

**¿Asimetrías en la definición de los Sistemas de Inteligencia Artificial en la AEAT?,
¿asimetrías también en la protección de los obligados tributarios?[1]**

Recientemente, participé en el foro de debate técnico AEAT/CDC sobre la utilización del *big data* y la IA por parte de la Administración tributaria, invitado por la AEDAF, junto a Daniel Gómez-Olano González.[2]

Para los neófitos, como era mi caso, es un foro donde la AEAT y el CDC, a través del IEF, invitan a las principales asociaciones profesionales de nuestro país para “debatir” determinados temas de interés e incentivar, en un ejercicio de exposición pública, la rendición de cuentas.

En este contexto, dos altos cargos de la AEAT nos explicaron *grosso modo* qué hacen con la IA y las garantías que aplican.

Me hizo gracia que ambos se habían descargado e impreso el libro que publiqué recientemente en la editorial Atelier (*Sistemas de inteligencia artificial en la Agencia Estatal de Administración Tributaria: Despliegue tecnológico y control jurídico*).[3] Esto pone de relieve que la AEAT está atenta a los avances en la investigación que arroja la academia; un colectivo, normalmente, tan denostado en esto de la transferencia del conocimiento.

Se plantearon diversas interpelaciones muy interesantes, principalmente por parte de una vocal del CDC y de la representación de la AEDAF, que dieron lugar a un corto (por la restricción de la moderación) pero interesante debate que me ha servido para reflexionar y traer a colación esta entrada.

En una de ellas, el Abogado del Estado y gran jurista, a mi parecer —sin ironía, de verdad—, nos explicó la importancia que tenía catalogar un *software* como IA o como mero sistema de tratamiento masivo de datos.

Cuestión sobre cuya importancia coincido plenamente con el Subdirector de Organización y Asistencia Jurídica de la AEAT. Ahora bien (aquí ya discreparíamos), creo que la definición que recoge la EIAAT, y que se defendió en el foro, es muy distinta de la reconocida por el RIA.

¿Servirá esta entrada (si cae en manos de algún responsable de la reelaboración de la EIAAT 2026) para eliminar la definición restrictiva contenida en las estrategias de la AEAT de 2024 y 2025?

Ojalá.

¿Qué consideramos exactamente Inteligencia Artificial (IA)?

Definir en la práctica lo que es un sistema de inteligencia artificial no es un mero ejercicio semántico, ya que de esta definición depende qué sistemas estarán sujetos a controles en función de su nivel de riesgo y cuáles operarán en la sombra.

Al analizar la estrategia de la AEAT frente al nuevo Reglamento de Inteligencia Artificial (RIA) de la Unión Europea, emerge una preocupante discrepancia conceptual.

Esta discrepancia no es una cuestión menor. Las razones son puramente garantistas: en teoría, de la catalogación jurídica de un *software* como SIA depende la aplicación inmediata de garantías reforzadas y la aplicación de la Estrategia de IA de la AEAT.

Si un sistema se etiqueta legalmente como IA, entra en el radio de acción de las obligaciones de transparencia, supervisión humana, explicabilidad y evaluación de impacto y sesgos con un carácter reforzado, distinto del tratamiento masivo de datos. Si, por el contrario, se camufla bajo el paraguas del *software* tradicional, elude este escrutinio reforzado.

Para entender el conflicto, debemos acudir al artículo 3.1 del RIA, que establece una definición amplia y tecnológicamente neutra.

Según el RIA, un sistema de IA es un sistema basado en una máquina, diseñado para funcionar con distintos niveles de autonomía y que, para objetivos explícitos o implícitos, infiere de la información de entrada la manera de generar resultados (como predicciones, recomendaciones o decisiones) que pueden influir en entornos físicos o virtuales.

Frente a esta concepción integradora —orientada a los efectos del sistema y a su capacidad para inferir e influir—, la postura de la Administración tributaria española resulta, cuando menos, llamativa por su afán delimitador. ¿Por qué la AEAT se resiste a aplicar este marco conceptual a sus herramientas analíticas más potentes? La respuesta apunta a un uso estratégico de la discrecionalidad interpretativa para eludir la estricta gobernanza comunitaria.

1. Desde las exclusiones estratégicas al escudo del “determinismo”, ¿una definición “a medida” de la AEAT?

Para comprender la brecha existente, es esencial analizar cómo la AEAT ha articulado su propia visión en la Estrategia de Inteligencia Artificial (EIAAT). A diferencia del enfoque funcional y amplio del legislador europeo, la AEAT ha optado por una interpretación restrictiva que pone el foco en las técnicas específicas de programación utilizadas, limitando la IA a sistemas que apliquen estrategias de aprendizaje automático o estrategias basadas en la lógica y el conocimiento.

