

Intervenciones experienciales en procesos de aprendizaje automático

David Serra Navarro, Universitat de Girona, David.serranavarro@udg.edu,
<https://orcid.org/0000-0001-7062-647X> //Recepción: 19/07/2023,
Aceptación: 24/11/2023, Publicación: 08/02/2024

Resumen

Este ensayo pretende focalizarse sobre la apertura de nuevos procesos creativos vinculados con las herramientas de la IA, mostrándose una metodología de trabajo que más allá de sus resultados en el ámbito gráfico y las nuevas industrias creativas que se derivan, nos invita a desarrollar un marco exploratorio entre humano y máquina, pero también abrir en clave experimental y pedagógica una vía para dialogar en el entrenamiento de modelos vinculado al aprendizaje automático de la IA.

Palabras clave

Diseño gráfico; inteligencia artificial; aprendizaje automático; modelos de difusión; entrenamiento experiencial

Experiential interventions in machine learning processes

Abstract

This essay aims to focus on the opening of new creative processes linked to AI tools, showing a work methodology that beyond its results in the graphic field and the new creative industries that arise, invites us to develop an exploratory framework between human and machine, but also open in an experimental and pedagogical way to dialogue in the training of models linked to AI machine learning.

Keywords

Graphic design; artificial intelligence; machine learning; diffusion models; experiential training

Introducción

La actual irrupción de la inteligencia artificial (IA) sobre la sociedad, la tecnología y la economía implica una profunda reflexión de cómo debemos tratar el impacto real de estas herramientas en un futuro que ya está en la esquina. Algunos de los aspectos clave que conlleva la IA se relacionan directamente con nuevos procesos de automatización, control de la eficiencia, cambios en el mercado laboral, mayor experiencia de usuario, e inevitablemente como abordar su vertiente ética (Cukier et al., 2021). Este ensayo pretende focalizarse sobre la apertura de nuevos procesos creativos vinculados con las herramientas de la IA, mostrándose una metodología de trabajo que más allá de sus resultados en el ámbito gráfico sobre las nuevas industrias creativas que se derivan, nos invita a desarrollar un diálogo exploratorio entre humano y máquina.

En los diferentes posicionamientos para contextualizar la IA, podemos identificar una serie características que se nos presentan asociadas a mejoras como: el acceso de la información, la traducción y la comunicación (Vargas-Sierra, 2020), la automatización de tareas rutinarias, la educación personalizada (Sánchez y Lama, 2007) o aplicaciones de ayuda a profesores y alumnado (Kiecza, 2022). No obstante, pese a este aparente potencial para democratizar el conocimiento, también plantea desafíos y riesgos. La brecha digital, la privacidad de los datos y los sesgos algorítmicos son algunos de los problemas que deben tratarse para garantizar que la democratización de este conocimiento vinculado al tratamiento de los datos, sea inclusiva y equitativa (Flores-Vivar y García-Peñalvo, 2023). Es este punto, en la superposición de información que nutre el aprendizaje de la máquina, momento previo en el que los algoritmos interpretaran resultados para su posterior validación humana-diseñadora, es donde iniciamos la metodología de trabajo.

Contexto: la IA y los modelos de difusión

A modo de esquema general se establecen diferentes niveles para clasificar la IA, un punto de partida podría ser describirla como la máquina que es capaz de imitar el razonamiento humano. Remontándonos en la historia, este concepto ya se nos relaciona con el autómatas rudimentario *Ars Magna* que ideó Ramón Llull en el siglo XIII, considerado como el primer intento de utilización de medios lógicos para producir conocimiento, ideas que posteriormente en el siglo XVII Leibniz retomaría con su libro *Ars combinatoria*, o que en su contraposición se encadenan con la negación de René Descartes en la que apuntaba la imposibilidad que una

