has already tested and implemented ensemble methods in these models, where appropriate.

TREASURY INSPECTOR GENERAL
FOR TAX ADMINISTRATION

U.S. DEPARTMENT OF THE TREASURY

WASHINGTON, D.C. 20024

May 19, 2025

MEMORANDUM FOR: COMMISSIONER OF INTERNAL REVENUE

FROM: Diana M. Tengesdal
Acting Deputy Inspector General for Audit

SUBJECT: Final Audit Report – The IRS Could Leverage Examination Results in Artificial Intelligence Examination Case Selection Models and Improve Processes to Evaluate Performance (Audit No.: 2024308019)

This report presents the results of our review of the effectiveness of the Large Business and International and Small Business/Self-Employed Divisions' artificial intelligence models in selecting returns and issues for examination. This review is part of our Fiscal Year 2025 Annual Audit Plan and addresses the major management and performance challenge of *Managing Inflation Reduction Act Transformation Efforts.*

Management's complete response to the draft report is included as Appendix VII. If you have any questions, please contact me or Matthew A. Weir, Assistant Inspector General for Audit (Compliance and Enforcement Operations).

Table of Contents

Background

The federal government recognizes the potential of artificial intelligence (AI) to increase efficiency and improve government services. Key AI terminologies are defined as follows:

- **AI** – A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems use machine and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action.[1]

  A common example of AI is for customer service assistance in which questions are answered by a chat bot as opposed to a human interacting with the caller.

- **AI model** – A component of an AI system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.[2]

- **Machine-learning** – Application of AI that is characterized by providing systems the ability to automatically learn and improve on the basis of data or experience, without being explicitly programmed.[3]

The Presidential Executive Order, *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*, encouraged federal agencies to continue to use AI, when appropriate, to benefit the American people. Agencies must therefore design, develop, acquire, and use AI in a manner that fosters public trust and confidence. Additionally, agencies shall be transparent in disclosing relevant information regarding their use of AI to appropriate stakeholders, including Congress and the public, to the extent practicable and in accordance with applicable laws and policies.[4] Furthermore, each federal agency is to annually prepare an inventory of its non-classified and non-sensitive use of AI. Other Executive Orders addressing AI have been issued with the latest order addressing the topic of removing barriers to American leadership in the development and use of AI.[5]

The Department of the Treasury's August 2022 report on its inventory of AI use included the following three AI projects at the Internal Revenue Service (IRS):

- Large Partnership Compliance (Large Business and International (LB&I) Division).
- Line Anomaly Recommender (LB&I Division). This model produces tax noncompliance risk related scores for each return line-item and an overall risk score for the entire return.

[1] 15 U.S.C. § 9401(3).

[2] National Telecommunications and Information Administration, *Artificial Intelligence Accountability Policy Report* (March 2024).

[3] 15 U.S.C. § 9401(11).

[4] Executive Order 13960, *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government* (December 3, 2020), which expounded upon Executive Order 13859, *Maintaining American Leadership in Artificial Intelligence* (February 11, 2019).

[5] Executive Order 14179, *Removing Barriers to American Leadership in Artificial Intelligence* (January 23, 2025).

- Small Business/Self-Employed (SB/SE) Division Issue Recommender. This AI use consists of two AI models: individual return classification and individual return selection.

These AI uses fall under the general category of AI models. Furthermore, the three AI uses are considered unsupervised machine-learning.

There are several methods to train machine-learning algorithms, including supervised and unsupervised.

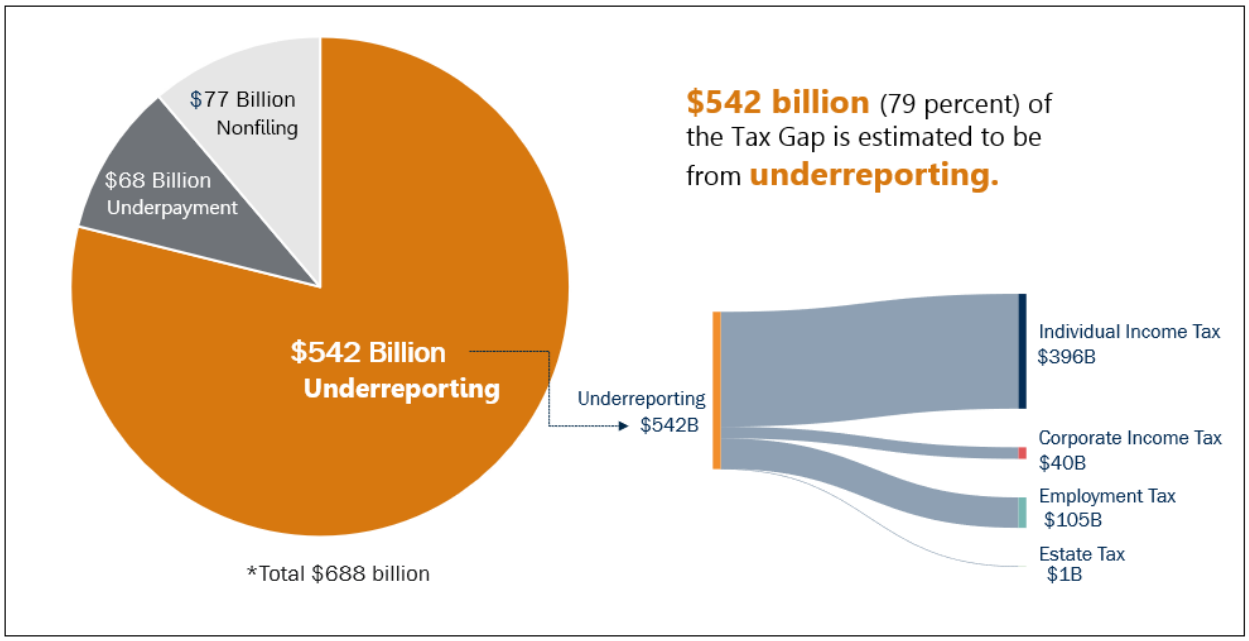
- Under the supervised machine-learning method, data scientists present an algorithm with labeled input data. For example, a supervised learning model can predict how long a commute will be based on the time of day, weather conditions, and so on. But first, one will have to train the model by giving it examples of different commute times (labels) and the conditions that produced those commute times.
- Under the unsupervised machine-learning method, data scientists present an algorithm with unlabeled data and allow the algorithm to identify structures in the inputs without a preconceived idea of what to expect. An unsupervised learning model works on its own to discover the inherent structure of unlabeled data.

The scope of this review was limited to understanding how the IRS is using AI technologies in the three projects listed above and their impact on return selection and issues selected for examination. According to the IRS, as part of its efforts to address the Tax Gap, it may use AI technologies to support examination processes when they provide the most effective approach to identify noncompliance while minimizing burden on compliant taxpayers.

Figure 1 summarizes the most significant components of the gross Tax Gap (*i.e.*, the difference between what taxpayers owe annually and what they pay voluntarily and timely). The IRS most recently estimated the Tax Gap to be $688 billion annually, of which $542 billion (79 percent) is estimated to be from underreporting.[6]

[6] According to the IRS, the Tax Gap projections do not fully represent noncompliance in some components of the tax system, particularly related to corporate income tax, income from flow-through entities, foreign or illegal activities, digital assets, and Coronavirus Disease 2019 Pandemic credits, because data are lacking.