Esta restricción técnica sirve como base para establecer un listado de exclusiones explícitas. Según la propia política de la Agencia, quedan deliberadamente fuera de su concepto de IA tecnologías tan potentes como:

- El tratamiento masivo de datos.
- El análisis de redes o grafos.
- Los sistemas de análisis de riesgos.
- La robotización o automatización de procesos.

Para justificar por qué excluye estas potentes herramientas de su política de IA, la AEAT utiliza un doble argumento.

En primer lugar, sostiene que estos sistemas no son IA “siempre y cuando su funcionamiento sea determinista y se base en reglas establecidas por humanos”.

La narrativa oficial defiende que, al obedecer instrucciones y parámetros prefijados por funcionarios, estas herramientas no estarían haciendo uso de las capacidades analíticas, predictivas o generativas que la Agencia considera exclusivas y “propias” de la IA.

En segundo lugar, la AEAT minimiza el grado de autonomía de estas herramientas de control. Argumenta públicamente que en la fase de selección de contribuyentes no se usa IA, sino “herramientas de análisis masivo” que se limitan al tratamiento automatizado de datos para “ayudar a los funcionarios”, quienes siempre deben realizar una valoración humana antes de tomar la decisión de iniciar una inspección.

Al afirmar que el sistema solo asiste, la Administración intenta diluir el concepto de “influencia sustancial” en la toma de decisiones que el RIA utiliza para modular la catalogación del riesgo de un SIA.

Esta conclusión no es una simple discrepancia técnica, sino que parece una definición construida con la intención (o al menos con el efecto) de establecer una frontera artificial.

Esta frontera permite a la AEAT excluir del perímetro formal de la IA a un número considerable de sus sistemas analíticos más avanzados e intrusivos, especialmente aquellos dedicados a la vigilancia, el control y la instrucción de los procedimientos.

Se niega la etiqueta de IA a sistemas que en la práctica cruzan millones de datos, realizan evaluaciones complejas o analizan redes de vínculos, como es el caso de su gran sistema de riesgos HERMES, el proyecto de minería de datos MIDAS o la herramienta de análisis de redes NIDEL.

Pensemos, por ejemplo, en el sofisticado sistema de detección de “falsos no residentes”. Esta herramienta cruza más de 70 fuentes de datos distintas e inconexas (consumo, redes familiares, patrimonio, etc.) para inferir la probabilidad de que un gran patrimonio esté simulando su residencia fiscal fuera de España. Bajo el prisma europeo, es evidente que el sistema infiere y predice. Para la AEAT, mientras el algoritmo siga reglas humanas predefinidas, es un mero “tratamiento masivo de datos” que escapa al escrutinio de la IA.

El peligro de esta definición “a medida” es monumental: al exigir técnicas de autoaprendizaje explícitas y excluir el análisis avanzado de riesgos bajo la excusa del determinismo, la AEAT corre el grave riesgo de dejar fuera de las garantías y de la gobernanza específica de la IA a sistemas que tienen una influencia sustancial y directa en los derechos fundamentales de los ciudadanos.

Como vengo defendiendo en anteriores trabajos, se publicita así la IA “amable” (como los asistentes virtuales), mientras el verdadero núcleo del control algorítmico opera en una zona gris bajo la inofensiva etiqueta de *big data*.

2. ¿Inferencia vs. determinismo?

La colisión entre la postura de la AEAT y el RIA de la Unión Europea no es una simple discrepancia técnica, sino una profunda brecha conceptual sobre lo que significa que una máquina “razone”. A mi juicio, esta divergencia se resume en tres diferencias estructurales fundamentales:

2.1. El falso escudo del “determinismo” (inferencia compleja vs. reglas estáticas)

- **La postura de la AEAT.** La AEAT excluye expresamente de su definición de IA a los sistemas de tratamiento masivo de datos, análisis de redes o sistemas de análisis de riesgos argumentando que “funcionan de manera determinista, basándose en reglas fijadas por humanos”. Bajo esta lógica, si un sistema obedece a parámetros predefinidos (por ejemplo, el sistema HERMES o TESEO), se considera un mero tratamiento automatizado que no hace uso de capacidades analíticas propias de la IA.
- **La realidad del RIA.** La norma no excluye a los sistemas deterministas. A nivel computacional, casi todo el *software* es determinista, incluidos muchos modelos entrenados de IA. El RIA solo deja fuera de su concepto a los sistemas de procesamiento básico o a la heurística clásica (reglas estadísticas muy simples y estáticas que no aprenden de los datos). Sin embargo, la ley europea reconoce explícitamente como IA a los enfoques basados en la lógica y el conocimiento, como los sistemas expertos.

