

XVIII Jornada MEDES 2024 de la Fundación Lilly: «La segunda revolución digital de la comunicación científica. Desafíos para la medicina en español»

La IA aplicada a la información biomédica en español solo será efectiva si aumentan los corpus de calidad en este idioma



De izda. a dcha.: José Antonio Sacristán, director de la Fundación Lilly; Asunción Gómez Pérez, académica de número de la Real Academia Española y catedrática de Inteligencia Artificial y vicerrectora de Investigación, Innovación y Doctorado de la Universidad Politécnica de Madrid; Eva Ortega-Paño, secretaria general de Investigación del Ministerio de Ciencia, Innovación y Universidades; Pilar Rico Castro, codirectora de la jornada MEDES 2024 y jefa de la Unidad de Acceso Abierto de FECYT; y Elea Giménez Toledo, codirectora de la jornada MEDES 2024, coordinadora de la plataforma ES-CIENCIA (CSIC) y consejera científica de la Fundación Lilly

- **Los corpus —las grandes cantidades de textos con los que se entrenan los modelos— no deben tener sesgos, ni emplear lenguaje tóxico que incite al odio o sea ofensivo**
- **Además, deben ser anonimizados, recoger información veraz, estar actualizados en lo que respecta a reglamentos y legislaciones, contar con licencias claramente identificadas y con una complejidad adecuada**

Madrid, 10 de julio de 2024.- Fomentar los corpus de información biomédica en español, de calidad, accesible y reutilizable, albergados en plataformas de *software libre* y con modelos de datos federados es una de las aportaciones más importantes que hace la ciencia

NOTA DE PRENSA

abierta. Esta es una tarea imprescindible en la segunda revolución digital, donde las inteligencias artificiales desarrollan labores que tradicionalmente hacían los humanos. Así lo han planteado los participantes en la XVIII Jornada MEDES 2024, titulada «**La segunda revolución digital de la comunicación científica. Desafíos para la medicina en español**» y organizada por la Fundación Lilly en el marco de los Cursos de Verano Complutense de El Escorial.

«Los grandes modelos de lenguaje, esos gigantescos corpus de información sobre los que se aplica la lingüística computacional para desarrollar herramientas de procesamiento de lenguaje natural, son la clave para enfrentar de manera responsable los retos que presenta la inteligencia artificial en términos de propiedad intelectual y de control de sesgos en la comunicación biomédica en español», explica **Pilar Rico Castro, codirectora de esta jornada MEDES y jefa de la Unidad de Acceso Abierto de la Fundación Española para la Ciencia y la Tecnología (FECYT)**.

Por su parte, **Elea Giménez Toledo, codirectora de la jornada MEDES 2024, coordinadora de la plataforma ES-CIENCIA (CSIC) y consejera científica de la Fundación Lilly**, señala la necesidad de disponer de corpus científicos en español, que sirvan para entrenar los modelos, desarrollar aplicaciones y cuyos usos sean respetuosos con la propiedad intelectual. Además, se ha referido a las dos velocidades en el uso de los textos legibles por las máquinas: mientras las grandes compañías toman todo tipo de materiales para alimentar sistemas conversacionales, por ejemplo, las restricciones de uso para los investigadores son notorias.

El lenguaje es el *software* de los humanos

Durante la jornada, **Asunción Gómez Pérez, académica de número de la Real Academia Española y catedrática de Inteligencia Artificial y vicerrectora de Investigación, Innovación y Doctorado de la Universidad Politécnica de Madrid**, ha presentado el lenguaje como el *software* de los humanos, con el que nos comunicamos sin restricciones. A partir de nuestra lengua, generamos el lenguaje de las máquinas.

«Las máquinas las programan humanos y el conocimiento sobre el que operan es conocimiento humano. Por ello, nuestra responsabilidad es la de seguir creando el mejor conocimiento biomédico en español posible y albergarlo en aquellos espacios en los que nuestras máquinas puedan procesarlo mejor que nosotros para hacer análisis a gran escala y generar nuevas evidencias», comenta **Pilar Rico Castro**.

En cuanto a la adopción de la inteligencia artificial generativa para la comunicación biomédica, **Asunción Gómez** asegura que el modelo nativo de lenguaje para el español, además de entrenarse con textos en español de calidad, debe ser sostenible en el tiempo, respetar los derechos de autor y seguir las normas lingüísticas.

En este sentido, se recomienda evitar los dialectos digitales del español, es decir, la transformación de nuestro idioma debida a la irrupción de las tecnologías en la comunicación. Aunque pueden facilitar la comunicación rápida y eficiente en ciertos

NOTA DE PRENSA

contextos, estas variaciones lingüísticas también pueden presentar desafíos en términos de comprensión y preservación de la lengua, ya que pueden incluir cambios en el vocabulario, la gramática, la ortografía o el estilo de comunicación.

Por otro lado, la académica subraya que, para obtener información de alta calidad, los corpus no deben tener sesgos, ni emplear lenguaje tóxico que incite al odio, violencia verbal o sea ofensivo. También deben ser anonimizados, recoger información veraz, estar actualizados en lo que respecta al reglamento europeo de IA y otras legislaciones, contar con licencias claramente identificadas y con una complejidad adecuada.

La tecnología y el arte para humanizar la comunicación médico-paciente

Durante la jornada, también se han presentado herramientas clave para humanizar la medicina y transmitir a los pacientes la información sobre su enfermedad de la manera más clara y rigurosa posible. En este sentido, y en el contexto de la segunda revolución tecnológica, se ha expuesto el beneficio del lenguaje claro, la infografía y la medicina gráfica —que incluye formatos como el cómic, la ilustración o la novela gráfica—.

Presentar información rigurosa mediante recursos accesibles y sencillos de entender permite una democratización del conocimiento médico y, por ende, el empoderamiento del paciente, quien cada vez más solicita ser participe de todo el proceso médico y, para ello, necesita información y, sobre todo, entenderla.

Fundación Lilly: Ciencia, Medicina y Humanismo

La Fundación Lilly tiene como objetivo contribuir al desarrollo de la sanidad española en beneficio de la salud de los ciudadanos; impulsar la CIENCIA y la investigación, así como la promoción de la cultura y la divulgación científica entre la sociedad; favorecer el desarrollo de la MEDICINA a través de la generación del conocimiento biomédico, su difusión, la formación de los profesionales y la educación médica; y fomentar los valores fundamentales del HUMANISMO en el ámbito de la sanidad. En el área de Ciencia, anualmente convoca los Premios de Investigación Biomédica, las Citas con la Ciencia y apoya diversas acciones de divulgación científica. En el de Medicina, destacan iniciativas como la Cátedra de Educación Médica, MEDES – MEDicina en ESpañol o su programa de gestión sanitaria. La iniciativa Medicina Centrada en el Paciente y las actividades en torno a la figura del médico y humanista español Andrés Laguna o la difusión del legado de Santiago Ramón y Cajal centran el área del humanismo médico.

fundacionlilly.com –