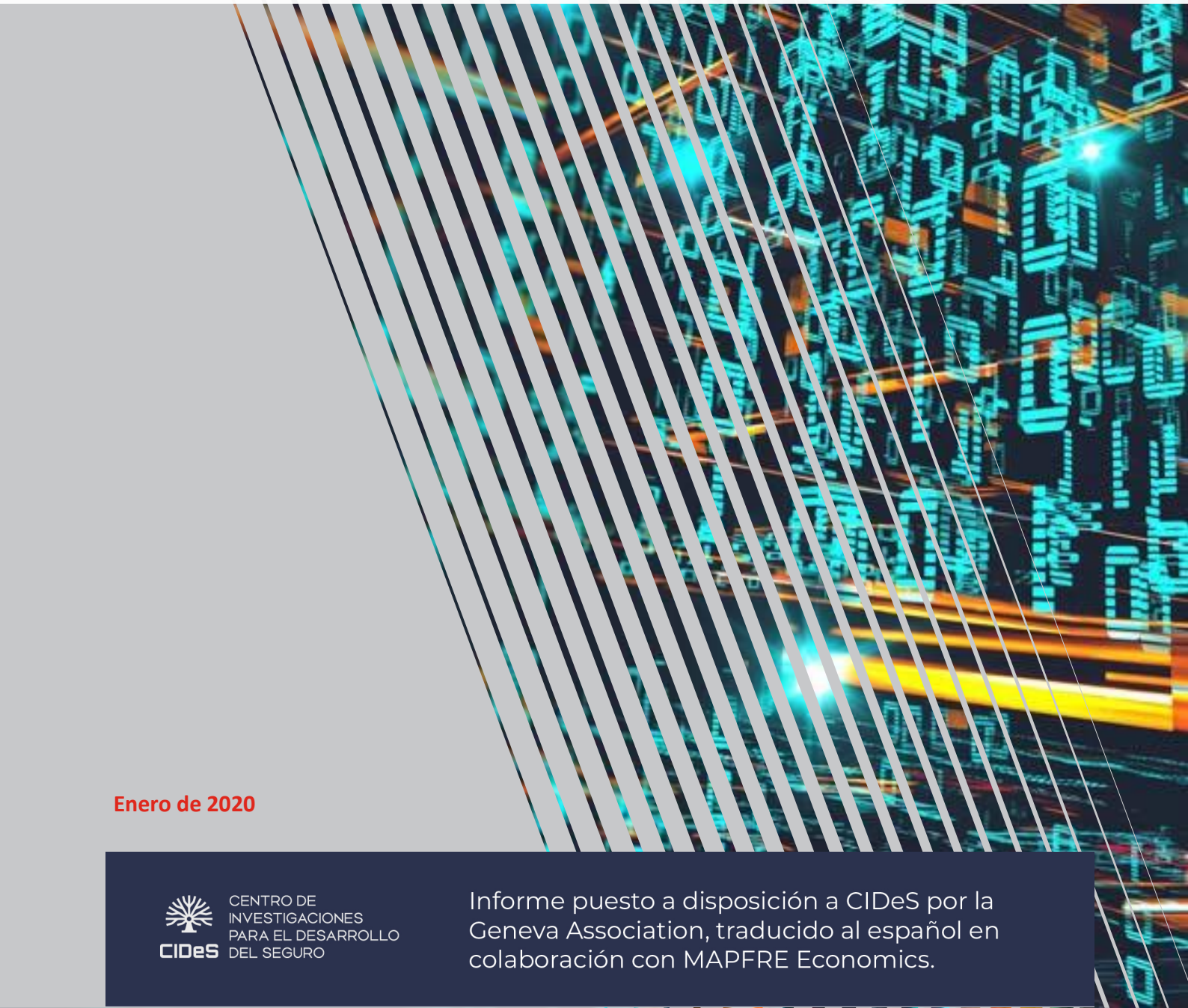
THE GENEVA ASSOCIATION

INSURANCE FOR A BETTER WORLD

Cómo promover la inteligencia artificial responsable en los seguros



Enero de 2020

CENTRO DE INVESTIGACIONES PARA EL DESARROLLO DEL SEGURO
CIDeS

Informe puesto a disposición a CIDeS por la Geneva Association, traducido al español en colaboración con MAPFRE Economics.

The Geneva Association

The Geneva Association se creó en 1973 y es la única asociación mundial de empresas de seguros; nuestros miembros son consejeros delegados (CEO) de aseguradoras y reaseguradoras. Sobre la base de una investigación exhaustiva realizada en colaboración con nuestros miembros, instituciones académicas y organizaciones multilaterales, nuestra misión es identificar e investigar las tendencias clave que posiblemente configuren o influyan en el sector asegurador en el futuro, con especial énfasis en lo que está en juego para el sector; ofrecer recomendaciones para el sector y para quienes elaboran políticas; crear un espacio de trabajo para nuestros miembros, los responsables públicos, el mundo académico y las organizaciones multilaterales y no gubernamentales para analizar estas tendencias y propuestas; y llegar a los líderes de opinión y a las instituciones influyentes para mostrar el valor del seguro en la comprensión de los riesgos y en la construcción de economías y sociedades más prósperas y, en consecuencia, de un mundo más sostenible.

The Geneva Association— Asociación internacional para el estudio de la economía de los seguros. Talstrasse 70, CH-8001 Zúrich
Correo electrónico: secretariat@genevaassociation.org | Tel: +41 44 200 49 00 | Fax : +41 44 200 49 99

Créditos de las fotografías:
Portada - Carlos Castilla, Shutterstock.

Enero de 2020
Cómo promover la inteligencia artificial responsable en los seguros

Publicado por The Geneva Association— Asociación internacional para el estudio de la economía de los seguros, Zúrich.

Cómo promover la inteligencia artificial responsable en los seguros

Benno Keller

Contenido

Agradecimientos

Esta publicación es un producto de la línea de trabajo sobre Nuevas tecnologías y datos de The Geneva Association, patrocinada por Thomas Buberl, director general del Grupo AXA. Estamos especialmente agradecidos a los miembros del grupo de trabajo creado para respaldar nuestras actividades de investigación sobre Nuevas tecnologías y datos, a saber: Paul DiPaola y Evan Hughes (AIG), Henning Schult (Allianz), Ashley Howell (Aviva), Patricia Plas, Chaouki Boutharouite y Dora Elamri (AXA), Atsushi Izu (Dai-ichi Life), Bruno Scaroni (Generali), Chris Reid (Intact Financial), Hiroki Hayashi (Nippon Life) y Lutz Wilhelmy (Swiss Re).
También queremos agradecer a los siguientes especialistas, que se prestaron a participar en un taller dedicado al tema de este informe: Ahed Abdelky (Allianz), Marcin Detyniecki y Sarah El Marjani (AXA), Alberto Branchesi (Generali), Bo Gong (Ping An Technology Research Institute) y Achraf Louitri (Intact Financial). Por último, nos gustaría agradecer a Michele Loi, de la Universidad de Zurich, por sus valiosas ideas y aportes.

Prólogo

La inteligencia artificial (IA) ya determina nuestra vida cotidiana, desde controlar las noticias y los anuncios que vemos en internet hasta seleccionar las rutas que tomamos con el GPS. Sectores enteros, como la educación, la medicina, el derecho y las finanzas, se ven influidos por ella.

En el sector de los seguros, la IA puede reducir costos, agilizar los procesos de tramitación de los siniestros y aumentar la eficacia de las interacciones con los clientes. Los productos como los seguros paramétricos basados en la IA pueden ayudar a asegurar a más personas. La IA también puede mejorar las capacidades de prevención, gestión y mitigación de riesgos y reducir problemas endémicos del sector, como el riesgo moral y el fraude.

Sin embargo, las ventajas de la IA se disiparán si los clientes no confían en que las aseguradoras usen la tecnología y los datos de forma responsable.

Este informe elaborado por The Geneva Association analiza los principios que consideramos más importantes para alcanzar los niveles necesarios de confianza de los clientes: transparencia, explicabilidad y equidad. Esperamos que este análisis, y nuestras correspondientes recomendaciones para las aseguradoras, contribuyan al uso responsable de la IA en los seguros y a la materialización de todos sus beneficios para la sociedad.

Jad Ariss
director general
The Geneva Association

1. Resumen ejecutivo

Desde el final del período conocido como el 'Invierno de la IA', comprendido entre los años setenta y fines de los noventa, caracterizado por contratiempos y decepciones, la inteligencia artificial (IA) ha logrado un avance extraordinario. Hoy en día, los sistemas de IA se utilizan comercialmente en un campo de aplicaciones cada vez más amplio.

Muchas aseguradoras están implementando sistemas inteligentes que automatizan tareas rutinarias o ayudan a la toma de decisiones humanas a lo largo de toda la cadena de valor de los seguros. Estos sistemas combinan nuevos tipos de algoritmos de aprendizaje con el análisis de datos procedentes de nuevos tipos de fuentes de datos, como los datos de los medios de comunicación en línea y los datos de la internet de las cosas (IoT) (The Geneva Association 2018).[1] En el futuro, los sistemas inteligentes tomarán decisiones normalizadas, de forma autónoma, en una cantidad cada vez mayor de ámbitos.

