

Agentes Autónomos Descentralizados: Convergencia de Grandes Modelos de Lenguaje y Blockchain para la Toma de Decisiones Automatizadas, Auditables y Colaborativas

Jorge Polanco Roque

[\[jorge.polanco@tec.mx\]](mailto:jorge.polanco@tec.mx)

Abstract. La conjunción de la inteligencia artificial (IA), especialmente a partir de modelos de lenguaje de gran tamaño (LLMs), y la tecnología de cadena de bloques (blockchain) está impulsando una transformación de las estructuras de gobernanza y de los procesos de toma de decisiones en entornos distribuidos. En este contexto, surgen los Agentes Autónomos Descentralizados (AADs) como unidades de software capaces de razonar sobre datos registrados en contratos inteligentes, coordinarse con otros actores de la red y desencadenar acciones tanto on-chain como off-chain sin requerir intermediarios centralizados. Este trabajo propone un marco conceptual para el desarrollo de AADs, subrayando los aspectos tecnológicos críticos —que abarcan infraestructura en la nube, conexiones con proveedores para conexión a Blockchains, diseño de contratos inteligentes y oráculos— y la integración con modelos de IA que proveen capacidades cognitivas avanzadas. Asimismo, se describe un estudio de caso centrado en votaciones descentralizadas y se discuten retos técnicos, regulatorios y éticos, junto con líneas de investigación futura y tendencias emergentes.

Keywords: Agentes Autónomos Descentralizados, Blockchain, IA, LLM, Smart Contracts, Gobernanza Descentralizada, Oráculos, Cloud Computing, Web 2.0, Web 3.0.

1 Introducción

La proliferación de tecnologías como la IA y blockchain ha posibilitado la construcción de ecosistemas digitales con un alto grado de autonomía y fiabilidad. La IA, impulsada por métodos de aprendizaje profundo y de aprendizaje automático tradicional, ha incrementado sus capacidades gracias a los denominados modelos de lenguaje de gran tamaño, que permiten extraer patrones y razonar de manera contextual a partir de un vasto conjunto de datos [4]. Al mismo tiempo, la tecnología blockchain ofrece un entorno distribuido, inmutable y trazable para la ejecución de transacciones y el despliegue de aplicaciones descentralizadas, lo que reduce la dependencia de entidades centrales e incrementa la confianza y la transparencia [2][3].

En la convergencia de estas dos corrientes tecnológicas se sitúan los Agentes Autónomos Descentralizados. Dichos agentes adquieren un rol activo en la red, recabando información on-chain y off-chain para procesarla mediante diversos tipos de algoritmos—particularmente, mediante LLMs— y emitir decisiones o recomendaciones que pueden plasmarse directamente en contratos inteligentes. Esto abre un abanico de oportunidades en la gobernanza distribuida, la gestión de cadenas de suministro, el arbitraje de disputas, la detección de fraude y la orquestación de recursos en organizaciones virtuales. Al mismo tiempo, surgen importantes preguntas acerca de la escalabilidad de estas soluciones, la responsabilidad legal en entornos completamente descentralizados y la forma en que se garantizará la privacidad de los datos personales [6].

Este artículo se estructura en nueve secciones para analizar la convergencia entre inteligencia artificial y blockchain en la construcción de Agentes Autónomos Descentralizados. Tras esta introducción, la segunda sección expone los fundamentos de ambas tecnologías, con énfasis en los modelos de lenguaje de gran tamaño y su papel en la toma de decisiones en entornos distribuidos.

En la tercera sección, se presenta un marco conceptual para el desarrollo de AADs, detallando la infraestructura necesaria, la conexión con blockchain mediante proveedores como Infura, la gobernanza descentralizada y la integración con oráculos. La cuarta sección introduce un estudio de caso sobre votaciones descentralizadas, donde un AAD con capacidades cognitivas avanzadas supervisa la emisión de votos y detecta irregularidades.

La quinta sección explora mejoras tecnológicas y propuestas algorítmicas para optimizar el desempeño de los AADs, incluyendo mecanismos de consenso basados en reputación, Zero-Knowledge Proofs y estrategias de escalabilidad mediante sharding temático. En la sexta sección, se analizan los desafíos normativos y éticos, con un enfoque en la responsabilidad legal de las decisiones automatizadas y su compatibilidad con marcos regulatorios. Finalmente, la séptima sección presenta conclusiones y futuras líneas de investigación, abordando el impacto potencial de los AADs en la gobernanza descentralizada y la economía digital.

