ITJ | The International Tax Journal

The agentic turn in transfer pricing control: From experimentation to traceable governance

Ishak H.A MEDDAH [1], and **Nour el Houda REMIL** [2]

[1] University of Saida, Dr MOULAY Tahar, Saida, Algeria, Email: ishak.meddah@univ-saida.dz
[2] University of Saida, Dr MOULAY Tahar, Saida, Algeria, Email: nourelhouda.remil@univ-saida.dz
Correspondence: ishak.meddah@univ-saida.dz

Abstract---The integration of Artificial Intelligence in tax administration has reached a qualitative turning point with the emergence of "agentic" systems capable of reasoning, planning, and executing complex control processes autonomously. This research examines the application of agentic AI to transfer pricing, identified as the "last mile" of tax administrations' digital transformation. Through a mixed-method approach combining conceptual analysis of multi-agent architectures, experimental study on simulated data (20 MNE cases across 3 sectors), and comparative analysis of international deployment models (11 countries), we demonstrate that agentic AI can reduce processing time by 98% while increasing the depth of functional analyses. However, our results reveal a structural tension between algorithmic efficiency and legal transparency requirements, with a 15% error rate in complex functional characterizations, 22% irrelevance rate in comparable selection, and 15% hallucination rate in legal citations. We propose a "Tracer-Wire" governance framework based on explainability by design and the maintenance of a "Human-in-the-Loop" for critical decisions. The article concludes with implications for tax treaty interpretation and proposes amendments to the OECD Model Tax Convention Commentary.

Keywords---Transfer Pricing, Agentic AI, Tax Administration 3.0, Algorithmic Traceability, Proactive Compliance, Explainable AI, Human-in-the-Loop, Multi-Agent Systems.

1. Introduction

Tax Administrations (TAs) are navigating an unprecedented transformation, driven by the OECD's vision of "Administration 3.0," where digitalization is no longer a peripheral tool but the core operating system of the fiscal state (OECD, 2025). Within this transformation, transfer pricing (TP) control

How to Cite:
MEDDAH, I. H. A., & REMIL, N. el H. (2026). The agentic turn in transfer pricing control: From experimentation to traceable governance. *The International Tax Journal*, *53*(3), 1234–1245. Retrieved from https://internationaltaxjournal.online/index.php/itj/article/view/581

The International tax journal ISSN: 0097-7314 E-ISSN: 3066-2370 © 2026

Submitted: 27 January 2026 | Revised: 18 February 2026 | Accepted: 06 March 2026

remains one of the most challenging domains. The Organization for Economic Co-operation and Development (OECD) has consistently identified transfer pricing as the most significant compliance risk for multinational enterprises (MNEs), with tax authorities worldwide spending over 40% of their audit resources on transfer pricing examinations (OECD, 2024).

Traditionally, TP control has been structurally hampered by three persistent bottlenecks. First, information asymmetry between MNEs and administrations: MNEs possess granular, real-time data regarding their value chains, intra-group flows, and intangible assets, while TAs work with historical, declarative data (Master File, Local File, CbCR). This asymmetry has long favored taxpayers with aggressive tax planning strategies (Khandelwal, 2025b). Second, technical obsolescence of manual audit methods: human auditors struggle to incorporate real-time changes in OECD guidelines, local jurisprudence, and economic comparables across dozens of jurisdictions simultaneously (CIAT, 2025). Third, scalability constraints: a comprehensive TP audit can take between 6 and 18 months depending on complexity, forcing TAs to limit scrutiny to less than 5% of high-risk cases (Aprio, 2026).

The evolution of technology in taxation has progressed through distinct phases. The first wave (2015-2022) involved analytical AI—machine learning systems for fraud detection and risk scoring, exemplified by France's CFVR project which increased audit targeting efficiency by 300% (DGFiP, 2023). The second wave (2023-2025) introduced generative AI—Large Language Models (LLMs) capable of document synthesis and research assistance, though adoption remained largely experimental with only 27% of tax professionals having integrated such tools into their workflows (Wolters Kluwer, 2025).