El uso de la IA en los seguros puede generar beneficios económicos y sociales que van más allá de las aseguradoras y sus clientes al mejorar la agrupación de los riesgos (*risk pooling*) y potenciar la reducción, mitigación y prevención de los riesgos.

Para promover la adopción de los sistemas de IA y aprovechar estas ventajas, las aseguradoras deben ganarse la confianza de sus clientes utilizando la nueva tecnología de forma responsable.

1 Por lo tanto, los sistemas inteligentes en sentido amplio son tecnologías que ayudan o sustituyen a la toma de decisiones humanas (Autoridad Monetaria de Singapur 2018 y BaFin 2018).

En los últimos años, se ha desarrollado un intenso debate sobre lo que implica el uso responsable de la IA, y en los últimos 12 a 24 meses han proliferado las directrices éticas publicadas por organizaciones gubernamentales y no gubernamentales. La cuestión sobre qué significa el uso responsable de la IA y cómo debe garantizarse es objeto de un debate permanente.

Un análisis de las directrices para el uso ético de la IA sugiere que existe una convergencia mundial sobre cinco principios básicos: (1) *transparencia y flexibilidad*, (2) *equidad,* (3) *seguridad*, (4) *responsabilidad y* (5) *privacidad* (Jobin et. al. 2019).[2] Sin embargo, existen diferencias importantes en la forma de interpretar estos principios, así como en los requisitos necesarios para su cumplimiento. Además, todavía existe una gran incertidumbre sobre cómo deben aplicarse las directrices y principios éticos en un contexto específico (Jobin et. al. 2019).

La respuesta definitiva a estas preguntas seguirá siendo incierta. Sin embargo, este informe pretende contribuir a identificar y analizar las principales compensaciones que surgen al aplicar los principios básicos de la IA responsable en los seguros. Por ejemplo, mejorar la interpretabilidad de los modelos complejos suele tener como contrapartida menores beneficios, y los modelos excesivamente complejos pueden tener beneficios adicionales limitados.

Aunque los cinco principios básicos de la IA responsable son fundamentales, este informe se centra en los dos principios que plantean cuestiones particularmente complejas en el ámbito de los seguros: 1) la transparencia y la explicabilidad y 2) la equidad.[3]

La aplicación del principio de equidad en el ámbito de los seguros requiere compensaciones que no suelen plantearse en otros sectores. Atenuar el sesgo y la discriminación es especialmente difícil en el sector de los seguros, donde existen normas de equidad diferentes y excluyentes entre sí.

Garantizar una evaluación equilibrada de los beneficios y riesgos de los usos específicos de la IA requiere una asignación clara de las funciones y responsabilidades dentro de una organización, así como conocimientos y experiencia en la ciencia de datos, la ciencia actuarial, la gestión de riesgos y la protección de datos.

Según nuestra investigación, el informe concluye con tres recomendaciones para que las aseguradoras promuevan el uso responsable de la IA en sus organizaciones: 1) establecer directrices y políticas internas, 2) adoptar una estructura de gobernanza adecuada para abordar los riesgos relacionados y 3) desarrollar e implementar programas de capacitación integrales para empleados y representantes. Aunque existen muchos modelos diferentes de gobernanza, las aseguradoras pueden basarse en los marcos actuariales y de gestión de riesgos existentes.

2 Jobin et. al. (2019) uso de los términos 'transparencia', 'justicia e imparcialidad', 'no maleficencia', 'responsabilidad y rendición de cuentas' y 'privacidad' para los cinco principios básicos.

3 El énfasis en la transparencia, la explicabilidad y la equidad no significa que los principios de seguridad, responsabilidad y privacidad tengan menos importancia. De hecho, la seguridad de los sistemas de IA y la preservación de la privacidad de los clientes pueden considerarse la base del uso correcto de la IA. Para tener acceso a un análisis de las cuestiones de privacidad que surgen con el análisis de los macrodatos (*big data*) en el sector de los seguros, consulte The Geneva Association 2018.

2. Beneficios de la IA en los seguros

Hoy en día, la IA puede realizar tareas especialmente útiles en los seguros. Por ejemplo, los avances en el procesamiento del lenguaje natural permiten a los sistemas inteligentes 'hablar' e interactuar con los seres humanos. Las aseguradoras utilizan cada vez más agentes conversacionales (por ejemplo, *chatbots*) que pueden identificar y responder a consultas complejas de los clientes y están disponibles 24 horas al día, los 7 días de la semana (consulte el recuadro 1).

Recuadro 1: Agentes conversacionales en los seguros

Los agentes conversacionales se utilizan en varias líneas de negocios y en diferentes partes de la cadena de valor. Permiten a los clientes interactuar con su aseguradora las 24 horas al día, los 7 días a la semana, a través del chat en línea, un canal preferido sobre todo por los clientes más jóvenes.

Mediante técnicas de aprendizaje avanzado para clasificar las entradas o *input* en un lenguaje natural, los agentes conversacionales pueden identificar y responder consultas complejas de los clientes, por ejemplo, relacionadas con los seguros de salud y de automóvil. Como resultado, pueden mejorar considerablemente la experiencia del cliente, así como aumentar la eficiencia de la aseguradora al procesar grandes cantidades de solicitudes de los clientes. Además, los agentes conversacionales permiten a las aseguradoras entablar nuevos tipos de comunicación con sus clientes.

Asimismo, pueden desempeñar distintos tipos de funciones, desde guiar a los usuarios hacia la información que necesitan y orientarlos en los trámites relacionados con los seguros (por ejemplo, presentar una reclamación) hasta procesar incluso las transacciones comerciales. No solo benefician a los clientes, sino también a los agentes de servicio, que pueden centrarse en tareas en las que el criterio humano es clave.

Para dar respuestas personalizadas a las consultas de los usuarios, los agentes conversacionales pueden utilizar información interna del producto, así como información no personal de terceros, por ejemplo, conocimientos relacionados con la salud.

Las aseguradoras han diseñado agentes conversacionales para garantizar la confiabilidad y la privacidad del usuario.

Además, los sistemas inteligentes pueden 'ver' y reconocer objetos en imágenes y extraer información relacionada. Este tipo de visión artificial permite a las aseguradoras automatizar tareas rutinarias manuales y cognitivas, por ejemplo, extraer datos de documentos escritos e imágenes para utilizarlos en el proceso de reclamaciones o suscripción de pólizas (consulte el recuadro 2).

Recuadro 2: La visión artificial en los seguros

Diversas aseguradoras utilizan la visión artificial para automatizar tareas rutinarias en la suscripción de pólizas y gestión de siniestros extrayendo información de documentos e imágenes mediante técnicas de aprendizaje automático y aprendizaje profundo.

Por ejemplo, la visión artificial se utiliza en documentos existentes para validar y verificar la información facilitada por el cliente durante el proceso de suscripción de pólizas. Esto incluye la verificación de las imágenes facilitadas por los clientes en las solicitudes de seguro de automóvil, por ejemplo, indicando el tipo de automóvil e identificando a los clientes mediante su matrícula. De este modo, la visión artificial puede ayudar a garantizar el nivel adecuado de cobertura, así como a detectar fraude en los seguros.

En las reclamaciones, la visión artificial se utiliza para validar la autenticidad de las imágenes suministradas por los clientes y extraer información de documentos, por ejemplo, formularios de accidentes, que sirven de base para clasificar las reclamaciones y automatizar los procesos de reclamación.

Por ejemplo, una aseguradora puede utilizar el aprendizaje automático y profundo para entender el tipo de documento y extraer información de documentación médica (incluidas facturas, recetas y otros documentos) fotografiados y presentados por los asegurados de planes de seguro médico. El sistema identifica el tratamiento médico y el diagnóstico, extrae todos los datos de la factura médica (importe, fecha, número de IVA, código fiscal, número de recibo) y, en cuestión de segundos, coteja la información con la cobertura de seguro correspondiente del asegurado.

La visión artificial también se utiliza en nuevos tipos de datos. Por ejemplo, una aseguradora utiliza el aprendizaje profundo para reconocer y medir el tamaño de las plantaciones de diferentes cultivos como trigo, maíz y arroz a partir de imágenes satelitales, y las utiliza como referencia para ofrecer seguros de cosechas. La visión artificial también se utiliza para mejorar la respuesta ante catástrofes identificando y localizando a los clientes afectados por desastres naturales.

Los sistemas inteligentes se distinguen por detectar patrones y correlaciones en datos complejos de un modo que es muy difícil, por no decir imposible, para el ser humano. Los patrones identificados son la base de tareas analíticas como la clasificación, la regresión y la agrupación, que desempeñan un papel importante en el modelo de negocio de los seguros. En comparación con los enfoques de modelización tradicionales utilizados en los seguros (como los modelos lineales generalizados), los sistemas inteligentes tienen el potencial de ofrecer predicciones mucho más acertadas, porque pueden aprender relaciones no lineales complejas entre las variables. Hoy en día, estas capacidades de los sistemas inteligentes se utilizan para ayudar a la toma de decisiones humanas (consulte el recuadro 3).

