



**LA SECCIÓN DE CIENCIAS MATEMÁTICAS DE LA
REAL ACADEMIA DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES DE ESPAÑA**

se complace en invitarle a la sesión científica pública

“Inteligencia artificial interpretable”

impartida por el

Prof. D. Pedro Larrañaga Múgica

Universidad Politécnica de Madrid

Miércoles, 5 de junio de 2024
18.00 horas

Retransmisión en directo en



Calle Valverde, 22
28004 Madrid

Resumen de la conferencia

La creciente difusión y popularidad de los sistemas inteligentes basados en el denominado aprendizaje automático plantea la disyuntiva de su uso indiscriminado frente a la necesidad de comprender el funcionamiento interno de los mismos. La explicación *post hoc*, a la que habitualmente se recurre ante los modelos de caja negra, no resulta suficiente para modelizaciones y decisiones en el ámbito científico ni tampoco en situaciones de alto riesgo. En tales casos se requiere que el humano sea capaz de interpretar tanto la salida del sistema como el proceso interno seguido por el mismo hasta alcanzar dicho resultado.

En esta conferencia analizaremos las potencialidades en interpretabilidad del paradigma conocido como red Bayesiana enfatizando su transparencia y versatilidad para llevar a cabo distintos tipos de razonamiento, como por ejemplo el predictivo, diagnóstico, intercausal, contrafáctico, o abductivo. Se mostrarán diversos casos de uso reales tanto de modelización como de optimización en dominios como la neurociencia computacional y la industria.

Pedro Larrañaga es Catedrático de Ciencias de la Computación e Inteligencia Artificial en la Universidad Politécnica de Madrid y Director de la *ELLIS Unit* Madrid. Coautor de más de 200 artículos científicos en revistas del *Journal Citation Reports* y director/codirector de 36 tesis doctorales, ha participado en más de 100 proyectos de investigación tanto públicos como privados. Ha sido gestor del Plan Nacional en Tecnologías de la Información y miembro de la CNEAI. Es miembro electo de la *European Association for Artificial Intelligence*, de la *Academia Europaea*, de la *Asia-Pacific Artificial Intelligence Association* y de la Academia de las Ciencias, las Artes y las Letras del País Vasco. Es también *ELLIS Fellow* e *IEEE Fellow*. Ha recibido el Premio Nacional Arimel en Ciencias de la Computación y el Premio de la Asociación Española para la Inteligencia Artificial. Sus intereses investigadores se centran en las redes Bayesianas, la selección de variables y los algoritmos de estimación de distribuciones con aplicaciones en neurociencia computacional e industria.