2 Fundamentos de Blockchain, Votaciones Descentralizadas e IA

La tecnología blockchain se fundamenta en un registro inmutable y distribuido, donde los nodos validan transacciones mediante protocolos de consenso criptográficos. Este paradigma, introducido con Bitcoin [2], evolucionó con redes como Ethereum, que incorporan contratos inteligentes para la ejecución programática de transacciones [1].

Aunque el consenso inicial se basaba en prueba de trabajo (Proof-of-Work), alternativas como la prueba de participación (Proof-of-Stake) han mejorado la escalabilidad y reducido el consumo energético [5]. No obstante, los costos de transacción y la latencia han impulsado soluciones de segunda capa, como sidechains y rollups, que optimizan la eficiencia de la red principal [4].

Paralelamente, la inteligencia artificial ha avanzado desde modelos de aprendizaje automático tradicionales hasta arquitecturas neuronales profundas capaces de procesar grandes volúmenes de datos. En este contexto, los modelos de lenguaje de gran tamaño han demostrado un desempeño sobresaliente en comprensión y generación de texto, síntesis de información y razonamiento contextual. Estas arquitecturas —como GPT-4— se entrenan con conjuntos masivos de datos en infraestructuras escalables y pueden ajustarse para tareas específicas de todo tipo.

La convergencia entre IA y blockchain radica en la capacidad de los modelos de lenguaje para automatizar y optimizar procesos en redes descentralizadas, mientras la blockchain garantiza transparencia, auditabilidad e integridad de los datos [3]. Para lograr esta integración, parte del procesamiento de IA se delega a infraestructuras en la nube o entornos de computación distribuida, mientras la blockchain registra los resultados esenciales, asegurando inmutabilidad y consenso.

Un campo donde esta integración cobra especial relevancia es el de las votaciones descentralizadas. La inmutabilidad y trazabilidad de la blockchain permiten un registro confiable de cada voto y garantizan la transparencia en los resultados. Sin embargo, la incorporación de agentes basados en modelos de lenguaje introduce una capa de interpretación y análisis que amplía las capacidades de supervisión y detección de irregularidades en el proceso electoral de una “DAO”, por ejemplo.

Las DAOs (Organizaciones Autónomas Descentralizadas) son estructuras de gobernanza basadas en contratos inteligentes, donde las decisiones se toman colectivamente mediante un sistema de votación con tokens. Su transparencia e inmutabilidad han permitido su adopción en sectores como finanzas descentralizadas (DeFi), inversión colectiva y gestión de comunidades, eliminando intermediarios y fortaleciendo la autonomía organizativa.

En este contexto, los LLM pueden actuar como supervisores inteligentes, analizando dinámicas de participación y detectando anomalías en tiempo real. Un modelo entrenado y/o contextualizado en gobernanza descentralizada puede evaluar si una propuesta contradice decisiones previas, identificar estrategias de manipulación por grupos minoritarios o detectar patrones de votación atípicos. Además, estos agentes pueden generar reportes en lenguaje natural, facilitando la toma de decisiones informadas por parte de la comunidad.

Más aún, los LLM pueden desempeñar un papel activo en la votación, operando como representantes algorítmicos dentro de una DAO. Un agente de este tipo podría analizar el historial de gobernanza, evaluar propuestas según criterios predefinidos y emitir votos en representación de usuarios que deleguen su decisión en la IA. Esto resulta particularmente útil en entornos donde la toma de decisiones requiere análisis técnico o económico complejo, permitiendo que un Agente Autónomo Descentralizado vote con base en información estructurada y en tiempo real.

Al combinar el razonamiento avanzado de los LLM con la seguridad y descentralización de la blockchain, es posible construir sistemas de gobernanza más eficientes y autónomos. En lugar de depender exclusivamente de reglas predefinidas o de la intervención manual de los votantes, las DAOs pueden evolucionar hacia modelos de decisión más dinámicos, donde la IA no solo detecte irregularidades, sino que también proponga estrategias de mitigación, facilite la deliberación colectiva y participe activamente en la toma de decisiones.