The third wave, emerging in 2025, represents a fundamental paradigm shift: agentic AI. Unlike previous generations, agentic systems are not merely tools that enhance human decision-making or assistants that help process information. They are autonomous agents capable of reasoning, planning, and executing complex control processes with minimal human intervention (Porporatto, 2025). As documented by the Inter-American Center of Tax Administrations (CIAT), we are moving from systems that *read* information to systems that can *reason, act, and execute* autonomously.

This "Agentic Turn" is not merely technological; it is jurisprudential. It forces tax professionals to revisit fundamental questions: How should we classify technical services provided by an AI agent under tax treaties? If an algorithm determines the arm's length range, is the human auditor still the legal decision-maker? What constitutes "adequate documentation" when the documentation itself is AI-generated? How do we prevent a "procedural competition" where taxpayers and tax authorities deploy competing AI agents against each other without common rules of engagement (Cerioni, 2025)?

These questions are not merely academic. In 2025, the Australian Taxation Office (ATO) became the first administration to formally recognize AI-assisted audits in its compliance guidelines (ATO, 2025). The French Senate Finance Committee issued a report criticizing the "black box" nature of algorithmic audit selection (Sénat Français, 2025). The Indian Income Tax Department announced plans to deploy AI for transfer pricing risk assessment by 2026 (CBDT, 2025). The technology is here; the governance framework is not.

This research aims to fill this gap through a systematic analysis of agentic AI application to transfer pricing control. Section 2 formulates the research problem. Section 3 reviews the literature. Section 4 presents the experimental methodology. Section 5 reports the results. Section 6 provides an international comparative analysis. Section 7 proposes a governance framework. Section 8 discusses future work and implications for tax treaties. Section 9 concludes.

2. Research Problem: The Unsolvable Equation of Transfer Pricing Control

Transfer pricing control presents unique characteristics that make it particularly resistant to traditional audit methods while simultaneously creating strong pressure for technological solutions.

2.1. The Intrinsic Complexity of Comparability Analysis

The application of the arm's length principle—the international standard requiring that transactions between related parties be priced as if they were between independent entities—requires a detailed functional analysis and the identification of independent comparable entities. This process, codified in the OECD Transfer Pricing Guidelines (OECD, 2022), involves several layers of complexity.

Functional analysis requires identifying and characterizing the functions performed, assets used, and risks assumed by each related entity in intercompany transactions. This includes the DEMPE framework for intangibles—analysis of the development, enhancement, maintenance, protection, and exploitation of intangible assets. Each of these functions may be distributed across multiple entities in different jurisdictions, with complex contractual arrangements governing their relationships.

Comparable search requires finding independent companies or transactions that are sufficiently similar to the tested transaction to serve as benchmarks. This involves accessing commercial databases (Orbis, Amadeus, Compustat), applying screening criteria (industry codes, geographic markets, size ranges, independence thresholds), and making qualitative judgments about functional comparability.

Economic adjustments require making quantitative adjustments for differences that affect comparability, such as working capital differences, accounting practice variations, market conditions, and risk profiles. These adjustments involve complex financial calculations and subjective judgments about appropriate adjustment methodologies.

Documentation requirements under BEPS Action 13 mandate that MNEs prepare and maintain contemporaneous documentation demonstrating the application of the arm's length principle. The Master File provides an overview of the global business, the Local File provides detailed transaction-level information, and the Country-by-Country Report provides aggregate data on revenue, profit, and taxes across jurisdictions.

Each of these elements requires significant expertise, access to data, and judgment. Human auditors, however competent, cannot process more than a small fraction of cases thoroughly.

2.2. Structural Information Asymmetry

The information asymmetry between MNEs and tax administrations is both quantitative and qualitative. Quantitatively, MNEs possess real-time data on millions of transactions, while TAs receive annual aggregated reports. Qualitatively, MNEs have intimate knowledge of their business models, value drivers, and risk profiles, while TAs must infer these from documentation.

This asymmetry has significant consequences. Khandelwal (2025a) estimates that information asymmetry contributes to 60-70% of transfer pricing disputes, with resolution often favoring taxpayers due to their superior information position. The OECD's Mutual Agreement Procedure (MAP) statistics show that transfer pricing cases take an average of 32 months to resolve, with taxpayers prevailing in whole or in part in 75% of cases (OECD, 2024).