Figure 1: Gross Annual Tax Gap Summary for Tax Year (TY) 2021[7]



*Source: IRS Publication 5869, Federal Tax Compliance Research: Tax Gap Projections for TYs 2020 and 2021 (October 2023). *Off due to rounding.*

Noncompliant taxpayers can undermine public confidence in the fairness and integrity of the federal tax system, encouraging more noncompliance.[8] To address underreporting by taxpayers, the IRS selects tax returns for examination. According to IRS policy, the primary objective in selecting returns for examination is to promote the highest degree of voluntary compliance on the part of taxpayers. This requires:

1) Exercising professional judgment in selecting enough returns from all classes to assure all taxpayers of equitable consideration.
2) Using available experience and statistics indicating the probability of substantial error.
3) Making the most efficient use of examination staffing and other resources.

Examination no-change rates

The IRS uses the no-change rate, (*i.e.*, the percentage of examinations resulting in no change to the tax return), to provide insight into the effectiveness of tax return selection. This performance metric can signal that current return selection processes and resources are potentially inefficient and burden compliant taxpayers. Figure 2 shows that a high percentage of returns examined resulted in no change to tax liability, particularly among non-individual tax return examinations.

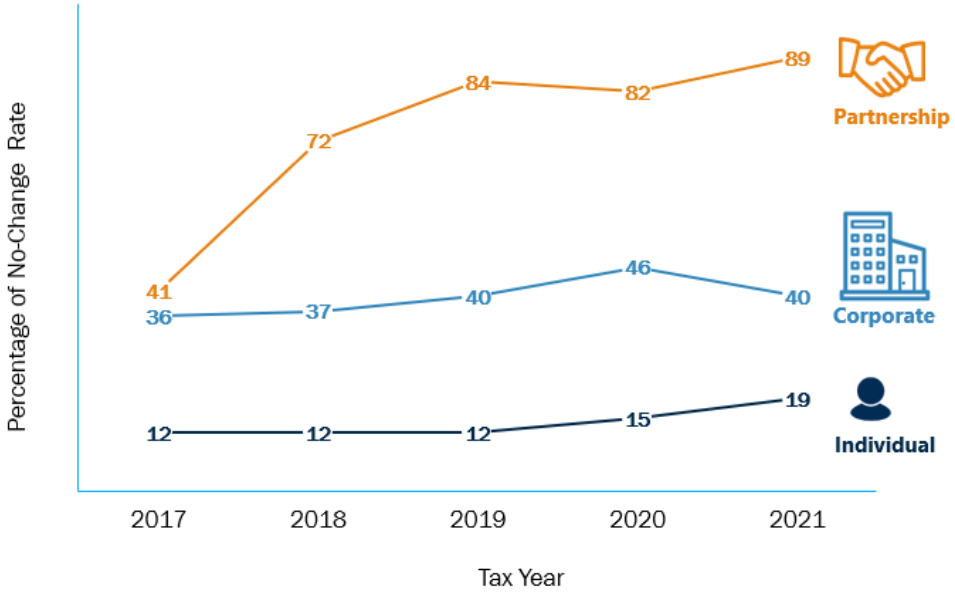
The high no change rate in large partnership audits has been a challenge for the IRS's compliance efforts in part due to the fact that many partnerships have multiple tiers of connected partnerships with thousands of partners, and it is challenging for the IRS to trace

[7] See Appendix VIII for a glossary of terms.

[8] Congressional Research Service, *Federal Tax Gap: Size, Contributing Factors, and the Debate Over Reducing It* (October 2023).

transactions through the tiers of partners to the ultimate partner who is at issue in the examination.[9]

Figure 2: IRS Examination No-Change Rate Percentage by Return Type for TYs 2017 Through 2021[10]



Source: Analysis of IRS Fiscal Year 2023 Data Book, Table 17.

The no-change rate is a concern to stakeholders. During a February 2024 congressional hearing, the IRS Commissioner stated that AI solutions currently implemented at the agency could reduce no-change audits that burden taxpayers and cost the IRS resources.[11]

Inflation Reduction Act (IRA) impact

In addition to the IRS's annual appropriation, in August 2022, Congress enacted the IRA, giving the IRS funding to improve the administration of the tax system and services provided to taxpayers. The IRS initially received $79.4 billion from the IRA. However, as of March 2025, Congress subsequently reduced IRA funding to $37.6 billion.[12]

---

[9] GAO, GAO-23-106020, *IRS Audit Processes Can Be Strengthened to Address a Growing Number of Large, Complex Partnerships* (July 2023).

[10] The data are as of September 30, 2023. According to IRS management, no-change examinations (examinations where no adjustment to tax liability is made) tend to close more quickly than examinations resulting in changes. Consequently, additional examination closures after September 30, 2023, may result in lowering the no-change percentages. The IRS stated that the Data Book Table 17 shows there are a substantial number of open examinations for TYs 2020 and 2021 for all three taxpayer segments, along with many large corporate and partnership examinations still open for TYs 2018 and 2019. As such, the no-change rates as presented in Figure 2 may decrease over time. Also, the partnership and corporate no-change rates represent returns with all asset sizes.

[11] U.S. House of Representatives, Committee on Ways and Means, Hearing on IRS Oversight (February 15, 2024).

[12] The Fiscal Responsibility Act of 2023 (Pub. L. No. 118-5, 137 Stat. 10) rescinded $1.4 billion; The Further Consolidated Appropriations Act, 2024 (Pub. L. No. 118-47, 138 Stat. 460) rescinded $20.2 billion; and the Full-Year Continuing Appropriations and Extensions Act, 2025 (Pub. L. No. 119-4) rescinded another $20.2 billion. Each of these rescissions reduced the amount of enforcement funding.

The IRA Strategic Operating Plan outlines how the IRS will use IRA funds.[13] The IRS intended to use IRA resources to strengthen enforcement of complex partnerships, large corporations, and high-income, high wealth individuals who do not pay their taxes. Strategic Operating Plan Objective No. 4 states "*deliver cutting-edge technology, data, and analytics to operate more effectively*." The IRS planned to:

- Harness data and analytics to drive operations and decision-making.
- Apply enhanced analytics capabilities to improve tax administration.
- Use new analytics models, such as compliance risk analytics, to reduce burden on compliant taxpayers by improving case selection.
- Regularly evaluate and improve data, tools, and governance processes to help ensure that models are working as intended and not subject to unobserved biases.[14]

However, given the recent governmentwide cost cutting efforts (*e.g.*, IRA funding recission, hiring freeze, and anticipated reduction in force) it is now unclear if the IRS will be able to pursue and achieve the above plans.

Results of Review

The IRS Undertook Initiatives to Improve Tax Return Classification and Issue Selection Using Artificial Intelligence

The IRS recognized that AI could improve return selection and enhance tax compliance prior to the December 2020 Presidential Executive Order. The IRS integrated statistical and machine-learning techniques into return selection processes with the goal of increasing efficiency and effectiveness. The IRS took actions to revamp how it selects returns and identifies issues for examination for individual, corporation (total assets $10 million to $250 million), and large partnership returns.[15] These actions included exploring unsupervised models, which are trained on current return data and do not rely on historical examination results. The following provides an overview of each AI model and methodologies, with details in Appendices II through V.

[13] *IRS Inflation Reduction Act Strategic Operating Plan FY 2023 – FY 2031* (April 2023).

[14] GAO, GAO-21-519SP, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (June 2021), cited identifying potential bias, inequities, and other societal concerns resulting from the AI system as a key AI performance practice.

