

# GENERACIÓN DE DATOS SINTÉTICOS ORIENTADA A LA DETECCIÓN DEL LA/FT MEDIANTE SIMULACIÓN BASADA EN AGENTES

Rincón Gómez, Yinneth Zulay<sup>1</sup> - Melgarejo Rey, Miguel Alberto<sup>2</sup>

## RESUMEN

*La investigación en Colombia para la Prevención del Lavado de Activos y Financiación del Terrorismo en adelante por sus siglas (LA/FT) enfrenta dificultades importantes como: el acceso a datos reales es limitado debido a las normas de protección de datos personales (Habeas data) por temas de confidencialidad y regulación. Esta restricción complica el desarrollo y la validación de herramientas analíticas en contextos reales. Ante este problema, la generación de datos sintéticos se presenta como una alternativa útil para entrenar y evaluar sistemas de detección. En este trabajo se propone un modelo de generación de datos sintéticos que utiliza una simulación basada en agentes para reproducir comportamientos transaccionales y patrones que pueden surgir en contextos asociados al riesgo de LA/FT. La asociación de reglas de inferencia difusa permite representar la incertidumbre presente en los perfiles sociodemográficos, en los comportamientos financieros y en las tipologías de riesgo. La propuesta busca convertir conocimiento experto en un modelo computacional capaz de generar datos sintéticos con características similares a las que podrían observarse en redes transaccionales reales. Estos datos pueden utilizarse en entornos controlados para probar y ajustar algoritmos de analítica aplicados a la detección del LA/FT.*

**Palabras clave:** lavado de activos y financiación del terrorismo, datos sintéticos sin semillas, lógica difusa, simulación basada en agentes

## SYNTHETIC DATA GENERATION ORIENTED TO LA/FT DETECTION THROUGH AGENT-BASED SIMULATION

### ABSTRACT

*Research in Colombia for the prevention of Money Laundering and Terrorist Financing, hereinafter LA/FT, faces important difficulties such as limited access to real data due to personal data protection rules, confidentiality and regulation. This restriction complicates the development and validation of analytical tools in real contexts. In response, synthetic data generation is presented as a useful alternative to train and evaluate detection systems. This paper proposes a synthetic data generation model that uses agent-based simulation to reproduce transactional behaviors and patterns that may arise in contexts associated with LA/FT risk. The association of fuzzy inference rules makes it possible to represent uncertainty in sociodemographic profiles, financial behaviors and risk typologies. The proposal seeks to convert expert knowledge into a computational model capable of generating synthetic data with characteristics similar to those that could be observed in real transactional networks.*

**Keywords:** anti money laundering, seedless synthetic data, fuzzy logic, fuzzy agent-based simulation

---

<sup>1</sup> Universidad Distrital Francisco José de Caldas (Colombia). [yzrincong@udistrital.edu.co](mailto:yzrincong@udistrital.edu.co)

<sup>2</sup> Universidad Distrital Francisco José de Caldas (Colombia). [mmelgarejo@udistrital.edu.co](mailto:mmelgarejo@udistrital.edu.co)

## Introducción

La validación de los datos para la detección del LA/FT enfrenta una criticidad por la escasa disponibilidad de datos transaccionales abiertos o accesibles, así mismo la economía se ve constantemente amenazada por actividades delictivas complejas. Para contrarrestar estas amenazas, las instituciones financieras han adoptado sistemas de monitoreo basados en analítica de datos; sin embargo, el desarrollo y validación de estos algoritmos enfrentan un obstáculo crítico: la inaccesibilidad a datos transaccionales reales debido a las regulaciones de protección de datos personales y secreto bancario (*habeas data*). Como señala Ding (2025), esta limitación crea un entorno de investigación restringido donde la reproducibilidad científica es casi inexistente.

Ante esto, la generación de datos sintéticos ha surgido como una solución proyectada. No obstante, la mayoría de los enfoques actuales dependen de datos reales previos para entrenar modelos generativos, como redes neuronales generativas adversarias (GAN) (Assefa et al., 2020). Esta investigación propone una ruptura paradigmática mediante la generación de datos sintéticos. El enfoque se fundamenta en la intercalación del conocimiento experto y las tipologías de LA/FT hacia un entorno de simulación computacional.

La propuesta técnica integra la simulación basada en agente implementados en la plataforma NetLogo (Wilensky, 1999). La simulación permite que la interacción de agentes opere bajo reglas de comportamiento heterogéneas (Wilensky, 1999). Por su parte, los sistemas difusos aportan un mecanismo viable para modelar los perfiles de riesgo, donde la frontera entre una transacción legítima y una sospechosa no es binaria, sino gradual (Zadeh, 1965). Este diseño no solo busca replicar datos, sino generar un laboratorio donde se puedan probar tipologías de lavado de activos antes de que ocurran en el mundo real, cerrando así la brecha entre la teoría basada en expertos y la validación de algoritmos de detección (Lopez Rojas et al., 2016).

El problema radica en la irregularidad de información y las barreras que impiden el avance de la analítica avanzada en la detección del LA/FT. Mientras que las organizaciones criminales y delictivas evolucionan en sus tácticas de ocultamiento, la comunidad científica se encuentra estancada por la imposibilidad de acceder a bases de datos transaccionales reales debido a las regulaciones de protección de datos (*habeas data*) y el secreto bancario (Ding, 2025; López Rojas et al., 2016).

## Antecedentes

La investigación en la detección del LA/FT ha transitado históricamente por tres grandes etapas de recolección y uso de datos. En una primera etapa, el análisis se centraba en datos reales anonimizados proporcionados por instituciones bajo acuerdos de confidencialidad estrictos. Autores como Ngai et al. (2011) demostraron que, si bien estos datos son de alta fidelidad, la imposibilidad de compartirlos frenaba el avance de la comunidad académica.

El reporte de operaciones sospechosas se apoyó en enfoques heurísticos y sistemas de detección de casos en entornos financieros, orientados a identificar patrones inusuales de manera rápida dentro del proceso de monitoreo (Le-Khac y Markos, 2009). En los primeros

trabajos sobre detección del LA/FT se emplearon enfoques clásicos basados en heurísticas, máquinas de soporte vectorial y árboles de decisión para discriminar operaciones sospechosas y evaluar niveles de riesgo (Le-Khac y Markos, 2009; Tang y Yin, 2005; Wang & Yang, 2007).

En una segunda etapa, surgió el concepto de datos sintéticos, donde destaca el simulador *PaySim*, desarrollado por Lopez Rojas et al. (2016). Utiliza modelos de simulación basados en agentes calibrados a partir de un mes de registros de transacciones reales de una empresa de dinero móvil. Aunque este avance permitió la libre distribución de conjunto de datos para pruebas de algoritmos de detección, mantiene una dependencia crítica: requiere de un conjunto de datos real base para su configuración inicial, lo que en muchos contextos sigue siendo una barrera.

