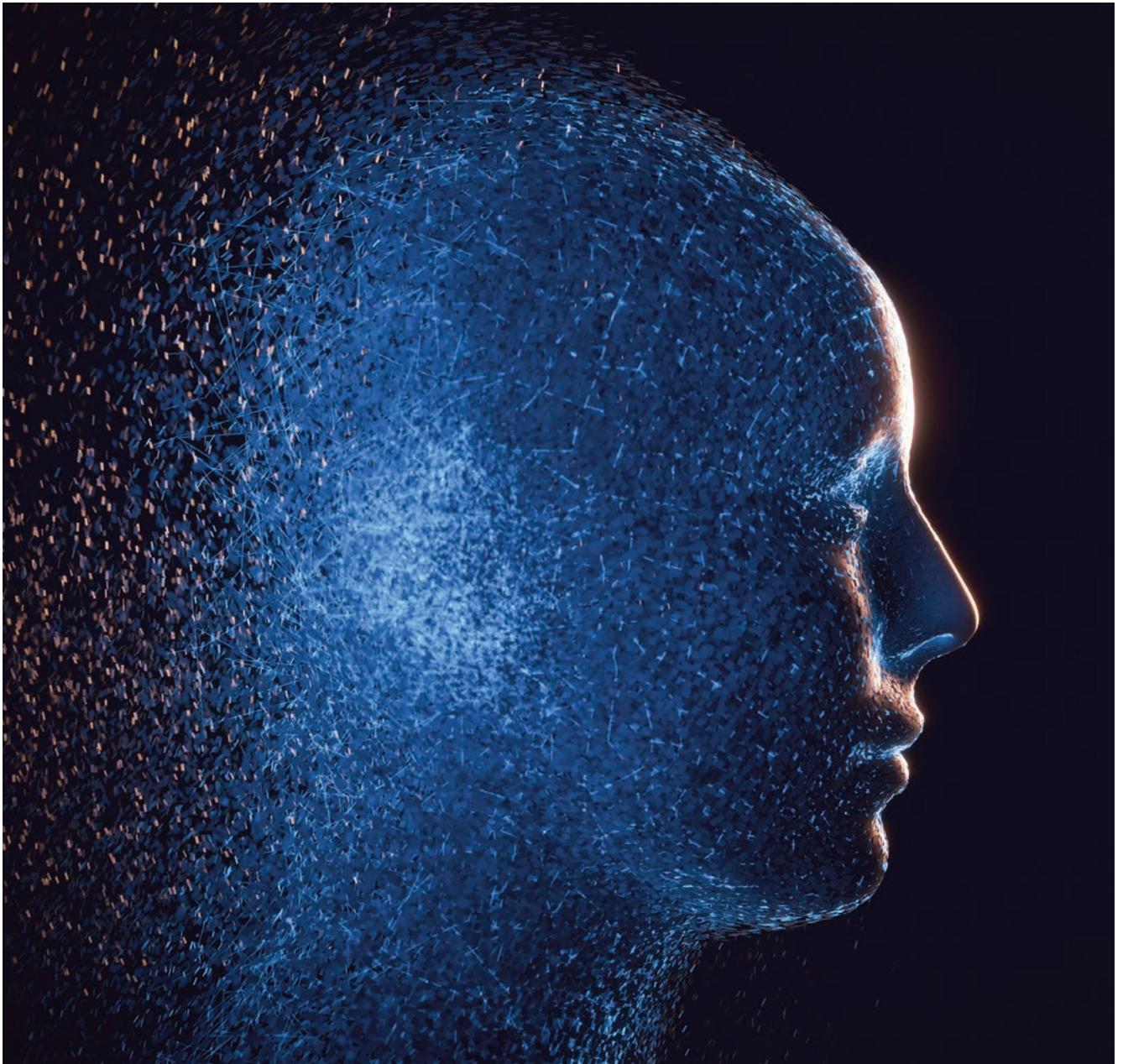


Conclusiones

“Con la programación adecuada, un ordenador puede convertirse en un teatro, un instrumento musical, un libro de consulta, un contrincante de ajedrez. Ninguna otra entidad en el mundo salvo el ser humano tiene una naturaleza tan adaptable y universal”.

Daniel Hillis⁷⁴



En este estudio se ha presentado la inteligencia artificial explicable (XAI), sus fundamentos, contexto y técnicas para mejorar la interpretabilidad de los modelos. Se han discutido los principales desafíos que enfrentan los modelos de inteligencia artificial en términos de interpretabilidad y cómo la tecnología puede ayudar a abordarlos, incluyendo un caso práctico desarrollado con ModelCraft™ para demostrar cómo se pueden emplear estas técnicas para entender y explicar los modelos de AI.

La disciplina de la AI, y dentro de ella la XAI, ha crecido en importancia a nivel mundial en los últimos años, cuando el desarrollo de tecnologías de AI de alto rendimiento se ha convertido en una prioridad para muchos sectores, desde la salud hasta la seguridad, pasando por los servicios financieros o la energía, entre otros. La interpretabilidad surge como una necesidad para entender y mejorar los modelos de AI, lo que reviste especial complejidad en el caso de determinadas técnicas.

Como se ha visto, los modelos de AI pueden enfrentar dificultades para explicar los resultados o la lógica detrás de sus decisiones. Esto se debe a que estos modelos utilizan técnicas de aprendizaje profundo y algoritmos complejos para aprender a partir de datos, los cuales a menudo son difíciles de interpretar, y esto plantea desafíos a la hora de evaluar los modelos de AI y la confiabilidad de sus resultados.

Por todo ello, el marco regulatorio sobre AI está evolucionando con rapidez, y se espera que las organizaciones se adapten a los nuevos requerimientos de transparencia, explicabilidad e imparcialidad en el uso de modelos de AI. Esto implica la necesidad de un enfoque integral que integre la interpretabilidad y explicabilidad en la organización y los procesos de cada compañía, abarcando técnicas de interpretabilidad, gestión del riesgo de modelo, colaboración interdisciplinaria y capacitación en XAI para los profesionales involucrados en el desarrollo y la aplicación de la AI, entre otros.

En conclusión, la interpretabilidad de los modelos de inteligencia artificial es un área de investigación emergente, y es esperable que continúe desarrollándose y creciendo en importancia a medida que los modelos de AI se vuelvan más complejos, la regulación siga proliferando, y su uso se extienda a más ámbitos de alta sensibilidad.

⁷⁴Daniel Hillis (n. 1956), inventor, empresario y científico estadounidense, pionero de la computación paralela y su uso en el campo de la inteligencia artificial, con más de 300 patentes publicadas.