BANCO DE INNOVACIÓN EN LAS ADMINISTRACIONES PÚBLICAS

TÍTULO

La RAE promueve el correcto uso de la lengua española en la Inteligencia Artificial con el Proyecto LEIA

INFORMACIÓN INICIAL:

PROBLEMA:

El español es hoy en día la tercera lengua más hablada en el mundo, tras el inglés y el chino. De los casi 600 millones de hablantes en todo el mundo, casi 493 millones lo hablan como lengua materna. Sin embargo, el inglés es el idioma que domina las infraestructuras lingüísticas disponibles para entrenar y desarrollar sistemas de tecnologías del lenguaje y, por tanto, la lengua de comunicación preferente de la inteligencia artificial (IA).

En los últimos años la inteligencia artificial ha experimentado una transformación significativa ya que ha pasado de un razonamiento sobre conjuntos de datos de números a un pensamiento en palabras sobre el lenguaje humano. Con la llegada de esta revolucionaria tecnología, la capacidad de "entender" textos ya no está exclusivamente reservada al cerebro humano, por lo que es clave pensar cómo entrenar estos algoritmos con las palabras adecuadas y la referencia de un diccionario pertinente, a fin de normativizar el buen uso de nuestro idioma en las máquinas.

El problema de los denominados "dialectos digitales" presentes por estas máquinas basadas en IA es que pueden llegar a crear jerigonzas, jergas distintas según el usuario y según el fabricante de esa IA, pudiendo dar lugar a una situación en la que las máquinas "hablen" un español incorrecto, sin cumplir las reglas elaboradas por la Real Academia Española (RAE). Además, la lengua que utiliza la IA es decisiva para garantizar algunos de los principios en que ha de basarse esta tecnología: accesibilidad, igualdad y no discriminación y la protección general de los derechos individuales y colectivos de la ciudadanía.

En cualquier caso, la IA tiene un gran potencial tanto a nivel tecnológico, económico y social y una gran incidencia en actividades cotidianas (*chatbots*, asistentes virtuales, traducción simultánea, buscadores de internet, etc.) y el sector público juega un papel fundamental en el impulso de su desarrollo y en su integración en el tejido productivo, la economía y la sociedad.

Ante la importancia que supone posicionar el idioma español en estas tecnologías en favor de la comunidad hispanohablante, la Real Academia de la Lengua Española (RAE), cuyo papel en la normalización del lenguaje es esencial, viene trabajando desde los últimos años en proyectos orientados a desarrollar un corpus único de la lengua española para la inteligencia artificial. Uno de ellos es el proyecto LEIA (Lengua Española en la Inteligencia Artificial), liderado por la RAE, junto con las 22 academias de la lengua española en el resto del mundo, cuyo objetivo es proporcionar a las grandes plataformas digitales globales los recursos de un diccionario y una gramática única para el despliegue de sus servicios sobre un lenguaje común.

Imagen 1. **Logotipo del Proyecto LEIA**

Fuente: RAE.

SOLUCIÓN GLOBAL:

El proyecto LEIA, liderado por la RAE, nace de la preocupación de ver cómo crecía el número de máquinas hablantes en el mundo y de la percepción de que hablaban según el criterio de su fabricante, que podía ser distinto, dependiendo de la ubicación geográfica o de la cultura de esa empresa.

Su propósito es enseñar a las máquinas a hablar un correcto español en los medios tecnológicos y trabajar en el desarrollo de herramientas y aplicaciones que ayuden a los humanos a adquirir un buen uso del español. Además, para garantizar el cumplimiento de estos dos objetivos, el proyecto incluye la creación de un certificado que permita verificar que las empresas e instituciones se comprometen a esforzarse por emplear un español correcto en sus plataformas y tecnologías.

