Seminario 'Huella digital ¿servidumbre o servicio?

Explotación de datos y verdad científica

(Síntesis de la sesión del 21 de enero de 2021)

El pasado 21 de enero, el comité de expertos del Seminario *La huella digital: ¿servidumbre o servicio?* celebró su octava sesión por videoconferencia. Sin apartarse del enfoque interdisciplinar propio del Seminario, y con las diferencias de enfoque que ello implica, en esta sesión se ha debatido sobre la explotación de grandes datos y su relación con la verdad científica.

La sesión tuvo inicio con la ponencia de **Alfredo Marcos**, Catedrático de filosofía de la ciencia de la Universidad de Valladolid, y continuó con los comentarios de **Sara Lumbreras**, investigadora en el Instituto de Investigación Tecnológica de la Universidad Pontifica Comillas y **Moisés Barrio**, Letrado del Consejo de Estado y Director del Diploma de alta Especialización en Legal Tech y transformación digital (DAELT) de la Escuela de Práctica Jurídica de la Universidad Complutense de Madrid. A continuación, tuvo lugar un amplio debate en el que participaron todos los presentes (lista de participantes adjunta).

Haciendo referencia a los debates de sesiones anteriores y a la temática general del seminario, Alfredo Marcos toma como referencia la posición de John Dewey sobre la experiencia humana interactiva: dejamos nuestra huella sobre las tecnologías digitales, y este proceso deja huella en nosotros. La técnica no es neutral en términos absolutos: es necesaria y forma parte de la vida humana. En cada etapa de desarrollo, elegimos entre líneas de desarrollo tecnológico, unas mejores que otras. Cada nueva tecnología cambia nuestras capacidades, pero también cambia nuestras necesidades: servidumbre y servicio van de la mano, y hay servidumbres que no son aceptables. Para el ponente, la respuesta está en una actitud de serenidad y de cierto desprendimiento ante lo tecnológico, rechazando aquellos desarrollos inspirados en hipótesis y objetivos que tienen efectos devastadores para la vida humana.

Ontología y denominación de la Inteligencia Artificial

En 1956, el informático John McCarthy acuñó el término "Inteligencia Artificial" (IA) en la famosa conferencia de Dartmouth. Se ha cumplido sobradamente lo que aquella visión profética suponía en términos de potencia de cálculo, clasificación y análisis. Pero ¿existe realmente algo que se pueda llamar en rigor "inteligencia artificial"? En la visión de Alfredo Marcos, lo que la IA tiene de inteligente se lo pone el ser humano (no es artificial), y lo que tiene de artificial no es inteligente. Simular la inteligencia no es ser inteligente. En todo caso, habría que hablar de *sistemas de inteligencia artificial*, de los que forman parte quienes diseñan y utilizan las máquinas y sus programas.

Varios autores denuncian hoy un "mito de la IA" y las teorías de ciencia ficción sobre la "IA fuerte" apartan el foco del debate de la realidad actual, en la que se están delegando realmente funciones a las máquinas, a veces sin los necesarios controles.

El término IA puede ser desorientador. Se han propuesto distintos términos para sustituirlo, como el de *inteligencia asistida* o *ampliada*, o el de *inteligencia artificial centrada en humanos*. Alfredo Marcos propone hablar de *sistemas de control delegado* (CoDe). Sara Lumbreras prefiere hablar de *sistemas de apoyo a la decisión*. Otros participantes en el seminario rechazan entrar en un debate de terminología y piden que se hable de hechos.

Pero la cuestión no es solo de lenguaje, insisten los filósofos, sino de concepción de la realidad: la máquina, sin persona humana para interpretarla, no es más que materia sometida a cambios físicos. La confusión a la que puede llevar el término IA ha sido alimentada por los medios de comunicación, que han contribuido a la mitificación de la tecnología basándose en las promesas de la IA, y no tanto en sus aplicaciones actuales.

Si se toma el concepto de inteligencia como capacidad de entender y resolver problemas, se debe aceptar que el término aplicado a una máquina no es más que una aproximación: una máquina no es capaz de entender, ni de percibir problemas; las problemáticas que ha de resolver son siempre del hombre.