máquina pudiera ser inteligente (González, 2011). Actualmente la IA es un término tan amplio, ya sea como sistema informático o conjunto de ideas, que nos es necesario comprender que no se trata únicamente de un sistema de lógica, sino de una tecnología que imita las funciones cognitivas que los humanos asocian con otras mentes humanas (Estupiñán et al., 2021); de esta manera encontramos complejos algoritmos que tienen la capacidad de aprender, y que no son simplemente órdenes programadas para realizar tareas específicas. Precisamente esta capacidad nos sitúa en subconjunto de IA en el que las personas proporcionan datos de entrenamiento a las máquinas, nos estamos refiriendo a la noción aprendizaje automático, más conocido como *Machine Learning* (ML); un método de análisis de macrodatos (*Big Data*) que automatiza la construcción de modelos analíticos, y que deviene la rama de la IA basada en la idea de que los sistemas pueden aprender de datos, identificar patrones y tomar decisiones con una mínima intervención humana. Ya en un paso más avanzado dentro del ML, encontraríamos el subconjunto del *Deep Learning* (DL), un aprendizaje profundo que consiste en entrenar el algoritmo basado en la estructura de una red neuronal artificial con una gran cantidad de datos para que aprenda siguiendo los patrones en que el ser humano aprende.

Nuestro interés está ubicado en la zona ML, y especialmente en las plataformas de modelos generativos de imágenes basados en modelos de difusión, lo que significa que su función es generar datos similares a los utilizados en su entrenamiento.

Centrándonos en la generación de imágenes para su uso en el proceso estratégico del diseño, entendido como una relación de fuerzas cambiante a lo largo de la búsqueda de una solución (Morales, 2022:21), observamos el reciente auge de plataformas como Bluewillow, Dalle-2, Craiyon, Firefly, Midjourney... en las que el usuario realiza una petición escrita (*prompt*), lenguaje natural, y obtiene una serie de resultados-imágenes. De esta manera, prestaremos atención sobre una herramienta *open-source*, Stable Diffusion¹, ya que su comportamiento se basa en los modelos de difusión que son el marco de trabajo. Los modelos de difusión, que eclosionaron en 2020, son un tipo de modelo generativo de datos diseñados para aprender la estructura latente de un conjunto de datos y poder definir la forma en que los puntos de datos que se difunden a través del espacio latente. Esto quiere decir que al modelo

1. <https://stability.ai/blog/stable-diffusion-public-release>

lo entrenamos con imágenes/datos donde ensuciamos ciertos píxeles, con manchas, desenfoques (*blur*) o distorsiones que agregamos a propósito y le pedimos como salida la imagen correcta (la imagen original sin ruido). Entonces, la red neuronal del modelo aprende a quitar el ruido, es decir, transformar esas manchas (ruido) en la imagen original (Bagnato, 2022). Concretamente, Stable Diffusion trabaja con el modelo ML llamado *Latent Diffusion*, entrenado para eliminar el ruido de imágenes borrosas paso a paso con ruido gaussiano. Esta compleja ingeniería informática, descrita minuciosamente en trabajos de diferentes autores (Rombach et al., 2022; O'Connor, 2022; Ho et al. 2020) no es el objetivo de este ensayo. La reflexión troncal de esta propuesta es aproximarnos a las bases de una metodología creativa basada en el ML donde el aprendizaje sea recíproco entre humano y máquina, un punto híbrido en donde los resultados surjan de diálogo-proceso entre diseñadora y modelo de difusión.

Previo sobre preadizaje automático (ml)

Para poder comprender la intervención gráfica planteada nos es necesario explicar las acciones básicas que configuran el entrenamiento estándar para permitir a las máquinas aprender y tomar decisiones basadas en datos, sin la necesidad de una programación para cada tarea. Según extraemos del propio ChatGPT 3.5, para iniciar el proceso de aprendizaje automático debemos seguir la siguiente secuencia de pasos:

1. Recopilación de datos: Se recopilan datos relevantes para el problema que se desea resolver. Estos datos pueden ser numéricos, textuales, imágenes, etc. Cuanta más información de calidad se tenga, mejor será el aprendizaje.
2. Preprocesamiento de datos: Los datos recopilados a menudo contienen ruido, valores atípicos o pueden estar en formatos no adecuados para el modelo de aprendizaje automático. En esta etapa, se realiza un procesamiento para limpiar, normalizar y transformar los datos en un formato adecuado para su uso posterior.
3. Selección del algoritmo/modelo: Existen diferentes tipos de algoritmos de aprendizaje automático, como el aprendizaje supervisado, no supervisado, por refuerzo, entre otros. Se elige el algoritmo o modelo que mejor se ajuste al problema y los datos disponibles.
4. Entrenamiento del modelo: En esta etapa, se utiliza una parte de los datos recopilados (conjunto de entrenamiento) para entrenar el modelo. El modelo ajusta sus parámetros internos de acuerdo con los patrones y características presentes en los datos de entrenamiento. Du-

rante el entrenamiento, el modelo realiza iteraciones para mejorar su capacidad para hacer predicciones o tomar decisiones.