Además, la participación de estos agentes refuerza la confianza en los procesos electorales, ya que su supervisión sigue reglas codificadas en contratos inteligentes, auditable por la comunidad y sin intervención de una autoridad central. Esto no solo mejora la resistencia de las votaciones frente a manipulaciones, sino que también permite la adaptación dinámica de las reglas de gobernanza conforme evoluciona la comunidad, consolidando así un modelo más robusto y participativo.

La integración de modelos de lenguaje de gran tamaño con blockchain redefine la gobernanza descentralizada al combinar automatización, transparencia y análisis avanzado en la toma de decisiones. Esta sinergia no solo fortalece la resiliencia de los sistemas de votación descentralizados, sino que también permite la evolución hacia modelos de gobernanza más dinámicos, eficientes y adaptativos, reduciendo riesgos de manipulación y mejorando la toma de decisiones en comunidades autónomas.

3 Arquitectura de los Agentes Autónomos Descentralizados

El concepto de Agentes Autónomos Descentralizados surge como un intento de unificar la lógica de los agentes autónomos de IA y la ejecución segura en blockchain en un único marco de referencia. Un AAD se concibe como una entidad de software que posee llaves criptográficas para firmar transacciones, extrae datos tanto de la cadena (on-chain) como de fuentes externas (off-chain), y razona sobre dichos datos mediante algoritmos de inteligencia artificial, con el objetivo de desplegar acciones automatizadas.

En términos de arquitectura, la infraestructura para un AAD puede organizarse en capas que facilitan su diseño y despliegue:

En la **Capa de Datos y Conectividad**, el AAD se vincula con proveedores de nodos, como Infura, para interactuar con la red principal sin necesidad de mantener un nodo completo, lo que reduce la complejidad de configuración y mantenimiento. Además, esta capa integra servicios de almacenamiento como InterPlanetary File System (IPFS) o bases de datos en la nube para manejar datos masivos que no se almacenan directamente en la blockchain. El AAD también establece enlaces con oráculos, que suministran información externa, por ejemplo, datos de mercado o identidad digital, y permiten la verificación de eventos del mundo real.

En la **Capa de Inteligencia Artificial**, se emplean modelos de lenguaje de gran tamaño para procesar y razonar sobre grandes volúmenes de información. Estos modelos se entrenan en infraestructuras de computación escalables (GPU, TPU) y luego se implementan en servidores o contenedores que puedan comunicarse con la capa on-chain mediante interfaces de programación. El LLM puede especializarse en la detección de fraude, el análisis de transacciones financieras, la clasificación de propuestas en la DAO o cualquier otra tarea requerida por la organización descentralizada.

En la **Capa de Blockchain y Contratos Inteligentes**, se definen las reglas de operación de la red y las condiciones en las que el AAD puede tomar acciones específicas. Los contratos inteligentes, escritos en lenguajes como Solidity (en el caso de Ethereum), recogen la lógica necesaria para transferir activos, restringir comportamientos indebidos y gestionar eventos relevantes para el AAD. De esta manera, la validación de cada acción del agente se registra en la red, generando transparencia y facilitando el escrutinio público o privado.

En la **Capa de Gobernanza Descentralizada**, la comunidad o los participantes que ostenten tokens de gobernanza votan las actualizaciones de políticas, los parámetros de IA y las potenciales sanciones a comportamientos maliciosos. El AAD puede tener derecho de voto si la comunidad así lo dispone, o puede asumir un papel de auditor a fin de identificar anomalías y proponer sanciones. Este mecanismo de participación abierta y verificable crea incentivos para la contribución honesta y dificulta la manipulación por parte de uno o pocos actores.

Finalmente, en la **Capa de Interfaz y Aplicaciones Híbridas**, los usuarios finales, desarrolladores o empresas interactúan con el AAD y con la red. Esta capa puede incluir tableros de control, formularios de votación o servicios de suscripción a eventos. Combina tecnologías web 2.0 tradicionales (por ejemplo, servidores en AWS, front-ends JavaScript) con el acceso a la blockchain a través de librerías como web3.js o ethers.js, con la finalidad de presentar la información de la forma más accesible posible. Esta arquitectura escalable y modular se sustenta en la sinergia entre la nube y la cadena de bloques, aprovechando la potencia de cómputo fuera de la cadena para entrenar y ejecutar la IA, mientras la blockchain garantiza la inmutabilidad, la trazabilidad y la gobernanza colectiva [7].