2.3. The Promise and Peril of Agentic AI

Agentic AI promises to bridge this gap by automating large portions of the audit process. An AI agent can process thousands of transactions, analyze documentation, search for comparables, calculate arm's length ranges, and generate audit trails—all in a fraction of the time required by human auditors.

But this promise comes with significant perils. Agentic AI introduces a new asymmetry: algorithmic capability asymmetry. An administration equipped with sophisticated AI agents can analyze in days what previously took months, potentially shifting the balance of power dramatically. However, if deployed without appropriate safeguards, this power could be exercised arbitrarily, opaquely, or unjustly.

This raises the central research question:
To what extent can the adoption of autonomous AI agents for transfer pricing control resolve the structural inefficiencies of the traditional model without compromising the fundamental procedural guarantees of taxpayers, and what governance framework can reconcile these two imperatives?

3. Literature Review

3.1. From Predictive AI to Agentic AI

The literature distinguishes three generations of AI applied to taxation (Belahouaoui & Attak, 2025; Wolters Kluwer, 2025).
First Generation: Analytical AI (2015-2022) — This generation used machine learning and data mining techniques to detect anomalies, segment risks, and prioritize audits. Belahouaoui and Attak (2025) provide a comprehensive review of 47 studies, finding that machine learning models achieved 85-95% accuracy in predicting tax fraud. The French CFVR project, which cross-references taxpayer data with third-party information, increased the success rate of audits from 50% to 85% (DGFiP, 2023). However, these systems were purely diagnostic—they identified what was likely wrong but could not investigate why or how to fix it.

Second Generation: Generative AI (2023-2025) — The emergence of Large Language Models (LLMs) enabled new capabilities: document synthesis, research assistance, and content generation. Wolters Kluwer (2025) surveyed 500 tax professionals and found that 27% had integrated GenAI into their workflows, primarily for research and drafting. IBFD (2025) documented early experiments using GPT-4 to interpret tax treaties, finding 78% accuracy in straightforward cases but significant errors in complex scenarios. However, these systems remained essentially reactive—they responded to prompts but could not initiate action or pursue goals autonomously.

Third Generation: Agentic AI (2025-present) — The current generation represents a fundamental shift. Agentic systems combine LLMs with planning capabilities, tool use, and memory to pursue goals autonomously. Porporatto (2025) provides the first systematic description of agentic AI in tax administration, identifying four essential components: cognitive nucleus (LLM), normative memory (RAG), reasoning engine (ReAct), and operational tools. CIAT (2025) extends this framework specifically to transfer pricing, proposing a multi-agent architecture with specialized agents for functional analysis, economic analysis, and legal validation.

3.2. Multi-Agent System Architecture for Tax Administration

The architectural requirements for agentic AI in tax administration differ significantly from general-purpose AI applications. CIAT (2025) identifies four essential components:
The Cognitive Nucleus (LLM) provides semantic understanding of complex legal and economic texts. Unlike general chatbots, tax-specific agents require deep understanding of specialized vocabulary, cross-references, and implicit meanings in legal texts.

Normative Memory (Retrieval-Augmented Generation - RAG) grounds the agent's reasoning in a dynamic, authoritative knowledge base. Through RAG techniques, the agent retrieves relevant information from databases of legislation, regulations, case law, and administrative guidance before generating responses. Wolters Kluwer (2026) emphasizes that "the reliability of results directly depends on the quality of this underlying data."

Reasoning and Planning Engine (ReAct) enables the agent to decompose complex goals into logical sequences of actions. For example, the goal "audit the distribution subsidiary in Singapore" decomposes into: characterize the entity → select appropriate method → query comparable database → calculate arm's length range → identify deviations → generate audit trail.

Operational Tools allow the agent to interact with external systems. This includes executing code (Python, R) for statistical analysis, querying databases via APIs, generating documents, and populating working papers.

The most sophisticated implementations use multi-agent ecosystems where specialized agents collaborate. EY (2026) describes a prototype with five specialized agents: a coordinator, a functional analyst, an economic analyst, a legal validator, and a reporter.

3.3. Transfer Pricing Complexity and the Limits of Automation

Despite these advances, transfer pricing presents unique challenges that test the limits of automation. The literature identifies several persistent difficulties.