5. Validación y ajuste del modelo: Una vez entrenado, el modelo se evalúa mediante datos que no se utilizaron en el entrenamiento (conjunto de validación). Se analiza el rendimiento del modelo y se realizan ajustes para mejorar su precisión y generalización.
6. Evaluación del modelo: Después de validar y ajustar el modelo, se evalúa su rendimiento con datos adicionales (conjunto de prueba) que no se utilizaron en ninguna etapa anterior. Esto proporciona una medida objetiva de qué tan bien el modelo puede generalizar y realizar predicciones precisas en nuevos datos.
7. Implementación y despliegue: Una vez que el modelo ha sido entrenado y evaluado de manera satisfactoria, se puede implementar en producción para realizar predicciones o tomar decisiones en tiempo real sobre nuevos datos.

Es importante tener en cuenta que el proceso de aprendizaje automático puede variar según el tipo de algoritmo y el problema específico que se quiera abordar. Además, la calidad de los datos, la selección del algoritmo y los ajustes son aspectos críticos que influyen en el rendimiento y el éxito del modelo de aprendizaje automático. (ChatGPT 3.5, 2023).

Metodología creativa: ml + learning by doing

La exploración de la IA para aprovechar su potencial creativo aplicado al diseño nos abre múltiples posibilidades, tales como en una primera instancia la generación automática de contenido visual, la optimización de diseño, el reconocimiento y clasificación de imágenes, la personalización y recomendaciones al usuario, la edición y mejora automática de imágenes, y finalmente la característica que nos ocupa: El diseño generativo. En este punto que implica el desarrollo pautado ML, y en el que proponemos intervenir, cabe señalar que durante el entrenamiento el aprendizaje del significado de las imágenes/datos es un proceso iterativo y depende en gran medida de la calidad y cantidad de datos etiquetados disponibles, así como de la elección de un modelo adecuado y técnicas de extracción de características, además, de los avances en el uso de redes neuronales profundas para aprender y comprender el contenido de las imágenes.

Aferrándonos a la afirmación de Joseph Weizenbaum (1976) de que ninguna máquina debería nunca tomar decisiones de forma completamente autónoma o dar consejos que requieran, entre otras cosas, de la sabiduría, producto de experiencias

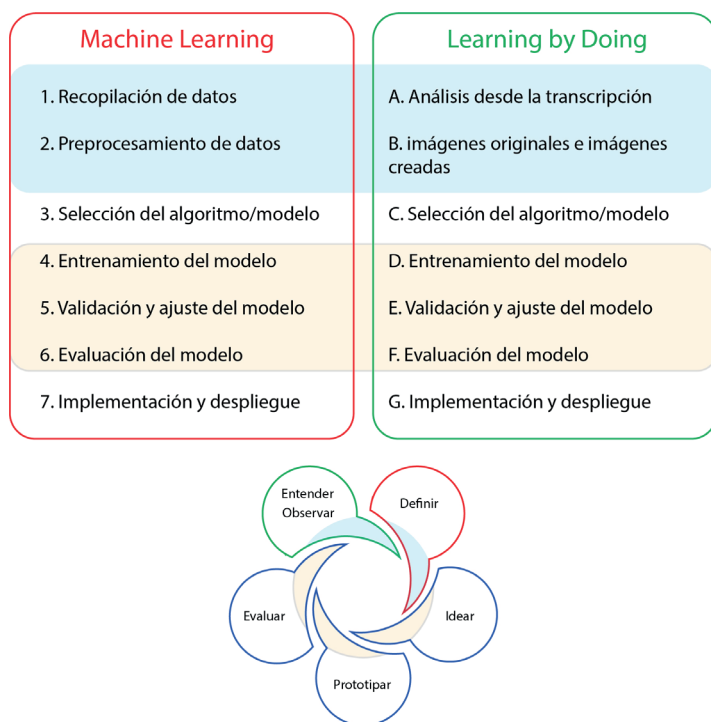


Figura 1. Representación de correlaciones en el proceso de entrenamiento ML+LBD (+Design thinking). Realización propia.