[15] Form 1040, *U.S. Individual Income Tax Return*; Form 1120, *U.S. Corporation Income Tax Return*; and Form 1065, *U.S. Return of Partnership Income*.

The SB/SE Division's Form 1040 return classification model automated the classification function saving examination resources

In 2017, the SB/SE Division's Form 1040 return classification project in coordination with the Research, Applied Analytics & Statistics office (RAAS), began developing models that learn relationships between individual tax return line-items to detect potential noncompliant returns. The model replaced the prior manual classification process and uses machine-learning techniques. The model uses the total Form 1040 population of any given tax year, except returns suspected of identity theft. According to IRS management, automating the classification function has allowed the IRS to reassign 14 classifiers to the examination function and they estimate the increased examinations could generate approximately $26.1 million in additional tax assessments yearly.

The Form 1040 return classification AI model learns relationships between tax return line-items.

The Form 1040 return selection model development work started around January 2020 and is in pilot phase testing. Figure 3 provides a timeline summarizing the development and deployment of the Form 1040 return classification and selection models.

Figure 3: Timeline of the Form 1040 Return Classification and Selection Models

Source: SB/SE Division and Research, Applied Analytics, and Statistics management.

According to IRS management, the IRS awarded two contracts totaling $7.9 million with a performance period from July 2019 to September 2024. See Appendix II for detailed information on the Form 1040 return classification model.

IRS management stated that the Form 1040 return selection model development work started around January 2020. The model uses the same data that are used for training the Form 1040 return classification model. [redacted]

[redacted][16] According to the SB/SE Division, the goal is to select returns for examination that have the highest potential for noncompliance. Pilot test results found the model would yield a higher dollar tax assessment per examination hour but also result in higher no-change rates. As such, IRS management has not made any deployment decisions on the model. See Appendix III for detailed information on the Form 1040 return selection model.

The National Research Program (NRP) is a comprehensive effort by the IRS to measure compliance for different types of taxes and various groups of taxpayers. The examination of taxpayer returns coordinated by the NRP provide the IRS with a statistically valid representation of the compliance characteristics of taxpayers. In comparison to the AI models, the IRS currently uses the Discriminant Index Function (DIF) return selection model, which is dependent upon NRP studies of examination data.[17] [redacted]

[redacted]

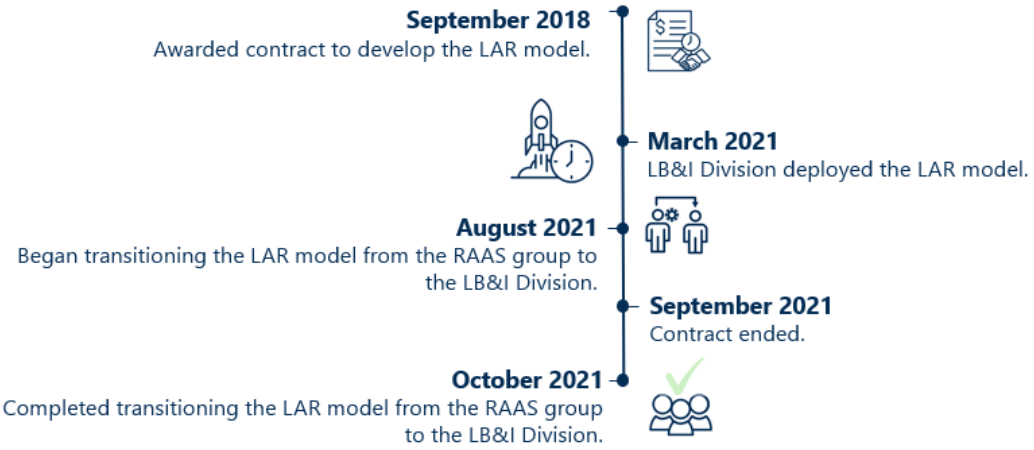
Data are not available to evaluate the LB&I Division's Corporate Line Anomaly Recommender and Large Partnership Compliance Return Selection Models' performance

In March 2021, the LB&I Division deployed the Line Anomaly Recommender (LAR) model to analyze Forms 1120, *U.S. Corporation Income Tax Return*, with total assets between $10 million and up to $250 million for new examination employees to work. The LAR model is used to identify returns that deviate from standards or expectations through an unsupervised statistical approach with output subject to human review. Human classifiers review and assess risks of the model output. Returns that are classified high risk for noncompliance are made available to the field, contingent on the demand from available field resources. Figure 4 is a timeline summarizing the development and deployment of the LAR model.

[16] [redacted]

[17] The DIF model is a mathematical technique used to computer-score income tax returns as to examination potential. The higher the score the higher the examination potential. [redacted]

Figure 4: Timeline of the LAR Model



Source: LB&I Division and RAAS management.

The RAAS group, relying on the LB&I Division's subject matter experts and an external contractor, developed the LAR model. The IRS spent $3.9 million between September 2018 and September 2021 on two contracts to develop and manage the model. See Appendix IV for detailed information on the LAR model, including trends in examination no-change rates for Form 1120 returns with total assets between $10 million and up to $250 million. LB&I does not have specific data regarding the LAR model examination results.

The Large Partnership Compliance (LPC) model was developed to bring attention to and increase compliance efforts for some of the largest and most complex partnership returns. Development began in October 2020 using a contractor tasked with developing a data science driven model and machine-learning services to improve workload inventory and case selection for large partnership compliance. LB&I Division management stated that machine-learning technology was applied to identify potential compliance risks in the areas of partnership tax, general income tax and accounting, and international tax in a taxpayer segment that historically has been subject to limited examination coverage.[18] Figure 5 summarizes the timeline of the LPC model development and deployment.

[18] IRS, press release No. IR-2024-09, *IRS Ramps Up New Initiatives Using Inflation Reduction Act Funding to Ensure Complex Partnerships, Large Corporations Pay Taxes Owned, Continues to Close Millionaire Tax Debt Cases* (January 2024).

Figure 5: Timeline of the LPC Model

Source: LB&I Division management.

According to IRS management, with the help of AI, return selection involved groundbreaking collaboration among experts in data science and tax enforcement. The large partnership taxpayers' behaviors are continuously changing, so the LB&I Division wanted to establish machine-learning to become more forward-looking in evaluating data. See Appendix V for detailed information on the LPC model. Results on LPC model examinations are not available as these examinations have not closed.

Additional Data and Techniques Could Be Used to Improve Return and Issue Selection Models

The IRS's current statistical and machine-learning classification and selection models are driven by current return data and do not always incorporate examination results, which could identify new areas of noncompliance. One of the IRS's IRA Strategic Operating Plan initiatives is to "*employ centralized, analytics-driven, risk-based methods to aid in the selection of compliance cases*." The IRS anticipated increasing audits on large corporations and partnerships but given the impact to IRS resources from governmentwide cost cutting efforts, its vision is not likely to materialize. For example, according to IRS management, they had planned to nearly triple audit rates on large corporations with assets of more than $250 million in TY 2026 as compared to TY 2019. Also, it was the IRS's intent to increase audit rates nearly 10-fold on large complex partnerships with assets of more than $10 million, going from 0.1 percent in TY 2019 to 1 percent in TY 2026.

IRS management stated they will continually update analytic models as they receive more data and learn more about noncompliance and the efficacy of compliance treatments. The IRS planned to establish a structure for incorporating feedback and ensuring that the analytics it uses continue to evolve.