Cuando una herramienta cruza millones de datos aplicando miles de reglas complejas para deducir una puntuación de riesgo fiscal, no está haciendo un filtro básico; está realizando una inferencia compleja basada en lógica, lo cual encaja de lleno en la definición de IA del RIA.

Un caso paradigmático de esta disonancia es TESEO. La AEAT lo clasifica oficialmente como una mera “herramienta de visualización” determinista. Sin embargo, cuando TESEO aplica complejos algoritmos matemáticos (como el cálculo de modularidad o la distancia de complicidad) para descubrir “marañas societarias” ocultas y comunidades de fraude que un humano jamás podría ver a simple vista en un mar de datos, no está simplemente “dibujando”; está infiriendo estructuras subyacentes. Negarle la condición de IA es ignorar su verdadera naturaleza analítica.

Es importante destacar que, a nivel computacional, casi todo el *software* (incluyendo un modelo de *Machine Learning* entrenado y congelado) es “determinista” (mismas entradas producen mismas salidas). La AEAT parece confundir “determinista” con “ejecución simple de reglas” o “*software* tradicional”. El RIA no excluye los sistemas deterministas; excluye los sistemas simples que no realizan inferencias complejas. El debate real no es determinismo vs. no determinismo, sino cómo se genera la lógica y cuál es su complejidad.

2.2. El foco en el “cómo” frente al “qué”, la restricción técnica frente a la amplitud funcional

- **La postura de la AEAT.** La estrategia de la AEAT adopta una visión restrictiva enfocada en técnicas concretas. Al exigir que el sistema aplique explícitamente estrategias de aprendizaje automático (*Machine Learning*) para no ser considerado un mero programa convencional, traza una frontera artificial que excluye herramientas analíticas muy avanzadas e intrusivas orientadas al control tributario.
- **La realidad del RIA.** El legislador optó por una definición funcional y tecnológicamente neutra. Al RIA no le importa tanto la etiqueta de la técnica de programación utilizada, sino la capacidad del sistema para inferir cómo generar resultados (predicciones, decisiones o recomendaciones) a partir de los datos que recibe.

Además, el RIA desmonta el gran mito técnico: la capacidad de autoaprendizaje o adaptación continua tras su despliegue es completamente opcional. Un sistema que se entrena, se “congela” y luego opera de forma estática para generar inferencias **sigue siendo IA**.

Pensemos en el sistema aduanero DOBLING, entrenado para reconocer patrones en los Documentos Únicos Administrativos (DUA) y predecir de forma autónoma si son fraudulentos, llegando a paralizar envíos. O en los modelos predictivos específicos de la Agencia para detectar potenciales no declarantes en el IRPF. Independientemente de si su código de *Machine Learning* se autoajusta a diario o se mantiene estático tras su entrenamiento, encuadran en la definición funcional del RIA. Ello evidencia que la restricción técnica de la AEAT no tiene amparo en el estándar del art. 3.1 del RIA.

2.3. La negación de la autonomía decisional, la minimización de la influencia sustancial

- **La postura de la AEAT.** Para evitar clasificar sus potentes sistemas de control como IA, la AEAT minimiza su nivel de independencia. Argumenta públicamente que, en la fase de selección de contribuyentes para inspección, sus herramientas de análisis masivo de riesgos simplemente “ayudan a los funcionarios”, limitándose a un tratamiento automatizado de datos que siempre debe ser valorado y decidido por un humano.
- **La realidad del RIA.** El Reglamento no exige que una máquina tome decisiones de forma 100% independiente y sin humanos para ser considerada IA. Solo requiere que el sistema esté diseñado para funcionar con “distintos niveles de autonomía” o con “cierto grado de independencia de acciones”. Un sistema que procesa autónomamente millones de registros y bases de datos para generar una

lista priorizada de contribuyentes sospechosos (como hace el sistema HERMES) está operando con un alto grado de independencia inferencial.

Aunque un inspector humano firme la decisión final de investigar, el algoritmo ejerce una influencia sustancial y determinante en esa decisión, lo que desbarata la excusa de que son meras “herramientas de ayuda” sin autonomía.

Aquí entra en juego un factor jurídico clave que la AEAT obvia: el sesgo de confirmación y automatización. Como examinamos en la anterior entrada, consideremos los sistemas ROI (Registro de Operadores Intracomunitarios) o IVAN (Devoluciones a no residentes).