La eficacia de los sistemas de detección del LA/FT depende, en gran medida, de su capacidad para distinguir entre transacciones legítimas y operaciones potencialmente ilícitas. No obstante, en el ámbito de la investigación este problema presenta una dificultad metodológica particular: los datos transaccionales reales suelen ser altamente confidenciales, lo que limita la disponibilidad de conjuntos públicos y restringe la comparación rigurosa entre modelos en aspectos como eficiencia, efectividad, desbalance de clases, deriva conceptual e interpretabilidad (Jensen, 2023; Altman et al., 2023).

### **Objetivos**

- Diseñar un modelo de generación de datos sintéticos para la detección del Lavado de Activos y la Financiación del Terrorismo (LA/FT), mediante la simulación basada en agentes.
- Identificar las tipologías de LA/FT y los comportamientos transaccionales asociados, con el fin de definir las variables, umbrales y reglas de monitoreo que regirán el comportamiento esperado de los agentes.
- Diseñar la arquitectura de simulación en la plataforma Netlogo definiendo las reglas y protocolos de los agentes al igual la dinámica de un entorno financiero.
- Modelar el comportamiento transaccional sospechoso mediante sistemas de inferencia difusa.
- Generar un conjunto de datos sintéticos a partir del modelo diseñado, contrastando descriptivamente los perfiles transaccionales resultantes.

### **Marco referencial**

En la práctica, la innovación en la prevención y detección del LA/FT ha avanzado en dos direcciones principales. La primera corresponde a la incorporación de nuevas tecnologías como la analítica avanzada, la automatización y la inteligencia artificial en los sistemas de monitoreo transaccional. La segunda se orienta a la construcción de entornos reproducibles que permitan experimentar y comparar métodos cuando los datos reales no pueden hacerse públicos. Desde el plano regulatorio y operativo, existe consenso en que estas tecnologías pueden mejorar la calidad y la eficiencia de las medidas *Anti-Money Laundering (AML/CFT)*, siempre que su adopción se enmarque en un enfoque responsable y basado en riesgos (Grupo de Acción Financiera Internacional [GAFI], 2021).



En *Anti-Money Laundering* (AML), la simulación a gran escala ha cobrado fuerza especialmente en problemas donde la estructura de red resulta determinante. Un trabajo influyente en esta línea es “*Scalable Graph Learning for Anti-Money Laundering: A First Look*”, que introduce AMLSim como generador y simulador para construir grafos de transacciones sintéticas a gran escala y evaluar técnicas de aprendizaje en grafos aplicadas a AML (Weber et al., 2018).

### **Marco conceptual**

- **Agente (Agent):** Entidad computacional autónoma, con ubicación y capacidades sensoriales, que interactúa con otras entidades dentro de un entorno para alcanzar objetivos específicos. En NetLogo, los agentes (*turtles*) ejecutan comportamientos distribuidos (Wilensky & Rand, 2015).
- **Dato Sintético:** Información generada mediante modelos que no requieren una distribución de probabilidad de datos reales como entrada inicial, basándose en reglas lógicas y conocimiento experto para su construcción (Assefa et al., 2020).
- **Financiación del terrorismo:** La financiación del terrorismo consiste en proveer o recaudar fondos, bienes o activos, por medios directos o indirectos, con el propósito de facilitar la comisión de actos terroristas o de apoyar a terroristas y organizaciones terrorista (GAFI, 2025; Naciones Unidas, 1999).
- **Inferencia Difusa (Fuzzy Inference):** Proceso de mapeo de un conjunto de entradas dadas a una salida utilizando la teoría de conjuntos difusos, que permite capturar el razonamiento humano aproximado en sistemas de decisión (Ross, 2016).
- **Modelo Basado en Agentes:** Paradigma de modelado computacional que analiza el sistema desde la perspectiva de sus constituyentes individuales, permitiendo observar la emergencia de fenómenos globales a partir de reglas locales (Macal y North, 2010).
- **Tipología de Riesgo:** Clasificación de los métodos y tendencias utilizados por las organizaciones criminales para dar apariencia de legalidad a fondos ilícitos, sirviendo como base para el diseño de escenarios de detección (GAFI, 2023).

### **Metodología**

La metodología ODD (*Overview, Design concepts, Details*) es un protocolo estandarizado para describir modelos basados en agentes y otros modelos de simulación de manera clara, completa y replicable. Su estructura organiza la descripción del modelo en tres niveles: una visión general del propósito, las entidades, variables y escalas, así como del proceso y la secuencia temporal; los conceptos de diseño que explican la lógica del comportamiento del sistema; y los detalles de implementación, como la inicialización, los datos de entrada y los submodelos. En conjunto, ODD busca mejorar la transparencia del modelo, facilitar su comprensión y favorecer su reproducción o comparación en otros estudios (Grimm et al., 2020; Szangolies, 2024).

La investigación adopta el enfoque ODD con el fin de diseñar y desarrollar un modelo computacional como un generador de datos sintéticos para LA/FT donde se busca

identificar el problema, plantear el diseño, construir un modelo en NetLogo y evaluar su utilidad mediante los resultados obtenidos. Este método permite iterar entre la base de conocimientos (teoría y expertos) y el desarrollo del artefacto (modelo en NetLogo) como se observa en la Tabla 1.

**Tabla 1**

*Metodología para la generación de datos sintéticos mediante el protocolo ODD.*

Fase Metodológica	Componentes Esenciales
Fase 1: Definición del Modelo	Establecimiento del objetivo general de la simulación, revisión de la literatura económica sobre lavado de dinero y análisis del estado del arte en modelos basados en agentes.
Fase 2: Análisis del Sistema	Mapeo exhaustivo del entorno LA/FT (proceso del crimen financiero). Identificación cualitativa de los actores involucrados, sus roles y sus interacciones iniciales.
Fase 3: Diseño del Modelo	Traducción de conductas a variables cuantitativas. Formalización de las reglas de decisión difusas y estructuración del pseudocódigo de simulación para NetLogo.
Fase 4: Simulación y Calibración	Codificación del entorno en la plataforma NetLogo, calibración de parámetros iniciales mediante datos históricos y ejecución de escenarios de estrés transaccional ( <i>ticks</i> ).
Fase 5: Análisis de Resultados	Extracción masiva de datos (comportamientos emergentes), generación de indicadores clave de riesgo (KRI) y evaluación estadística de los falsos positivos/negativos.
Fase 6: Evaluación Final	Interpretación multidisciplinaria de los hallazgos a la luz de la normativa vigente (UIAF/GAFILAT), documentación de lecciones aprendidas y formulación de conclusiones.

