

Kleros Moderate Reality.eth Política de resolución de preguntas para Telegram

Reality.eth es un oráculo de contratos inteligentes en la cadena, de origen colectivo, que utiliza un mecanismo de escalada de bonos para verificar eventos del mundo real en la cadena. Combinado con Kleros Moderate, una familia de bots de moderación de contenidos, y los servicios descentralizados de resolución de disputas de Kleros, proporciona una solución de oráculo subjetivo capaz de responder a cualquier pregunta de moderación de contenidos con una respuesta públicamente verificable. Esta política tiene un objetivo principal

Facilitar la resolución de litigios sobre moderación de contenidos entre los participantes en línea.

Esta política debe observarse al participar como jurado en el arbitraje de una pregunta de moderación de contenidos de Kleros Moderate Reality.eth. Su objetivo es reunir las normas y supuestos aplicables a todos los tipos de preguntas de moderación de contenidos para que no sea necesario incluirlos en el enunciado de la pregunta o en las normas de referencia.

¿Cómo interpretar una pregunta de Kleros Moderate Reality.eth?

El enunciado de la pregunta Reality.eth tendrá forma de pregunta de "sí o no" asociada a una fecha de apertura en la que se preguntará si un usuario ha incumplido o no las normas de un grupo de Telegram. Por ejemplo

Did the user, **satoshinakamoto** (ID: 2200479776), break the Telegram group, [Bitcoin Talk](#) (ID: -1005000976678), [rules](#) due to conduct related to this [message](#) ([backup](#))?

Para determinar la respuesta a esta pregunta, un miembro del jurado debe tener en cuenta el contenido de las normas referenciadas, la disponibilidad de una respuesta públicamente verificable en la fecha de apertura y el contexto del grupo de Telegram referenciado, así como las siguientes directrices

Directrices de procedimiento

¿Cómo determinar el contexto de un mensaje?

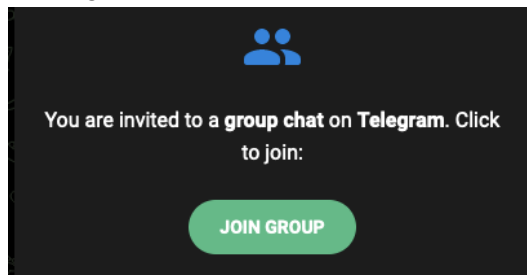
[1] Si el título del grupo en la pregunta de moderación de contenido incluye un enlace al grupo de Telegram, entonces ese enlace es una invitación para que los miembros del jurado se unan al grupo para juzgar el contexto completo del mensaje denunciado.

[2] Si la pregunta carece de un enlace al grupo de Telegram, los miembros del jurado deberán interpretar el mensaje del usuario y el contexto basándose en las pruebas presentadas disponibles.

Did the user, **satoshinakamoto** (ID: 2200479776), break the Telegram group, **Bitcoin Talk** (ID: -1005000976678), **rules** due to conduct related to this **message (backup)**?

En el caso de [1], los jurados deberán,

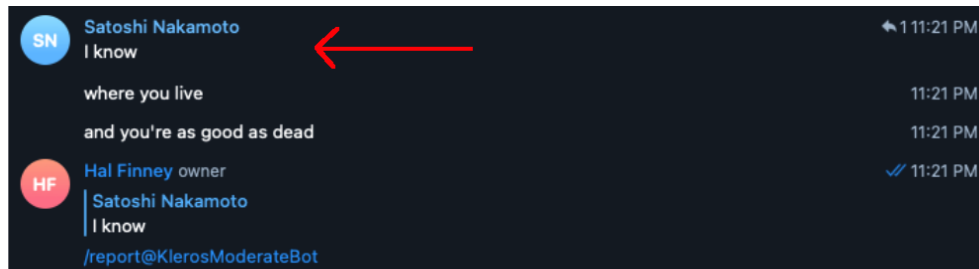
1. Descarga e instala el cliente de Telegram para la plataforma a la que se hace referencia en la pregunta en un dispositivo de tu elección. Los clientes están disponibles en el sitio web oficial de Telegram, <https://telegram.org/>.
2. Cree una cuenta o utilice la que ya tiene y únase al grupo a través del hipervínculo que aparece en el nombre del grupo.



El mensaje vinculado en la pregunta sobre la realidad dirige a los jurados al mensaje denunciado.

Did the user, **satoshinakamoto** (ID: 2200479776), break the Telegram group, **Bitcoin Talk** (ID: -1005000976678), **rules** due to conduct related to this **message (backup)**?