Recuadro 3: El uso de la IA para ayudar a tomar decisiones

La IA se utiliza para ayudar a los agentes de ventas a tomar decisiones y permitirles ofrecer al cliente un servicio más personalizado obteniendo información de los datos internos de los clientes, por ejemplo, datos sobre productos, siniestros y ubicación, así como de la interacción con los clientes. Los agentes reciben recomendaciones de oportunidades de venta cruzada y venta adicional o *upselling* y ofertas de productos relacionados, mientras que la decisión de dirigirse a los clientes con ofertas de productos específicos sigue siendo del agente. Estas herramientas han demostrado ser muy eficaces para aumentar la eficiencia del canal de ventas.

Una aseguradora puede utilizar el aprendizaje automático para predecir reclamaciones como base para determinar las tarifas adecuadas para clientes existentes y clientes nuevos en los seguros de automóvil.

Los algoritmos se aplican a los datos tradicionales utilizados en la suscripción de seguros de automóviles y facilitados por los clientes, incluidos el tipo y la marca del automóvil, el año y el historial de siniestros. En comparación con los modelos tradicionales de fijación de precios, los algoritmos mejorados aumentan considerablemente la precisión de las predicciones.

Las aseguradoras utilizan diferentes medidas para evitar el sesgo y garantizar la equidad en sus aplicaciones que ayudan a la toma de decisiones humanas (véase la sección Equidad).

Aunque los sistemas inteligentes todavía no son utilizados de forma generalizada por las aseguradoras para automatizar por completo la toma de decisiones, los avances en los algoritmos de aprendizaje podrían permitir, en el futuro, que estos sistemas automaticen la toma de decisiones estandarizada en cada vez más áreas con supervisión humana. Por ejemplo, la IA podría utilizarse para automatizar el enfoque de suscripción de seguros en áreas de riesgo estandarizadas y homogéneas (SCOR 2018). McKinsey prevé que la suscripción de seguros en forma manual se habrá erradicado para 2030, en relación con la mayoría de los productos para particulares y pequeñas empresas en seguros de vida, propiedad y accidentes (McKinsey 2018).

Los beneficios potenciales de esos usos de la IA van mucho más allá de las aseguradoras y sus clientes (véase la Figura 1). Por ejemplo, las aplicaciones de la IA pueden ayudar a ampliar la cobertura de los seguros a grupos de clientes nuevos y previamente no asegurados o infra-asegurados, o ampliar la variedad de riesgos cubiertos por el seguro. Al hacerlo, estos usos de la IA permiten ampliar el alcance de la agrupación de riesgos, que constituye el eje central de la función económica y social de los seguros. Al mismo tiempo, sin embargo, existe una gran preocupación dado que una mayor personalización de los seguros gracias a la IA pueda derivar en la exclusión de grupos específicos de clientes, por ejemplo, aquellos considerados de alto riesgo (véase la sección 'Equidad', pág. 12, y The Geneva Association 2018).

El uso de la IA también puede reducir el costo de la agrupación de riesgos automatizando tareas específicas, aumentando la precisión de las evaluaciones de riesgo o reduciendo el riesgo moral y la selección adversa. [4]

Además, el uso de la IA podría dar lugar a nuevos conocimientos sobre los riesgos que, a su vez, podrían ayudar a mitigarlos y prevenirlos. El uso de la IA podría promover la reducción del riesgo mediante una mejor correlación entre las primas y el riesgo. Además, disponer de mejores datos facilitaría el establecimiento de sistemas avanzados de gestión de riesgos y alerta temprana que permitan intervenir a tiempo para reducir las pérdidas. De este modo, el uso de la IA puede contribuir a ampliar la función de los seguros desde una mera protección frente al riesgo hacia la 'predicción y prevención' (The Geneva Association 2018).

Figura 1: Beneficios socioeconómicos de la IA para los seguros

Ampliar el alcance de la agrupación de riesgos

- Ampliar la cobertura de los seguros a segmentos de clientes nuevos y anteriormente no asegurados, facilitando el acceso a productos personalizados (por ejemplo, seguros de vida para personas con enfermedades preexistentes).
- Ampliar la variedad de riesgos para los que se dispone de cobertura de seguros mediante un mayor conocimiento de los riesgos (por ejemplo, los ciberriesgos).

Reducir el costo de la agrupación de riesgos

- Seguros más rentables gracias a la automatización de tareas específicas, mejores evaluaciones de los riesgos y reducción del riesgo moral y la selección adversa.

Mitigar y prevenir los riesgos

- Nuevos conocimientos sobre riesgos que ayudan a mitigarlos y prevenirlos
- Sistemas de alerta temprana que permitan reducir las pérdidas

Fuente: The Geneva Association

Sin embargo, los sistemas inteligentes plantean varios desafíos. Por ejemplo, necesitan grandes cantidades de datos para aprender, y su aprendizaje es tan bueno como los datos utilizados para entrenarlos, de modo que cualquier posible sesgo en los datos será aprendido por el sistema. [5] Además, debido a su complejidad, puede resultar difícil entender por qué un sistema ha tomado una determinada decisión, ya que estos sistemas son difíciles de interpretar. Al igual que los seres humanos, los sistemas inteligentes pueden cometer errores incluso cuando los datos no están sesgados. Los problemas conocidos en informática que pueden provocar errores son el sobreajuste (*overfitting*) o la maldición de la dimensionalidad, por ejemplo. [6] En tales situaciones, los patrones aprendidos a partir de los datos no pueden generalizarse.

La toma de decisiones humana no está exenta de sesgos y errores. Sin embargo, a diferencia de la toma de decisiones humana descentralizada, el uso de la IA para la toma de decisiones autónoma a gran escala supone que incluso un pequeño error sistemático puede tener consecuencias de gran alcance (Ruf et al. 2019a).

La preocupación por las consecuencias no deseadas y por el uso malintencionado de la IA ha desencadenado un intenso debate sobre su uso responsable. En la siguiente sección se analizan los principios clave para una IA responsable que han surgido de este debate y se identifican consideraciones importantes para la aplicación de esos principios en el sector de los seguros.

4 El riesgo moral se produce cuando las personas aumentan su exposición al riesgo porque saben que no asumen el costo de esos riesgos. La selección adversa se produce cuando las personas tienen mejor información que el asegurador sobre sus riesgos. En este caso, las personas de alto riesgo tienden a contratar seguros, mientras que los de bajo riesgo no lo hacen. Esto puede desencadenar una dinámica adversa en el mercado, con un aumento de las primas y una menor cobertura.

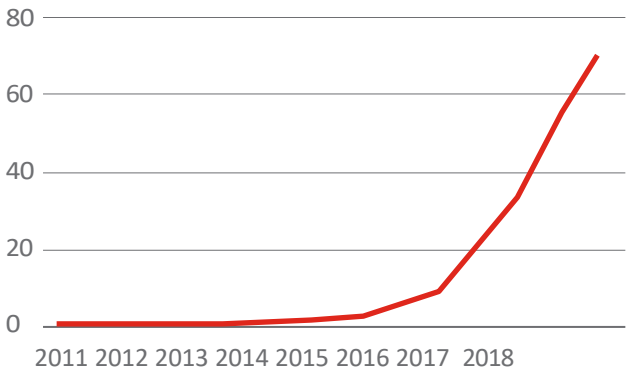
5 Un ejemplo ilustrativo de este efecto es Tay, un *chatbot* de Microsoft que rápidamente se volvió racista y sexista (Metz 2016).

6 Un ejemplo de sobreajuste fue el fracaso de Google Flu, un sistema para predecir síntomas de gripe basado en las búsquedas de las personas en Google (Lazer y Kennedy 2015).

3. La IA responsable en los seguros

En los últimos dos años han proliferado las directrices sobre la IA responsable publicadas por gobiernos, organizaciones internacionales, autoridades regulatorias, instituciones académicas, organismos industriales y empresas (véase la Figura 2). AlgorithmWatch ha clasificado más de 80 directrices de este tipo en su inventario mundial de directrices éticas sobre el uso de la IA.[7]

Figura 2: Cantidad de directrices éticas sobre la IA



Fuente: Jobin et. al. 2019

Aproximadamente un tercio de estas directrices se han publicado en los EE. UU., otro tercio en la UE (excepto el Reino Unido), y alrededor de una quinta parte en el Reino Unido y Asia (sobre todo en Japón).

Como compromisos voluntarios, dichas directrices postulan principios generales para el uso responsable de la IA.[8] Aunque existen diferencias sustanciales entre las directrices, parece existir una convergencia global hacia cinco principios básicos: seguridad, responsabilidad, privacidad, transparencia y equidad (Jobin et. al. 2019).[9]

Las directrices existentes no suelen analizar cómo abordar las compensaciones que surgen cuando estos principios se aplican en la práctica. [10] También cabe mencionar que las directrices publicadas se centran principalmente en cómo preservar los valores clave, más que en cómo la IA podría contribuir al desarrollo de estos valores (Jobin et. al. 2019).