According to the *Standards for Internal Control in the Federal Government*, monitoring assesses the quality of performance over time and promptly resolves the findings of audits and other

reviews. Corrective actions are a necessary complement to control activities to achieve objectives.[19]

Examination results could improve the Form 1040 return classification model

The Form 1040 return classification model identifies the top three issues for examination. However, the issues identified by the model are considered recommendations only. The examiner and their manager can add or remove issues based on their independent risk analysis. The timing of the Form 1040 return classification model run for each tax year generally does not allow immediate feedback on the productivity of the recommended issues. This is because most of the examinations are not completed by the time the model finishes its cycle run. Using TY 2021 as an example, the earliest examinations would be initiated in October 2022 with the last model cycle run in February 2023. Based on the timeline, the only examination results available would be examinations opened and closed between October 2022 and February 2023. Figure 6 provides an illustration of the model's timing and the start of examinations.

Figure 6: Timeline of the Form 1040 Return Classification Model Relative to When Examinations Are Started

Source: SB/SE Division management.

Notwithstanding the above limitations, there are opportunities for the IRS to use the examination results to continually refine the Form 1040 return classification model. The data needed to analyze examination covered issues and associated results are systemically available. For example, when examiners close a case, they prepare a Form 4549-A, *Report of Income Tax Examination Changes*, and detail the list of adjustments to taxable income by issue and the amount for each tax period. The IRS could review and analyze closed examination performance results to assess areas to improve the Form 1040 return classification model. For example, the IRS could:

- Review the productivity of model suggested issues (*e.g.*, if it resulted in a tax adjustment). If a specific issue is consistently unproductive, refine the model to stop recommending that issue.

[19] GAO, GAO-14-704G, *Standards for Internal Control in the Federal Government* (September 2014). The *Standards for Internal Control in the Federal Government* known as the "Green Book" sets the standards for an effective internal control system for federal agencies. Internal control helps an entity run its operations efficiently and effectively, report reliable information about its operations, and comply with applicable laws and regulations.

- Review the productivity of issues not suggested by the model but added by the examiners and their manager. If an issue is consistently productive but the model did not select it, then refine the model to identify returns with those issues.
- Compare identified issues to the examiners' pursued issues. This would indicate differences between machine and human risk assessments and allow the IRS to explore the cause for differences and make timely model adjustments.
- Review the distribution of productive issues based on examinations between tax years as compared with the model output. Multi-tax year trends may reveal useful information, especially when there have not been major tax law changes between the years.

Examination results could improve LAR and LPC return selection models

The results from Form 1120 and large partnership return examinations could similarly be useful in refining LAR and LPC return selection models, respectively. According to the LB&I Division, the examination cases that closed in Fiscal Year (FY) 2023 all averaged more than one year to complete, so it does not lend to immediate feedback on the LAR model.

However, the IRS could still analyze the results of the Form 1120 return issues and the associated productivity (*e.g.*, if the issues resulted in an additional tax assessment). Then, it could compare the results to the LAR model's established relationship between return line-items and the reasonableness of the assigned risk score and establish a feedback loop to ensure that the LAR model achieves the desired results.

According to LB&I Division management, prior to deploying the LAR model, they conducted multiple LAR model testing using historical examination results. After the March 2021 deployment, [redacted] It is important for the LB&I Division to continue to refine the LAR model based on examination results.

The LPC model could also benefit from consideration of examination experience. The productivity of ongoing large partnership examinations could shed light on the effectiveness of the data science aspect of the LPC model and the input of subject matter experts. According to IRS management, they are starting to collect information on LPC model examinations. We acknowledge that these large partnership returns are complex, and the examination could take multiple years to complete. As of April 2024, there were 82 TY 2021 large partnership returns identified as high risk by the LPC model that were subsequently selected for examination.

Overall, the IRS should review and analyze closed examination performance results to improve AI compliance models discussed previously.

Recommendation 1: The Chief Tax Compliance Officer should require the SB/SE and LB&I Divisions' Commissioners, in partnership with the Chief Data and Analytics Officer, where appropriate, to use governance processes to ensure that examination performance results are part of the monitoring and continuous refinement of AI classification (deployed) and return selection models (if deployed).

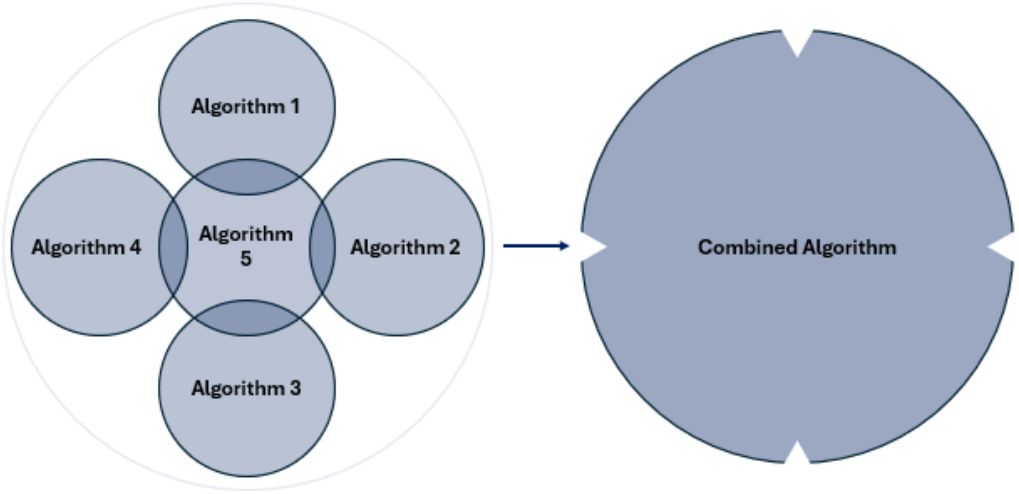
Management's Response: IRS management agreed with the recommendation, subject to staffing constraints and anticipated new Treasury guidance for AI governance. The Chief Data and Analytics Officer will verify that AI classification and return selection use cases go through the AI governance process and will establish a procedure for such use

cases to document how examination performance results are used for monitoring and regular refinement of AI models.

Ensemble machine-learning could be incorporated into compliance models

The IRS implemented machine-learning to improve return classification and selection. However, there are opportunities to potentially enhance the current models. Specifically, the IRS should consider evaluating ensemble machine-learning for improving the accuracy of identifying noncompliant taxpayers and narrowing the Tax Gap. Ensemble machine-learning can include using multiple models that result in a combined output that is potentially more accurate than a single model. Figure 7 provides an illustration of ensemble machine-learning (the number of individual algorithms or models is unlimited).

Figure 7: Illustration of Ensemble Machine-Learning



Source: TIGTA hypothetical example.

As illustrated in Figure 7, ensemble machine-learning is an approach that combines multiple machine-learning algorithms to improve predictive performance.

For example, the IRS could evaluate ensemble-machine learning for the following:

- Adding an algorithm to analyze return information for multiple consecutive years for unusual trends (particularly if there are no major tax law changes).
- Adding external data or information (*e.g.*, data by industry) to measure against industry benchmarks.
- Adding an algorithm to consider the potential additional tax assessment as part of the return selection models.