*Nota.* Elaboración propia

La formalización computacional justifica el uso de una Simulación Basada en Agentes para generar datos sintéticos que posean “realismo semántico” a partir de interacciones locales (Assefa, Dochev, & Eyal, 2020). El estudio adopta un enfoque de diseño y construcción de modelo orientado al desarrollo de un generador de datos sintéticos sin datos reales para LA/FT.

### **Propósito del modelo**

El modelo es diseñado para generar datos transaccionales sintéticos desde cero que representen comportamientos financieros asociados tanto a actividades legítimas como a tipologías de LA/FT, sin utilizar datos históricos reales como base de calibración. En lugar de depender de distribuciones aprendidas a partir de datos observados, el sistema parte del conocimiento experto formalizado en reglas de decisión difusas y en protocolos de interacción entre agentes.

## Desarrollo del modelo

El modelo se construyó en tres capas articuladas. Primero, las tipologías de LA/FT documentadas por la UIAF y GAFILAT (GAFILAT, 2016), se tradujeron en reglas de comportamiento para los agentes, definiendo quiénes participan en el esquema y qué acciones ejecutan. Sobre esa base, se implementó la simulación en NetLogo (Wilensky, NetLogo, 1999), donde los agentes interactúan en ciclos discretos que reproducen la dinámica transaccional del esquema. Finalmente, cada transacción generada es evaluada por un sistema de inferencia difusa tipo Mamdani, que asigna un puntaje de riesgo continuo (*risk\_score*) a partir de señales cualitativas como la frecuencia, la cercanía a zonas fronterizas y la consistencia del perfil del agente, tal y como se refleja la secuencia en la Tabla 2 y Tabla 3.

**Tabla 2**

*Esquema de desarrollo e implementación del modelo generador sintético basado en agentes y sistema difuso Mamdani*

Paso secuencial	Denominación del hito	Descripción de actividades e instrumentos técnicos
Paso 1	Conocimiento Experto	Recopilación, análisis e integración de señales de alerta e informes de tipologías emitidos por organismos reguladores (UIAF, GAFILAT, GAFI) para asentar las bases del comportamiento delictivo.
Paso 2	Formalización Lógica	Traducción de las tipologías cualitativas a reglas de negocio condicionales (Booleanas y de umbral) y asignación de roles matemáticos específicos para cada tipología de agente simulado.
Paso 3	Programación en NetLogo	Desarrollo del entorno computacional distribuido. Configuración del espacio (patches), temporizadores ( <i>ticks</i> ) y codificación de las rutinas de movimiento de capitales financieros entre nodos.
Paso 4	Modelado de Lógica Difusa	Implementación del algoritmo de inferencia difusa tipo Mamdani para calcular el nivel exacto de sospecha ( <i>risk_score</i> ) transaccional, superando la rigidez de los sistemas tradicionales basados en reglas fijas.
Paso 5	Generación de Datos Sintéticos	Ejecución de simulaciones a gran escala (Montecarlo) y exportación automatizada de archivos CSV contentivos de millones de transacciones enriquecidas con variables de riesgo para su análisis posterior.

*Nota.* Elaboración propia

**Tabla 3**

*Esquema de tipología simulada*

Agente / Nodo de Red	Etapas LA/FT Asociada	Acción y Comportamiento Programado	Parámetros de Control
Cientes Legítimos	Operación Económica Normal	Ejecución de depósitos, retiros y transferencias comerciales estándar dentro de los circuitos económicos lícitos.	Frecuencia salarial, consumo promedio, dispersión geográfica baja.
Mulas financieras ( <i>Smurfs</i> )	<i>Placement</i> (Colocación)	Fraccionamiento de grandes sumas de dinero en efectivo ilícito para realizar depósitos repetitivos por montos inferiores al umbral de reporte legal.	Probabilidad de depósito, capacidad máxima por <i>tick</i> , límite de alerta bancaria.
Bancos e Infraestructura	Ecosistema de Soporte (Intermediación)	Procesamiento de las operaciones transaccionales, acumulación de fondos en cuentas concentradoras y actualización de contadores de red.	Nivel de efectividad de SARLAFT, umbrales automáticos de bloqueo.
Agente Estructurador	<i>Layering and Integration</i> (Estratificación)	Agrupación de fondos limpios provenientes de múltiples mulas y ordenamiento de transferencias electrónicas internacionales de alta velocidad hacia paraísos fiscales o empresas fachada.	Destino final de fondos (Jurisdicciones no cooperantes), ratio de diversificación.

Nota. Elaboración propia

**Simulación basada en Agentes**

En la simulación se simplifica el esquema en cuatro etapas: (i) operaciones legítimas (transferencias locales), (ii) *placement* (depósitos en efectivo por mulas), (iii) *layering* (traslado por cheque hacia la cuenta concentradora), y (iv) integración (transferencias internacionales a beneficiarios) tal como se muestra en la Tabla 4.

**Tabla 4**

*Componentes principales del modelo basado en agentes*

Componente	Descripción
Dimensión espacial	Zonas fronterizas
Dimensión transaccional	Uso de efectivo, cheque y transferencias
Dimensión de perfil	Inconsistencias económicas frente al nivel transaccional
Agentes cliente	Realizan transacciones financieras
Agentes financieros	Procesan operaciones y aplican reglas de monitoreo

Agentes de riesgo	Representan comportamientos asociados a tipologías específicas
Variables de estado del cliente	Saldo disponible, frecuencia transaccional, monto promedio de transacción, perfil sociodemográfico sintético y nivel de riesgo calculado

Nota. Elaboración propia

La simulación opera en tiempo discreto mediante *ticks*, donde cada *tick* representa una hora; en los experimentos se simulan 240 *ticks* (10 días). Las distancias se expresan en Km y se derivan de una escala Km-por-patch.

**Tabla 5**

*Roles de agentes y relación con la tipología*

Rol del Agente	Descripción en Tipología	Atributo crítico en el modelo
Pitufio (Mule)	Persona que realiza múltiples depósitos en efectivo en zonas fronterizas.	Ubicación frontera; frecuencia de depósitos.
Estructurador (Launderer)	Dueño de la cuenta concentradora (Banco D) que envía giros internacionales.	Actividad declarada vs. giros.
Entidad Financiera (Bank)	Bancos A, B, C (receptores) y Banco D (emisor internacional).	Rol local vs. internacional.
Beneficiario (Sink)	Empresas en Asia con alertas o antecedentes.	Riesgo jurisdiccional; <i>flagged</i> .