7 https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/.

8 Las directrices también varían considerablemente según el grado de especificación. Por ejemplo, los Principios de Inteligencia Artificial de Microsoft consta de 51 palabras, mientras que las directrices de la Asociación de Normas IEEE ocupan más de 260 páginas.

9 Jobin et. al. (2019) uso de los términos 'transparencia', 'justicia y equidad', 'no maleficencia', 'responsabilidad y rendición de cuentas' y 'privacidad' para los cinco principios básicos.

10 Excepto las directrices publicadas por el Grupo de especialistas de alto nivel de la UE (Comisión Europea 2019), las directrices en general no reconocen la existencia de compensaciones al aplicar los principios para una IA responsable.

En el sector de los seguros, estos principios básicos no son nuevos y hace tiempo que desempeñan un papel importante. No hace falta decir que los productos de seguros deben ser seguros, deben proteger los datos de los clientes y las aseguradoras deben rendir cuentas a sus clientes. De hecho, varias leyes y regulaciones —incluidas las leyes de seguros, las leyes de protección de datos y privacidad, las leyes antidiscriminación y los requisitos de supervisión —regulan el comportamiento equitativo, transparente y responsable de las aseguradoras, así como la protección de la privacidad.

Sin embargo, el uso de la IA plantea algunas cuestiones complicadas. ¿Qué compensaciones surgen de la aplicación de los principios básicos? ¿Cómo pueden las aseguradoras promover y demostrar el cumplimiento de tales principios, y qué cambios, si existiesen, son necesarios para los mecanismos de gobernanza existentes y marcos de gestión de riesgos para este fin? Los reguladores de seguros también se plantean cada vez más este tipo de preguntas (véase, por ejemplo, BaFin 2018).

A continuación, se analizan en mayor detalle el principio de transparencia y explicabilidad y el principio de equidad, sobre la base de un análisis de ocho directrices publicadas por diversos agentes gubernamentales y no gubernamentales (consulte el Recuadro 4). El énfasis en la transparencia, la explicabilidad y la equidad no significa que la seguridad, la responsabilidad y la privacidad tengan menor importancia. Sin embargo, como veremos más adelante, el principio de transparencia y explicabilidad, y el principio de equidad plantean cuestiones específicas de los seguros. [11]

Recuadro 4: Ocho directrices

- **Directrices éticas para una IA digna de confianza** del Grupo de Especialistas de Alto Nivel sobre IA creado por la Comisión Europea (Comisión Europea 2019)
- **Ética cotidiana para la inteligencia artificial** de IBM (IBM 2019)
- **Diseño alineado desde el punto de vista ético** por el Instituto de Ingenieros Eléctricos y Electrónicos (IEEE 2019)
- **Principios de aprendizaje automático responsable** por el Instituto para el Aprendizaje Automático Ético (The Institute for Ethical ML 2019)
- **Principios de la IA** de Microsoft (Microsoft 2019)
- **Recomendación del Consejo de la OCDE sobre inteligencia artificial**, adoptada el 22 de mayo de 2019 (OCDE 2019)
- **Principios para promover la equidad, la ética, la rendición de cuentas y la transparencia (FEAT, por sus siglas en inglés) en el uso de la inteligencia artificial y el análisis de datos en el sector financiero de Singapur** por la Autoridad Monetaria de Singapur (Autoridad Monetaria de Singapur 2018)
- **Cómo evitar resultados discriminatorios en el aprendizaje automático** por el Foro Económico Mundial (Foro Económico Mundial 2018)

3.1. Transparencia y explicabilidad

La transparencia y la explicabilidad es un principio clave y fundamental en todas las directrices analizadas. [12] Se consideran importantes para generar confianza entre los clientes y otras partes interesadas. Las directrices suelen exigir que se capacite a las personas para entender las razones que subyacen a las decisiones y las consecuencias que las afectan. Algunas directrices también mencionan la importancia de la transparencia y la explicabilidad para que las personas puedan interponer recursos contra las decisiones que los afectan (Comisión Europea 2019). [13]

Ofrecer una explicación es especialmente importante cuando una decisión tiene un impacto significativo en la persona afectada. Por lo tanto, el nivel de explicabilidad depende, en gran medida, del contexto y de la gravedad de las consecuencias cuando un resultado es erróneo o inexacto (Comisión Europea 2019). La interpretación de los resultados algorítmicos —entendida como la 'capacidad de explicar o proporcionar el significado en términos entendibles para una persona' (Doshi-Velez y Kim 2017)— no solo es importante para brindar una explicación significativa a las personas afectadas, sino que también es indispensable para evaluar el rendimiento de los sistemas de IA y para su mejora continua, y por ende para una ciencia de datos fundamentada. Además, la interpretabilidad de los resultados algorítmicos ayuda a confiar en la capacidad del sistema para hacer predicciones precisas. Por ello, las aseguradoras deben esforzarse por mejorar la interpretabilidad de sus sistemas de IA, sobre todo si éstos tienen un impacto significativo en las personas.

Las directrices publicadas no especifican en detalle qué debe revelarse a quién, y existen diferencias sustanciales entre las directrices con respecto a lo que implica la transparencia. En general, sin embargo, cabe distinguir entre las revelaciones *a priori* y las explicaciones a posteriori.

11 La seguridad de los sistemas de IA y la preservación de la privacidad de los clientes pueden considerarse la base de un uso equitativo de la IA. Para un análisis de las cuestiones de privacidad que surgen con el uso de la analítica de *big data* en los seguros, véase The Geneva Association 2018.

12 Las directrices también utilizan los términos 'interpretabilidad', 'explicabilidad' o 'entendimiento'.

13 Véase la sección "Equidad", p. 12.

Todas las directrices analizadas sugieren algún tipo de información *a priori* para los usuarios. Por ejemplo, varias directrices hacen hincapié en que las personas deben ser conscientes de que están interactuando con un sistema de IA como un chatbot (por ejemplo, IBM 2019, OCDE 2019, Autoridad Monetaria de Singapur 2018). Varias directrices exigen que el uso de la IA en el proceso de toma de decisiones debe ser notificado a los usuarios (por ejemplo, OCDE 2019, Autoridad Monetaria de Singapur 2018 y Foro Económico Mundial 2018). Para promover la confianza de los usuarios, existen directrices que sugieren que estas divulgaciones deben incluir una descripción de las capacidades y el propósito de los sistemas de IA (Comisión Europea 2019) o que las organizaciones se comprometan a promover un entendimiento general de los sistemas de IA (OCDE 2019). Asimismo, las directrices analizadas exigen que las decisiones tomadas por los sistemas de IA sean explicadas a los usuarios con términos sencillos (OCDE 2019, Foro Económico Mundial 2018).

Algunas directrices establecen que dichas explicaciones *a posteriori* deben suministrarse a los usuarios que las soliciten (Autoridad Monetaria de Singapur 2018). Para permitir que las personas entiendan las decisiones que las afectan, algunas directrices exigen que se explique la lógica involucrada en la toma de decisiones (OCDE 2019, Foro Económico Mundial 2018).[14] Esto puede incluir explicaciones claras sobre qué datos se utilizan, cómo los datos afectan la decisión y las consecuencias de la decisión (Autoridad Monetaria de Singapur 2018, Comisión Europea 2019). Varias directrices insisten en que la información facilitada a las personas debe tener sentido para ellas.

Por ejemplo, brindar una explicación compleja del algoritmo casi no tiene sentido para las personas (Comisión Europea 2019). Para que tengan sentido, las explicaciones ofrecidas a los usuarios deben tener una finalidad clara. Wachter et al. 2017 identifica los siguientes objetivos de explicaciones significativas:

1. Informar y ayudar a la persona a entender por qué se ha tomado una determinada decisión.
2. Brindar los motivos para impugnar decisiones adversas
3. Entender qué podría modificarse para obtener un resultado deseado en el futuro.

Brindar explicaciones significativas es un desafío, ya que algunos algoritmos de 'caja negra' son complejos por naturaleza —el precio que se debe pagar por una mayor precisión— y, por lo tanto, son difíciles de interpretar y explicar. En estos sistemas no suele ser posible interpretar y explicar el papel de las distintas variables en general.[15]

En los últimos años, las ciencias de la computación han realizado esfuerzos considerables para superar los desafíos de interpretar y explicar los algoritmos de 'caja negra'. Por ejemplo, los enfoques de ingeniería inversa consisten en crear sustitutos algorítmicos interpretables[16] —una técnica reciente que se debe profundizar (Ruf et al. 2019a).

Los enfoques de diseño se centran en imponer ciertas restricciones a las predicciones (Guidotti et al. 2018 y Hall et al. 2017).[17] Cuando no sea posible dar una explicación sobre el papel de las distintas variables en una decisión individual, se pueden utilizar otros tipos de explicaciones (véase la Figura 3).