According to the Office of Management and Budget's (OMB) guidance, agencies should develop adequate infrastructure and capacity to sufficiently curate agency data for use in training, testing, and operating AI.[20] Further, the IRS's 2024 IRA Strategic Operating Plan annual update cites "*AI models and systems, as well as models developed using other advanced analytics and*

[20] OMB, Memorandum M-25-21, *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust* (April 2025).

statistical methods, are trustworthy and benefit tax administration" as one of the 2025 priority efforts.[21] According to IRS management, the AI Assurance Team completed a review of the LPC, LAR, and Form 1040 return classification models in August, September, and November 2024, respectively. IRS management stated that they have made significant progress in developing infrastructure and capacity relating to AI.

Taxpayers' perceptions of the risk of being caught violating tax laws can affect their decisions, even if they themselves are not examined.[22] As one research paper concludes, there is "*strong evidence that audits are a potent tool to foster voluntary compliance … suggest[ing] that the allocation of audit resources…ought to be modified to reflect this indirect effect on voluntary compliance*."[23] By using the ensemble machine-learning approach, the IRS could better identify tax returns in which taxpayers may have violated tax laws and further encourage voluntary compliance and improve examination productivity.

Recommendation 2: The Chief Tax Compliance Officer should require SB/SE and LB&I Divisions' Commissioners, in partnership with the Chief Data and Analytics Officer, to refine AI models discussed in this report by incorporating ensemble machine-learning when appropriate.

Management's Response: IRS management agreed with the recommendation. The IRS stated that it has already tested and implemented ensemble methods in these models, where appropriate.

The IRS Needs to Improve Processes to Evaluate Implemented Artificial Intelligence Models' Performance

The IRS does not have sufficient processes for evaluating whether implemented AI models perform better than prior processes or are achieving their intended objectives. For example, the IRS cannot readily demonstrate in real-world context if AI models are improving the IRS's examination compliance efforts overall.

According to the OMB's guidance, federal agencies are encouraged to better track and evaluate performance of their procured AI. Agencies are to conduct ongoing testing and validation on AI model performance and associated risk management measures in real-world conditions.

Furthermore, the Government Accountability Office (GAO) identified key practices to help ensure accountability and responsible AI use by federal agencies and other entities involved in the design, development, deployment, and continuous monitoring of AI systems.[24] Two key practices cover performance and monitoring:

[21] IRS, *2024 IRA Strategic Operating Plan Annual Update Supplement*, Publication 3744-A (Rev. April 2024).

[22] Congressional Budget Office, Publication 56422, *Trends in the Internal Revenue Service's Funding and Enforcement* (July 2020).

[23] Alan H. Plumley, *The Impact of the IRS on Voluntary Tax Compliance: Preliminary Empirical Results*, National Tax Association 95th Annual Conference on Taxation (November 14–16, 2002).

[24] GAO, GAO-21-519SP, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (June 2021).

- The need to define performance metrics that are precise, consistent, and reproducible. Moreover, a need to assess performance against defined metrics to ensure that the AI system functions as intended and is sufficiently robust.
- AI systems are dynamic, and performance can vary over time. Management should establish a monitoring framework to ensure that the AI system maintains its utility and stays aligned with objectives. This would involve monitoring performance, which includes tracking outputs generated from predictive models and performance parameters to determine if the results are as expected. For example, when there are major tax law changes, the models should be re-evaluated and updated as necessary.

Furthermore, according to the GAO, as part of the monitoring plan, entities should decide and document the range of model drift that is acceptable. Model drift refers to the changes in relationship between the data inputs and the prediction outputs. Model drift could cause performance degradation (*e.g.*, inaccurately predicting tax compliance due to lack of continuous machine-learning). Entities may need to retrain the component of the AI system if the model drift for each component is not within the acceptable range. The range should be established based on the nature, scope, and purpose of the component and the risks it poses.

During our review, IRS management noted the evaluation processes conducted included pilot programs using random sampling and controls (for the individual return classification and selection models), and back-testing of models using historical audit results. However, these processes were conducted pre-implementation. To follow the OMB's guidance, the IRS should create an ongoing process to monitor each model's performance and demonstrate to leadership and stakeholders whether AI models are improving examination performance and achieving the IRS's goals. IRS management agreed that it could do more to monitor performance.

Recommendation 3: The Chief Tax Compliance Officer should require SB/SE and LB&I Divisions' Commissioners, in partnership with the Chief Data and Analytics Officer where appropriate, to establish a measurement plan with appropriate metrics to monitor the Form 1040 return classification and LAR and LPC models to ensure that they are achieving the expected benefits and correct any model drifts.

Management's Response: IRS management agreed with the recommendation, subject to staffing constraints and anticipated new Treasury guidance for AI governance. The Chief Data and Analytics Officer will establish a procedure for AI use cases such as those in this report to document their performance measurement and monitoring plans.

Appendix I

Detailed Objective, Scope, and Methodology

The overall objective of this audit was to determine the effectiveness of the LB&I and SB/SE Divisions' AI models in selecting returns and issues for examination. To accomplish our objective, we:

- Interviewed LB&I and SB/SE Divisions' and RAAS function management regarding the genesis, objectives, development, methodology, machine-learning source data, testing/piloting, and oversight of the Form 1040 return classification and selection and the LAR and LPC models.
- Discussed with LB&I and SB/SE Divisions' and RAAS function management how the IRS plans to evaluate whether AI models are better or worse than prior processes.
- Obtained information on the external contractors involved in the Form 1040 return classification and selection, LAR and LPC models' development, and the associated costs.
- Reviewed the Form 1040 return classification and return selection models pilot test results.
- Reviewed the IRS's estimated additional annual tax assessments by reassigning Form 1040 return classifiers to conduct examinations.
- Obtained the IRS's corrective actions in response to the GAO's audit recommendations on the IRS's enforcement efforts of large partnerships.[1]
- Obtained information on the LB&I Division's data scientists staffing increase as a result of IRA funding.
- Reviewed guidance governing federal agencies' use of AI as promulgated by Presidential Executive Orders, the OMB, and the GAO.

Performance of This Review

This review was performed with information obtained from the IRS's RAAS function and the LB&I and SB/SE Divisions, which are all headquartered in Washington, D.C., during the period September 2023 through January 2025. We conducted this performance audit in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objective. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objective.

Data Validation Methodology

We performed tests to assess the reliability of published data reported in the IRS FY 2023 Data Book, Table 17, *Examination Coverage and Recommended Additional Tax After Examination, by*

[1] GAO, GAO-23-106020, *Tax Enforcement-IRS Audit Processes Can Be Strengthened to Address a Growing Number of Large, Complex Partnerships* (July 2023).

Type and Size of Return, Tax Years 2013–2021. Specifically, we checked the mathematical accuracy of the TYs 2017 through 2021 total number returns examined and closed and total number of returns examined with no change used to compute the no change examination rates for individual and corporate returns. We evaluated the data by (1) performing reconciliations for each tax year and (2) through discussions with agency officials knowledgeable about the data. We determined that the data were sufficiently reliable for the purposes of this report.

We also performed tests to assess the reliability of the 82 partnership returns selected for examination under the LPC model. We evaluated the data by (1) researching the Integrated Data Retrieval System to confirm the taxpayer identification number, taxpayer name, and existence of an examination indicator, and (2) interviewing agency officials knowledgeable about the data. We determined that the data were sufficiently reliable for the purposes of this report.

Internal Controls Methodology

Internal controls relate to management's plans, methods, and procedures used to meet their mission, goals, and objectives. Internal controls include the processes and procedures for planning, organizing, directing, and controlling program operations. They include the systems for measuring, reporting, and monitoring program performance. We did not assess internal controls because doing so was not applicable within the context of our objective. Our analysis was limited to reviewing the no-change rate by type of return.