Nota. Elaboración propia

**Tabla 6**

*Flujo por tick de la simulación basada en agentes para la tipología de transferencias electrónicas (LA/FT)*

Bloque Funcional del FIS	Variable / Métrica Evaluada	Tipo de Mapeo / Función Matemática	Impacto en el Modelo
Variables de Entrada ( <i>Fuzzificación</i> )	Monto de la transacción	Función de pertenencia trapezoidal y triangular (Bajo, Medio, Alto).	Alimentación multidimensional del motor difuso por cada TICK individual.
Variables de Entrada ( <i>Fuzzificación</i> )	Frecuencia transaccional	Función de pertenencia Gaussiana (Esporádica, Regular, Intensiva).	Alimentación multidimensional del motor difuso por cada TICK individual.
Variables de Entrada ( <i>Fuzzificación</i> )	Riesgo Geográfico	Función Lógica (Puntuación asignada según jurisdicción de origen/destino).	Alimentación multidimensional del motor difuso por cada TICK individual.

Motor de Inferencia ( <i>Mamdani</i> )	Base de Reglas Expertas (IF - THEN)	Evaluación simultánea mediante operadores lógicos difusos (Mínimo para AND / Máximo para OR). Ejemplo: IF Monto es Alto AND Frecuencia es Intensiva THEN Riesgo es Muy Alto.	Interconexión lógica de variables sin depender de umbrales rígidos lineales.
Variable de Salida ( <i>Defuzzificación</i> )	<i>risk_score</i>	Método del Centroide (Conversión del área difusa resultante a un valor escalar numérico continuo entre 0.0 y 1.0).	Generación del indicador cuantitativo final de alerta para la matriz de riesgos.

Nota. Elaboración propia

Para mejor facilidad de lectura del modelo, además de los roles descritos en la tabla 5, se presenta el siguiente Bloque Funcional del FIS de la Tabla 6 que sigue cada agente en la simulación. Este flujo resume qué acciones ejecuta cada actor por *tick* (hora) y en qué condiciones se activan eventos clave como depósitos, traslado por cheque, concentración de fondos y transferencias internacionales. Con esto se entiende mejor cómo la tipología se “traduce” a procesos operativos dentro del modelo basado en agentes, antes de aplicar el sistema difuso de puntaje de riesgo.

### Sistema de inferencia difusa para puntaje de riesgo

Para estimar el riesgo transaccional, el modelo incorpora un sistema de inferencia difusa tipo Mamdani, capaz de combinar señales cualitativas propias de la tipología sin recurrir a umbrales rígidos (Mamdani y Assilian, 1975). En cada *tick* de la simulación desarrollada en NetLogo (Wilensky, NetLogo, 1999), las transacciones generadas alimentan este sistema. En una primera etapa, se calculan las variables de estado como entradas nítidas (*crisp inputs*); posteriormente, estas se interpretan mediante etiquetas lingüísticas (*fuzzy inputs*) para obtener un puntaje final de riesgo (*risk\_score*) el cual se calcula utilizando funciones de pertenencia trapezoidales, triangulares y suaves (sigmoideal/gaussiana) para mapear las entradas numéricas a grados de pertenencia (Zadeh, 1965). La Tabla 7 sintetiza las etapas del lavado contempladas y el comportamiento esperado en cada una de ellas.

**Tabla 7**

*Señales de la tipología y sus variables del modelo*

Etapa del Lavado	Comportamiento del Agente	Variable de Estado (Crisp Input)	Variable Lingüística (Fuzzy Input)
Colocación	El Pitufo deposita efectivo en Bancos A, B, C.	<i>origin_dist_Km</i>	Nivel de Riesgo Geográfico
Colocación	El Pitufo realiza múltiples depósitos pequeños.	<i>amount_ratio, freq_count</i>	Monto y frecuencia

Estratificación	El Pitufo mueve fondos al Estructurador (Banco D) mediante cheques.	<i>tx_type (Efectivo/Cheque)</i>	Opacidad del Instrumento
Integración	El Estructurador envía transferencia internacional.	<i>jur_risk, flagged</i>	Riesgo Jurisdiccional
Transversal	Operación no coincide con perfil.	<i>profile_cons</i>	Consistencia de Perfil

Nota. Elaboración propia

El conjunto de datos integra variables estructurales, espaciales, temporales y conductuales. No se basa en registros reales ni en calibración histórica, sino en reglas expertas formalizadas mediante los sistemas difusos. La combinación de estas variables permite analizar tanto la dinámica del esquema como la capacidad emergente del sistema de riesgo.

Este modelo usa las siguientes entradas: distancia a frontera (*dist\_Km*), frecuencia en ventana (*freq\_count*), relación con umbral (*amount\_ratio*), consistencia de perfil (*profile\_cons*), riesgo jurisdiccional (*jur\_risk*), indicador flagged (0/1) y velocidad (*speed\_hours*).

Para asignar un puntaje de riesgo a cada transacción, se implementó un sistema de inferencia difusa tipo Mamdani. Su funcionamiento se apoya en reglas “SI – ENTONCES” evidenciado en la Tabla 8 que articulan señales cualitativas propias de la tipología, como la cercanía a la frontera, la frecuencia dentro de una ventana de tiempo, la estructuración próxima al umbral, la consistencia con el perfil y el nivel de riesgo del destino.

**Tabla 8**

*Base de reglas difusas (Mamdani) para risk\_score*

Regla	IF (antecedentes)	Operador	THEN (consecuente)	Interpretación tipológica (1 línea)
R1	<i>geo_border</i> $\wedge$ <i>f_high</i> $\wedge$ <i>a_structured</i>	AND = min	<i>risk = Critical</i>	Repetición cerca de frontera + estructuración cercana al umbral (patrón pitufo).
R2	<i>c_incons</i> $\wedge$ <i>j_high</i>	AND = min	<i>risk = Critical</i>	Operación no coherente con el perfil y destino/jurisdicción de mayor riesgo.
R3	<i>flagged</i>	– (directa)	<i>risk = High</i>	Beneficiario o contraparte con señales previas eleva el riesgo.
R4	<i>v_immediate</i> $\wedge$ <i>j_high</i>	AND = min	<i>risk = High</i>	Transferencia rápida hacia jurisdicción de riesgo (patrón acelerado).

R5	$a_{high}$	– (directa)	$risk = High$	Monto que supera el umbral de reporte (señal cuantitativa directa).
R6	$geo_{interior} \wedge c_{cons}$ $\wedge j_{low} \wedge (\neg flagged)$ $\wedge v_{normal}$	AND = min, NOT = (1 – flagged)	$risk = Low$	Condiciones compatibles con operación esperada (baja señal de riesgo).