La propia Agencia presume de que estos sistemas generan de forma autónoma propuestas predictivas de “aprobar”, “denegar” o “a estudiar” con una precisión anunciada superior al 95 % o incluso al 99 %. Ante una máquina dotada de esa aparente infalibilidad matemática que propone “denegar” una solicitud, ¿qué funcionario se atreverá a contradecirla sistemáticamente? La propuesta algorítmica, en la práctica, predetermina el cauce procedimental y tiende a convertirse en la decisión administrativa final.

Incluso los sistemas con “cierto grado de independencia”, que requieren alguna entrada humana o siguen instrucciones automatizadas, se consideran autónomos porque pueden generar resultados sin un control humano directo.

Dicho de otra forma, un sistema sí cumple el criterio de autonomía incluso si requiere datos de entrada proporcionados manualmente por un humano, siempre que el sistema sea capaz de generar el resultado por sí mismo, sin que un humano controle o especifique explícitamente cómo se produce esa salida.[4]

Por ejemplo, los asistentes virtuales de la AEAT, que utilizan PLN, reciben la entrada en lenguaje natural del usuario e infieren de forma autónoma su significado para generar, si fuera el caso, una respuesta coherente. Aunque la base de conocimiento pueda ser supervisada por humanos, el proceso de comprensión y generación de la respuesta específica a una consulta dada tiene ese grado de independencia.

3. ¿Una estrategia narrativa deliberada?

Desde mi punto de vista, esta “huida hacia delante” en la actuación de la AEAT se interpreta como una estrategia consciente de gestión de la narrativa y de la percepción pública en torno a la IA.

Consiste en abrazar y publicitar de forma muy visible la IA en aquellos ámbitos donde resulta “amable” y comunicacionalmente ventajosa (como en los asistentes virtuales para la información y asistencia al contribuyente), mientras se difumina o excluye por definición su papel en las áreas nucleares del poder de comprobación e investigación.

Esta asimetría constituye una “transparencia selectiva” que permite a la Agencia ocultar sus herramientas analíticas de control fiscal más potentes e intrusivas bajo etiquetas inofensivas como *big data*, “análisis de riesgos” o sistemas “deterministas”. De este

modo, la AEAT intenta controlar el perímetro de aplicación de su propia estrategia corporativa y de las futuras regulaciones europeas.

La consecuencia directa e institucional de forzar una definición restrictiva de la IA es la creación de una peligrosa “zona gris”.

Al negar oficialmente que sistemas con un impacto decisivo en la selección y control de contribuyentes (como los perfiladores HERMES o MIDAS, o el analizador de redes TESEO) sean Inteligencia Artificial, la AEAT logra sustraer estas potentes herramientas de su propio marco de gobernanza interna (la EIAAT) y de las exigencias del RIA.

Al excluir estos sistemas del perímetro formal de la IA, operan sin el escrutinio adicional y específico que justificaría su inmenso poder de influencia.

Esta evasión de los controles algorítmicos específicos podría generar un impacto sistémico que erosiona las garantías del Estado de Derecho. Cuando la AEAT se escuda en la semántica del “determinismo” o en la falta de “autoaprendizaje continuo” para ocultar la verdadera naturaleza inferencial de sus algoritmos de control, podría privar al contribuyente de su principal “escudo jurídico”.

Redefinir la IA a la baja en un documento estratégico interno no elimina el inmenso poder intrusivo de la máquina; simplemente nos alerta sobre el posible vaciamiento de los derechos que asisten a los obligados tributarios, de la exigencia de motivación real de los actos de gravamen y del acceso a una tutela judicial efectiva frente a la “caja negra institucional”.

La sofisticación en la lucha contra el fraude es un fin imperativo y legítimo, pero en un ordenamiento democrático no puede erigirse sobre una asimetría procesal inaceptable: una Administración tecnológicamente omnisciente frente a un contribuyente algorítmicamente ciego.

[1] Esta investigación es resultado de la Cátedra Jean Monnet EUTAXRIGHTS “The protection of taxpayers’ rights in the EU Law” (Project: 101175282: ERASMUS-JMO-2024-HEI-TCH-RSCH) financiada por la Unión Europea y del Proyecto de I+D+I, referencia: PID2022-139650OB-100: “Administración electrónica, inteligencia artificial y tributos”, financiado por MCIN/AEI/10.13039/501100011033/ y “FEDER, UE”.

[2] Celebrado en marzo de 2026 en el IEF.

[3] Disponible gratuitamente para su descarga en: <https://atelieropenaccess.com/products/sistemas-de-inteligencia-artificial-en-la-agencia-estatal-de-administracion-tributaria-despliegue-tecnologico-y-control-juridico>

[4] Apartado 18 de “Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)”.