Appendix II

The Small Business/Self-Employed Division's Form 1040 Return Classification Model

The SB/SE Division uses the Form 1040 return classification model to identify examination issues for DIF-selected returns.[1] DIF-selected returns are scored using the DIF model and delivered in descending score order for classification.

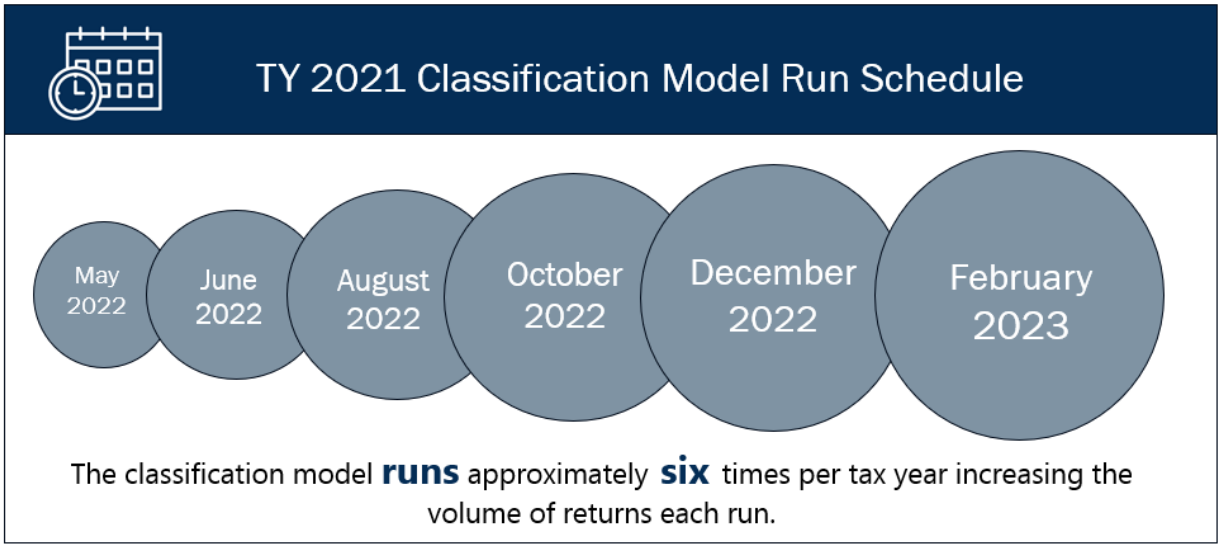
According to IRS management, DIF was implemented in 1969 and not based on AI. However, DIF is a supervised learning model based on regression model/discriminant analysis. The SB/SE Division chose to implement the Form 1040 return classification model to replace the prior manual classification process after testing historical data and conducting a statistically valid generalized control trial pilot (compared the return classification model's performance to manual classification). Classification is the process of screening returns to determine what issues on the tax return should be audited, if any, and the type of employee who should conduct the audit. According to IRS management, manual classification resulted in different outcomes when the same pool of returns was given to different employees to classify, and it takes examiners offline instead of examining returns.

The Form 1040 return classification model is generally trained on the total population of individual income tax returns each tax year and [redacted][2] Based on machine-learning, the model computes an expected amount for each filled line-item and calculates the difference between what the taxpayer reported and what the machine computes. Line-items with larger differences are assigned a higher risk score. The Form 1040 return classification model does not explicitly calculate a potential tax assessment amount, [redacted] Each tax year, the Form 1040 return classification model will run approximately six times. Figure 1 provides, for TY 2021 returns, the Form 1040 return classification model run schedule.

[1] According to the SB/SE Division's FY 2023 Field Examination performance data, about 29 percent of the closed audit cases were DIF-selected returns. The percentage of cases selected for examination by DIF can vary widely from year to year.

[2] [redacted]

Figure 1: TY 2021 Form 1040 Return Classification Model Run Schedule



Source: SB/SE Division and RAAS management.

The model learns as taxpayers file additional returns. After the last run, the scoring is finalized. For each selected Form 1040 return, the model's top three issues are forwarded with the tax return package to an examiner and serve as a guide when assessing the return. The examiner can disregard recommended issues and/or add other issues based on their independent risk analysis.

According to IRS management, automating the classification function allows the IRS to reassign 14 classifiers to the examination function, and it estimates that the increased examinations could generate approximately $26 million in additional tax assessments yearly. Furthermore, based on FY 2022 processing, the IRS estimated another $5.7 million to $8.2 million in additional tax assessments from repurposed resources after simplifying procedures for training return classification.

Appendix III

The Small Business/Self-Employed Division's Form 1040 Return Selection Model

The same contractor worked on both the Form 1040 return classification model and the selection model. Under RAAS supervision, the contractor wrote the computer code to develop and test the models. The Form 1040 return selection model is built on the same Form 1040 data used to train the classification model. [redacted]

The return selection model assigns a score to each return representing the total aggregated risk identified by the model. The model assumes that compliance risk for many of the line items is directional (*e.g.*, noncompliant taxpayers will underreport income and/or overreport expenses). [redacted]

Like the classification model, the return selection model is retrained as more returns are filed for a specific tax year and will learn more with each run. The return selection model follows the same run schedule as the classification model presented in Appendix II. The overall return score could change with each subsequent model run.

[redacted] The RAAS group acknowledged that the former Secretary of the Treasury's directive to not increase the audit rate relative to historical levels for households below $400,000 will affect the return selection model for certain taxpayers and must be resolved if SB/SE Division management decides to implement the Form 1040 return selection model.

Appendix IV

The Large Business and International Division's Line Anomaly Recommender Return Selection Model

The LB&I Division is using the LAR model in place of the Discriminant Analysis System (DAS) model for the purpose of selecting training returns. The current DAS model for selecting training returns in examination activity codes [redacted] (*i.e.*, the most recent tax year that the DAS model is based on). Activity codes are assigned to tax returns during processing to categorize types of returns based on specific elements such as the amount of total assets or income. According to IRS management, DAS is not an AI model; however, it is a supervised learning model.

A prior TIGTA audit analyzed the Form 1120 returns closed in the DAS workstream during FY 2015 through FY 2018 and found that nearly 50 percent were closed with no change to the tax return.[1]

According to LB&I Division management, they were looking for new opportunities to improve return selection. When the LAR model became available, and the testing of the model showed higher selection of non-compliance returns, they decided to use it instead of the DAS model for the three activity codes to select training returns. LB&I Division management stated that the LAR model is an addition to the LB&I Division's return selection catalog.

The source data used to train the LAR model includes [redacted]

[1] TIGTA, Report No. 2020-30-031, *The Large Case Examination Selection Method Consistently Results in High No-Change Rates* (June 2020).

Figure 1: [redacted]

The LAR model produces scores for each return line-item and an overall risk score for the entire return. The LAR model is relationship-based, so it looks at the relationship between the return line-items. It generates an estimated value for every single line-item on the return and then compares those values to what was reported on the return. The difference (anomaly) between machine-computed and return-reported data drives the risk score. [redacted]

The IRS typically runs the LAR model at least twice for each tax year to capture both calendar year and fiscal year filers.

To mitigate bias, all the returns within the specific activity code are used. The model does not select issues for examination. Even though the model identifies anomalies by Form 1120 line-items, LB&I Division management explained that the scoring does not translate into issues for examination. The model output is manually classified and/or secondary issue filters are used to identify high risk returns. The classifiers do not receive the LAR model scoring information.