Nota. Elaboración propia

Se corrieron tres (3) escenarios expuestos en la Tabla 9, para evaluar la sensibilidad del modelo ante cambios en el comportamiento de los depósitos de las mulas, el escenario base corresponde al experimento principal con el comportamiento de cada uno.

**Tabla 9**

*Matriz de escenarios: parámetro de estructuración y controles*

Variable	Escenario 0	Escenario 1	Escenario 2
<i>mule-deposit-prob</i>	0.35	0.20	0.50
<i>num-customers</i>	200	200	200
<i>Nummules</i>	20	20	20
<i>num-launders</i>	3	3	3
<i>num-beneficiaries</i>	40	40	40
<i>border-width</i>	6	6	6
<i>Km-per-patch</i>	5	5	5
<i>report-threshold</i>	10	10	10
<i>ticks-per-day</i>	24	24	24
<i>freq-window</i>	24	24	24
<i>max-deposits-per-tick</i>	3	3	3
<i>customer-tx-prob</i>	0.05	0.05	0.05
<i>max-wires-per-tick</i>	6	6	6
<i>max-ticks</i>	240	240	240

Nota. Elaboración propia

## Resultados

### Configuración del experimento

Dado el carácter estocástico del modelo, se realizaron 10 corridas por cada escenario, para un total de 30 ejecuciones. Esto permitió analizar la variabilidad de los resultados y seleccionar, en cada escenario, la corrida con mejor desempeño según el valor de AUC.

En cada escenario se tomó como referencia la corrida con mejor desempeño, definida a partir del mayor valor de AUC. Bajo este criterio, se seleccionaron *S0\_run04*, *S1\_run09* y *S2\_run02*, siendo *S1\_run09* el conjunto de datos principal para el análisis posterior. Entre las corridas analizadas, *S1\_run09* fue la que se tomó como referencia principal, porque mostró el mejor comportamiento dentro del escenario S1. En esa ejecución el modelo logró diferenciar mejor entre transacciones legítimas y transacciones asociadas a LA/FT, reflejado en el mayor valor de AUC observado.

En conjunto, este comportamiento indica que *mule – deposit – prob* influye de manera directa tanto en el volumen como en la composición de las transacciones generadas. En ese sentido, el parámetro actúa como un mecanismo de control sobre la intensidad de la fase de placement y sobre la forma en que esta se proyecta hacia las etapas posteriores del esquema.

La tabla 10 muestra, para cada escenario (S0, S1 y S2), la corrida con mejor desempeño entre las 10 réplicas realizadas, entendida como aquella que alcanzó el mayor valor de AUC. En la tabla 10, se presenta el número total de transacciones generadas, su distribución entre operaciones legítimas y operaciones asociadas a LA/FT, así como dos indicadores clave para valorar el comportamiento del puntaje de riesgo: el AUC (ROC), que permite apreciar su capacidad de separabilidad, y el porcentaje de transacciones LA/FT con *risk\_score*  $\geq 70$ , utilizado como una medida de cobertura de los casos clasificados con riesgo alto.

**Tabla 10**

*Resumen por escenario*

Escenario	<i>mule-deposit-prob</i>	Transacciones	Legítimas	LA/FT	AUC (ROC)	% LA/FT con score $\geq 70$
S0	0.35	11,419	2,424	8,995	0.720	43.9 %
S1	0.20	8,109	2,475	5,634	0.846	69.3 %
S2	0.50	13,852	2,485	11,367	0.608	21.5 %

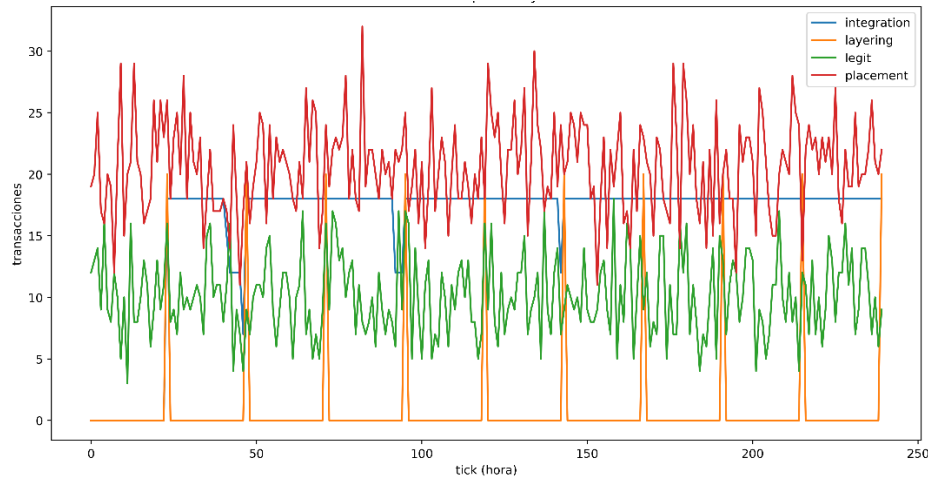
Nota. Elaboración propia

### Comportamiento temporal de la dinámica por tick

La Figura 2 presenta la evolución temporal de las transacciones por *tick* y por fase en los tres escenarios analizados. En el escenario base (S0) se aprecia una dinámica relativamente balanceada: la fase de *placement* concentra el mayor volumen de actividad, *legit* muestra oscilaciones moderadas, *integration* se mantiene en niveles estables y *layering* aparece en pulsos discretos y periódicos.

## Figura 2

### Transacciones por tick y escenario 0



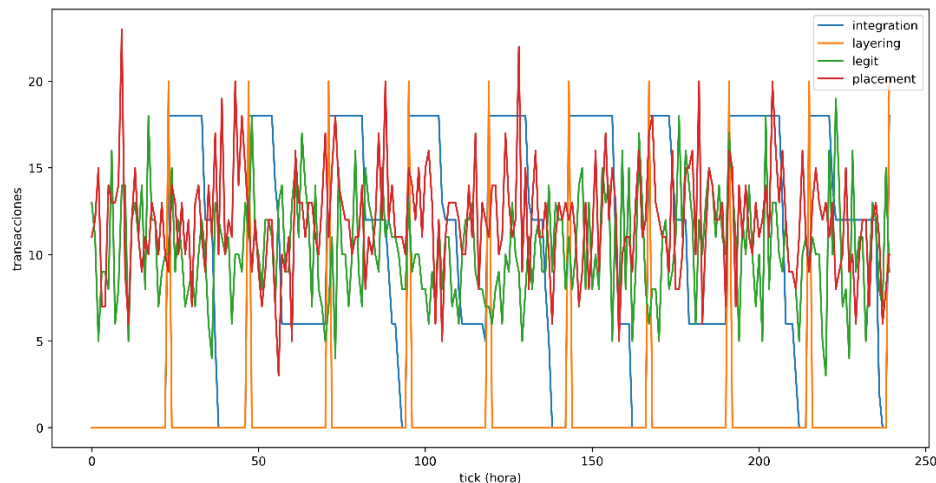
Nota. Elaboración propia

### Resultados del modelo por escenario

A diferencia de S0, el escenario S1 en la *Figura 3* presenta una distribución más equilibrada de la actividad entre sus fases constitutivas. Las transacciones legítimas y las correspondientes a *placement* permanecen activas durante prácticamente toda la simulación, con variaciones entre *ticks* que se inscriben dentro de rangos esperados. La fase *integration*, adquiere aquí un protagonismo más visible: aparece en bloques mejor delimitados y con cambios notorios en su intensidad, lo que contrasta con el papel más residual que desempeña en otros escenarios.