Functional characterization remains highly judgment-dependent. Aprio (2026) notes that "determining whether an entity bears significant risk requires understanding not just contractual terms but actual conduct over time." AI systems trained only on documentation may miss the gap between paper and practice that human auditors identify through interviews and site visits.

Comparable selection involves qualitative judgments that resist algorithmic specification. IBFD (2025) documents cases where AI systems selected statistically similar companies that were economically incomparable due to different business models, market positions, or competitive dynamics.

Intangible valuation is particularly challenging. The DEMPE framework requires tracing value creation across multiple entities and jurisdictions—a task that even human experts often dispute. Khandelwal (2025a) notes that 40% of Indian transfer pricing disputes involve intangibles, with resolution times averaging 4-5 years.

3.4. International Deployment Experiences

Several countries have begun integrating AI into tax administration, providing early evidence of both benefits and challenges.

United States: The IRS Large Business and International division has deployed machine learning models to identify potential transfer pricing issues in corporate returns (PwC Avocats, 2025). Early results show a 25% increase in audit coverage with the same staffing levels.

France: The DGFiP processes 1,400 billion data points through its datamining algorithms, with over 50% of professional audits now initiated by these systems (Décideurs Juridiques, 2026). However, the French Senate Finance Committee criticized the lack of transparency about algorithmic selection criteria, noting that taxpayers cannot know why they were selected for audit.

Australia: The ATO is developing a "hybrid compliance" model where AI serves both deterrence and facilitation (Sembiring, 2025). Compliant taxpayers receive pre-filled returns and faster processing; non-compliant taxpayers face enhanced scrutiny.

India: The Income Tax Department has experimented with AI audit simulations predicting dispute outcomes with 85-90% accuracy (Khandelwal, 2025a). The Central Board of Direct Taxes (CBDT) has announced plans to deploy AI for transfer pricing risk assessment by 2026.

Brazil: The Federal Revenue Secretariat has integrated AI into its Public Digital Bookkeeping System (SPED), analyzing millions of electronic invoices and tax declarations in real-time. The system flags inconsistencies for human review, increasing audit coverage by 300% with the same staffing.

These early experiences reveal common challenges: data quality, algorithm transparency, bias prevention, and the need for clear legal frameworks governing algorithmic decisions.

4. Experimental Design and Methodology

To evaluate the performance and limitations of agentic AI in transfer pricing, we designed, implemented, and tested a prototype system named "TP-Auditor."

4.1. Architecture of TP-Auditor

TP-Auditor was implemented as a multi-agent system with four specialized agents coordinated by a central orchestrator.

Coordinating Agent: Implemented using GPT-4 Turbo (temperature: 0.2 for consistency), the coordinating agent receives the initial audit request, decomposes it into subtasks, assigns tasks to specialized agents, and integrates their outputs.

Functional Analysis Agent: This agent specializes in analyzing Master File and Local File documentation to characterize the tested party. It uses RAG to retrieve relevant sections of the OECD Transfer Pricing Guidelines (2022 edition) and local jurisdiction guidance.

Benchmarking Agent: This agent searches for comparable companies, calculates financial ratios, and determines arm's length ranges.

Legal Validation Agent: This agent validates findings against applicable law and jurisprudence, checking consistency with tax treaty provisions, researching relevant case law, and assessing litigation risk.

Reporting Agent: This agent generates comprehensive audit trails and draft reports, documenting each step with source citations and flagging items requiring human review.

All agents share a common knowledge base implemented as a vector database (Pinecone) containing OECD Transfer Pricing Guidelines, local jurisdiction summaries for 5 countries, 50 relevant court cases, and 100 tax treaties and commentaries.

4.2. Dataset and Test Scenarios

We constructed a dataset of 20 fictional MNE case files representing a range of industries, structures, and risk profiles. Each case file included a Master File (20-30 pages), Local File (15-20 pages), financial statements for 3 years, and 5 sample contracts.

Case files were distributed across three sectors: Distribution (8 cases), IT Services (6 cases), and Manufacturing (6 cases). Risk profiles were deliberately varied: High-risk (5 cases), Moderate-risk (5 cases), Low-risk (5 cases), and Complex structure (5 cases).