En asociación con los grandes operadores tecnológicos (Telefónica, Microsoft, Amazon, Google, Twitter, Meta, entre otros), la RAE contribuye a través de LEIA a mantener la unidad de la lengua evitando que la IA de estos operadores tecnológicos hagan un mal uso de nuestra lengua aceptando corrupciones y formas incorrectas de expresión cuyo uso se expande rápidamente entre la población.

LEIA se enmarca dentro del Plan de Recuperación, Transformación y Resiliencia (PRTR) en su componente 16, Reforma 1 y en los objetivos 251 y 253 (del anexo de la Propuesta de Decisión de Ejecución del Consejo relativa a la aprobación de la evaluación del Plan de Recuperación, Transformación y Resiliencia) principalmente en la línea de ayudas a proyectos de I+D en inteligencia artificial y a la finalización de los mismos en el momento de cumplimiento acordado.

Asimismo, la Estrategia Nacional de Inteligencia Artificial (ENIA), propuesta dentro de la Agenda España Digital 2026, e ideada para orientar los planes sectoriales, estatales y estrategias regionales en materia de inteligencia artificial para el período 2020-2026, contempla en su línea de actuación 3.2 proveer infraestructuras tecnológicas habilitadoras. Y, más específicamente, en la medida 14 establece la "creación de recursos en la iniciativa de Lengua Española en la IA (LEIA) impulsada por la Real Academia de la Lengua Española y el Plan Nacional del Lenguaje".

El proyecto LEIA se incluye, además, en el Proyecto Estratégico para la Recuperación y la Transformación Económica (PERTE) "Nueva Economía de la Lengua", proyecto que tiene el objetivo de movilizar inversiones públicas y privadas para maximizar el valor del español y las lenguas cooficiales en el proceso de la transformación digital a nivel mundial. En el primer eje, que consiste en la creación de una base de conocimiento (corpus) en español y lenguas cooficiales, se incluye el apoyo al Proyecto Lengua Española e Inteligencia Artificial (LEIA) para crear un banco de datos de la lengua. Concretamente, se integra en el Proyecto Tractor 1: Creación de nuevos corpus amplios para la nueva economía de la lengua.

COSTE APROXIMADO:

Para la ejecución del proyecto el Gobierno ha concedido una subvención de concesión directa de 5 millones de euros con cargo a los fondos de la Unión Europea del Mecanismo de Recuperación y Resiliencia del Ministerio de Asuntos Económicos y Transformación Digital.

El importe de la subvención está destinado a cubrir los costes de las siguientes actuaciones:

1. Creación de un departamento técnico (550.000 €).
2. Observatorio de neologismos (1.050.000 €).
3. Creación y desarrollo de herramientas de verificación lingüística (1.100.000 €).
4. Creación y desarrollo de herramientas de respuesta automática (465.000 €).
5. Recopilación de material basado en la diversidad (470.000 €).
6. Desarrollo de medidas que garanticen la plena accesibilidad de las personas con discapacidad a las herramientas relacionadas con la IA (330.000 €).
7. Mejorar las herramientas tecnológicas que se emplean en los trabajos de regulación de la lengua (255.000 €).
8. Retos, becas y trabajo con las universidades (625.000 €).
9. Labores de digitalización de los fondos propios de la Real Academia Española y obras (155.000 €).

TERRITORIO:

Nivel nacional e internacional (el objetivo es abarcar a toda la comunidad hispanohablante).

PÚBLICO DESTINATARIO:

Usuarios finales, empresas y entidades públicas que hagan uso de tecnologías basadas en inteligencia artificial dentro de la comunidad hispanohablante.

ENTIDAD QUE LA HA LLEVADO A CABO:

El proyecto ha sido impulsado por la Real Academia Española (RAE), con el respaldo de la Asociación de Academias de la Lengua Española (ASALE).

Por su parte, la Real Academia Española (RAE) y el Ministerio de Asuntos Económicos y Transformación Digital, a través de la Secretaría de Estado de Digitalización e Inteligencia Artificial han suscrito un convenio para el impulso del proyecto.