4 Caso de uso: Creación Automatizada de DAOs mediante LLMs

La necesidad de diseñar contratos inteligentes adaptados a distintos modelos de gobernanza ha impulsado el desarrollo de plataformas que permitan crear DAOs de forma rápida y flexible. En este contexto, se ha concebido un proyecto en el que un pipeline de inteligencia artificial facilita la configuración y el despliegue de estos contratos, integrando tanto la experiencia de los usuarios como la generación automatizada de código mediante un LLM. El flujo de trabajo se basa en un árbol de decisiones, en el que cada rama responde a requisitos específicos de gobernanza. Por ejemplo, la plataforma puede solicitar información sobre los tipos de votación necesarios, los umbrales de quórum o los mecanismos de penalización por inactividad. Con base en esta información, se parametrizan las variables en un prompt interno que llama al LLM, en este caso GPT-4o, el cual genera el contrato inteligente alineado con las necesidades definidas.

La plataforma valida y compila el contrato resultante para verificar su coherencia y corregir posibles errores. Una vez obtenido el código definitivo, la plataforma ofrece la posibilidad de desplegar el contrato con un solo clic, conectándose a la red blockchain establecida. El usuario, por tanto, no requiere experiencia profunda en programación de contratos inteligentes ni en detalles de la infraestructura blockchain. El LLM actúa como un asistente cognitivo, traduciendo necesidades de gobernanza a instrucciones de código de manera transparente y auditable.

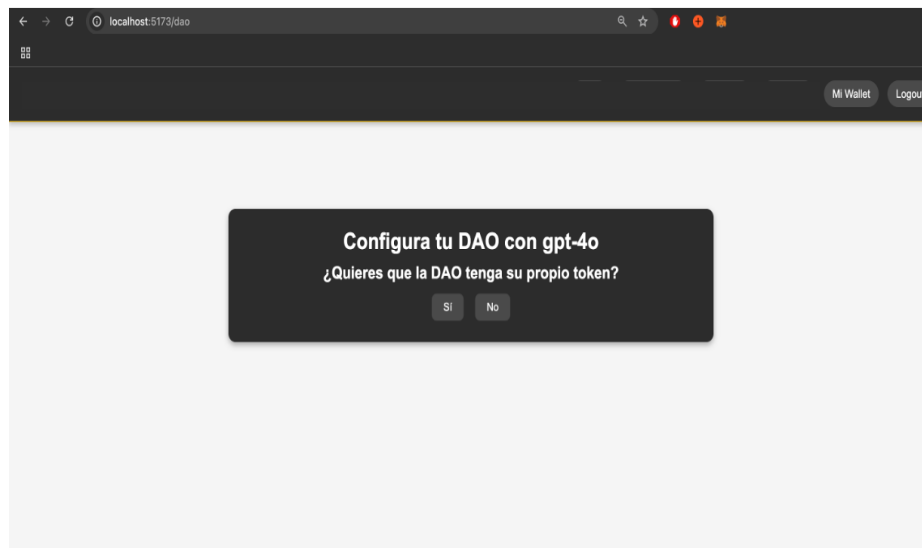


Fig. 1. La plataforma guía al usuario a través de un árbol de decisiones, parametrizando la generación automatizada del contrato inteligente según las respuestas.