Figura 3: Opciones de explicaciones a posteriori en función de la importancia del impacto

Importancia del impacto en los usuarios

Tipo de explicación	Descripción	Ejemplo
Lógica general de la toma de decisiones	Facilitar un entendimiento cualitativo de la relación entre las variables de las entradas y la predicción del modelo.	"La prima del seguro de un automóvil se basa en el año, el tipo de vehículo y una calificación de riesgo determinada en función de la velocidad, la aceleración y la fuerza de frenado".
Lógica de la decisión individual	Ofrecer un entendimiento cualitativo de los factores clave que impulsaron la decisión.	"Su prima aumentó porque aumentó su calificación de riesgo basada en la velocidad, la aceleración y la fuerza de frenado"
Explicación contrafactual	Explicación de la forma "Si X no hubiese ocurrido, Y no se habría producido".	"Su seguro de hogar no habría sido denegado si usted hubiese instalado persianas contra tormentas"
Certificación / auditoría independiente	Certificación por un organismo independiente	"Nuestro modelo de calificación ha sido homologado como preciso y equitativo"
Divulgación al público de los resultados de las auditorías independientes de algoritmos	Divulgación completa del informe de auditoría independiente	

Fuente: The Geneva Association

14 En la UE, suministrar 'información significativa sobre la lógica implicada' en las decisiones automatizadas es un requisito legal (Reglamento General de Protección de Datos).

15 El Reglamento General de Protección de Datos (RGPD) de la UE reconoce estos desafíos al afirmar que no siempre es posible explicar por qué un modelo ha generado un determinado resultado o decisión (y qué combinación de los datos ingresados ha contribuido a ello). Según el RGPD, en esas circunstancias pueden exigirse otras medidas de explicabilidad (por ejemplo, trazabilidad, auditabilidad y comunicación transparente sobre las capacidades del sistema), siempre que el sistema en su conjunto respete los derechos fundamentales.

16 Los algorítmicos sustitutos pueden entenderse como modelos de aproximación que imitan el resultado del algoritmo y se pueden interpretar fácilmente.

17 Por ejemplo, los seres humanos entienden fácilmente las relaciones entre variables que solo cambian en una dirección (el riesgo aumenta con la edad, por ejemplo) (Hall et. al. 2017). Las técnicas tradicionales de modelización utilizadas en los seguros, como los modelos lineales generalizados, suelen producir este tipo de relaciones y, por lo tanto, son relativamente fáciles de interpretar. Por tanto, imponer restricciones a un sistema de IA para que produzca las denominadas relaciones monótonas puede mejorar su interpretabilidad a costa de reducir su precisión.

Por ejemplo, se han propuesto explicaciones contrafactuales para abordar los derechos de información de las personas en virtud del RGPD (Wachter et al. 2017). Un ejemplo de explicación contrafractual: 'Le denegaron un préstamo porque sus ingresos anuales eran de 30.000 libras. Si sus ingresos hubiesen sido de 45.000 libras, le habrían otorgado un préstamo'. Sin embargo, las explicaciones contrafactuales de las decisiones complejas de la 'caja negra' no siempre son confiables y deben cumplir con ciertos criterios estadísticos (Laugel et al. 2019). Por lo tanto, debe vigilarse la calidad de las explicaciones contrafactuales ex post.

Conclusiones y debate

La interpretabilidad de los resultados algorítmicos es importante para fortalecer la confianza y el entendimiento de los clientes.
Sin embargo, suele ser difícil lograr una interpretabilidad perfecta, y mejorar la interpretabilidad de los modelos complejos puede ir en detrimento de su precisión. En estos casos, el desafío consiste en cómo crear beneficios económicos sin afectar la confianza de los clientes.

Se debe promover la aplicación de modelos que puedan interpretarse, en particular si sus resultados tienen un impacto significativo en los clientes. Cuando se utiliza para seleccionar riesgos y fijar precios, la confianza en los sistemas de IA se puede promover utilizando fuentes de datos que estén relacionadas con el riesgo asegurado de forma tal que sea entendible para los clientes de forma intuitiva (Christen et al. 2019).

Cuando sea difícil explicar los resultados algorítmicos de manera sencilla para los consumidores, existen otras medidas que pueden promover la confianza de los clientes, por ejemplo, la trazabilidad, auditabilidad y comunicación transparente sobre las capacidades de un sistema (Comisión Europea 2019). Las ventajas de los modelos demasiado complejos no siempre justifican una reducción de la interpretabilidad. Las aseguradoras deben desarrollar y aplicar las respectivas directrices y políticas internas para garantizar un enfoque coherente de la transparencia y la explicabilidad de los resultados algorítmicos.

3.2. Equidad

La equidad es otro principio básico y de suma importancia en las directrices para una IA responsable. La equidad está asociada a muchos valores diferentes, como la libertad, la dignidad, la autonomía, la privacidad, la ausencia de discriminación, la igualdad y la diversidad. A menudo, estos valores se deben interpretar en su contexto, incluido el contexto cultural. Por lo tanto, es imposible establecer una norma universal de equidad.

En términos generales, cabe distinguir entre una dimensión procedimental y una dimensión sustantiva de la equidad (véase, por ejemplo, Comisión Europea 2019).

Proceso equitativo: la dimensión procedimental implica que los consumidores reciban un trato equitativo a lo largo de todo el proceso. Un aspecto importante del trato equitativo es la posibilidad de que los clientes impugnen y reclamen una reparación efectiva contra las decisiones que los afectan (véase, por ejemplo. Comisión Europea 2019). [18] En los seguros existen requisitos de conducta de mercado para garantizar un trato equitativo a los clientes, independientemente de la tecnología utilizada. [19]

Decisiones equitativas: las decisiones deben ser equitativas en el sentido de que no discriminen ni perjudiquen de manera injusta a personas o grupos de personas (OCDE 2019, Comisión Europea 2019, Autoridad Monetaria de Singapur 2018). Por ello, la mayoría de las directrices hacen hincapié en la ausencia o minimización de sesgos injustos y discriminación de las decisiones basadas en la IA como elemento clave de la equidad. Algunas directrices también mencionan que la distribución equitativa y justa tanto de los beneficios como de los costos es una característica de las decisiones justas (por ejemplo, Comisión Europea 2019).

Equidad en la informática

En informática, el sesgo se refiere a un error sistemático que coloca a ciertos grupos de manera sistemática en ventaja comparativa y a otros en desventaja. Los seres humanos no están libres del sesgo, y el uso de la IA puede, de hecho, mejorar la equidad de las decisiones en determinadas circunstancias. Sin embargo, dado que los algoritmos de aprendizaje automático tienen el potencial de desplegarse a gran escala, incluso un sesgo sistemático mínimo puede afectar a una gran cantidad de personas (Ruf et al. 2019a). Para los científicos de datos, el desafío consiste por ende en identificar, medir y mitigar los posibles sesgos que podrían poner en desventaja sistemática a determinados grupos.

El sesgo puede afectar la toma de decisiones algorítmicas mediante datos en varios niveles (Barocas y Selbst 2016, Ruf et al. 2019b). Por ejemplo, los datos utilizados en los algoritmos de capacitación pueden ser el resultado de una recopilación de datos sesgados. El sesgo también puede producirse cuando un algoritmo se utiliza con datos nuevos que son muy diferentes de los datos ingresados durante la formación. Por último, el sesgo puede deberse al criterio humano a la hora de etiquetar los datos de la formación.

Además, la discriminación también puede producirse con datos precisos e imparciales. Por ejemplo, aunque no se tenga en cuenta explícitamente un atributo protegido como el sexo, la etnia o aspectos similares, dichos atributos pueden intervenir en la toma de decisiones a través de *proxis* o de una combinación compleja de ellos que se correlacionen con este atributo (discriminación indirecta).

---

18 La dimensión procedimental de la equidad está, por lo tanto, relacionada de manera estrecha con el principio de transparencia y explicabilidad antes mencionado.

19 El Principio Básico de Seguros 19 (ICP19) de la Asociación Internacional de Supervisores de Seguros (IAIS) establece que 'el supervisor exige que las aseguradoras y los intermediarios, en el ejercicio de la actividad aseguradora, traten a los clientes con equidad, antes de celebrar un contrato y hasta el momento en que se haya cumplido con todas las obligaciones derivadas de éste'.

En informática, recientemente se han desarrollado distintos enfoques para mitigar el sesgo. Los enfoques que se centran en los datos que se introducen (input) se basan en un mejor muestreo de los datos, mientras que los enfoques que hacen hincapié en los datos que se generan (output) se basan en eliminar la discriminación en el resultado del algoritmo.

Para identificar, medir y atenuar la discriminación y el sesgo de los sistemas de IA, es necesario definir matemáticamente los conceptos de equidad. Hasta la fecha, no existe un consenso sobre la formulación matemática de la equidad (Fiedler et al. 2016), y existen muchas definiciones diferentes e incompatibles entre sí. [20] Sin embargo, las diferentes definiciones pueden producir resultados totalmente distintos (Bellamy et al.2018), y es imposible atender al mismo tiempo todas las definiciones de equidad (Kleinberg et al. 2017).