Cuenta con el apoyo de Telefónica y la Fundación Endesa e, igualmente, se han sumado otras grandes empresas tecnológicas, como Microsoft, Amazon, Google, Twitter y Meta.

DESCRIPCIÓN DE LA POLÍTICA O PROGRAMA:

Para hacer efectivo este proyecto, el pasado mes de marzo de 2022 la Real Academia Española (RAE) y la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA) del Ministerio de Asuntos Económicos y Transformación Digital, suscribieron un Convenio para el desarrollo del proyecto LEIA, en el marco de la Estrategia Nacional de Inteligencia.

En el ámbito de las instituciones, se creó el Comisionado especial para la Alianza por la Nueva Economía de la Lengua por el Real Decreto 156/2022, de 1 de marzo, dentro de la Gobernanza de la Alianza para la nueva economía de la lengua. Este Comisionado es el encargado de impulsar y coordinar todos los proyectos relativos al PERTE de la Economía de la Lengua, entre ellos el Proyecto LEIA de la RAE, para así maximizar las sinergias y el impacto de las distintas iniciativas recogidas en el mismo.

Imagen 2. **Presentación del acto de la firma del Convenio para el desarrollo del Proyecto LEIA**

Fuente: RAE.

Para el impulso a la disponibilidad de recursos necesarios para la ejecución de este proyecto se aprobó el Real Decreto 632/2022, de 26 de julio, por el que se regula la concesión directa de una subvención a la Real Academia Española para la realización del proyecto Lengua Española e Inteligencia Artificial (LEIA).

La propuesta LEIA, objeto de financiación en el mencionado real decreto, persigue los siguientes objetivos (artículo 1):

a) Generar recursos lingüísticos, modelos, y poner todos los materiales, datos, corpus de entrenamiento (y todos aquellos datos necesarios), códigos fuente, y cualquier otro tipo de recursos generados con esta subvención, a disposición de las empresas y de la sociedad de manera accesible, abierta, gratuita

y en formatos adecuados, y para cualquier fin, comercial o no comercial, garantizando en todos los casos:

1. La reproducibilidad de los trabajos y actividades subvencionadas. Se documentarán todos los trabajos explicando y justificando las decisiones tomadas, y se elaborarán guías para que cualquier persona/empresa pueda reproducir los trabajos/actividades desarrolladas.
2. El correcto funcionamiento de los materiales creados, permitiendo que cualquier empresa o persona interesada en las tecnologías del lenguaje pueda utilizarlo para cualquier finalidad (comercial o no comercial).
3. Reutilización. Han de proporcionarse todos aquellos datos, corpus, información y cualquier material necesario para garantizar la reutilización.
4. La titularidad de los desarrollos y resultados obtenidos en ejecución de este proyecto corresponderá de forma conjunta y a partes iguales a la Real Academia Española y la Secretaría de Estado de Digitalización e Inteligencia Artificial.

Los materiales, titularidad de la Real Academia Española, que sean necesarios para garantizar la reproducibilidad, funcionamiento y reutilización de los resultados, así como todos los materiales generados y resultados obtenidos (entre los que se incluyen todos los datos, información, corpus, códigos fuente, modelos de lenguaje, etc.) se pondrán a disposición de las Administraciones públicas para su uso y libre acceso en formatos que garanticen la reproducibilidad, funcionamiento y reutilización.

Asimismo, dichos materiales y resultados se pondrán a disposición del público y de la industria a los efectos de reproducibilidad y reutilización, en los términos que se definan en el marco del Plan Nacional de Tecnologías del Lenguaje y bajo las condiciones impuestas por la Comisión de Seguimiento del Convenio entre la Secretaría de Estado de Digitalización e Inteligencia Artificial y la Real Academia Española, para el desarrollo del proyecto Lengua Española e Inteligencia Artificial (LEIA), en el marco de la Estrategia Nacional de Inteligencia Artificial, utilizándose formatos que garanticen la reproducibilidad, funcionamiento y reutilización. A tales efectos, esta Comisión de Seguimiento, acordará la forma en que el material generado será licenciado para su uso y reutilización. Se mencionará –en todo caso– la autoría de la Real Academia Española, al ser reutilizados bajo la licencia que la Secretaría de Estado de Digitalización e Inteligencia Artificial y Real Academia Española consideren.