Para los que critican el uso del término IA, dicha nomenclatura devalúa la inteligencia y genera una concepción dualista, como si en la condición humana, la inteligencia se pudiese separar de la persona. Además, al categorizar un sistema como "inteligente" se tiende a justificar la delegación de responsabilidad. Es decir, si es la "inteligencia" del algoritmo la que ha tomado una decisión, se tenderá a atribuirle unas consecuencias consideradas como inevitables. Si, al contrario, se comprende que la máquina no tiene

una experiencia real de los datos que se le han introducido, y que estos datos no tienen valor fuera de la inteligencia humana, se llega a la conclusión necesaria de que la IA no puede ser otra cosa que una ayuda, un dato adicional para tomar la decisión más acertada.

Tómese, por ejemplo, el caso del número: la máquina no es capaz de sumar dos más dos, ya que no existe en ella la noción de número o de cantidad. Hace unas décadas, los ingenieros aprendían a programar desde lo básico; así entendían que en una máquina sumar uno a un número dado se consigue al cambiar su último bit (si es 1 se pone 0 y si es 0 se pone 1, en el primer caso se cambia recursivamente el bit anterior). Esta operación tan básica toma infinitas formas complejas, pero es clave para comprender que en las máquinas no hay números, solo posiciones de memoria ordenadas que almacenan unos y ceros. Esa es la base sobre la que se ha construido el edificio de las ciencias de la computación. Pero los programadores y científicos de datos que actualmente utilizan los modelos de IA para la toma de decisiones no siempre han conocido el nivel más básico de estos modelos, sino que utilizan softwares que simplifican los procesos. Por ello les resulta a veces más difícil ver que dentro de la máquina, no hay realmente conocimiento, no hay conceptos. Tener esto claro es base para saber qué se puede esperar y qué se puede delegar a los modelos de IA. La IA no comprende los datos, sólo los manipula. Decir que los comprende, sería como defender que una cacatúa, que se limita a repetir unos sonidos que imitan el habla humana, es verdaderamente capaz de comunicar un conocimiento, un sentimiento o un pensamiento propio.

El modo de conocer de las máquinas

Más allá de la terminología, es evidente que existen unos sistemas, conocidos como IA, que se utilizan como herramientas de gestión. Para poder orientar las decisiones correctas en materia de automatización, es necesario comprender cómo funcionan estos sistemas, qué tipo de ayuda brindan en la toma de decisiones, y lo más importante: establecer sus límites de aplicación. En resumen, los algoritmos de IA ejecutan sus funciones detectando líneas de correlación entre enormes cantidades de datos. Pero no descubren relaciones causales, es decir: no permiten comprender ni explicar el fenómeno. Los procesos de la inteligencia humana se pueden categorizar en inducción y deducción, a las que Charles S. Peirce (1839-1914) ha añadido la categoría de

abducción, o sea el paso creativo de la formulación de conjeturas y de hipótesis. El algoritmo procede por inducción, no por deducción, y menos aún por abducción.

Los algoritmos funcionan gracias a cantidades ingentes de datos; en el histórico de datos se encuentran patrones que se aplican a nuevos escenarios sobre los que hay que decidir. No sería extraño que, en función de la composición de los datos utilizados, la máquina llegue en ciertos casos a resultados aparentemente correctos, pero realmente falsos. Es como si un alumno estuviese aprendiendo únicamente a través de ejemplos de problemas resueltos, en los que encuentra una lógica que posteriormente aplica para resolver nuevos problemas. Se produce una situación típica de problemas de la inducción, en los que todas las premisas son verdaderas, pero la conclusión puede ser falsa porque no se cumple la regla de estabilidad del contexto ("rebus sic stantibus"). Como en cualquier diagnóstico, hace falta el juicio clínico para saber si una conclusión es o no es válida. Por lo tanto, en una realidad cambiante, todos los modelos tienen un grado de incertidumbre. No es que un algoritmo sea capaz de predecir (hablar en futuro), sino que establece posibles modelos (habla en condicional). El algoritmo propone lo que algo podría ser, en caso de que se cumpliesen ciertas condiciones, basándose en la experiencia previa.