La equidad en los seguros

En el sector de los seguros, la forma de garantizar decisiones equitativas es especialmente intrincada y compleja en comparación con otros sectores. De hecho, como explicaremos más adelante, la ausencia de discriminación absoluta es imposible de conseguir, sobre todo en lo que respecta a la selección de riesgos y a las decisiones sobre los precios. Por el contrario, las aseguradoras deben elegir cómo discriminar (véase, por ejemplo, Lo y Christen 2019).

Minimizar la discriminación implica equilibrar las compensaciones entre los diferentes conceptos de equidad y plantea la cuestión de qué datos pueden utilizarse en el proceso de suscripción de pólizas y como base para seleccionar los riesgos y fijar las tarifas. Estas cuestiones se han debatido durante muchos años, y las aseguradoras están sujetas a requisitos de equidad sustantiva en virtud de la legislación vigente, que difieren según las jurisdicciones (consulte el recuadro 5). Estos requisitos de equidad suelen regular el tipo de información que una aseguradora puede utilizar en la toma de decisiones, por ejemplo, con respecto al uso de datos genéticos en seguros de vida y salud en algunas jurisdicciones (véase The Geneva Association 2017).

Los siguientes conceptos de equidad son particularmente relevantes en el ámbito de los seguros:

La equidad actuarial exige que los riesgos similares reciban un trato similar, de modo que la prima que paga una persona se corresponda con el riesgo real.

La ausencia de discriminación implica que la prima que paga una persona no se base en factores irrelevantes sobre los que la persona no puede influir, sobre todo si estos factores están relacionados con grupos socialmente protegidos. La ausencia de discriminación está relacionada con las nociones de equidad de grupo (el objetivo de que los grupos definidos por atributos protegidos como el género, la etnia, la orientación sexual, etc. reciban tratamientos o resultados similares) y las decisiones de trato desigual se basan (en parte) en un atributo sensible. La equidad de grupo es en gran medida coherente con la noción del impacto desigual, es decir, resultados que perjudican o benefician desproporcionadamente a personas con determinados atributos sensibles.

La *solidaridad o mutualización* se menciona a menudo como una característica clave de los seguros. La creciente individualización de los seguros, impulsada por los sistemas de IA y la analítica de datos, puede perjudicar a determinados grupos, por ejemplo, mediante el cobro de primas demasiado costosas o que se les niegue la cobertura (The Geneva Association 2018). Aunque estas consideraciones no son específicas del uso de los sistemas inteligentes, es posible que se acentúen.

A menudo no basta con eliminar los atributos sensibles de los datos para garantizar la ausencia de discriminación ('imparcialidad por desconocimiento'), ya que dichos atributos pueden detectarse fácilmente en *proxis* que se correlacionan con estos atributos (Pedreschi et al. 2008). La elección de la métrica adecuada para la ausencia de discriminación es particularmente compleja, ya que existen múltiples definiciones relacionadas que no pueden aplicarse al mismo tiempo. [21]

Recuadro 5: Equidad jurídica

La libertad de las aseguradoras para utilizar determinadas características en la selección de riesgos y la fijación de precios suele regirse por una combinación de legislación contra la discriminación, legislación sobre privacidad/protección de datos y legislación sobre seguros.

- Las leyes contra la discriminación prohíben el trato injusto de las personas en función de atributos sensibles en muchas jurisdicciones, por ejemplo, en EE. UU., las leyes federales prohíben la discriminación por motivos de raza, color, religión, nacionalidad, sexo, estado civil, edad y embarazo en muchas circunstancias. En la UE, la directiva de género prohíbe la desigualdad de trato por motivos de sexo.
- La legislación sobre la privacidad impone restricciones o prohíbe el procesamiento de determinados tipos de información sensible, por ejemplo, la relativa a datos genéticos.
- La legislación sobre seguros puede permitir el uso de parámetros sensibles si está justificado desde el punto de vista actuarial. Por ejemplo, en algunos países (incluido Estados Unidos), el género es un factor de suscripción admitido. La legislación sobre seguros también puede imponer restricciones a determinados usos de la información personal (por ejemplo, la prohibición de utilizar datos personales para optimizar precios en algunas jurisdicciones).

20 Narayanan 2018 proporciona 21 definiciones matemáticas de equidad extraídas de la literatura.

21 Entre las métricas concretas de ausencia de discriminación que se están debatiendo se incluyen la igualdad de probabilidades, la igualdad de oportunidades, la paridad demográfica (estadística) y la paridad de la tasa de predicción, entre otras (véase, por ejemplo, Ruf et al. 2019b).

Aunque hasta ahora no ha existido un debate generalizado sobre las posibles métricas que se deben utilizar para medir la asequibilidad y la exclusión, los reguladores sí imponen en algunos casos límites a la individualización de los seguros, por ejemplo, limitando la relación permitida entre la prima más alta y la más baja. La Asociación Holandesa de Seguros ha desarrollado un 'monitor de solidaridad' para evaluar la evolución del diferencial de las primas de seguros y la asegurabilidad individual a lo largo del tiempo.

La importancia relativa de los distintos conceptos de equidad varía según los distintos tipos de seguros y jurisdicciones. Por ejemplo, en muchas jurisdicciones, la solidaridad es considerada una característica esencial del seguro de salud.
Según la aplicación, se puede seleccionar un conjunto de métricas diferentes y supervisarlas en forma constante. Por ejemplo, con el objetivo de limitar los efectos adversos de los algoritmos de fijación de precios mejorados en los clientes, como el sesgo y las implicaciones para la asequibilidad, una aseguradora implementó un sistema integral de supervisión posterior. Los modelos son examinados por diversos equipos de diferentes áreas, y los resultados del sistema se someten a nueve pruebas diferentes para validar que los modelos se ajustan a las expectativas del equipo de modelización, los socios comerciales y los entes reguladores.
En los casos en que se requiere la aprobación reglamentaria de las tarifas, se ha dotado a los reguladores de herramientas específicas para hacer un seguimiento del rendimiento del sistema.

Conclusiones y debate

La equidad exige centrarse en reducir la discriminación y el sesgo. Por lo tanto, la imparcialidad está estrechamente relacionada con el principio de transparencia y explicabilidad, ya que hace hincapié en que las personas afectadas puedan entender y objetar los resultados algorítmicos. En el sector de los seguros, mitigar el sesgo y la discriminación es particularmente difícil, ya que existen conceptos y parámetros de la ausencia de discriminación diferentes y excluyentes entre sí.

Para supervisar y mitigar el sesgo es necesario cuantificar la equidad. Aunque los estudios académicos sobre los parámetros de equidad adecuados están en constante evolución y deben ser promovidos, las aseguradoras deben identificar definiciones de equidad específicas para cada uso de la IA. Dado que la equidad no es un concepto nuevo en el sector de los seguros, se deben aprovechar los marcos y las normas existentes (como la ética actuarial). Asimismo, se deben definir claramente las funciones y responsabilidades con respecto a la supervisión y mitigación del sesgo.

Dado que el sesgo puede influir en la toma de decisiones en distintas fases, es importante concientizar a los distintos niveles de gestión mediante programas educativos y de formación adecuados.

4. Recomendaciones

Para que los clientes de seguros aprovechen las ventajas potenciales de la IA, es fundamental que confíen en estos sistemas. El sector de los seguros puede facilitar esta confianza afirmando los principios básicos del uso responsable de la IA, como la seguridad y la privacidad, la transparencia y la explicabilidad, y la equidad, y promoviendo el uso responsable de la IA mediante las medidas que se describen a continuación.

1. Establecer directrices y políticas internas para el uso de la IA

Las directrices y políticas internas desempeñan un papel fundamental a la hora de aumentar la concientización sobre las compensaciones entre los beneficios y los riesgos en el uso de la IA en el sector de seguros. Por lo tanto, las aseguradoras deben desarrollar y adoptar directrices y políticas respectivas que incluyan principios para tratar las cuestiones de transparencia y explicabilidad e imparcialidad. En particular, las directrices deben ayudar a aclarar cómo deben evaluarse los beneficios y los riesgos del uso de la IA caso por caso.
Actuarios, gestores de riesgos, científicos de datos y responsables de la protección de datos deben trabajar en estrecha colaboración para desarrollar y aplicar tales directrices y políticas.

Al hacerlo, las aseguradoras pueden adoptar un enfoque centrado en los riesgos para la gobernanza de la IA, lo que supone prestar especial atención a los usos de los sistemas de IA que puedan tener un impacto significativo en las personas. La importancia del impacto se refiere a las consecuencias de las decisiones para las personas afectadas y depende de las circunstancias específicas en las que se utilice la IA.

Por ejemplo, es probable que los usos de la IA que automatizan tareas específicas, pero no cambian la lógica de la toma de decisiones (como la extracción de información relevante de documentos mediante la visión artificial) muestren un impacto de menor importancia. En cambio, cualquier uso de la IA que cambie la lógica de la toma de decisiones (es decir, que aplique un nuevo modelo a los datos existentes) puede mostrar un impacto más importante. El mayor nivel de importancia del impacto puede darse en los usos que cambian la lógica de la toma de decisiones basada en nuevas fuentes de datos.