## Figura 3

### Transacciones por tick y escenario 1

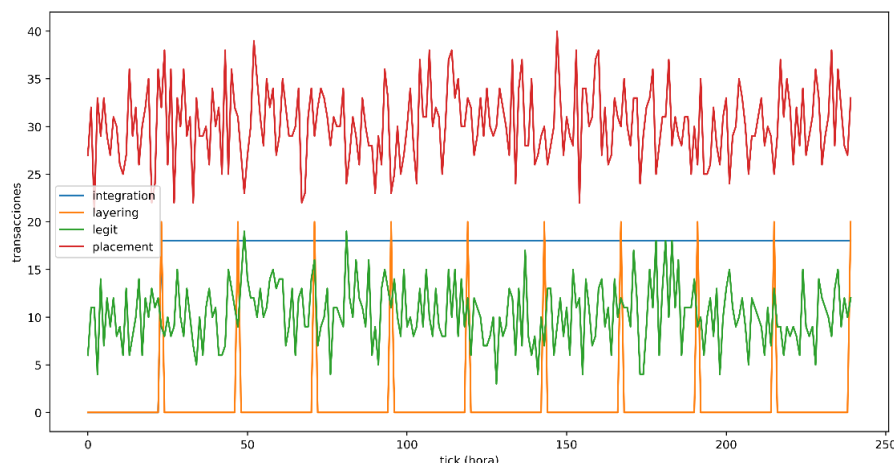


Nota. Elaboración propia

El escenario S2 como muestra la Figura 4 se caracteriza por una concentración de la actividad en la fase de *placement*, que impone su peso sobre las transacciones durante prácticamente toda la simulación con valores elevados y sostenidos. Las transacciones *legit* acompañan este comportamiento de forma continua, aunque con una variabilidad entre *ticks* más pronunciada que la observada en escenarios anteriores, lo que apunta a una mayor turbulencia en el componente legítimo del sistema.

**Figura 4**

*Transacciones por tick y escenario 2*



Nota. Elaboración propia

La Tabla 11 muestra que el escenario S1 presenta el mejor desempeño promedio ( $AUC = 0.828 \pm 0.017$ ) y la mayor proporción de transacciones laundering clasificadas con riesgo alto ( $62.4 \% \pm 3.3 \%$ ). En contraste, S2 reduce de manera importante la separabilidad ( $AUC = 0.579 \pm 0.016$ ) y disminuye la proporción de transacciones laundering con riesgo  $\geq 70$  ( $18.2 \% \pm 3.2 \%$ ). Por su parte, el escenario S0 se ubica en una posición intermedia ( $AUC = 0.678 \pm 0.021$ ).

**Tabla 11**

Indicadores de la mejor corrida por escenario.

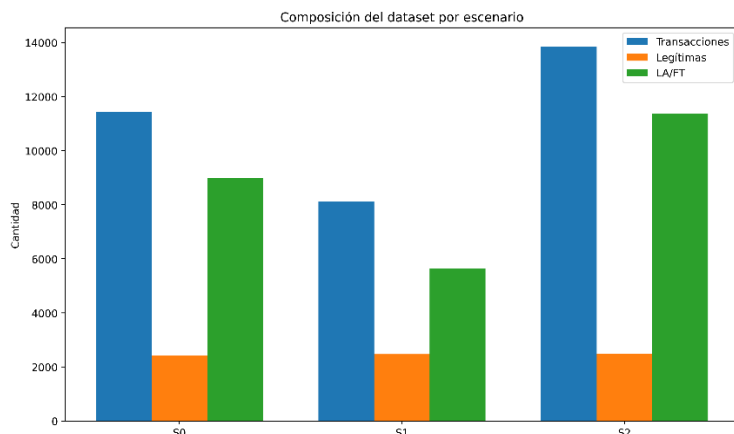
Esc.	AUC media $\pm$ sd	AUC mejor corrida	Riesgo medio laundering $\pm$ sd	% laundering $\geq 70$ media $\pm$ sd	% laundering $\geq 70$ mejor corrida
S0	0.678 $\pm$ 0.021	0.720 (corrida 04)	31.45 $\pm$ 35.77	38.3 % $\pm$ 4.2 %	43.9 % (corrida 04)
S1	0.828 $\pm$ 0.017	0.846 (corrida 09)	59.71 $\pm$ 26.63	62.4 % $\pm$ 3.3 %	69.3 % (corrida 09)
S2	0.579 $\pm$ 0.016	0.608 (corrida 02)	13.31 $\pm$ 26.66	18.2 % $\pm$ 3.2 %	21.5 % (corrida 02)

Nota. Elaboración propia

La Figura 5 muestra que el tamaño y la composición del conjunto de datos cambian entre escenarios, especialmente en el número de transacciones asociadas a LA/FT. Mientras las transacciones legítimas se mantienen en un rango muy similar en S0, S1 y S2, el volumen total y la cantidad de operaciones vinculadas al guion de LA/FT varían con mayor claridad. En particular, S2 presenta la mayor carga transaccional, seguido de S0, mientras que S1 genera el conjunto más pequeño. Esto sugiere que el parámetro evaluado afecta principalmente la intensidad de las transacciones sospechosas, sin alterar de forma importante la base de operaciones legítimas del modelo.

