

Inteligencia Artificial y Ciencia de Datos: ¿Para dónde apuntan?

Publicada el 3 de marzo de 2019

Escrita por John Atkinson Abutridy

Es preocupante que desde hace algún tiempo se ha ido generando una confusión creciente entre lo que es la Inteligencia Artificial (IA) y la Ciencia de Datos (Data Science), tanto en aspectos conceptuales como prácticos. Claramente todas las tecnologías de alguna forma están relacionadas en muchas aplicaciones, pero esto naturalmente no significa que sean lo mismo, tal como tratan de “vender” muchos profesionales y empresas. Más aún, cada día se están produciendo más estragos a nivel empresarial y laboral, con más del 90% de empresas que usando la “moda” (hype) para vender usan, aplican conceptos e incluso productos sin entenderlos realmente: empresas que supuestamente venden productos de IA pero en realidad son productos de data sciences, empresas que supuestamente venden productos de IA, que no tiene nada de IA, etc.

Siendo justos, me parece natural la confusión entre IA y Data Science ya que, a menos que alguien tenga conocimiento y/o experiencia muy especializada, el común de la gente no tiene porqué saberlo. Sin embargo, esto no tiene justificación cuando estas nuevas tecnologías se vician y mal usan para intentar vender y/o comunicar algo cool. Ejemplos hay muchos en el ámbito profesional/laboral, sin embargo, donde se me ha incrementado exponencialmente el número de canas ha sido recientemente al leer un anuncio de una empresa buscando “Especialista en IA” y en los requerimientos del cargo, se solicitaba “..conocimiento de metodologías scrum ... y estar certificado en Project Management...” (nada que ver!!), etc, y leyendo toda la lista, uno llega al punto final y aún no encuentra qué temas de IA debería conocer el candidato. En otros casos, el desconocimiento de lo que se publica es al menos más sutil: “.. con conocimientos de IA (chatbots)..” (sic), o incluso en perfiles de supuestos “científicos de datos” donde aparecen descripciones al estilo “Data Scientist (AI)” (sic), aludiendo que casi una cosa sería casi equivalente a la otra!!!. Pero definitivamente, un anuncio relacionado de trabajo que me ayudó a quitarme el hipo que tenía, es este: “...Chief Data Scientist que domine los conceptos teóricos de programación estructurada orientada a objetos...”(sic). Claramente, en estos casos, no sólo se evidencia una falta de comprensión total sino una mezcla de “peras con manzanas” para producir quizás algún efecto en la audiencia, o no sé qué...Sin embargo, esto es sólo la punta del iceberg, empresas que

"supuestamente" venden tecnologías/productos de IA en realidad o son cosas que no tienen que ver con IA (a propósito o por desconocimiento), o que en realidad son soluciones de analítica al estilo de data sciences.

Matrices más o menos, la IA es la ciencia de hacer que los computadores realicen tareas que si las hicieran los humanos serían catalogadas como inteligentes. Aunque existen teorías y modelos más profundos de lo que implica integralmente la "inteligencia", para propósitos de esta diferenciación, el foco será más práctico, y enfocado en lo cognitivo. La componente "inteligencia" de "Inteligencia Artificial" no viene de regalo, ni de puro nombre "rimbombante", sino como un conjunto de capacidades "combinadas" que permiten la realización de tareas o resolución de problemas (cognitivamente) complejos, de forma eficiente, muchas de las cuales están basadas en el uso intensivo de conocimiento (NO datos).

Estas capacidades usualmente incluyen aquellas que uno pensaría como similares en un ser humano: representación de conocimiento y razonamiento (Knowledge Representation ó KR), búsqueda de soluciones (Search), planificación de actividades (Planning), procesamiento de lenguaje natural (Natural Language Processing ó NLP), agencia ó inteligencia social (Multi-Agent Systems), robótica autónoma (Intelligent & Autonomous Robotics), visión artificial (Computer Vision), y aprendizaje automático (Machine Learning ó ML). Una característica común de antaño en muchas de estas capacidades es que están basadas en Conocimiento (knowledge-based) y NO basadas en datos (data-oriented) como ocurre en sistemas de procesamiento convencional de datos. Por otro lado, es importante destacar que no todo lo que tiene alguna palabra "parecida" al ámbito de IA, es IA. Por ejemplo, áreas como la robótica "a secas" han existido hace mucho tiempo y convencionalmente involucran un proceso de automatización más mecánico, y que es suficiente para varios ambientes productivos "estáticos". Sin embargo, cuando se requiere tareas de automatización en ambientes de producción dinámicos (ej. ensamble automotor, etc), claramente necesitamos robótica más "inteligente". Así también, visión por computador inicialmente no tiene que ver con IA, sino más bien con métodos geométricos, análisis tradicional de imágenes, etc. Otra historia es que las tareas de visión por computador se realicen con técnicas de IA y que generen una solución actual no sólo factible, sino también más eficiente y robusta que lo tradicional, y así sucesivamente, se puede seguir con varios ejemplos.