Del mismo modo, las aplicaciones en el compromiso con el cliente pueden tener menor importancia que las aplicaciones que se utilizan para determinar los pagos a los asegurados. Las aplicaciones en materia de suscripción y fijación de precios pueden presentar los niveles más altos de importancia, en particular si pueden dar lugar a la exclusión de clientes. [22]

La figura 4 ofrece una clasificación ilustrativa de los casos de uso en función de su naturaleza y posición en la cadena de valor. En la práctica, cada caso de uso debe evaluarse en función de las ventajas individuales.

22 Las directrices de la UE sobre la toma de decisiones individuales automatizadas y la elaboración de perfiles mencionan, como ejemplo de un impacto significativo, la denegación automática de una solicitud de crédito en línea. Según estas directrices, las decisiones de presentar publicidad selectiva basada en perfiles no tendrán, en general, un impacto significativo similar sobre las personas. La fijación de precios diferenciados basada en datos personales podría tener un impacto significativo si genera la imposibilidad de pagar.

Figura 4: Posible clasificación de las aplicaciones de la IA y su importancia (a título ilustrativo)

Cambio en la lógica y nuevas fuentes de datos	• Asesoramiento automatizado o *robo-advice* mediante fuentes de datos externas	• Optimización de precios con datos sobre el estilo de vida • Precios de todos los algoritmos con datos sobre el estilo de vida	
Cambio en la lógica de la toma de decisiones	• Segmentación de clientes y publicidad dirigida	• Algoritmos de fijación de precios con datos tradicionales	• Detección de fraudes • Clasificación automatizada de siniestros
Sin cambios en la lógica o los datos	• Agente conversacional (chatbot)	• Visión artificial para extraer información de documentos	• Visión artificial para extraer información de documentos
	Compromiso de los clientes	**Suscripción de pólizas / fijación de precios**	**Reclamaciones**

■ Importancia alta ■ Importancia media ■ Importancia baja

2. Adoptar estructuras de gobernanza adecuadas

Garantizar un uso responsable de la IA requiere una estructura de gobernanza adecuada que asigne responsabilidades claras a personas, comités o departamentos que tengan las competencias necesarias para la toma de decisiones, las aptitudes y conocimientos necesarios, así como los procesos asociados, incluidos los desencadenantes y la escalabilidad.

Existen muchos modelos de gobernanza diferentes, cada uno con sus propias ventajas e inconvenientes. La elección de un modelo organizacional dependerá de la estructura y la cultura existentes en la empresa.

A continuación, se enumeran algunos aspectos importantes que deben tenerse en cuenta a la hora de elegir un modelo organizacional adecuado:

- ***Modelo centralizado versus modelo descentralizado***

La responsabilidad del uso equitativo de la IA puede asignarse a un equipo central (por ejemplo, bajo la dirección de un responsable de ética de datos) o delegarse a la empresa local. Aunque un modelo centralizado puede tener la ventaja de facilitar un planteamiento coherente en toda la organización, existe el riesgo de que ese equipo parezca demasiado alejado de las necesidades de la empresa y los clientes.

- ***Confianza en las estructuras de gobierno existentes frente a la creación de estructuras nuevas***

La supervisión del uso equitativo y responsable de los sistemas de IA podría delegarse en un marco existente, como el marco de gestión de riesgos, el marco de gobernanza informática o el marco de cumplimiento. Otra alternativa es establecer un mecanismo de supervisión nuevo y exclusivo, perfectamente integrado en el marco de gestión de riesgos existente.

- ***Conocimientos y experiencia adecuados***

Es fundamental garantizar niveles adecuados y diversidad de competencias, conocimientos y experiencia (incluida la ciencia de datos, los conocimientos actuariales, jurídicos y normativos) en las decisiones clave sobre el uso de los sistemas de IA.

3. Desarrollar e implementar programas de capacitación interna

Por último, para garantizar un uso responsable de la IA es necesario que las distintas áreas y cargos directivos sean conscientes de los beneficios y riesgos que conlleva. Para aumentar la concientización, las aseguradoras deben desarrollar e implementar programas de capacitación integrales sobre los beneficios y riesgos de la IA, así como las respectivas directrices y políticas internas.
Lo ideal sería que estos programas de capacitación estuviesen dirigidos a empleados de todos los niveles de gestión y áreas donde se toman decisiones, incluidos los representantes y otros empleados de atención al cliente.

Glosario

Inteligencia artificial: Rama de la informática que se ocupa de la simulación informática de conductas inteligentes. Por lo general, el término se utiliza para referirse a la capacidad de una máquina para imitar el comportamiento inteligente (humano). comportamiento (humano) (https://www.merriam-webster.com/ dictionary/artificial%20intelligence).

Sesgo: Consiste en un error sistemático que coloca a ciertos grupos en una ventaja sistemática y a otros en una desventaja sistemática.

Macrodatos *(Big data)*: Un activo de información de gran volumen, velocidad y variedad que exige formas rentables e innovadoras de tratamiento de la información para mejorar el conocimiento y la toma de decisiones. Los macrodatos pueden evaluarse mediante cinco parámetros "V": volumen, velocidad, variedad, veracidad y variabilidad (Swan 2015). Algunos comentaristas han agregado visualización y valor a esos parámetros (Devan 2016). Otras definiciones hacen hincapié en la complejidad de los macrodatos. El Instituto Nacional de Estándares y Tecnología (NIST, por sus siglas en inglés) define los macrodatos como datos que superan la capacidad y las aptitudes de los métodos y sistemas actuales.

Visión artificial: El campo de estudio que rodea la forma en que las computadoras ven y entienden las imágenes y los videos digitales. La visión artificial abarca todas las tareas realizadas por los sistemas biológicos de visión, como "ver" o percibir un estímulo visual, entender lo que se ve y extraer información compleja en un formato que puede utilizarse en otros procesos. Este campo interdisciplinar simula y automatiza estos elementos de los sistemas de visión humana mediante sensores, computadoras y algoritmos de aprendizaje automático. La visión artificial es la teoría en la que se basa la capacidad de los sistemas de IA para ver y entender el entorno que los rodea (https://deepai.org/machine-learning- glossary-and-terms/computer-vision).

Maldición de la dimensionalidad: Fenómenos estadísticos que se producen al clasificar, organizar y analizar datos de alta dimensión (con cientos o miles de variables) que no se producen en espacios de baja dimensión, en concreto la cuestión de la escasez de datos y la 'cercanía' de los datos.[23]

Datos: Hechos y estadísticas recopilados para referencia o análisis (Swan 2015).

Protección de datos: Técnicamente hablando, el proceso de salvaguardar información importante frente a casos de corrupción o pérdida.[24] La jurisprudencia europea tiende a tratar la protección de datos como una manifestación del derecho a la privacidad. Aunque los conceptos de protección de datos y privacidad se entremezclan, también existen diferencias en cuanto a su alcance, ya que el ámbito de la protección de datos es más amplio que el de la privacidad (Kokott, J. y Sobotta, C. 2013).

Ciencia de datos: La extracción de conocimientos a partir de datos. La ciencia de datos emplea técnicas y teorías de las matemáticas, la estadística, la informática y la tecnología de la información —por ejemplo, el aprendizaje automático— para descubrir patrones en los datos a partir de los cuales se pueden desarrollar modelos predictivos (Swan 2015).

Aprendizaje profundo: Técnica de aprendizaje automático que construye redes neuronales artificiales para imitar la estructura y el funcionamiento del cerebro humano (https://deepai. org/machine-learning-glossary-and-terms/deep-learning).

Impacto desigual: El hecho de que los resultados perjudiquen o beneficien de manera desproporcionada a las personas con un determinado atributo sensible.

Trato desigual: Decisiones basadas (parcialmente) en un atributo sensible.

Explicabilidad: Véase interpretabilidad.

23 https://deepai.org/machine-learning-glossary-and-terms/curse-of-dimensionality
24 http://searchstorage.techtarget.com/definition/data-protection

Información: Datos aportados o aprendidos sobre algo o alguien. Tanto los datos como la información pueden servir de base para un razonamiento o cálculo. Aunque antes se distinguía más entre los datos como hechos y estadísticas subyacentes y la información como conocimiento extraído de estos hechos y estadísticas, las definiciones se han asemejado bastante y a menudo se utilizan como sinónimos (Swan 2015).

Privacidad de la información: El interés de las personas en ejercer el control sobre el acceso a datos sobre ellas mismas (Stanford Encyclopedia of Philosophy 2014). Véase la sección 'Transparencia y explicabilidad'.

Internet de las cosas (IoT): Definida por la Unión Internacional de Telecomunicaciones (UIT) como 'una infraestructura global para la sociedad de la información que permite servicios avanzados interconectando cosas (físicas y virtuales) sobre la base de las tecnologías de la información y la comunicación interoperables existentes y en evolución' (IoT-GSI 2015).

Interpretabilidad: La capacidad de explicar o proporcionar el significado en términos entendibles para un ser humano (Doshi- Vélez y Kim 2017).