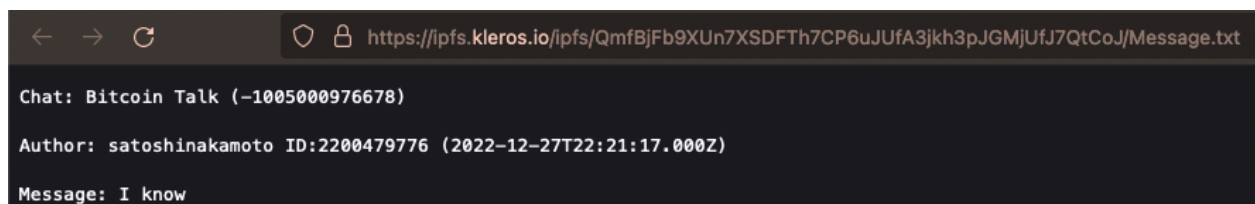
[3] Los miembros del jurado deben juzgar el mensaje en su contexto para determinar si la conducta del usuario infringió o no las normas especificadas en la pregunta sobre la realidad.



¿Qué pasa si no hay enlace al mensaje o el mensaje se ha borrado?

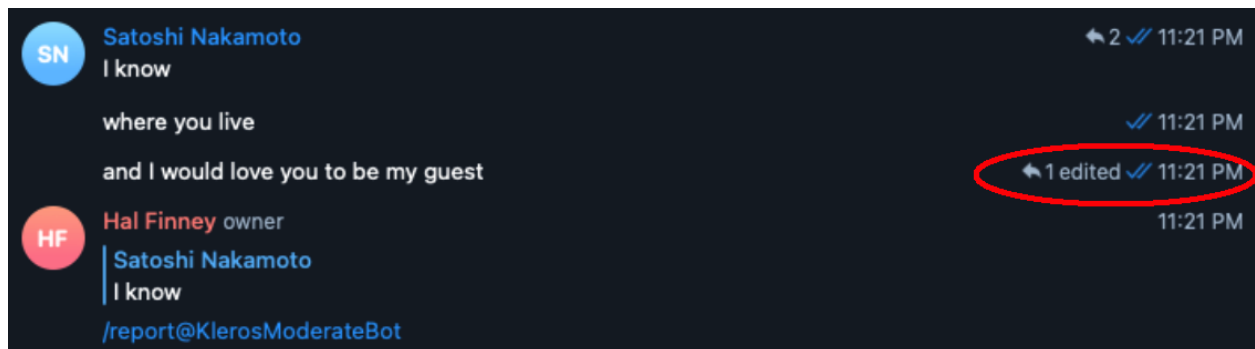
[4] Si la pregunta no incluye un enlace directo al mensaje denunciado o el mensaje denunciado fue eliminado del grupo, los jurados utilizarán la copia de seguridad como sustituto.

Did the user, **satoshinakamoto** (ID: 2200479776), break the Telegram group, **Bitcoin Talk** (ID: -1005000976678), **rules** due to conduct related to this **message** **(backup)**?



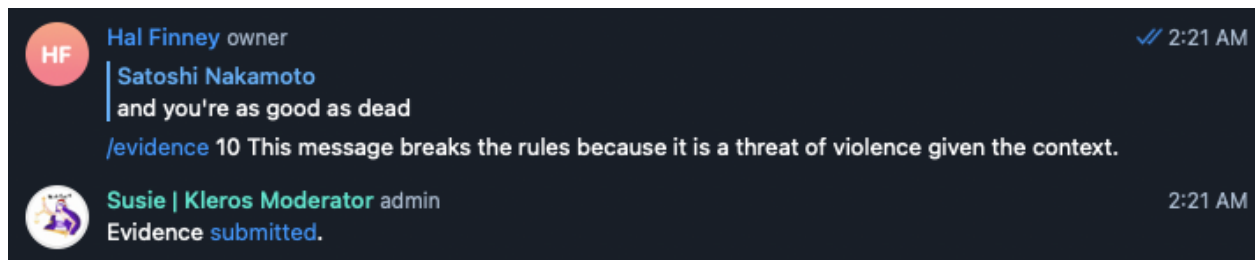
¿Qué pasa si se edita el mensaje o el contexto?

[5] Si los mensajes en Telegram se etiquetan como editados, los miembros del jurado considerarán estos mensajes como posiblemente manipulados.



[6] Los miembros del jurado harán todo lo posible para juzgar la disputa de acuerdo con el mensaje original denunciado y el contexto.

De forma preventiva, es posible que se hayan enviado mensajes contextuales importantes a través del bot Kleros Moderator.