Cada una de estas capacidades constituye un área de trabajo independiente pero obviamente relacionada, de la IA, cuya complejidad es tal, que generalmente los científicos se especializan y profundizan en una de ellas, y las combinan a medida que poseen más experiencia. Sutilmente, he recalcado la naturaleza “integrada” o combinada de estas áreas debido a que sistemas prácticos de IA realmente inteligentes de verdad que aporta a un negocio, generalmente unen varias capacidades para resolver un problema en forma eficiente. Esto no sólo es importante por su relevancia práctica en los negocios, sino que se sustenta en la naturaleza concurrente en cómo operan los procesos cognitivos en nosotros, los seres humanos. ¿O acaso Ud. si habla, no puede caminar? ¿O cuando razona no puede aprender? ¿O cuando se mueve no puede interactuar con otros? ¿O cuando habla no puede ver?, etc. ¿Se da cuenta? La inteligencia es mucho más que hacer que un computador realice una tarea simple!! (bueno, algunos no pueden caminar y mascar chicle al mismo tiempo, pero eso es otra historia).

Cada una de las áreas de la IA ha generado desde sus inicios un sinnúmero de avances y aplicaciones, prácticamente desde los años 40. Sin embargo, desde mediados de los años 90, muchas de ellas han comenzado a impactar masivamente en los negocios. Una de ellas, conocida como minería de datos, es una tecnología que combina otras áreas existentes para extraer, inferir y/o descubrir patrones accionables en los datos para tomar mejores decisiones. Para que sea posible, la minería de datos "pidió prestado" y adaptó técnicas provenientes desde aprendizaje automático (ML), estadística, visualización, bases de datos, y computación de alto rendimiento (HPC).

Como parte de los requerimientos esperables en varias áreas de negocio, muchas de las tareas de “minado” de datos son más bien de naturaleza analítica en vez de descubrimiento de conocimiento per se, por lo que sólo se utiliza un conjunto muy reducido de técnicas de ML. Por otro lado, algunos algoritmos de minado de datos, debieron ser adaptados para abordar problemas en donde no sólo disponemos de muchos datos sino que su dinámica es alta, entre otras características. Interesantemente, esta problemática no es exclusiva de los negocios sino de cualquier problema científico-tecnológico, por lo que para abordarlo efectivamente generalmente se debería seguir el “método científico”, comenzando desde la exploración de

datos y planteo de las hipótesis, hasta llegar al desarrollo de los modelos e interpretación que dará la solución final al problema original.

Debido a esto, en los últimos años, en un intento por estandarizar y formalizar este proceso, se evolucionó a un área conocida como ciencia de datos (data sciences) que en términos simples es el área que permite resolver problemas analíticamente complejos utilizando mecanismos de inferencia de datos (ej. Estadísticos, Basados en ML, etc), desarrollando algoritmos y tecnologías asociadas. Su objetivo último es utilizar los datos para producir formas creativas de generar valor a un negocio.

¿Se da cuenta? Por el lado que se mire, la IA no tiene nada que ver con data sciences. Una historia diferente es que data sciences “pida prestado” una parte de ML desarrollado por IA para hacer su “pega” (ML data del año 1959!!), por lo que están conectados de alguna forma, como podrían estarlo obviamente otras tecnologías y/o aplicaciones. Un detalle importante a destacar es que la ciencia de los datos sin el uso de ML podría igual hacer su “pega”, tal como se ha hecho con herramientas estadísticas tradicionales y sus aplicaciones en SPSS, Matlab, etc. De hecho, por ejemplo, una tarea de predicción utilizando métodos de regresión tradicionales permite resolver un problema, sin el uso de ML, y mantiene su paraguas de data sciences, por los pasos que implica llegar al resultado.

Entonces, ¿Una gran parte de las empresas (y muchos profesionales) que dice que vende/hace IA realmente no lo hace y vende otra cosa (ej. Data Sciences), por desconocimiento?

Para responder lo anterior me remontaré a mi época de estudiante de pregrado en los 80s. Un día estaba en la cola para almorzar en el casino de la Universidad, y de aburrido, me puse a ver cuanto anuncio y publicidad había pegado en las paredes. Uno de estos anuncios me llamó profundamente la atención, pues marca un patrón muy transversal en la sociedad que vivimos, especialmente cuando se trata de modas y hypes. En este anuncio había la imagen de excremento y moscas alrededor, y abajo una leyenda que decía: “Coma excremento, millones de moscas no pueden estar equivocadas”.

Fuente: <https://www.linkedin.com/pulse/inteligencia-artificial-y-ciencia-de-datos-para-d%C3%B3nde-john/>