Aunque en muchas ocasiones la IA funciona de forma correcta y supone una herramienta efectiva, su modo de proceder puede presentar diversos problemas típicos. Entre estas problemáticas está un posible *sobreajuste*. Este se da cuando los datos que se proporcionan no son suficientes para que se pueda generalizar; entonces el algoritmo se fuerza y extrae patrones que no son extrapolables. Es como si el estudiante tuviera muy pocos ejemplos para hacer generalizaciones y comprender todo un nuevo contexto. Este error, difícil de detectar, es salvable cuando se puede acceder y auditar las reglas que ha derivado el algoritmo. Pero ello no se da en todos los casos, a menudo se trabaja con una "caja negra": no sólo la IA no comprende la conclusión a la que ha llegado, sino que tampoco permite comprender cómo ha llegado. Es decir, aunque sea posible examinar el código inicial con el que trabaja, no por ello se comprende el camino seguido por el algoritmo para tomar una decisión, eso permanece oculto, a menudo por razones comerciales o de propiedad intelectual. En estos casos la confianza en la IA como herramienta para la toma de decisiones se hace mucho más compleja, pues ¿cómo podemos confiar en algo que no somos capaces de comprender? ¿en algo que nos oculta su modo de proceder? A estas preguntas de algunos participantes, otros contestan que, a pesar de muchos siglos de investigación psicológica y neurológica cada vez más

profunda y detallada, tampoco se entiende completamente cómo procede el cerebro humano, y ello, de por sí, no quita legitimidad a las decisiones humanas.

Otro de los problemas recurrentes de la IA es el del *sesgo algorítmico*. El problema del sesgo aparece cuando el algoritmo se basa en variables consideradas inadecuadas. El ejemplo más conocido de problemas con el sesgo algorítmico es el de COMPAS (Perfiles de Gestión de Delincuentes Correccionales para Sanciones Alternativas, por sus siglas en inglés), un algoritmo de tipo caja negra utilizado en los Estados Unidos para predecir la reincidencia criminal en los presos. Este algoritmo, desarrollado por Northpoint, Inc. se ha utilizado durante años para decidir si a los presos se les concede o no la libertad condicional. Después de un análisis detallado de sus predicciones, la agencia de noticias ProPublica encontró que el algoritmo utilizaba la raza de los presos como una de las variables principales. Así, los presos afroamericanos, sin importar su historial, recibían predicciones peores que los blancos. Esto tenía origen en la base de datos que se había utilizado para nutrir el algoritmo, en la que los presos de color tenían una peor tasa de reincidencia. Todo esto deriva en un caso de injusticia racial.

Ejemplos como este hay muchos. La solución a estos problemas quizá podría llegar si los algoritmos permitiesen ver las reglas sobre las que infieren, si dejasen de ser cajas negras. Si las variables son transparentes y la aplicación de la IA es clara, se corre menos peligro de caer en este tipo de problemas, tanto en el sobreajuste como en el sesgo algorítmico. Pero no es fácil aplicar esta idea. En la experiencia real de los "data scientists", sólo los modelos complejos, que el usuario no puede comprender, son los que tienen posibilidad de funcionar de forma adecuada en los problemas reales. Por otro lado, dicen los operadores, si intentamos auditar y hacer explicables los sistemas, estos pierden su efectividad y su razón de ser, ya que se perdería todo el tiempo que se pretende ahorrar. Los modelos de IA se desarrollan para cumplir unas tareas y si se intentan hacer completamente transparentes, auditables o explicables, carecerían de sentido para cumplir dichos objetivos.

El debate no aporta sobre esto ninguna respuesta definitiva. Si no se pueden auditar y hacer transparentes muchos procesos algorítmicos existentes, debería ser posible, sin embargo, proyectar nuevos modelos que busquen ser comprensibles desde el principio. Para ello, los desarrolladores deberían tomar en cuenta variables transparentes y, para conseguir un diseño verdaderamente adecuado al contexto, contar con equipos interdisciplinares. Esto ya es un hecho en muchos campos de aplicación de la IA: por ejemplo, a la hora de desarrollar algoritmos de diagnóstico médico, a nadie se le ocurre

contar únicamente con ingenieros, sino con médicos especialistas en la rama en cuestión. Los patrones en los datos son sólo correlaciones, es el experto quien es capaz de relacionarlas con fenómenos conocidos y nuevas hipótesis comprobables. La transparencia y la inteligibilidad del proceso no sólo se refiere al algoritmo; ante todo, se trata de que sean claros los fines que se pretende alcanzar y los datos con los que se trabaja. La oscuridad que se mantiene en torno a la IA tiene su origen, más que en la necesidad de los procesos en sí, en estrategias comerciales de carácter monopolista – las de unas pocas grandes empresas de Estados Unidos – o en estrategias políticas – en opinión del ponente, las del gobierno chino - .