humanas, así como de tener en cuenta valores humanos, la metodología ML+LBD que se propone no es otra que el usuario y/o participantes puedan integrarse en la resolución del problema (imagen) alimentando con inputs experienciales cuya función sea modificar el resultado de los algoritmos y contribuir a mejorar la interpretación/comportamiento de la máquina. En *Experiential Learning* (1984), David Kolb ya nos describía el papel de la experiencia en el aprendizaje, estructurando en un ciclo circular cuatro etapas clave: Experiencia concreta, experimentación activa, observación reflexiva y conceptualización abstracta. En este ciclo que puede iniciarse en cualquier etapa, la importancia se centra en la retroalimentación y la reflexión en el proceso, y es el mismo proceso de reflexión que convierte la experiencia en educación experiencial (Joplin, 2008). Tomando este marco conceptual, donde el usuario y sus acciones permanecen en revisión continuada emulando el paralelismo de un proceso de calidad que se inserta en la rueda VSMA (Verificación, Seguimiento, Modificación, Acreditación), se propone hibridar parte del proceso técnico ML con la acción humana y su intervención en el proceso de diseño, y en cierta manera participar en la obtención/creación de datos para realizar el entrenamiento.

En la siguiente figura vemos como la aplicación sistemática del método Design Thinking (empatizar, definir, idear, prototipar y testear) mantiene la misma secuencia lógica que el funcionamiento ML, pero además, como valor añadido, si el usuario interviene en el proceso de la recopilación como generador de datos, como interpretador de códigos gráficos y cocreador de significados, los resultados que obtendremos implicarán adquisición de conocimiento en un marco exploratorio sobre los límites del ML.

Con el fin de clarificar con datos visuales la metodología y sus *outputs* inmediatos, tomaremos de ejemplo un caso práctico: Aprendizaje usuario/ML de una miniatura del manuscrito *Martirologio de Usuardo*² (1400) con la posible finalidad de expandir³ el gótico Internacional en clave contemporánea como modelo entrenado ML, para posibilitar su difusión y democratización cultural de este patrimonio en la red. Recordemos que la difusión es un proceso para convertir muestras de una distribución de ruido gaussiano en muestras de una distribución arbitraria y más compleja, como la de imágenes naturales (Somepalli et al., 2023:6049). En otras palabras, el funcionamiento muy simplificado del entrenamiento del popular *open-source diffusion model* Stable Diffusion es que a partir de cientos de miles de imágenes⁴ se añade ruido gaussiano a lo largo de una serie de pasos hasta que las imágenes pierden su identidad (*forward diffusion*). Cuando la intención es que se nos genere una imagen nueva (*reverse diffusion*) se trata de invertir el proceso anterior.

Cabe contextualizar que existen diferentes técnicas para condicionar la aleatoriedad en la generación de imágenes marcadas por un *prompt*, ya sea que el origen sea simplemente una imagen (*image-to-image*) donde se traduce el origen a una imagen de destino conservando ciertas propiedades visuales de la original, o controlando la desviación a través de las semillas (*seed*)⁵, que son un parámetro para conservar aspectos centrales de la composición y/o estilo, y poder modificar de for-

2. El *Martirologio de Usuardo* es publicado en PL 123, 452-992, y PL 124, 1-860, según la edición de Jean-Baptiste du Sollier, *Martyrologium Usuardi Monachi* (Amberes, 1714).

3. Véase plataforma abierta de descargas de modelos entrenados: <https://civitai.com/>

4. Las cantidades ingentes de datos que utiliza Stable Diffusion provienen de la empresa sin ánimo de lucro LAION, que, gracias a un rastreo de imágenes en Internet, ha logrado crear una base de datos en la que las fotografías han quedado asociadas a una etiqueta. <https://laion.ai/>

5. <https://getimg.ai/guides/guide-to-seed-parameter-in-stable-diffusion>



Figura 2. Patrimonio visual: *Martirologio de Usuardo* (1400). Ejemplo de obtención de registros no literales a través del análisis y dibujo(s) de una miniatura que se incluirán en el entrenamiento ML.

ma más detallada características visuales. En este sentido, la práctica desarrollada está ubicada en los métodos para entrenar modelos, entre los que destacan Dreambooth, Embeddings y Lora (*Low-Rank Adaptation*)⁶. Este último, el que se ha utilizado, resulta especialmente singular ya que requiere menor capacidad de memoria y permite cierta agilidad en el proceso. Sintéticamente podemos concluir que únicamente con un rango entre 15 y 30 imágenes de muestra es posible generar un modelo entrenado que acote las infinitas preferencias de salida con la finalidad de controlar los resultados. Así pues, los tres factores para esta metodología vendrán determinados en la combinación de: la creación de muestras, la gestión de los *prompts* vinculados a datos de LAION, y finalmente las decisiones humanas para seleccionar las semillas válidas.