## Figura 5

### Composición del conjunto de datos por escenario



Nota. Elaboración propia

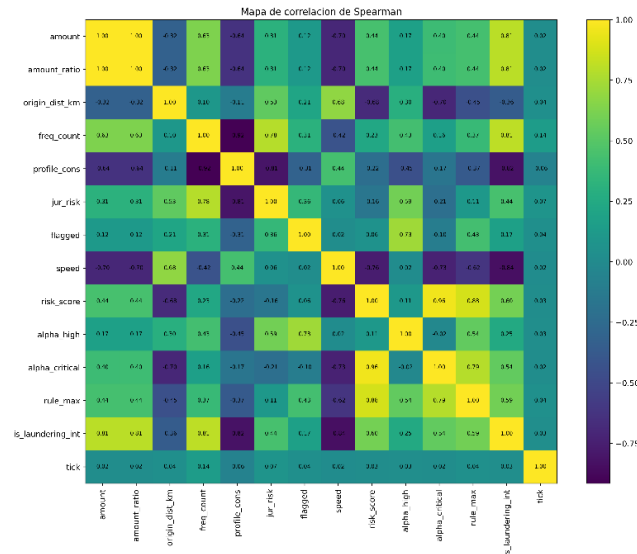
### Validación del conjunto de datos sintético

La validación del conjunto de datos sintético se orientó a establecer si la salida generada por el modelo conserva coherencia interna, estructura temporal y capacidad de diferenciación entre transacciones legítimas y transacciones asociadas a LA/FT. En este sentido, la validación no buscó demostrar que los datos sintéticos fueran una copia de datos reales, sino evaluar si el generador produjo un conjunto de datos plausible, interpretable y útil para análisis posteriores.

En el *dataset* de referencia *S1\_run09* se registraron 8,109 transacciones, de las cuales 69.48 % correspondieron a eventos asociados a LA/FT y 30.52 % a transacciones legítimas. Además, la distribución por fases mostró una estructura no uniforme: *placement* 35.31 % de las operaciones, *integration* 31.71 %, *legit* 30.52 % y *layering* 2.47 %. Estos valores indican que el archivo conserva una organización interna reconocible y consistente con la lógica del modelo.

**Figura 6**

Mapa de correlación entre variables cuantitativas y la etiqueta binaria.



Nota. Elaboración propia

En el análisis de correlación como se muestra en Figura 6 *el*, *risk\_score* se asoció fuertemente con *alpha\_critical* ( $\rho = 0.960$ ) y *rule\_max* ( $\rho = 0.880$ ), mientras que *is\_laundering* mostró asociaciones positivas con *freq\_count* ( $\rho = 0.809$ ) y *amount\_ratio* ( $\rho = 0.805$ ).

## Discusión

El modelo propuesto demostró que la integración de la simulación basada en agentes y los sistemas de inferencia difusa permite generar perfiles transaccionales coherentes con las tipologías de LA/FT sin depender de datos históricos. Al interpretar el comportamiento del sistema, se evidenció que la configuración de la fase de *placement* (depósitos de las mulas) actúa como un regulador estructural del riesgo. El desempeño superior del escenario S1 indica que un volumen transaccional controlado permite al sistema difuso aislar con mayor precisión las señales de alerta y estructurar mejor el ciclo de lavado. Por el contrario, escenarios de saturación (como S2) demuestran que un exceso de operaciones ilícitas inunda la red, lo que degrada la capacidad discriminadora del modelo y dificulta la priorización del riesgo.

Al contrastar estos resultados con el estado del arte, este enfoque representa un avance metodológico frente a herramientas analizadas previamente. Mientras que simuladores de referencia, como PaySim o SynthAML, mantienen una dependencia crítica hacia conjuntos de datos reales para su calibración inicial, el generador aquí propuesto logra construir redes transaccionales funcionales basándose puramente en el conocimiento heurístico de expertos. Esto tiene una implicación práctica profunda para la comunidad científica: permite superar el obstáculo del secreto bancario y las normativas de protección de datos (*habeas data*), ofreciendo un entorno experimental libre de restricciones de confidencialidad para entrenar y evaluar algoritmos analíticos.

Asimismo, las pruebas estadísticas aplicadas confirman que la simulación no produce secuencias estocásticas aisladas, sino un flujo financiero con memoria local y estructura sistémica. Las asociaciones fuertes entre variables como la frecuencia, la consistencia del perfil y la estructuración del monto hacia el riesgo transaccional validan computacionalmente las señales de alerta de la tipología del "pitufeo". De esta forma, el sistema logra traducir adecuadamente el conocimiento experto cualitativo en umbrales matemáticos operativos.

A pesar de estos aportes, el modelo presenta limitaciones inherentes a su diseño que deben ser consideradas. En primer lugar, el motor de inferencia difusa Mamdani requiere de una parametrización manual exhaustiva. Esta dependencia metodológica puede introducir sesgos subjetivos por parte del experto que diseña las reglas, restringiendo la flexibilidad del sistema para ajustarse de forma automática ante variaciones muy sutiles u ocultamientos avanzados en casos extremos. En segundo lugar, al concentrarse exclusivamente en una tipología específica de transferencias electrónicas, la capacidad generalizadora del conjunto de datos es limitada. La representación de escenarios criminales más amplios dependerá de la capacidad de integrar dinámicas adaptativas que superen la ejecución de guiones fijos.

## Conclusiones

En esta investigación se desarrolló un generador de datos sintéticos para una tipología de transferencias electrónicas asociada a LA/FT, combinando un modelo de simulación basada en agentes implementado en NetLogo con un sistema de inferencia difusa tipo Mamdani para asignar un puntaje de riesgo por transacción. Este enfoque permitió construir una base de datos sintética sin depender de históricos reales, lo que resulta especialmente valioso en contextos donde existen restricciones de acceso, confidencialidad o disponibilidad de información. Además, el modelo logró producir escenarios con tamaños y composiciones diferentes, lo que confirma que el generador responde a cambios en sus parámetros y no genera siempre una salida fija.

Los resultados permiten concluir que el modelo produce conjuntos de datos con una estructura interna clara y coherente. En el *dataset* de referencia *S1\_run09*, por ejemplo, se generaron 8,109 transacciones, de las cuales 69.48 % correspondieron a eventos asociados a LA/FT y 30.52 % a transacciones legítimas. A esto se suma que la distribución por fases tampoco fue uniforme: 35.31 % de las operaciones quedaron en placement, 31.71 % en integration, 30.52 % en legit y 2.47 % en layering.

También se concluye que el conjunto de datos presenta variación y heterogeneidad suficientes para fines analíticos. Esto se ve con claridad en el comportamiento del *risk\_score*. En *S1\_run09*, la media global del puntaje fue de 43.39 con una desviación estándar de 45.35. Sin embargo, al separar por tipo de transacción, la diferencia es mucho más clara: las transacciones legítimas tuvieron un riesgo medio de 0.00, mientras que las asociadas a LA/FT alcanzaron un riesgo medio de 62.45 y una mediana de 93.57.