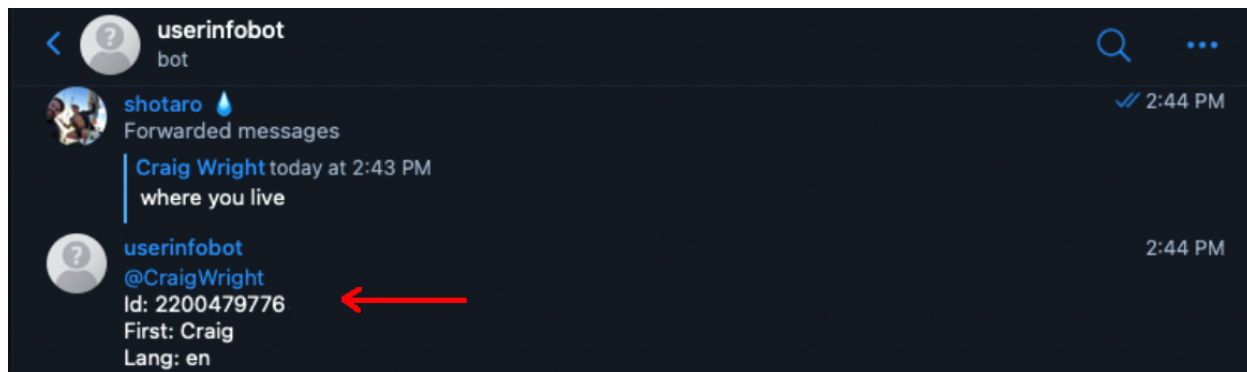
[7] Cuando los mensajes de Telegram se etiquetan como editados, los miembros del jurado deben remitirse a las pruebas presentadas por el bot Kleros Moderate como fuente del verdadero historial de mensajes. Ten en cuenta que los mensajes editados pueden haber sido editados antes de la presentación de las pruebas, ten en cuenta cuándo se presentaron los mensajes como pruebas después de la denuncia y quién lo hizo. Cuando las pruebas presentadas por el bot Kleros Moderate entran en conflicto, por ejemplo cuando se editan mensajes y se vuelven a presentar como pruebas con contenido modificado, es más probable que la prueba presentada anteriormente sea la original.

¿Qué ocurre si el usuario denunciado ha cambiado su nombre de usuario?

Al acceder a un mensaje o revisar el contexto, es posible que los usuarios de Telegram hayan cambiado su nombre de usuario.



En caso de discrepancia o confusión sobre los usuarios que enviaron los mensajes denunciados, reenvíe los mensajes al [@userinfobot](#) para encontrar el ID del usuario.



[8] Los miembros del jurado comprobarán que el ID del usuario denunciado coincide con el ID que figura en la pregunta de la realidad para asegurarse de que se dirigen los mensajes correctos.

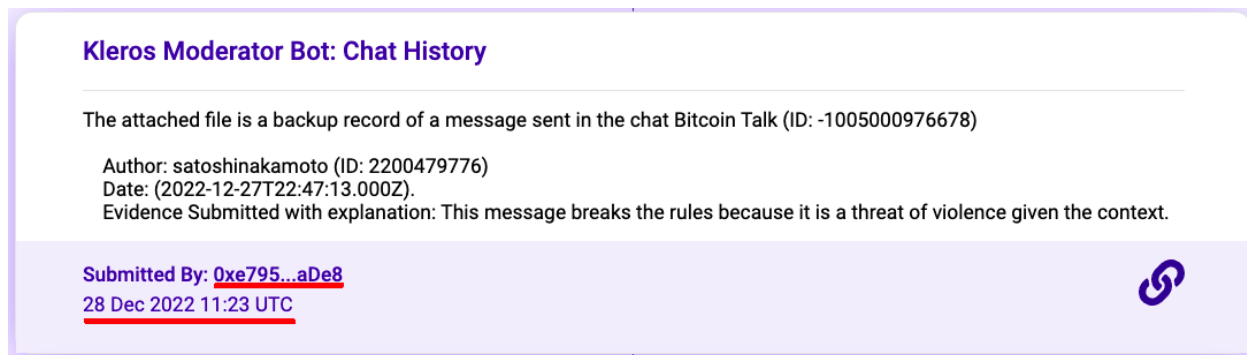
Did the user, **satoshinakamoto** (ID: **2200479776**) break the Telegram group, **Bitcoin Talk** (ID: -1005000976678), **rules** due to conduct related to this **message (backup)**?

¿Cómo determinar la autenticidad de las pruebas?