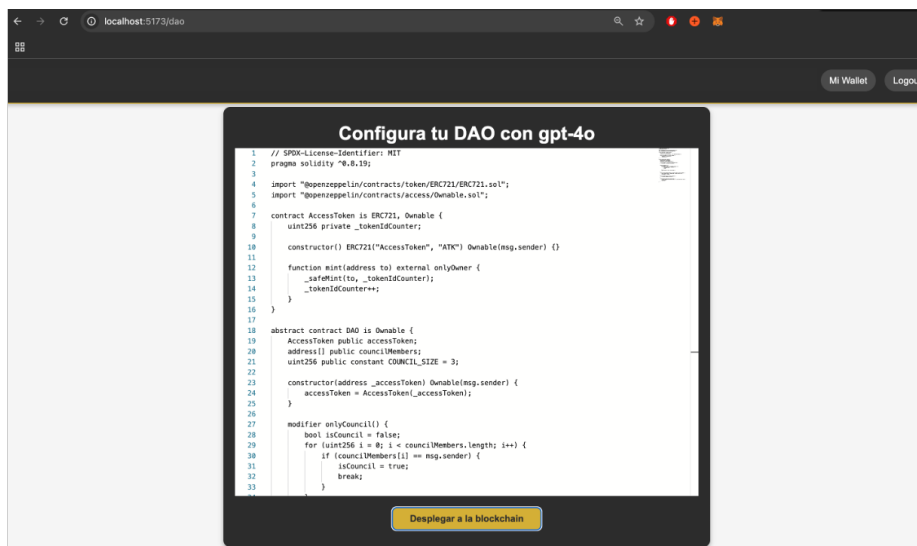


Fig. 2. Generación automatizada del contrato inteligente para la DAO. Tras definir los parámetros de gobernanza mediante GPT-4o, la plataforma muestra el código del contrato en Solidity, listo para su validación y despliegue en la blockchain con un solo clic.

En escenarios similares, las capacidades de los LLMs resultan útiles para automatizar procesos que suelen requerir intervención humana intensiva. Una aplicación análoga se presenta en la supervisión de votaciones descentralizadas, donde el agente IA se encarga de identificar comportamientos atípicos o indicios de fraude al procesar grandes volúmenes de datos transaccionales en tiempo real. De igual forma, en la detección temprana de anomalías financieras, estos modelos pueden analizar historiales de transacciones y comparar patrones observados con perfiles de actividad habitual, generando alertas on-chain cuando las operaciones superan determinados umbrales de riesgo. Otro caso práctico consiste en la gestión de cadenas de suministro, en la que un AAD monitorea el flujo de productos, validando documentos de procedencia y coordinando condiciones logísticas a través de contratos inteligentes.

La interacción entre un LLM y la blockchain proporciona una infraestructura potente para la automatización de procesos en la creación y gestión de DAOs. Mientras la red garantiza confianza distribuida y seguridad criptográfica, el modelo de lenguaje facilita la interpretación de requisitos y la generación de código. Sin embargo, depender exclusivamente de estos sistemas automatizados puede presentar riesgos en términos de validación, supervisión y adaptabilidad. Para mitigar estas limitaciones, es fundamental incorporar sistemas de validación híbridos, donde los contratos inteligentes generados sean revisados antes del despliegue y donde mecanismos como human-in-the-loop permitan la intervención humana en decisiones críticas. Este enfoque equilibra la eficiencia de la automatización con la necesidad de control y supervisión, asegurando que la descentralización no comprometa la seguridad ni la coherencia de la gobernanza.

5 Propuestas de Mejoras Tecnológicas y Algorítmicas

La implementación práctica de Agentes Autónomos Descentralizados que emplean modelos de lenguaje de gran tamaño enfrenta retos en confiabilidad, privacidad y escalabilidad.

Para abordarlos, existen diversas estrategias que permiten llevar estos agentes a entornos de producción de manera más segura y eficiente:

Un primer paso es **ampliar los mecanismos de consenso** más allá de la prueba de participación (Proof-of-Stake), incorporando métricas de reputación o confiabilidad generadas por la IA. Este enfoque requiere que cada nodo aporte un historial de comportamiento (p. ej., participación en votaciones, precisión en la detección de fraude o calidad de datos en procesos de aprendizaje federado). El LLM evalúa este historial y genera una puntuación de reputación dinámica, la cual se integra en el consenso para penalizar nodos maliciosos y recompensar contribuciones positivas.

En cuanto a la **privacidad**, el uso de cifrado homomórfico y Zero-Knowledge Proofs (ZKPs) [7] permite que los AADs trabajen con datos sensibles sin descriptarlos. Por ejemplo, el agente podría validar la legitimidad de una transacción o marcarla como fraudulenta mediante pruebas criptográficas que demuestren la corrección de su análisis sin revelar su contenido. Aunque esta aproximación eleva la complejidad y el costo computacional, resulta esencial en aplicaciones donde la confidencialidad es prioritaria.