En conjunto, el trabajo confirma dos aportes principales. El primero es que la simulación basada en agentes constituye una alternativa viable para generar datos sintéticos coherentes con una tipología específica de LA/FT. El segundo es que el componente difuso

Mamdani funciona como un mecanismo interpretable para asignar riesgo, aunque su desempeño depende de una calibración cuidadosa de reglas y funciones de pertenencia, especialmente cuando se analizan escenarios extremos.

### **Trabajos futuros**

Como trabajo futuro, se propone ampliar el modelo en varias direcciones que permitan aumentar su realismo, su capacidad explicativa y su utilidad analítica. En primer lugar, sería pertinente recalibrar las funciones de pertenencia y las reglas difusas para representar con mayor precisión los casos extremos, especialmente aquellos en los que las transacciones sospechosas se acercan de forma más fina a los umbrales de alerta. En segundo lugar, sería valioso incorporar nuevas tipologías de LA/FT que amplíen la diversidad del conjunto de datos y eviten que la simulación dependa de un solo guion transaccional.

En una línea más avanzada, el modelo podría evolucionar hacia el estudio de dinámicas no lineales y adaptativas, de manera que los agentes no solo ejecuten reglas fijas, sino que modifiquen su comportamiento en respuesta al monitoreo, a la presión regulatoria o a la detección de patrones de control. Esto permitiría simular trayectorias criminales más realistas, donde los actores ajustan sus estrategias para evadir la supervisión y generan comportamientos emergentes más complejos.

### **Referencias**

- Altman, E., Blanuša, J., von Niederhäusern, L., Egressy, B., Anghel, A., & Atasu, K. (2023). Realistic Synthetic Financial Transactions for Anti-Money Laundering Models. *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*. 36, 29851-29874. Curran Associates, Inc.
- Assefa, S. A., Dochev, D., & Eyal, I. (2020). Generating synthetic data in finance: Opportunities, challenges and pitfalls. *Proceedings of the First ACM International Conference on AI in Finance* (págs. 1-8). New York, NY: Association for Computing Machinery.  
<https://doi.org/10.1145/3383455.3422554>
- Ding, N. F. (2025). AML-CFSim: An agent-based simulation model for anti-money laundering from cyber fraud crimes. *Expert Systems with Applications*, 285.  
<https://doi.org/10.1016/j.eswa.2025.127995>
- Fan, J. (2025). Deep Learning Approaches for Anti-Money Laundering on Mobile Transactions: Review, Framework, and Directions. *arXiv preprint arXiv:2503.10058*.  
<https://doi.org/10.48550/arXiv.2503.10058>
- Financial Action Task Force. (2025). *FATF glossary*. FATF. <https://www.fatf-gafi.org/en/pages/fatf-glossary.html>
- Goodfellow, I. J. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 27, (págs. 2672-2680). <https://proceedings.neurips.cc/paper/5423-generative-adversarial-nets>
- Grimm, V., Railsback, S. F., Vincenot, C., Berger, U., Gallagher, C., DeAngelis, D., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M., Stillman, R. A., Thiele, J. C., & Ayllón, D. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *Journal of Artificial Societies and Social Simulation*, 23(2), 7. <https://doi.org/10.18564/jasss.4259>

- Grupo de Acción Financiera Internacional. (2021). Opportunities and challenges of new technologies for AML/CFT. FATF/OECD. <https://www.fatf-gafi.org/publications/financial-technologies-and-crime/documents/opportunities-challenges-new-technologies-aml-cft.html>
- Grupo de Acción Financiera Internacional. (2023). *International standards on combating money laundering and the financing of terrorism & proliferation: The FATF recommendations*. FATF/OECD.. <https://www.fatf-gafi.org/content/fatf-gafi/en/publications/Fatfrecommendations/Fatf-recommendations.html>
- Grupo de Acción Financiera de Latinoamérica [GAFILAT]. (2016). *Recopilación de tipologías regionales GAFILAT (2009–2016)*. GAFILAT. [https://biblioteca.gafilat.org/wp-content/uploads/2016/11/Recopilacion\\_de\\_Tipologias\\_Regionales\\_GAFILAT\\_2009-2016.pdf](https://biblioteca.gafilat.org/wp-content/uploads/2016/11/Recopilacion_de_Tipologias_Regionales_GAFILAT_2009-2016.pdf)
- Jensen, R. I. (2023). A synthetic data set to benchmark anti-money laundering methods. *Scientific Data*, 10, 661. <https://doi.org/10.1038/s41597-023-02569-2>
- Le-Khac, N.-A., & Markos, S.-M.-T. (2009). A heuristics approach for fast detecting suspicious money laundering cases in an investment bank. *International Journal of Computer and Information Engineering*, 3, 2742.
- Lopez Rojas, E. A., Elmir, A., & Axelsson, S. (2016). PaySim: A financial mobile money simulator for fraud detection. *28th European Modeling and Simulation Symposium (EMSS)* (págs. 249-255). Larnaca: Dime University of Genoa.
- Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3), 151-162. <https://doi.org/10.1057/jos.2010.3>
- Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1), 1-13. [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2)
- Ngai, E. W., Hu, Y., Wong, Y. J., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559-569. <https://doi.org/10.1016/j.dss.2010.08.006>
- Organización de las Naciones Unidas (ONU). (1999). *International Convention for the Suppression of the Financing of Terrorism*. Naciones Unidas.
- Ross, T. J. (2016). *Fuzzy Logic with Engineering Applications* (Vol. 4). Hoboken, NJ: John Wiley & Sons.
- Szangolies, L. (2024). Visual ODD: A standardised visualisation illustrating the narrative of agent-based models. *Journal of Artificial Societies and Social Simulation*, 27(4), 1. <https://doi.org/10.18564/jasss.5450>
- Tang, J., & Yin, J. (2005). Developing an intelligent data discriminating system of anti-money laundering based on SVM. *Proceedings of the 2005 International Conference on Machine Learning and Cybernetics*, (págs. 3453-3457).
- Wang, S.-N., & Yang, J.-G. (2007). A money laundering risk evaluation method based on decision tree. *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, (págs. 283-284).
- Weber, M., Chen, J., Suzumura, T., Pareja, A., Ma, T., Kanezashi, H., Kaler, T., Leiserson, C. E., & Schardl, T. B. (2018). *Scalable graph learning for anti-money laundering: A first look*. arXiv. <https://doi.org/10.48550/arXiv.1812.00076>

Wilensky, U. (1999). NetLogo [Software]. Center for Connected Learning and Computer-Based Modeling, Northwestern University. <http://ccl.northwestern.edu/netlogo/>

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge, MA: MIT Press.

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems*. 32, págs. 7333-7343. Red Hook, NY: Curran Associates, Inc.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.  
[https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)