La Real Academia Española se asegurará de que los resultados del trabajo desempeñado por las empresas a las que se subcontrate permitan el libre acceso y la reutilización en los términos previstos en esta cláusula.

b) Conocer el estado actual del uso de la lengua española en la red.
c) Concienciar de la necesidad de un buen uso del español en los nuevos medios y la tecnología para seguir garantizando la unidad del español en todas sus variedades.
d) Alcanzar un posicionamiento igual al del inglés.
e) Usar la IA para mejorar el uso del español y su procesamiento automático.
f) Acercar el proyecto LEIA a la industria tecnológica mundial.
g) Contribuir al desarrollo del I+D+i español.
h) Proporcionar recursos técnicos y documentales e ideas para mejorar el uso del español en la IA y en la tecnología en general.
i) Crear un sello digital que certifique que las herramientas tecnológicas y de IA usan el español de una manera apropiada.

Las actuaciones que se llevarán a cabo para alcanzar el objetivo del proyecto son las siguientes (artículo 6 y Anexo I del Real Decreto):

1. **Creación de un departamento técnico dedicado específicamente al proyecto LEIA.**
2. **Publicación del observatorio de neologismos.** El léxico del español no es sólo el que se recoge en el diccionario. Hay muchas palabras que, por distintos motivos, no están, o no pueden estar, presentes en él, como los neologismos o los tecnicismos. Para tener una visión completa del estado del español en la red es necesario desarrollar un observatorio donde por medio de la IA se recopilen a tiempo real y de forma automática estas palabras. Para ello, se requieren herramientas que permitan extraer y clasificar automáticamente los términos en cuestión de los textos.
3. **Creación y desarrollo de herramientas de verificación lingüística.** Una de las herramientas más demandadas es un corrector o verificador automático. Para conseguirlo, son necesarios distintos módulos, como un analizador sintáctico o un lexicón exhaustivo.
4. **Creación y desarrollo de herramientas de respuesta automática.** La RAE posee una gran cantidad de consultas ya resueltas que se pueden utilizar para dar respuestas automáticas. Para ello, es necesario organizar y clasificar adecuadamente dicho material y crear herramientas que permitan explotarlo óptimamente.

5. **Recopilación de material basado en la diversidad.** El español es un conjunto de distintas variedades. Crear herramientas que permitan recopilar y acceder a toda esa información y estudiarla puede servir para aprovechar toda esa riqueza lingüística, pero también para preservar la unidad e investigar los puntos en común de las distintas variedades.
6. **Desarrollo de medidas que garanticen la plena accesibilidad de las personas con discapacidad a las herramientas relacionadas con la IA.** La IA no puede ser un obstáculo para nadie. La RAE quiere explorar todos los recursos y herramientas que garanticen que ninguna persona quede excluida de los avances en esta área, principalmente los relacionados con la lengua española.
7. **Mejorar las herramientas tecnológicas que se emplean en los trabajos de regulación de la lengua.** Es necesario optimizar las herramientas que permiten difundir y aplicar las normas lingüísticas.
8. **Retos, becas y trabajo con las universidades.** En el ecosistema universitario reside mucho talento. La RAE quiere acercarse a los investigadores y alumnos de las universidades para escuchar sus ideas y para hacerles partícipes de este proyecto. En esta línea, se quiere organizar retos (*hackathons*), ofrecer becas a grupos de investigación y supervisar trabajos de los alumnos.
9. **Labores de digitalización de los fondos propios de la Real Academia Española y obras.** La RAE atesora fondos y obras que la digitalización ahora permite sacar a la luz de una manera más eficaz.