Para **escalar** estas operaciones, la fragmentación de la red en subcadenas especializadas o sharding temático [4] distribuye la carga de cómputo. Una subcadena puede dedicarse al entrenamiento y la ejecución del modelo IA, mientras la cadena principal registra únicamente el hash de los resultados y las actualizaciones clave. Esto aligera la congestión y reduce los costos de transacción, pero exige un protocolo de interoperabilidad multicadena, que permita a los AADs leer y escribir en las distintas subcadenas y, a su vez, publicar en la red principal los resultados finales de su procesamiento.

Un ejemplo práctico de este flujo incluye la fase de entrenamiento y evaluación en una subcadena, seguida de la generación de una prueba criptográfica que describe los resultados, y finaliza con el registro en la cadena principal, donde un contrato inteligente solo acepta los resultados acompañados de dicha prueba. En el caso de votaciones, la misma lógica puede garantizar la validez de las decisiones tomadas por el LLM, exigiendo evidencias de su análisis on-chain para salvaguardar la transparencia del proceso.

La **auditoría algorítmica** constituye otro pilar fundamental. La DAO puede configurar contratos inteligentes de auditoría que obliguen al AAD a proporcionar explicaciones resumidas de las decisiones más relevantes, ya sea la suspensión de un voto masivo o la detección de comportamientos anómalos.

Si la comunidad considera insuficientes las justificaciones, puede votar sanciones o la revocación de los privilegios del agente. Este mecanismo no solo promueve la responsabilidad del AAD, sino que también aporta claridad y confianza a los participantes.

La integración técnica entre IA y blockchain requiere librerías o APIs que faciliten la comunicación bidireccional. La IA debe consultar datos on-chain con baja latencia, lo que implica la adopción de servicios como Infura o nodos locales configurados de forma óptima. Por otra parte, los resultados de la IA tienen que transformarse en transacciones firmadas por el propio agente, para lo que se requiere un módulo criptográfico de alta seguridad, ya sea implementado en enclaves de hardware o en contenedores con medidas de protección reforzadas. En conjunto, la incorporación de mecanismos de reputación IA, cifrado homomórfico, ZKPs y sharding temático permite construir AADs escalables y seguros. La auditoría algorítmica, sumada a la gobernanza descentralizada, refuerza la confianza en la toma de decisiones automatizada e impulsa nuevos modelos de colaboración y gobernanza en la economía digital.

6 Perspectiva Normativa y Ética

La creciente autonomía de los Agentes Autónomos Descentralizados conlleva desafíos significativos en el ámbito legal y ético, pues redefine la manera en que se asigna la responsabilidad, se salvaguarda la privacidad y se gestionan los sesgos inherentes a los modelos de inteligencia artificial.

Un primer punto de tensión surge en torno a la **asignación de responsabilidad** cuando las decisiones son tomadas por un AAD que, en sentido estricto, no se halla bajo el dominio de una autoridad central. Los marcos legales vigentes, concebidos para organizaciones con personalidad jurídica y jerarquías definidas, carecen de pautas claras para determinar quién responde en caso de que las acciones de un AAD ocasionen daños o vulneren derechos. Las DAO pueden implementar mecanismos internos de resolución de disputas y votaciones correctivas, pero el valor legal de tales resoluciones ante las jurisdicciones tradicionales es incierto [3]. En la práctica, se abre la posibilidad de que los tribunales requieran identificar a los desarrolladores o promotores de la DAO, extendiendo la responsabilidad a quienes hayan facilitado la puesta en marcha del agente autónomo. Esto contrasta con la intención de descentralizar la gobernanza y reducir la dependencia de figuras centrales, generando un vacío regulatorio que exige la adaptación de los marcos normativos.

Un segundo punto delicado está relacionado con la **protección de datos personales** y la privacidad. La inmutabilidad de la cadena de bloques desafía la aplicación de leyes como el Reglamento General de Protección de Datos (RGPD) en Europa, que establece derechos como el de supresión o rectificación de la información [6].

