

[04] ¿EN QUÉ CASOS TIENE SENTIDO USAR TÉCNICAS BASADAS EN INTELIGENCIA ARTIFICIAL?

La Inteligencia Artificial puede ser una buena herramienta para el progreso social, pero existe un peligro real de uso inapropiado, prematuro o malintencionado. Dado que no siempre es recomendable, es importante analizar su idoneidad en distintos ámbitos.

En todo caso, habría que distinguir entre la IA basada en el conocimiento y la IA basada en datos.¹ La IA basada en el conocimiento, que se empezó a usar a finales de los setenta, intenta modelar el conocimiento humano en base al diseño de reglas y algoritmos. En ella, los expertos diseñan las soluciones a partir del análisis de los problemas. La IA basada en datos, en cambio, también conocida comúnmente como aprendizaje automático, trabaja de forma ascendente a partir de grandes cantidades de datos y ha aparecido como una gran novedad ya entrado el siglo XXI. En este caso, podríamos decir que los expertos van experimentando con distintos esquemas para ver si con alguno de ellos, una ingente cantidad de datos les resuelve el problema.

En lo que sigue, nos centraremos en esta última, la Inteligencia Artificial basada en datos y aprendizaje automático, por ser la que genera más expectativas y más interés social. En este caso, se sabe y se ha demostrado que las aplicaciones del aprendizaje profundo a dominios que involucran comportamientos regidos por reglas y cuestiones humanas, como la toma de decisiones financieras, la gestión de recursos humanos o la aplicación de la ley, son problemáticas desde un punto de vista humanista¹.

Y es que estos sistemas, basados habitualmente en redes neuronales profundas,² son masivamente heurísticos,³ con un número ingente de parámetros que hay que ajustar durante su etapa de aprendizaje. El hecho de que este proceso de aprendizaje sea subóptimo y que utilice grandes conjuntos de datos que inevitablemente son sesgados, lleva a que estos sistemas sean menos fiables de lo que cabría esperar,⁴ con errores en sus resultados y decisiones.

En este contexto, y una vez aceptado el inherente margen de error de estos sistemas, podemos distinguir distintos tipos de usos:

- Aplicaciones no críticas para casos concretos, con post-supervisión: se trata de sistemas para aplicaciones específicas, que requieren que la persona interesada revise los resultados y corrija posibles errores. Un ejemplo típico serían los sistemas de traducción automática, ciertamente útiles cuando conllevan una revisión sistemática del texto traducido.
- Aplicaciones críticas para casos concretos, con post-supervisión: Como ejemplo tenemos los sistemas de ayuda al diagnóstico médico y de interpretación

de imagen médica. Existe un muy elevado consenso entre los expertos en la necesidad de post-supervisión, de manera que el diagnóstico final sea responsabilidad del experto.⁵

- Aplicaciones no críticas que funcionan en promedio y sin post-supervisión: En este caso, podemos citar el ejemplo de los sistemas publicitarios que nos envían mensajes con recomendaciones personalizadas de todo tipo, basadas en nuestro comportamiento anterior. Son sistemas éticamente discutibles, invasivos para las personas, pero que pueden ser beneficiosos para las empresas anunciantes: aunque no incluyan post-supervisión, los errores individuales no impiden que, en promedio, estos sistemas acaben generando un aumento global de ventas.
- Aplicaciones críticas que funcionan en promedio y sin post-supervisión: Como ejemplo, podríamos mencionar los sistemas personalizados de propaganda electoral fragmentada que intentan enviar mensajes electorales a medida, en concordancia con los intereses concretos de cada persona y con lo que desearía que le prometiesen.⁶ Como en el caso anterior, a pesar de sus errores pueden ser efectivos porque consiguen más votos de los que hacen perder. Pero son ética y democráticamente inaceptables. La investigación de la empresa Cambridge Analytica obligó a su cierre.

En resumen, los sistemas de Inteligencia Artificial con post-supervisión pueden ser útiles en muchos casos, como ayuda a la toma de decisiones de los expertos y en nuestra vida diaria (sistemas de reconocimiento de voz y música, sistemas aconsejan determinados productos, etc.). Pero su post-supervisión por parte de humanos es esencial. Por lo que respecta a las aplicaciones que funcionan en promedio y sin post-supervisión (críticas o no) deberían ser objeto de una regulación que garantice los derechos de todas las personas.

Notas:

1. Luca Steels y Ramón López de Mántaras (2018), "The Barcelona declaration for the proper development and usage of artificial intelligence in Europe", IOS Press: <https://content.iospress.com/articles/ai-communications/aic180607>
2. Véase la respuesta a la pregunta 2
3. Las técnicas heurísticas son métodos prácticos para la resolución de problemas que no garantizan un resultado óptimo ni perfecto, pero que alcanzan soluciones aproximadas aceptables y suficientes. Habitualmente, los sistemas heurísticos incluyen un conjunto de parámetros que hay que ajustar experimentalmente para lograr que su comportamiento sea adecuado. A este proceso de ajuste se le denomina "afinado" ("tuning" en inglés) por su similitud con el método de puesta a punto de automóviles, durante el cual se ajustan manualmente multitud de variables o parámetros. Los sistemas masivamente heurísticos se caracterizan por depender del ajuste de millones o miles de millones de parámetros. Sobre el ajuste de parámetros en sistemas heurísticos, véase por ejemplo: Steven P. Coy et al. (2000), "Using Experimental Design to Find Effective Parameter Settings for Heuristics", Journal of Heuristics, Volume 7: <http://yalma.fime.uanl.mx/~roger/work/teaching/mecbs5122/Papers/DOE/joh-2001-coy-heuristic%20doe.pdf>
4. Véase la respuesta a la pregunta 3
5. Emma Beede (2020), "Healthcare AI systems that put people at the center": <https://www.blog.google/technology/health/healthcare-ai-systems-put-people-center/> - Véase también: Will Douglas (2020), "Google's medical AI was super accurate in a lab. Real life was a different story": <https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/>
6. Véase, por ejemplo, el caso de la empresa Cambridge Analytica: Pere Brunet (2018): <https://virvigblogs.cs.upc.edu/2018/05/11/segmentar-el-missatge/>