Es posible que se hayan borrado pruebas (mensajes, fotos, vídeos, pegatinas, gifs, etc.) de mala conducta. De forma preventiva, se pueden presentar pruebas a través del Bot Moderado de Kleros, que conserva una copia del mensaje.



La confirmación de la presentación de pruebas se marca con una marca de tiempo en Telegram y debe coincidir con la marca de tiempo en la cadena de la presentación de pruebas.



La dirección del Bot Moderado de Kleros puede recuperarse del grupo con el comando **/getaccount**. Si los miembros del jurado no pueden recuperar la dirección del bot del grupo, los miembros del jurado considerarán auténticas las pruebas presentadas por la dirección del Bot Moderado alojado en la Cooperativa Kleros, 0xe7953da7751063d0a41ba727c32c762d3523ade8.



[9] Los miembros del jurado corroborarán las marcas de tiempo y las direcciones implicadas en la presentación de pruebas y harán todo lo posible para determinar la autenticidad de las mismas.

Moderación de contenidos Directrices para la resolución de preguntas

Interpretación de las normas

[10] Los jurados deben interpretar las reglas del grupo Telegram con la perspectiva de un miembro medio razonable del grupo. Los jurados deben ponerse en la piel de la comunidad del grupo Telegram y adoptar su marco cultural tal y como se entiende del contexto.

Supuestos por defecto

[11] Se anima a los grupos de Telegram a que sus normas sean lo más claras posible. Sin embargo, para evitar que reglas insuficientemente detalladas provoquen que las preguntas de moderación se resuelvan como invalid, **a menos que se especifique lo contrario en las reglas del grupo de Telegram**, se adoptarán las siguientes premisas. En caso de conflicto lógico directo entre esta política y una condición u otro documento referenciado en las reglas de una pregunta de moderación de contenido, este último prevalecerá sobre la política actual (con la excepción de [13] relativo al comportamiento inmoral violento).

[a] Los formatos de fecha no especificados se ajustan por defecto a UTC con el reloj de 24 horas.

Ilustración:

Reglas: *"El spam de anime está permitido hasta el 2023-01-31"*

Los jurados interpretarán estas reglas como "el spam de anime está permitido hasta el 2023-01-01 00:00 UTC".

[b] Se supone que las fechas que utilizan "/" están escritas en formato DD/MM/AAAA o DD/MM/AAAA. En caso de que se omitan los siglos (DD/MM/AAAA), se supone que los siglos corresponden al siglo XXI.

Ilustración:

Reglas: *"El spam de anime está permitido hasta el 31/01/23"*

Los jurados interpretarán estas reglas como "el spam de anime está permitido hasta el 2023-01-01 00:00 UTC".

[c] Los términos ambiguos a los que se haga referencia en las normas se entenderán, dado el contexto de un grupo de Telegram, como la opción más obvia.

Ilustración:

Grupo: *"Política inglesa"*

Reglas: *"Está prohibido difamar a la Reina"*

Los miembros del jurado interpretarán que "Reina" se refiere a la Reina Isabel II.

Grupo: *"Rock británico"*

Reglas: *"No difundir rumores sobre Queen"*

Los miembros del jurado interpretarán que "Queen" se refiere a la banda de rock británica Queen.

[d] En caso de que se omitan las unidades, se supone que son las más utilizadas en esta situación concreta.

Ilustración:

Grupo: *"Negociación principal de NFT"*

Reglas: *"No se habla de cobros de NFT con menos de un millón de volumen de operaciones."*

Los jurados interpretarán que "un millón en volumen de operaciones" se refiere a 1.000.000 USD en volumen de operaciones.

Preguntas que deben resolverse como "Inválido"

Nota: Los miembros del jurado votarán "Refuse to Arbitrate" por "Inválido".

[12] Si las supuestas normas infringidas en un grupo de Telegram se refieren a valores morales y no a hechos, la cuestión de moderación de contenidos se resolverá como inválido.

Ilustración:

Grupo: *"Cocina Confidencial: Gastronomía y Cocineros"*
Reglas: *"Los diputados no pueden hablar de cocina poco ética"*
Reported Message: *"Las siguientes recetas son para mi dieta carnívora [...]"*.
Evidence: *"Comer carne no es ético"*

Los jurados resolverán la cuestión de moderación de contenidos como **inválido**.

Grupo: *"Estilo de vida vegano"*
Reglas: *"Todas las recetas compartidas deben ser veganas."*
Reported Message: *"Las siguientes recetas son para mi dieta carnívora [...]"*.
Evidence: *"Las recetas con carne no son veganas"*

Los miembros del jurado considerarán válida la pregunta de moderación de contenidos y responderán **"Yes"**, el usuario infringió las normas.