Como observamos en la siguiente serie de imágenes, el proceso de obtención de datos se realiza inicialmente mediante la transcripción de las miniaturas, información aprendida en el proceso mimético e información que se implementará en el propio entrenamiento conjuntamente con otras imágenes construidas desde la interpreta-

ción de los propios usuarios. Señalar que quién dota de significado la imagen es el usuario, pese a los básicos condicionamientos de los *prompts*, por esta razón el entrenamiento que se muestra no basa toda la información extraída del *Big Data*, sino en la “semilla crítica” que ya empieza en el algoritmo de inteligencia humana (IH) que delimita una orientación en los posibles resultados. Es decir, condicionar lo aleatorio con datos experienciales. Como señala Simon Grennan (2022), la relación entre lo que se observa y lo que se dibuja representa una comprensión única del mundo, y la naturaleza creativa de dibujar permite la exploración mental, mediando ideas y creación de representaciones (Springer, 2023:118).

De esta manera, lo que se sugiere no es simplemente generar un modelo⁷ para subirlo en la red, sino que cada modelo se haya realizado desde el trabajo previo de análisis de la imagen y experiencial, con criterios humanos, y en este caso

7. Véase modelo preliminar Textual Inversion desarrollado en el proceso Martirologi: <https://civitai.com/user/weareartists470>. Stable Diffusion Textual Inversion es una técnica que le permite agregar nuevos estilos u objetos a sus modelos de texto a imagen sin modificar el modelo subyacente. Funciona definiendo una nueva palabra clave que representa el concepto deseado y encontrando el vector de incrustación correspondiente dentro del modelo de lenguaje.

6. <https://stable-diffusion-art.com/lora/>



Figura 3. Muestra de resultados obtenidos aplicando un modelo entrenado con los valores derivados de la metodología relatada.

el componente pedagógico reside en la acción de los participantes en el propio proceso. Hecho que retroalimenta la conciencia sobre el funcionamiento específico de la IA en cuestión de generación de imágenes, pero que también puede ayudar a entender las limitaciones asociadas a modelos preestablecidos que circulan de forma masiva entre las comunidades de usuarios IA.

Conclusiones

La metodología de trabajo relatada, un híbrido entre aprendizaje automático y aprendizaje experiencial mediante la interpretación humana, nos abre un marco exploratorio y de investigación que nos sugiere nuevas formas de interrelacionarnos como diseñadores con las herramientas IA. Es importante señalar que la representación adecuada del dibujo en formato numérico (vectores o matrices) es fundamental para su uso efectivo en el entrenamiento de modelos de *Machine Learning*, y que crear datos sintéticos basados en dibujos pueden utilizarse como una forma de retroalimentación en entornos de aprendizaje por refuerzo. Tanto en los

resultados mostrados, como en el material previo a esta contribución, podemos intuir diferentes niveles de aplicabilidad, pero en esencia estamos delante un planteamiento pedagógico, con tintes de educación experiencial (Roberts, 2012), para tratar de enfrentarnos a la IA desde diferentes posicionamientos, ya sea a modo profesional o como docente, y de forma transversal entre posibles contribuyentes de sentido. Esta metodología, experimental y vinculada al funcionamiento de los modelos de difusión, pretende poner en valor el proceso de aprendizaje humano en relación a los algoritmos, ya que la colaboración entre la creatividad humana y la creatividad de la IA puede dar lugar a resultados aún más sorprendentes y significativos. Tendremos que estar atentos a los avances de ML, pero también en como poder proyectar con estas herramientas sin perder la capacidad humana de sentir emociones, comprender el contexto cultural y social, y percibir la belleza y la originalidad en diferentes formas de expresión artística; todo esto sin que la máquina se nos apropie del sentido crítico.