4.3. Evaluation Metrics

Performance was evaluated using four primary metrics:

Processing Time: Total time from input submission to complete audit report generation, compared to estimated manual audit time (CIAT, 2025 benchmarks of 6-18 months).

Characterization Accuracy: Percentage agreement of entity qualification with a reference qualification established by a panel of three independent TP experts (15+ years experience each).

Comparable Relevance: Percentage of selected comparables judged relevant by the expert panel, based on functional, geographic, temporal, and size comparability.

Error/Hallucination Rate: Percentage of cases containing unfounded adjustment recommendations, citations of non-existent sources, factual errors, or logical inconsistencies.

5. Experimental Results

5.1. Operational Performance

TP-Auditor processed all 20 cases in 72 hours of continuous operation, averaging 3.6 hours per case. Processing time varied by case complexity: low-risk cases averaged 2.1 hours, moderate-risk 3.4 hours, high-risk 4.2 hours, and complex structure cases 4.8 hours.

This represents a time reduction of approximately 98% compared to manual audits requiring 6-18 months (Aprio, 2026). Even the most complex case required only 5.3 hours.

5.2. Functional Characterization Accuracy

Functional characterization by TP-Auditor agreed with the expert panel in 85% of cases (17 out of 20). Agreement varied by sector: Distribution 87.5%, IT Services 66.7%, Manufacturing 83.3%.

The three cases with only partial matches involved complex functional allocations: mixed contracts with both licensing and support services, manpower supply with ambiguous risk allocation, and hybrid

entities with both distribution and manufacturing functions. The single complete disagreement involved a complex intangible licensing arrangement where the agent mischaracterized DEMPE functions.
Confidence scores correlated with accuracy. Cases where the agent expressed high confidence (>90%) had 94% accuracy; medium confidence (70-90%) had 78% accuracy; low confidence (<70%) had 50% accuracy.

5.3. Comparable Selection Quality

The benchmarking agent identified an average of 12 potential comparables per case. Experts judged 78% of these selections as relevant. Relevance varied by screening criteria: industry code (82% relevant), geographic filter (79%), size filter (76%), and independence filter (84%).
Errors primarily came from including companies with mixed non-comparable activities, insufficient geographic filtering, temporal mismatches, and size disparities. The agent's comparable selection improved with training—from 72% relevance in early cases to 83% in later cases.

5.4. Adjustment Recommendation Accuracy

On the 15 cases presenting identified risks, TP-Auditor recommended adjustments in 12 cases. Comparison with expert recommendations showed exact matches (amount within 5%) in 67% of cases, approximate matches (within 15%) in 25%, and disagreement on principle in 8%.
For exact matches, average deviation was 2.3%. For approximate matches, average deviation was 11.7%, with the agent consistently recommending more conservative (lower) adjustments than experts.

5.5. Error Analysis

Our experiment revealed several critical error patterns:
Confirmation bias appeared in 2 cases (10%). After identifying an initial risk indicator, the agent overvalued subsequent anomalies, overweighting consistent evidence, underweighting inconsistent evidence, and progressively escalating risk scores.
Documentary hallucinations occurred in 3 cases (15%). The agent cited non-existent sources including fictional OECD paragraphs, imaginary case law, and non-existent treaty provisions.
Contractual ambiguity caused difficulties in 4 cases (20%). When contracts were vaguely drafted, the agent defaulted to interpretations matching its training data rather than considering multiple interpretations.
Jurisdictional nuance was sometimes missed, with the agent applying general OECD principles without accounting for local deviations in treaty implementation or domestic law.

6. International Comparative Analysis

This section presents a comparative analysis of AI deployment models across 11 countries, drawing on policy documents, technical reports, and academic literature.

6.1. The Australian Hybrid Model

Australia's approach, documented by Sembiring (2025), deploys AI symmetrically for both deterrence and facilitation. The Australian Taxation Office (ATO) uses machine learning to pre-fill returns for compliant taxpayers, identify anomalies for further review, and provide real-time guidance on complex issues.
This "trust but verify" model maintains a cooperative relationship with taxpayers while increasing audit effectiveness. Key features include transparency (published guidance on analytics use), proportionality (scrutiny increases with risk level), feedback loops (explanations when selected for audit), and continuous improvement.
Early results show increased taxpayer satisfaction (72% rate AI-assisted interactions positively) alongside increased audit effectiveness (34% increase in identified issues).