Aprendizaje automático: Técnica o subdisciplina de la IA que proporciona a los sistemas la capacidad de aprender de forma automática y, a partir de la experiencia o los ejemplos, mejorar sin ser programados de manera explícita. El aprendizaje automático se centra en el desarrollo de programas informáticos capaces de acceder a los datos y utilizarlos para aprender por sí mismos. [25]

Sobreajuste (*overfitting*): En estadística, un análisis o modelo que se corresponde con demasiada exactitud o precisión con un conjunto concreto de datos y que, por lo tanto, es posible que no pueda ajustarse a datos adicionales o predecir observaciones futuras de manera fidedigna. [26]

Información o datos personales: Información o datos que están vinculados o pueden vincularse a personas individuales (Stanford 2014). En el Reglamento General de Protección de Datos europeo, los datos personales se definen como 'cualquier información relativa a una persona que puede ser identificada, directa o indirectamente, en particular mediante referencia a un identificador como un nombre, un número de identificación, datos de localización, identificador en línea o a uno o más factores específicos de la identidad física, fisiológica, genética, mental, económica, cultural o social de esa persona'. Esto significa que, en muchos casos, los identificadores en línea, como la dirección IP y las cookies, pasarán a considerarse datos personales si pueden vincularse (o son susceptibles de vincularse) al interesado sin esfuerzos indebidos.

Privacidad: No existe una definición aceptada a nivel mundial del concepto de privacidad. En este informe definimos la privacidad como el 'uso adecuado de los datos personales'.

25 http://www.expertsystem.com/machine-learning-definition/
26 https://www.lexico.com/en/definition/overfitting

Referencias

AXA. 2019. Autores: Ruf, B., M. Hirot. 2019a. *Regulating machine learning: Where do we stand? State of the art and challenges ahead. https://axa-rev-research.github.io/static/AXA_WhitePaper_RegulatingML.pdf*

AXA. 2019. Autores: Ruf, B., and V. Grari. 2019b. *Understanding and mitigating bias*. https://axa-rev-research.github.io/static/AXA_Booklet_Bias.pdf

Barocas, S., y A. D. Selbst. 2016. Big data's disparate impact. *California Law Review* 104 (3). http://dx.doi.org/10.15779/Z38BG31

Bellamy, R.K.E. et al. 2018. *AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias*. https://arxiv.org/abs/1810.01943

BaFin. 2018. *Big data meets artificial intelligence—Challenges and implications for the supervision and regulation of financial services.* Federal Financial Supervisory Authority. July 2018. https://www.bafin.de/SharedDocs/Downloads/EN/dl_bdai_studie en.html;jsessionid=1DE4EFB708FA602B6829FC41B0FB7432.1_cid372?nn=9866146

Christen, M. et al. 2019. *Big data ethics recommendations for the insurance industry. A consolidation report outlining the results of the NRP 75 project: Between solidarity and personalisation—Dealing with ethical and legal big data challenges in the insurance industry.* www.nfp75.ch/ SiteCollectionDocuments/NFP75-Website-Publikation- Christen.pdf

DeVan, A. 2016. *The 7 V's of Big Data.* https://www. impactradius.com/blog/7-vs-big-data/

Doshi-Velez, F., y B. Kim. 2017. *Towards a rigorous science of interpretable machine learning.* Cornell University research paper. https://arxiv.org/abs/1702.08608

European Commission. 2019. *Ethics guidelines for trustworthy AI.* High-Level Expert Group on Artificial Intelligence. April 2019.

Everyday ethics for artificial intelligence. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf

Fiedler, S. A., C. Scheidegger, and S. Venkatasubramanian. 2016). *On the (im)possibility of fairness.* Cornell University research paper. https://arxiv.org/abs/1609.07236

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Gianotti , and D. Pedreschi. 2018. *Estudio de métodos para explicar los modelos de caja negra.* ACM Computing Surveys. Febrero de 2018. https://dl.acm.org/citation.cfm?doid=3271482.3236009

Hall, P., S. Ambati y W. Phan. 2017. *Ideas on interpreting machine learning.* https://www.oreilly.com/radar/ideas-on-interpreting-machine-learning/

Asociación Internacional de Supervisores de Seguros (IAIS). 2019. *Insurance Core Principles and Common Framework for the Supervision of Internationally Active Insurance Groups 2019 (ICP19).*

Iniciativa sobre normas mundiales para la internet de las cosas (IoT-GSI). 2015. http://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx

IEEE. 2019. *Ethically aligned design.* Primera edición. https:// ethicsinaction.ieee.org/

The Institute for Ethical ML . 2019. *The Responsible Machine Learning Principles. A practical framework to design AI responsibly.* https://ethical.institute/principles.html

Jobin, A., M. Iencya y E. Vayena. 2019. *Inteligencia artificial: The global landscape of ethics guidelines.* Preprint version. Health Ethics and Policy Lab, ETH Zurich.

Kleinberg, J., S. Mullainathan y M. Raghavan. 2017. Compensaciones inherentes a la determinación equitativa de las calificaciones de riesgo. In *Innovations in Theoretical Computer Science.* ACM.

Kokott, J. y C. Sobotta. 2013. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. In *International Data Privacy Law,* 3(4). https://academic.oup.com/idpl/article/3/4/222/727206/The-distinction-between-privacy-and-data

Laugel, T., M. Lesot, C. Marala y M. Detyniecki. 2019. Issues with post-hoc counterfactual explanations: A discussion. *ICML Workshop on Human in the Loop Learning* (HILL 2019).

Lazer, D., and R. Kennedy. 2015. What we can learn from the epic failure of Google Flu trends. *Wired Opinion*. 10 de enero de 2015. https://www.wired.com/2015/10/can-learn- epic-failure-google-flu-trends/

Loi, M., y M. Christen. 2019. *Insurance discrimination and fairness in machine learning: An ethical analysis.* 17 August 2019. SSRN paper. http://dx.doi.org/10.2139/ssrn.3438823.

McKinsey & Company. 2018. *Insurance 2030 The impact of AI on the future of insurance.*

Metz, R. 2019. Why Microsoft accidentally unleashed a neo-nazi sexbot. *MIT Technology Review*, 24 March 2016. https://www.technologyreview.com/s/601111/why-microsoft-accidentally-unleashed-a-neo-nazi-sexbot/

Microsoft. 2019. *Microsoft AI Principles.* https://www.microsoft.com/en-us/ai/our-approach-to-ai

Monetary Authority of Singapore. 2018. *Principios para promover la equidad, la ética, la responsabilidad y la transparencia (FEAT) en el uso de la inteligencia artificial y el análisis de datos en el sector financiero de Singapur.* 12 November 2018. https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

Narayanan, A. 2018. Tutorial: 21 fairness definitions and their politics. *Conference on Fairness, Accountability, and Transparency*, February 2018.

OCDE. 2019. *Recommendations of the Council on Artificial Intelligence*, adopted on May 22, 2019.

Pedreschi, D., S. Ruggieri,y F. Turini. 2008. Discrimination-aware data mining. En *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data* Mining. Las Vegas, 2008.

SCOR. 2018. *The Impact of Artificial Intelligence on the (Re) Insurance Sector*. SCOR, 2018.

Stanford Encyclopedia of Philosophy. 2014. Privacy and Information Technology. November 2014.

Swan, M. 2015. Philosophy of big data: Expanding the human-data relation with big data science services. *2015 IEEE First International Conference on Big Data Computing Service and Applications*.

The Geneva Association. 2017. *Genetics and life insurance—A view into the microscope of regulation.* Author: Ronald Klein. Junio de 2017.

The Geneva Association. 2018. *Big data and insurance: Implications for innovation, competition and privacy.* Author: Benno Keller. March 2018.

The Geneva Association. 2019. *The role of trust in narrowing protection gaps.* Author: Kai-Uwe Schanz. November 2019.

Wachter, S., B. Mittelstadt, and C. Russell. 2017. *Explicaciones contrafácticas sin abrir la caja negra: Automated decisions and the GDPR.* Cornell University. https://arxiv.org/abs/1711.00399

World Economic Forum. 2018. *How to prevent discriminatory outcomes in machine learning.* White Paper. Global Future Council on Human Rights 2016-2018. March 2018. http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf

El uso de la inteligencia artificial (IA) en los seguros puede aportar beneficios económicos y sociales al reducir los costos de los seguros y ayudar a aumentar la cantidad de asegurados. Para este informe, The Geneva Association analizó dos de los cinco principios fundamentales identificados para el uso responsable de la IA 1) transparencia y explicabilidad y 2) equidad, que plantean cuestiones particularmente complejas en el sector de los seguros. Con este análisis y una serie de recomendaciones para las aseguradoras, pretendemos contribuir al uso responsable de la IA en el sector de los seguros y a la materialización de los beneficios para la sociedad.

The Geneva Association

Asociación internacional para el estudio de la economía de los seguros
Talstrasse 70, Zúrich, Suiza
Tel: +41 44 200 49 00 | Fax: +41 44 200 49 99

secretariat@genevaassociation.org