Si bien la cadena de bloques garantiza la trazabilidad y la verificación de registros, esta característica choca con la posibilidad de eliminar completamente un dato o revertirlo de manera forense. Para abordar estas exigencias, se investigan técnicas criptográficas de “borrado selectivo” o chameleon hashes, que permiten enmendar transacciones bajo consenso de la red, aunque su adopción masiva conlleva profundos cambios en el diseño de los protocolos y puede diluir, en cierta medida, el principio de inmutabilidad. Otra línea de desarrollo se centra en la disociación y el cifrado homomórfico, para que el contenido sensible nunca se exponga en texto plano, aunque estos métodos suelen incrementar la complejidad de la red y los costos de cómputo.

En cuanto a la **ética de la IA**, la naturaleza de los modelos de lenguaje de gran tamaño implica el riesgo de reproducir sesgos de los datos de entrenamiento, lo que puede derivar en discriminaciones o falsos positivos y perjudicar a ciertos sectores de la población. Los AADs que incorporen IA en su lógica decisoria enfrentan la necesidad de asegurar que dichos modelos cumplan criterios de justicia y equidad. Una manera de mitigar estos riesgos es la **auditoría algorítmica**, en la que se somete el comportamiento del modelo a evaluaciones continuas por parte de la comunidad o de equipos especializados. Sin embargo, la responsabilidad de corregir dichos sesgos recae en un colectivo difuso: los propietarios de tokens de gobernanza, los desarrolladores que implementan las actualizaciones y, en última instancia, la comunidad de nodos. Si el AAD asume funciones críticas —por ejemplo, supervisar transacciones financieras o decidir políticas de un fondo de inversión—, estos sesgos podrían tener consecuencias sociales y económicas de gran alcance.

La **transparencia** de las operaciones constituye otro factor clave en la legitimidad de los AADs. La capacidad de un agente para firmar transacciones y desencadenar acciones en la blockchain sin una supervisión humana explícita puede generar preocupación si su lógica interna se percibe como una “caja negra”. Para enfrentar esta inquietud, las DAO pueden exigir que el agente ofrezca descripciones mínimamente explicables de su razonamiento en los casos con mayor impacto. Herramientas de explicabilidad (por ejemplo, mecanismos de extracción de reglas o resúmenes de atención en modelos de lenguaje) permiten a los participantes comprender las motivaciones del AAD y, de ser necesario, responder con votaciones que limiten su autonomía o exijan reentrenamientos. No obstante, esta aspiración de transparencia debe equilibrarse con la protección de datos y la propiedad intelectual, pues exponer todos los detalles del modelo podría revelar información confidencial o estratégica.

Otro eje de tensión concierne a la **compatibilidad con los sistemas legales existentes**. Muchas jurisdicciones no reconocen la figura de un agente autónomo sin personalidad jurídica, lo que complica su participación en contratos legalmente vinculantes. La figura del “contrato inteligente” en sí misma plantea cuestiones acerca de la ejecutabilidad legal de los acuerdos, especialmente si las cláusulas se ejecutan de manera automática sin posibilidad de reclamación ante un órgano jurisdiccional.

Algunas iniciativas exploran la creación de organizaciones híbridas, donde la DAO se registra bajo formas legales tradicionales, lo que habilitaría una responsabilidad más clara en caso de disputas o demandas.

Finalmente, la **responsabilidad en la actualización y el mantenimiento** de estos agentes abre interrogantes relacionados con la continuidad de su desarrollo y la legitimidad de los cambios introducidos. Un AAD basado en IA no es estático: sus modelos pueden requerir reentrenamientos y/o contextualizaciones nuevas, ajustes de parámetros o cambios en el set de datos que se considera confiable. La DAO debe definir procesos transparentes para aprobar cada actualización y asegurar que los equipos técnicos cumplan con requisitos de calidad y validación. Si los procedimientos resultan ambiguos o demasiado restrictivos, se arriesga la obsolescencia del agente; en cambio, si se otorga un margen de maniobra excesivo, la gobernanza colectiva puede perder control en la gobernanza.