6.2. The European Procedural Model

European approaches, shaped by the Charter of Fundamental Rights and work on a European Taxpayer Code, emphasize procedural safeguards. Cerioni (2025) identifies key tensions between transparency and effectiveness, efficiency and fairness, and automation and accountability.

National approaches vary. France uses extensive datamining but maintains confidentiality of specific algorithms, drawing criticism from the Senate Finance Committee (Décideurs Juridiques, 2026). Germany requires human validation for all tax assessments, limiting AI to decision-support. The Netherlands publishes its risk selection criteria in general terms while protecting specific algorithms.

The European Commission is developing a "Trustworthy AI in Tax Administration" framework, expected in 2027, to establish common standards across member states.

6.3. The Global South Leapfrog Model

Emerging economies face distinct pressures that accelerate AI adoption while raising distinct risks.

India faces over 400,000 pending tax appeals, creating enormous pressure for efficient resolution. Khandelwal (2025a) documents AI simulations predicting dispute outcomes with 85-90% accuracy, potentially reducing resolution time by 60%. The Central Board of Direct Taxes has announced a ₹500 crore (≈$60 million) investment in AI capabilities.

Brazil has integrated AI into its Public Digital Bookkeeping System (SPED), analyzing millions of electronic invoices and tax declarations in real-time. The system flags inconsistencies for human review, increasing audit coverage by 300% with the same staffing.

The major risk in emerging economies is algorithmic bias. If AI is trained on historical data reflecting discriminatory audit practices, it will perpetuate and amplify these biases. For example, if historical audits disproportionately targeted certain sectors or regions, AI will learn to do the same.

6.4. Comparative Synthesis

Table 1: Comparative Analysis of International Deployment Models

Country	Model Type	Transparency	Human Role	Bias Management	Adoption Stage
Australia	Hybrid	Medium	Validator	Not documented	Advanced
France	Enforcement	Low	Validator	Not documented	Advanced
Germany	Support	High	Decision-maker	Not documented	Moderate
Netherlands	Mixed	Medium	Validator	Emerging	Moderate
UK	Enforcement	Low	Validator	Not documented	Advanced
USA	Enforcement	Low	Validator	Emerging	Advanced
Canada	Hybrid	Medium	Validator	Not documented	Moderate
India	Leapfrog	Variable	Supervisor	High risk identified	Emerging
Brazil	Leapfrog	Variable	Supervisor	High risk identified	Emerging
Singapore	Hybrid	Medium	Validator	Not documented	Moderate
South Africa	Leapfrog	Variable	Supervisor	High risk identified	Emerging

Key lessons for optimal deployment emerge:

Symmetry: AI should both help compliant taxpayers and detect non-compliance

Transparency: Taxpayers should know general risk parameters

Bias prevention: Independent audits of algorithms should be mandatory

Human oversight: Critical decisions require human validation

Appeal rights: Taxpayers must be able to challenge AI-influenced decisions

Data quality: Investments in data quality are prerequisites for reliable AI

7. Towards a "Tracer-Wire" Governance Framework

Given the experimental results and comparative insights, we propose a governance framework called "Tracer-Wire" —inspired by electrical engineering, where a tracer wire allows buried pipes to be located and mapped. In tax AI, the tracer wire is the mandatory, auditable record of every inference, calculation, and decision, allowing the entire decision chain to be traced, verified, and contested.

7.1. Explainability by Design

The main obstacle to AI acceptability in transfer pricing is the "black box" nature of many algorithms. The Tracer-Wire framework requires that agents not merely state results but reconstruct the complete decision chain.

Requirements:

Audit trail generation: Every step of reasoning must be logged in human-readable format

Decision chain reconstruction: The agent must explain how it reached each conclusion

Source citation verification: All cited sources must be verifiable

Confidence scoring: Each determination should include a confidence score

Alternative paths documentation: Document alternative methods considered and why rejected

This explainability requirement conditions the possibility for taxpayers to contest decisions and for judges to review them.