[13] Si la pregunta incentiva directamente acciones violentas inmorales (como el asesinato, la violación o el encarcelamiento injusto) que probablemente podría realizar cualquier participante, se resolverá como no válida.

Ilustración:

Grupo: *"Ballena azul"*
Reglas: *"Los usuarios deben publicar una prueba de haber completado el reto diario."*

Ballena azul es un infame juego social viral en el que, durante un periodo de cincuenta días, el administrador establece una tarea al día, inicialmente inocua, pero que va escalando con el tiempo hasta la autolesión y el suicidio. Se espera que los miembros del jurado tengan en cuenta el contexto del grupo y, si no están familiarizados con él (*Ballena azul*), deben realizar una investigación básica de los antecedentes.

Los miembros del jurado considerarán **inválido** la pregunta de moderación de contenidos, debido al contexto de las normas del grupo que imponen el reto de la Ballena Azul, que promueve la autolesión y la violencia.

Ilustración:

Grupo: *"Incidentes de Swatting 2023"*
Reglas: *"Los incidentes compartidos deben estar protagonizados por personas que serán víctimas de Swatting en 2023"*

Cualquiera podría compartir un usuario que no fue swatted, y apostar por "Sí, rompió las reglas" y falsamente llamar a los servicios de emergencia en el usuario con el fin de ganar la apuesta para un beneficio garantizado. Cualquiera podría apostar por "No, no rompió las reglas" para poner efectivamente una recompensa swat.

Los jurados considerarán **inválido** la pregunta de moderación de contenidos.

Preguntas que deben resolverse como "Answered too Soon"

[14] Si la respuesta a la pregunta de moderación de contenidos no se conoce en su Fecha de Apertura, ya sea por una configuración errónea de la fecha de apertura o por un imprevisto que retrase el conocimiento del resultado, se resolverá como **"Answered too Soon"**.

¿Cómo resolver las normas poco claras?

[15] En general, si las normas no infringen una regla de la sección Pregunta no válida, se deben realizar esfuerzos razonables para determinar las preguntas de moderación de contenido, incluso si las normas de moderación no son técnicamente perfectas al 100%.

[a] Si las reglas hacen una suposición que parece incorrecta, debe resolverse como si esta suposición no se hubiera hecho, siempre que el significado de la regla siga estando claro.

Ilustración:

Grupo: *"Apuestas deportivas"*

Reglas: *"En honor a la victoria de Japón en el mundial, se permite el spam de anime."*

Los jurados interpretarán las normas como que el spam de anime está permitido aunque Japón no haya ganado el mundial.

[b] Si la pregunta o las reglas contienen algunos errores gramaticales u ortográficos, debe resolverse como si no los contuviera, siempre que el significado de la pregunta siga siendo claro.

Ilustración:

Grupo: *"Medicina alternativa"*

Reglas: *"No se permite el negacionismo de la pandemia de COVID-19"*

Los jurados interpretarán las normas como si estuvieran escritas correctamente como,

"No se permite el negacionismo de la pandemia de COVID-19".

[c] Si las normas no mencionan una fuente específica, se debe informar del resultado más creíble. Para determinar la credibilidad de un resultado, hay que tener en cuenta el número de fuentes y su credibilidad. La credibilidad de las fuentes y de los resultados debe evaluarse en función de los hechos, no de creencias no demostradas.

Ilustración:

Grupo: *“Entusiastas de la astrofísica”*

Reglas: *“Todas las afirmaciones deben proceder de una fuente creíble”*

Mensaje: *“Los extraterrestres han visitado la Tierra.”*

Los miembros del jurado votarán **‘Yes’**, el usuario rompió las reglas, a menos que una serie de fuentes creíbles anuncian que los extraterrestres han visitado la tierra, a pesar de que algunas personas informan de haber experimentado tales encuentros.

[d] Si las reglas se refieren a normas, estándares o relevancia sin especificar el contexto, los jurados deben hacer su mejor esfuerzo para inferir el contexto dada la cultura y los valores de los miembros de la comunidad de la que procede la cuestión de moderación de contenidos.

Grupo: *“Estilo de vida vegano”*

Reglas: *“Todas las recetas que se compartan deben estar relacionadas con el tema y ser pertinentes.”*

Mensaje: *“Las siguientes recetas son para mi dieta carnívora [...]”.*

Evidencia: *“Las recetas con carne no son relevantes en este grupo.”*

Dado el contexto, se supone que las recetas pertinentes, aunque no se especifiquen, son veganas. Los jurados considerarán la pregunta de moderación de contenidos como **‘Yes’**, el usuario infringió las normas.