The LAR model was transitioned from RAAS to the LB&I Division in October 2021, and that division currently manages the model for each tax year. According to the LB&I Division, it mitigated potential risks of model drift by switching to a tax year-based model.[2] The LAR model

[2] Model drift refers to the changes in the relationship between the data inputs and the prediction outputs. Model drift could result in performance degradation. Source: GAO, GAO-21-519SP, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (June 2021).

is always trained on the most recently filed tax year returns. No specific taxpayer population is targeted when the entire Form 1120 population is used within each activity code.

Prior to deploying the LAR model, the IRS did not conduct a randomized controlled test, which according to IRS personnel is the highest standard for testing. However, the IRS performed testing using past examination results. [redacted]

- [redacted]
- [redacted].

Appendix V

The Large Business and International Division's Large Partnership Compliance Return Selection Model

The LB&I Division's definition of a "large" partnership is not based on any single characteristic. Figure 1 shows the Large Partnership characteristics. [redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]

According to LB&I Division management, the first 82 large partnership returns identified as high risk by the LPC model and currently under examination are TY 2021 returns. On average, these returns have more than $10 billion in assets. Each year, the IRS creates a new model using prior year tax return data. In April 2023, the LB&I Division's Risk Identification Control Board approved the TY 2021 LPC model for production. The TY 2021 LPC model was built off the original LPC model contract work, which was first developed using TY 2019 return data. Prior to deployment, the IRS tested the model to evaluate performance but did not conduct a pilot.

According to IRS management, the TY 2021 large partnership return selection process answered two questions:

1. Who is in the large partnership population?
2. What is the risk with each large partnership?

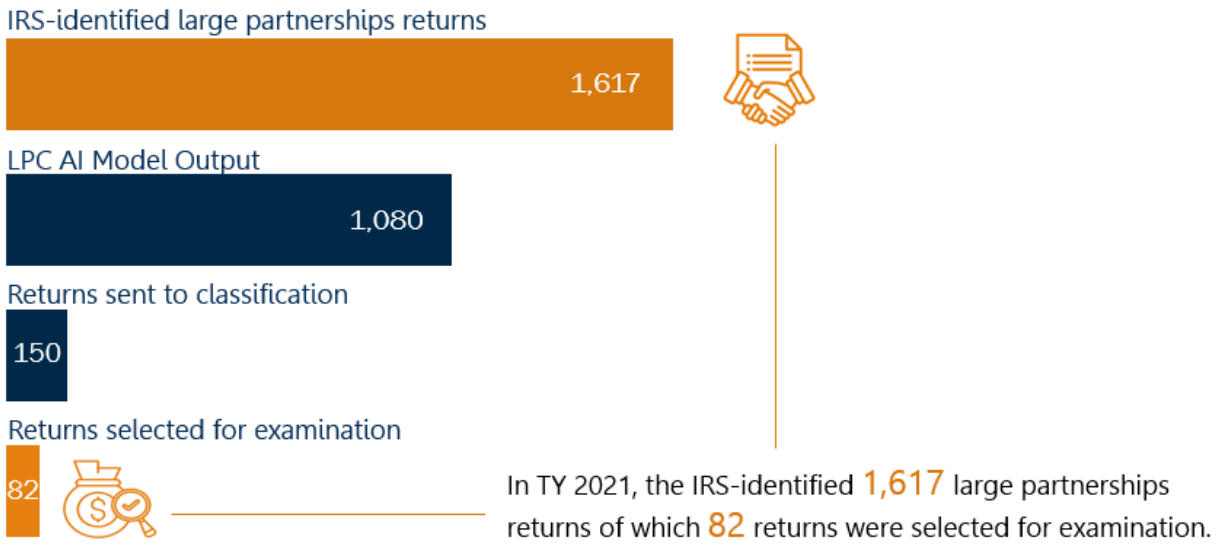
[redacted] According to the IRS, it had 1,617 potential large partnerships for TY 2021.

The IRS used the LPC model to answer the second question regarding risk. The LPC model consisted of two aspects:

Data science – The LPC model used machine-learning to analyze the complete TY 2021 large partnership return population. The input data included all the associated supporting tax forms and schedules. The LPC model looked for outliers and anomalies (*e.g.*, high expenses compared to the filing population).

[redacted]

Figure 2: TY 2021 Large Partnership Return Counts[1]



Source: LB&I Division management.

The most recent contract for the LPC model was for a period of performance from September 2022 through December 29, 2024, for $4.7 million. Thereafter, the LB&I Division assumed the responsibility of maintaining the LPC model. The contractor was hired because the LB&I Division did not have sufficient staffing to do modeling development. With IRA funding, the LB&I Division hired additional data scientists to perform the work in-house. However, with recent events, it is uncertain how the reduced staffing will impact the LPC model.

In July 2023, the GAO issued a report on the IRS's enforcement efforts of large partnerships.[2] The GAO made four recommendations, including improving the design and testing of its models, as well as developing guidance to define and establish measures to track large and complex partnership audits. According to IRS management, they are taking actions in response to the GAO's recommendations. For example, they are reviewing a sample of returns that the LPC model identified as not having any risks. Also, it plans to incorporate feedback from the recently completed classification of TY 2021 LPC returns in developing the next iteration of the LPC model. Furthermore, the LB&I Division is in the early stage of working with RAAS to develop partnership segments to better define characteristics of large, complex partnership entities. The IRS disagreed with the GAO's recommendation to create new examination activity codes to track resources used and examination results.

[1] Among the 82 returns selected for examinations, 3 returns are not in the 150 returns sent to classification.

[2] GAO, GAO-23-106020, *Tax Enforcement-IRS Audit Processes Can Be Strengthened to Address a Growing Number of Large, Complex Partnerships* (July 2023).

Appendix VI

Individual Return Examination Activity Codes

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

Appendix VII

Management's Response to the Draft Report

RESEARCH, APPLIED ANALYTICS AND STATISTICS

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, DC 20224

April 30, 2025

MEMORANDUM FOR DANNY VERNEUILLE
ACTING DEPUTY INSPECTOR GENERAL FOR AUDIT

FROM: Reza Rashidi RezaRashidi
Acting Chief Data and Analytics Officer

SUBJECT: Draft Audit Report – The IRS Could Leverage Examination Results in Artificial Intelligence Examination Case Selection Models and Improve Processes to Evaluate Performance (Audit # 2024308019)

Thank you for the opportunity to review your report titled *The IRS Could Leverage Examination Results in Artificial Intelligence Examination Case Selection Models and Improve Processes to Evaluate Performance*.

The IRS Office of Research, Applied Analytics, and Statistics (RAAS) is the Service's centralized research and analytics organization. One of RAAS's roles is to develop models to assist in selecting cases for examination. In this role RAAS works closely with partners in SB/SE and LB&I to develop, test, deploy, and refine models.

The IRS is committed to continuously improving the case selection process. This includes making use of well-established processes to evaluate model performance. This begins with extensive model testing during the development phase, including simulated testing on historical examination results, conducting pilot examinations in randomized controlled trials, and other methods as appropriate. All of this is done prior to widespread operational deployment to ensure any new model demonstrates improvement compared to current benchmarks. After a model is deployed, RAAS continues to monitor and evaluate model performance in concert with partners across the IRS to ensure models perform as expected and meet the needs of stakeholders.