7 Conclusiones y Futuras Líneas de Investigación

La construcción de Agentes Autónomos Descentralizados con modelos de lenguaje de gran tamaño en redes blockchain configura un nuevo paradigma de gobernanza y automatización que trasciende la lógica habitual de sistemas centralizados. Al permitir la participación directa de la IA en la ejecución de contratos inteligentes y la toma de decisiones colectivas, se abren vías innovadoras para la gestión de recursos, la supervisión de procesos y la detección temprana de anomalías. Esta convergencia de IA y blockchain, no obstante, exige equilibrios técnicos, normativos y éticos que se han discutido a lo largo del presente documento.

Por un lado, la **optimización técnica** demanda tanto soluciones de escalabilidad (sharding temático, rollups, sidechains) como la adopción de técnicas criptográficas avanzadas (cifrado homomórfico, Zero-Knowledge Proofs) para proteger la privacidad y respaldar el procesamiento seguro de datos sensibles. A su vez, la construcción de módulos de reputación y mecanismos de auditoría algorítmica fortalece la confiabilidad y la transparencia de los AADs, evitando la dependencia de una autoridad central y promoviendo la participación de la comunidad en la validación de acciones y resultados.

En el plano **legal y ético**, la autonomía creciente de los AADs invita a redefinir los esquemas de responsabilidad y protección de derechos. Es esencial aclarar los alcances de la “personalidad virtual”, la vinculación legal de contratos inteligentes y el tratamiento de datos almacenados en una cadena inmutable. Asimismo, mitigar sesgos en la IA y asegurar la explicabilidad de las decisiones se vuelve crítico para la confianza y la aceptación social de estos sistemas, sobre todo cuando desempeñan funciones sensibles o de alto impacto.

De cara al futuro, se vislumbran múltiples líneas de investigación y desarrollo:

1. **Arquitecturas Multicadena y Aprendizaje Federado:** Profundizar en protocolos que conecten diversas redes orientadas a tareas específicas (por ejemplo, entrenamiento de IA, almacenamiento masivo, ejecución de contratos de gobernanza), para escalar las operaciones y reducir costos.
2. **Modelos de Gobernanza Reputacional:** Diseñar mecanismos de consenso dinámicos basados en indicadores de contribución y fiabilidad, calculados por la IA a partir de historiales de comportamiento de los nodos.
3. **Explicabilidad y Auditoría Algorítmica:** Consolidar metodologías para que los AADs ofrezcan descripciones claras de sus procesos de inferencia y razonamiento, en especial ante decisiones conflictivas, y posibilitar la participación activa de la comunidad en la corrección de sesgos.
4. **Identidad Digital y Protección de Datos:** Integrar soluciones como Self-Sovereign Identity (SSI) y técnicas de borrado selectivo para conciliar la inmutabilidad de la cadena con las demandas normativas y el derecho al olvido.
5. **Interacción con la Economía Real:** Evaluar el rol de los AADs en sectores como logística, finanzas tradicionales, seguros o administración pública, y estudiar cómo se adaptan los procedimientos legales para acoger a estos agentes en la práctica.

En síntesis, los AADs combinados con LLMs constituyen un avance significativo en la convergencia de la inteligencia artificial y la tecnología blockchain, al posibilitar la toma de decisiones autónoma bajo mecanismos de gobernanza colectiva y alta auditabilidad. La naturaleza descentralizada de estos entornos, unida a la creciente sofisticación de los agentes cognitivos, plantea retos y oportunidades que invitan a un debate multidisciplinario, donde confluyan la ingeniería, el derecho, la economía y la ética. Conforme se consoliden estas soluciones y se establezcan los marcos regulatorios apropiados, es probable que los AADs desempeñen un papel creciente en la configuración de la próxima generación de sistemas de gestión y organización social.

Referencias

1. Buterin, V. (2014). Ethereum White Paper. Recuperado de <https://ethereum.org/en/whitepaper/>
2. Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Recuperado de <https://bitcoin.org/bitcoin.pdf>
3. Tapscott, D., & Tapscott, A. (2016). Blockchain Revolution: How the Technology Behind Bitcoin and Other Cryptocurrencies is Changing the World. Penguin.
4. Wang, S., Ouyang, L., Yuan, Y., Ni, X., Han, X., & Wang, F. Y. (2019). Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(11), 2266–2277.
5. Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper*, 1–32.
6. Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). Blockchain Technology Overview. National Institute of Standards and Technology (NIST).
7. Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. *2015 IEEE Security and Privacy Workshops*, 180–184.