Por último, en el marco de la colaboración entre la RAE y las operadoras tecnológicas, estas son algunas de las acciones en las que ya se han producido avances:

- Creación de una herramienta basada en IA para conocer el estado del español en internet. Herramienta creada conjuntamente con Amazon Web Services (AWS), gracias a la cual se permite examinar, en muy poco tiempo y mediante sistemas de inteligencia artificial, decenas de miles de documentos de internet a la vez con el fin de valorar el estado del español en el mundo. El desarrollo de esta herramienta permitirá acciones futuras como: análisis de la claridad del lenguaje administrativo, comparación de calidad del español por épocas, o la detección de errores comunes en asistentes de voz y otros dispositivos de IA, entre otros.
- Acuerdo entre la RAE y Microsoft para potenciar el buen uso del español en las aplicaciones de Microsoft 365, el buscador Bing, entre otros productos de la compañía, orientados a mejorar la revisión ortográfica y gramatical o crear modelos de traducción y transliteración automáticas basados en tecnologías de inteligencia artificial. Asimismo, se contribuirá a entrenar las aplicaciones

de inteligencia artificial de la compañía, poniendo el foco en evitar sesgos del lenguaje y en mejorar las habilidades de procesamiento del lenguaje natural. Para ello, la RAE ha facilitado el acceso al *Diccionario de la lengua española* (con cerca de 375 000 conjuntos de datos) a Microsoft, junto a las correspondientes actualizaciones anuales (del orden de entre 1000 y 2000 entradas).

- Junto a Telefónica, se han producido avances como el lanzamiento de Living App de la RAE de Movistar Plus para consultar definiciones o aprender más acerca del idioma y también la creación del juego de la RAE, disponible en el dispositivo de Movistar Home. Como novedad, Telefónica ha presentado el prototipo LEIA-X, una extensión para navegadores Chrome que usa IA para mejorar la comprensión del español destacando la acepción o significado más adecuado según el contexto.

OBSTÁCULOS SUPERADOS:

Uno de los principales obstáculos a superar es el establecimiento de mecanismos de asesoramiento que permitan analizar y verificar el desarrollo y la implementación confiable de tecnologías basadas en IA, así como de un marco ético y normativo que permita desarrollar una IA fiable y alineada con los principios éticos y sociales.

En este sentido, el reto pasa por concienciar sobre el desarrollo ético de la IA y establecer controles sobre esta tecnología para evitar casos en los que haya discriminación o miradas sesgadas, es decir, diseñar una IA que no proyecte ni aumente los sesgos en el uso del género o en los estereotipos y que cuente con una perspectiva humanista que garantice los derechos individuales y colectivos de los ciudadanos.

IMPACTO:

Con el proyecto LEIA se espera el siguiente impacto:

- Situar a España como un país puntero en inteligencia artificial, liderando, a nivel mundial, el uso de la lengua española en la IA.
- Promover la creación de empleo cualificado, estimulando tanto el talento español, como atrayendo talento global; e incorporando la IA como factor para mejorar la productividad de la empresa española y de la Administración pública.

CALENDARIO DE IMPLANTACIÓN Y REFERENCIA TEMPORAL:

El proyecto LEIA fue ideado y presentado en noviembre de 2019 en el XVI Congreso de la Asociación de Academias de la Lengua Española (ASALE). Desde entonces se han producido avances en su implantación

gracias a los acuerdos, convenios y colaboraciones con diferentes entidades y, su inclusión como eje tractor en el marco de la política impulsada por el PERTE de la Nueva Economía de la Lengua.