Referencias bibliográficas

- Bagnato, Juan Ignacio (2022). *Creación de imágenes Stable Diffusion*, en: <<https://www.aprendemachinelearning.com/crea-imagenes-stable-diffusion-con-inteligencia-artificial-en-tu-ordenador/>> (Consultado: 24 de enero de 2023).
- ChatGPT. (2023, Apr 10, 17:17 GMT). *Prompt: pasos para un entrenamiento de modelo de difusión*. ChatGPT 3 - open.ai - Jan. 9
- Cukier, Kenneth; Mayer-Schönberger; Viktor, & De-Vericourt, Francis. (2021). *Framers. La virtud humana en la era digital*. Turner Publicaciones S.L. <https://bit.ly/3OWbocD>
- Estupiñán Ricardo, Jesús; Leyva Vázquez, M. Y.; Peñafiel Palacios, A. J.; & El Assafiri Ojeda, Y. (2021). Inteligencia artificial y propiedad intelectual. *Revista Universidad y Sociedad*, 13(S3), pp. 362-368.
- Flores-Vivar, Jesús-Miguel & García-Peñalvo, Francisco-José. (2023). Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de Calidad (ODS4). *Comunicar: Revista científica iberoamericana de comunicación y educación*, (74), pp. 37-47.
- González, Rodrigo (2011). *Descartes: las intuiciones modales y la inteligencia artificial clásica*. Alpha (Osorno), (32), pp. 181-198.
- Grennan, Simon (2022). *Thinking About Drawing. An Introduction to Themes and Concepts*. London. Bloomsbury Visual Arts.
- Ho, Jonathan; Jain, Ajay; & Abbeel, Pieter. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, (33), pp. 6840-6851.
- Joplin, Laura (2008). *On Defining Experiential Education*. In K. Warren, D. Mitten, & T. A. Loeffler (Eds.), *The Theory and Practice of Experiential Education*, pp. 16-23.
- Kiecza, Daniel. (2022, March 16). Classroom: Practice sets: A more personal path to learning. *The Keyword Google*. <https://bit.ly/3PNzF5G>
- Kolb, David A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Saddle River, NJ: Prentice-Hall.
- Morales, Luis Rodríguez. (2022). Raciocinio en diseño. *un año de diseñarte, mm1*, (24), pp. 12-23.
- O'Connor, Ryan (2022). *Introduction to Diffusion Models for Machine Learning*. AssemblyAI. Disponible en <<https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/>> (Consultado: 12 de diciembre de 2022).
- Rombach, Robin; Blattmann, Andreas; Lorenz, Dominik; Esser, Patrick and Ommer, Björn (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695.
- Roberts, Jay W. (2012). *Beyond learning by doing: Theoretical currents in experiential education*. Routledge.
- Sánchez Vila, Eduardo; Lama Penín, Manuel. (2007). Monografía: Técnicas de la Inteligencia Artificial Aplicadas a la Educación. Inteligencia Artificial. *Revista Iberoamericana de Inteligencia Artificial*, 11(33), pp. 7-12.
- Somepalli, Gowthami; Singla, Vasu; Goldblum, Micah; Geiping, Jonas, & Goldstein, Tom (2023). Diffusion art or digital forgery? investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6048-6058.
- Springer, Leonardo. (2023). *The value of sketching in teaching graphic design: Developing skills in a higher education institution*. grafica, 11(21), pp. 113-119.
- Usuardo., Giffart, P. François., Du Sollier, J. Baptiste., Universidad Complutense (Alcalá de Henares). (1718). *Usuardi San-Germanensis Monachi Martyrologium sincerum: ad autographi in San-Germanensi Abbatia servati fidem editum, [et] ab observationibus R.P. Sollerii ... vindicatum*. Parisiis: apud Petrum Franciscum Giffart
- Vargas-Sierra, Chelo. (2020). La estación de trabajo del traductor en la era de la inteligencia artificial. Hacia la traducción asistida por conocimiento. *Pragmalingüística*, (28), pp. 166–187. <https://doi.org/10.25267/Pragmalinguistica.2020.i28.09>
- Weizenbaum, Joseph. (1976): *Computer Power and Human Reasoning: From Judgment to Calculation*, San Francisco, W. H. Freeman and Company.