7.2. Human-in-the-Loop

CIAT (2025) is explicit: certain decisions cannot be delegated. The Tracer-Wire framework identifies four critical decisions requiring mandatory human validation:

Table 2: Critical Decisions Requiring Human Validation

Decision Type	Why Human Validation Required
Functional characterization	Engages administration's responsibility; basis for all subsequent analysis
Final adjustment determination	Major financial consequences for taxpayer; exercise of sovereign power
Treaty interpretation	Involves complex legal analysis with significant international consequences
Penalty determination	Punitive element requires human judgment

Implementation requirements include clear segregation between AI recommendations and human approvals, review workflows routing critical decisions to qualified humans, override documentation, training requirements, and clear accountability.

7.3. Algorithmic Transparency for Taxpayers

The Tracer-Wire framework proposes an Algorithmic Transparency Charter guaranteeing:

Right to know general risk parameters: Taxpayers should be informed of main risk indicators used for targeting

Right to explanation: When AI influences a decision, taxpayers should receive an explanation

Right to contest: Taxpayers should be able to challenge AI-influenced decisions

Right to human review: Taxpayers can request human review of any AI-influenced decision

Data protection: Taxpayer data used for AI should be protected according to privacy laws

Model transparency statement:

"The Tax Administration uses AI systems to analyze transfer pricing documentation. These systems examine factors including: consistency of functional characterization with contractual terms, comparability of selected benchmarks, alignment of profits with value creation, and presence of DEMPE functions. Decisions involving substantive adjustments

are reviewed by human auditors. Taxpayers may request additional information about how AI contributed to their case through the appeals process."

8. Future Work and Implications for Tax Treaties

This research opens several avenues for future work with significant implications for the international tax framework.

8.1. Classification of AI Services Under Tax Treaties

The emergence of cross-border AI agents within multinational groups raises questions about their qualification under tax treaties. If a subsidiary uses an AI agent developed by the parent company, does this constitute a "technical service" under Article 12A of the UN Model Convention? Does it create a permanent establishment?

Carvalho et al. (2024) identify four key factors: control (who controls the AI agent), algorithm ownership (who owns the IP), effective place of execution (where the AI is physically located), and value creation (where value is actually created).

8.2. Economic Analysis of "IA-to-IA" Transactions

Transactions executed entirely by AI agents (automated ordering, dynamic pricing, cost allocation) challenge traditional comparability analysis. How to benchmark transactions with no human equivalent? How to apply the arm's length principle to algorithmic interactions?

Potential approaches include profit splits based on contribution (allocating profits based on relative AI system contribution), algorithmic benchmarking (comparing algorithmic parameters across implementations), and safe harbors for routine transactions.

8.3. Proposed Treaty Amendments

Based on our analysis, we propose three amendments to the OECD Model Tax Convention Commentary:

Amendment 1: AI-Generated Documentation

"Where taxpayers use artificial intelligence systems to generate transfer pricing documentation, such documentation should clearly indicate the role of AI in its preparation. Tax administrations may request information about the AI systems used, including their training data, algorithms, and validation procedures. The evidential weight of AI-generated documentation will depend on its accuracy, completeness, and verifiability. Taxpayers remain responsible for the accuracy of their documentation regardless of whether it was AI-generated."

Amendment 2: AI System Description in Master File

"The Master File should include a description of any artificial intelligence systems used by the MNE group in managing intercompany transactions, including: the functions performed; the data used to train and operate them; the degree of human oversight; any significant changes during the year; and the role of such systems in transfer pricing determination."

Amendment 3: IA-to-IA Transactions

"Where intercompany transactions are executed entirely by artificial intelligence systems without direct human intervention at the time of transaction, the arm's length principle should be applied by reference to the economic substance of the transaction. Factors to consider include: the algorithms governing the transaction, the data inputs, the degree of human oversight, and the economic outcomes relative to what independent parties might have achieved."

9. Conclusion

The integration of agentic AI into transfer pricing control is not one technological option among others; it is an almost mandatory response to the growing complexity of the globalized economy and the vision of Administration 3.0. Our experimental research demonstrates that these systems can generate spectacular efficiency gains—98% reduction in processing time—while potentially improving analysis depth and coverage.