The report focuses on three modeling programs: The Issue Recommender classification and case selection models (developed for SB/SE and focusing on Form 1040 individual returns), the Line Anomaly Recommender model (developed for LB&I and focusing on Form 1120 corporate returns), and the Large Partnership Compliance model (developed for LB&I and focusing on Form 1065 partnership returns). The AI use cases

in all three of these programs are subject to oversight under the IRS AI governance process.

As noted in your report, the AI governance process was established via interim guidance in May 2024. In accordance with that process, the AI use cases in deployment by the three programs were reviewed by the AI Assurance Team (AIAT) in calendar year 2024. For the AIAT reviews, representatives from the three programs provided evidence of the procedures currently in place for performance evaluation and ongoing monitoring and improvement. The use cases were recommended for approval by the AIAT following completion of their reviews, and in November 2024, the use cases were approved for continued use by a board of IRS senior executives.

In response to new executive orders and changes to governmentwide requirements for AI governance policies, the IRS issued new interim guidance in March 2025 which substantially modified the AI governance process. One change implemented by that guidance is the suspension of AIAT reviews for AI use cases. The IRS currently awaits additional guidance regarding new policies and priorities for AI governance within the Department of Treasury and will revise the AI governance process as necessary to align with new guidance received. As the AI governance process evolves, the IRS will continue to collect appropriate documentation from AI use cases, including documentation of performance evaluation, model effectiveness, and ongoing monitoring and improvement efforts, consistent with applicable law and governmentwide guidance. These activities align with Recommendation 1 and Recommendation 3 in the report, which focus on establishing governance processes and measurement plans to monitor and analyze the performance of these models.

The remaining recommendation in the report, Recommendation 2, focuses on incorporating ensemble methods when appropriate. Multiple factors must be considered (including model overfitting, data reliability and availability, and other factors) when weighing any benefits gained from using ensemble methods. RAAS regularly considers and tests ensemble methods during model development. (See, for example, the attached paper *A Semi-Supervised Approach to Anomaly Detection for Tax Compliance*, which describes research from the Issue Recommender program on an ensemble model for Form 1040 return scoring.) We employ ensemble methods only when it is feasible and beneficial to do so.

We thank you for acknowledging the important work that was done by RAAS, SB/SE, and LB&I during the development of these models to ensure they met expected performance standards before deployment. We also thank you for allowing us the opportunity to explain how our work already aligns with the recommendations in the report. We agree, as your report states, that “given the recent governmentwide cost cutting efforts (e.g., IRA funding recission, hiring freeze, and anticipated reduction in force) it is now unclear if the IRS will be able to pursue and achieve [their] plans” to “evaluate and improve data, tools, and governance processes to help ensure that models are working as intended,” among other goals described in the Inflation Reduction Act Strategic Operating Plan. However, the IRS is committed to exploring

new ways to ensure our programs are improving examination compliance efforts overall.

Our corrective action plan for the recommendations is attached. If you have any questions, please contact me at [redacted] or a member of your staff may contact [redacted] at [redacted]

Attachment

RECOMMENDATION #1:
The Chief Tax Compliance Officer should require the SB/SE and LB&I Divisions' Commissioners, in partnership with the Chief Data and Analytics Officer, where appropriate, to use governance processes to ensure that examination performance results are part of the monitoring and continuous refinement of AI classification (deployed) and return selection models (if deployed).

CORRECTIVE ACTION:
The IRS agrees with this recommendation, subject to staffing constraints and anticipated new Treasury guidance for AI governance. The Chief Data and Analytics Officer will verify that AI classification and return selection use cases go through the AI governance process and will establish a procedure for such use cases to document how examination performance results are used for monitoring and regular refinement of AI models.

IMPLEMENTATION DATE:
June 15, 2026

RESPONSIBLE OFFICIAL:
Chief Data and Analytics Officer

RECOMMENDATION #2:
The Chief Tax Compliance Officer should require SB/SE and LB&I Divisions' Commissioners, in partnership with the Chief Data and Analytics Officer, to refine AI models discussed in this report by incorporating ensemble machine-learning when appropriate.

CORRECTIVE ACTION:
The IRS agrees with this recommendation. The IRS has already tested and implemented ensemble methods in these models, where appropriate.

IMPLEMENTATION DATE:
Implemented

RESPONSIBLE OFFICIAL:
Chief Data and Analytics Officer

RECOMMENDATION #3:
The Chief Tax Compliance Officer should require SB/SE and LB&I Divisions' Commissioners, in partnership with the Chief Data and Analytics Officer where appropriate, to establish a measurement plan with appropriate metrics to monitor the Form 1040 return classification and LAR and LPC models to ensure that they are achieving the expected benefits and to correct any model drifts.

TIGTA 2024308019

CORRECTIVE ACTION:
The IRS agrees with this recommendation, subject to staffing constraints and anticipated new Treasury guidance for AI governance. The Chief Data and Analytics Officer will establish a procedure for AI use cases such as those in this report to document their performance measurement and monitoring plans.

IMPLEMENTATION DATE:
June 15, 2026

RESPONSIBLE OFFICIAL:
Chief Data and Analytics Officer

Appendix VIII

Glossary of Terms

Term	Definition
Activity Code	A code that identifies the type and condition of returns selected for audit.
Data Science	Per the U.S. Census Bureau, a field of study that uses scientific methods, processes, and systems to extract knowledge and insights from data.
Discriminant Analysis System	A computer model developed to score Forms 1120 as to examination potential. Generally, the higher the score, the greater the audit potential.
Fiscal Year	Any yearly accounting period, regardless of its relationship to a calendar year. The federal government's fiscal year begins on Oct. 1 and ends on Sept. 30.
Integrated Date Retrieval System	IRS computer system capable of retrieving or updating stored information. It works in conjunction with a taxpayer's account records.
Internal Revenue Manual	The primary source of instructions to employees relating to the administration and operation of the IRS. The Manual contains the directions employees need to carry out their operational responsibilities.
Machine-Learning Algorithm	A set of rules or processes used by an AI system to conduct tasks. Most often to discover new data insights and patterns, or to predict output values from a given set of input variables.
National Research Program	Provides a statistically valid random sample of filed returns representative of the compliance characteristics of taxpayers. Returns in this program are assigned to examiners as quickly as possible, and surveys before or after assignment are limited.
Tax Year	A 12-month accounting period for keeping records on income and expenses used as the basis for calculating the annual taxes due. For most individual taxpayers, the tax year is synonymous with the calendar year.
Total Positive Income	The sum of all positive amounts shown for the various sources of income reported on the individual tax return and, therefore, excludes losses.

Appendix IX

Abbreviations

AI	Artificial Intelligence
DAS	Discriminant Analysis System
DIF	Discriminant Index Function
EITC	Earned Income Tax Credit
FY	Fiscal Year
GAO	Government Accountability Office
IRA	Inflation Reduction Act
IRS	Internal Revenue Service
LAR	Line Anomaly Recommender
LB&I	Large Business and International Division
LPC	Large Partnership Compliance
NAICS	North American Industry Classification System
OMB	Office of Management and Budget
RAAS	Research, Applied Analytics & Statistics
SB/SE	Small Business/Self-Employed Division
TIGTA	Treasury Inspector General for Tax Administration
TY	Tax Year

To report fraud, waste, or abuse,
contact our hotline on the web at www.tigta.gov or via e-mail at oi.govreports@tigta.treas.gov.

To make suggestions to improve IRS policies, processes, or systems affecting taxpayers, contact us at www.tigta.gov/form/suggestions.

Information you provide is confidential, and you may remain anonymous.