En este sentido, según se expone en el Real Decreto 632/2022, de 26 de julio, por el que se regula la concesión directa de una subvención a la Real Academia Española para la realización del proyecto LEIA, el calendario de implantación se ha desarrollado en las siguientes nueve etapas:

1. Creación de un departamento técnico (sep-22 a dic-24).
2. Observatorio de neologismos (sep-22 a mar-24).
3. Creación y desarrollo de herramientas de verificación lingüística (sep-22 a dic-24).
4. Creación y desarrollo de herramientas de respuesta automática (sep-22 a dic-24).
5. Recopilación de material basado en la diversidad (ene-23 a dic-24).
6. Desarrollo de medidas que garanticen la plena accesibilidad de las personas con discapacidad a las herramientas relacionadas con la IA (abr-23 a dic-24).
7. Mejorar las herramientas tecnológicas que se emplean en los trabajos de regulación de la lengua (may-23 a dic-24).
8. Retos, becas y trabajo con las universidades (sep-22 a dic-24).
9. Labores de digitalización de los fondos propios de la Real Academia Española y obras (sep-22 a dic-24).

La fecha estimada para su finalización es el 31 de diciembre de 2024 y la duración máxima es de tres años.

DOCUMENTACIÓN DE CONSULTA Y APOYO:

- ¿Qué es LEIA? https://www.rae.es/noticia/que-es-leia#:~:text=El%20proyecto%20Lengua%20Espa%C3%B1ola%20e,el%20apoyo%20inicial%20de%20Telef%C3%B3nica.
- La inteligencia artificial también a la RAE https://www.lagaceta.com.ar/nota/993893/sociedad/inteligencia-artificial-tambien-rae.html
- Real Decreto 632/2022, de 26 de julio, por el que se regula la concesión directa de una subvención a la Real Academia Española para la realización del proyecto Lengua Española e Inteligencia Artificial (LEIA), en el marco del Plan de Recuperación, Transformación y Resiliencia. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-14118
- Resolución de 11 de marzo de 2022, de la Secretaría de Estado de Digitalización e Inteligencia Artificial, por la que se publica el Convenio con la Real Academia Española, para el desarrollo del

proyecto Lengua Española e Inteligencia Artificial, en el marco de la Estrategia Nacional de Inteligencia. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-4496

- Real Decreto 156/2022, de 1 de marzo, por el que se modifican el Real Decreto 139/2020, de 28 de enero, por el que se establece la estructura orgánica básica de los departamentos ministeriales, y el Real Decreto 403/2020, de 25 de febrero, por el que se desarrolla la estructura orgánica básica del Ministerio de Asuntos Económicos y Transformación Digital. https://www.boe.es/buscar/doc.php?id=BOE-A-2022-3292
- El Gobierno y la Real Academia Española unen fuerzas para impulsar el proyecto LEIA y potenciar el español en el desarrollo de la inteligencia artificial https://www.rae.es/noticia/el-gobierno-y-la-real-academia-espanola-unen-fuerzas-para-impulsar-el-proyecto-leia-y
- La RAE y AWS presentan una herramienta basada en inteligencia artificial para conocer el estado del español en Internet https://www.rae.es/noticia/la-rae-y-aws-presentan-una-herramienta-basada-en-inteligencia-artificial-para-conocer-el
- La RAE y Microsoft firman un acuerdo para potenciar el buen uso del español en la inteligencia artificial https://www.rae.es/noticia/la-rae-y-microsoft-firman-un-acuerdo-para-potenciar-el-buen-uso-del-espanol-en-la
- La RAE y Telefónica avanzan en su colaboración en inteligencia artificial con retos lingüísticos para dispositivos del hogar e investigación académica https://www.rae.es/noticia/la-rae-y-telefonica-avanzan-en-su-colaboracion-en-inteligencia-artificial-con-retos
- Telefónica presenta LEIA-X, la nueva extensión para mejorar la comprensión lectora basada en la inteligencia artificial https://www.telefonica.com/es/sala-comunicacion/blog/telefonica-presenta-leia-x-la-nueva-extension-para-mejorar-la-comprension-lectora-basada-en-la-inteligencia-artificial/