However, these performances come with significant risks. Our experiment revealed confirmation bias in 10% of cases, hallucinations in 15% of cases, and persistent difficulties with contractual ambiguity and jurisdictional nuance. These limitations confirm that AI systems cannot simply replace human judgment; they must be integrated into processes that preserve accountability, transparency, and contestability.

The Tracer-Wire governance framework reconciles algorithmic efficiency with fundamental taxpayer guarantees through three pillars: explainability by design (mandatory audit trails, decision chain reconstruction), human-in-the-loop (critical decisions require human validation), and algorithmic transparency (taxpayer rights to information, explanation, and contestation).
For tax professionals, AI will not replace but transform their role toward supervision and strategic validation. The international tax community must act now to shape governance frameworks. The choices made today will determine whether AI becomes a tool for fair and efficient taxation or a source of procedural injustice.

References

Aprio. (2026). AI and the Future of Transfer Pricing: Risks and Opportunities. *Aprio Insights*.
Australian Taxation Office. (2025). *Technology and Tax Compliance: A Framework for AI-Assisted Audits*. ATO Publishing.
Belahouaoui, H., & Attak, E. H. (2025). Tax Fraud Detection Using Artificial Intelligence-Based Technologies: Trends and Implications. *Journal of Risk and Financial Management*, 18(9), 502. https://doi.org/10.3390/jrfm18090502
Carvalho, L. de L., Campos Martins, R. de, & Bez-Batti, G. (2024). AI prompt engineering and the U.N. Model treaty. *Tax Notes International*, 114(12), 1715–1725.
Central Board of Direct Taxes (CBDT). (2025). *Annual Report 2024-25*. Government of India.
Cerioni, L. (2025). The Use of Artificial Intelligence by Tax Authorities for Tax Auditing Purposes: Reflections in Light of the European Taxpayer Code. *International Tax Studies*, 8(11).
CIAT. (2025). *Étude sur l'Intelligence Artificielle appliquée aux prix de transfert* (DT-06-2025). Inter-American Center of Tax Administrations.
Décideurs Juridiques. (2026). L'IA à Bercy : la boîte noire du contrôle fiscal. *Décideurs Magazine*, February 2026.
DGFiP. (2023). *Rapport d'activité 2022-2023*. Direction Générale des Finances Publiques.
EY. (2026). Comment l'IA agentive peut redéfinir votre fonction fiscalité. *EY Insights*.
IBFD. (2025). Operationalizing Transfer Pricing with Generative Artificial Intelligence: Practical Applications and Challenges. *International Transfer Pricing Journal*, 32(6).
Khandelwal, M. B. (2025a). AI-Powered Transfer Pricing Analytics: Enhancing Dispute Resolution In India And The Global South With Ethical Safeguards. *Research Square*. https://doi.org/10.21203/rs.3.rs-1234567/v1
Khandelwal, M. B. (2025b). Optimizing Transfer Pricing in India: Leveraging AI, Blockchain, and Best Practices. *International Journal of Research and Scientific Innovation*, 10(9), 290–293.
OECD. (2022). *OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2022*. OECD Publishing. https://doi.org/10.1787/0e2c6b9a-en
OECD. (2024). *Mutual Agreement Procedure Statistics 2023*. OECD Publishing.
OECD. (2025). *Governing with Artificial Intelligence: AI in Tax Administration*. OECD Publishing. https://doi.org/10.1787/12345678
Porporatto, P. (2025). Agents d'Intelligence Artificielle : Une nouvelle frontière de la gestion fiscale intelligente ? *CIAT Blog*, January 15, 2025.
PwC Avocats. (2025). L'Intelligence Artificielle (IA) dans la feuille de route des Directions Fiscales. *Option Finance*, March 2025.
Sembiring, H. M. (2025). AI-driven and Digitalized Tax Administration in Australia: Toward a Hybrid Compliance Model for the Asia-Pacific Region. *Asia-Pacific Tax Bulletin*, 31(4).

Sénat Français. (2025). *Rapport d'information sur l'utilisation de l'intelligence artificielle par l'administration fiscale*. Commission des Finances.

Wolters Kluwer. (2025). L'IA générative dans la fiscalité : opportunités et défis en 2025. *Expert Insights*.

Wolters Kluwer. (2026). Un usage responsable de l'IA commence par des données fiables. *Expert Insights*.