Sobre el uso actual de la IA

El planteamiento de una postura ética ante el desarrollo y aplicación de los modelos de IA se apoyará en el desarrollo por parte de las administraciones públicas de unas prácticas y de una legislación que cubran los aspectos aquí mencionados. Desde esta visión, Moisés Barrio propone un planteamiento positivo del uso de la IA en un "Estado algorítmico de derecho", en el que se adoptarían las herramientas tecnológicas de automatización solo cuando éstas supongan una mejora y cuando no afecten los fundamentos de legitimidad de la democracia.

El debate sobre el uso de la IA, aunque en ocasiones parezca un proyecto a futuro, es una cuestión actual. Diariamente se delegan diversas tareas y funciones a los modelos de IA. Tal es el caso de la bolsa, en la que se ha delegado ya casi todas las decisiones de inversión a los algoritmos, o lo que ocurre con cada persona y su teléfono móvil en el que ya se han delegado decisiones como qué ruta tomar en el coche o cuál es el mejor restaurante al que ir. Es difícil llevar la contraria a la IA a la hora de decidir en muchos aspectos, esto es algo que ya se puede ver en el día a día. De ahí que sea primordial una formación personal en virtudes, algo que va mucho más allá que el cumplir con una deontología, es decir con unas normas que usualmente las personas no hacen propias, sino que las aplican sin intervención de su libertad y conciencia.

Mediante una ética de las virtudes, y no una deontología de normas para un usuario pasivo, se puede salvar a la sociedad de la servidumbre de la tecnología y comprender el verdadero servicio de las herramientas como la IA. Como se ha dicho ya, toda tecnología cambia nuestras capacidades, pero también nuestras necesidades, así: servidumbre y servicio van de la mano. Pero es necesario analizar esto de forma

dinámica: el servicio y la servidumbre no siempre inciden simultáneamente y, en muchos casos, quienes se benefician no son los mismos que los que soportan las servidumbres. Un "estado algorítmico de derecho" supone también un acceso de todos a los medios digitales, como se vio en la sesión anterior, y ello está lejos de darse en la sociedad actual. Es necesario por consiguiente aplicar un discernimiento moral realista, que no toma como necesario todo lo que es técnicamente posible y busca, sin embargo, poner los medios disponibles al servicio del desarrollo humano.

Es decir que es necesario un juicio crítico que desmitifique la IA, que no la subestime ni la sobreestime, ya que no podemos renunciar a la misma: hay que seleccionar aquellas tecnologías que mejoran en lugar de socavar la vida humana.

Listado de asistentes:

1. **Albert Cortina**, Abogado, Experto en Transhumanismo Director del Estudio DTUM
2. **Alfonso Carcasona**, Consejero Delegado, AC Camerfirma
3. **Alfredo Marcos Martínez**, Catedrático de Filosofía de la Ciencia, Universidad de Valladolid
4. **Ángel Gómez de Agreda**, Coronel Jefe, Área de Análisis Geopolítico DICOES/ SEGENPOL
5. **Ángel González Ferrer**, Director Ejecutivo Centro Cultura Digital Consejo Pontificio para la cultura del Vaticano
6. **Carolina Villegas**, Investigadora de la Cátedra Iberdrola de Ética Financiera y Empresarial, Universidad Pontificia de Comillas
7. **David Roch Dupré**, Profesor de la Universidad Pontificia Comillas
8. **Diego Bodas Sagi**, Lead Data Scientist – Advanced Analytics, Mapfre España
9. **Domingo Sugranyes**, Director del Seminario de Huella Digital
10. **Esther de la Torre**, Responsible Digital Banking Manager, BBVA
11. **Francisco Javier López Martín**, Ex-Secretario General , CCOO de Madrid
12. **Gloria Sánchez Soriano**, Directora de Asesoría Jurídica de Tecnología, Costes y Transformación de grupo Santander
13. **Guillermo Monroy Pérez**, Profesor, Instituto de Estudios Bursátiles

14. **Javier Camacho Ibáñez**, Director de Sostenibilidad Ética y profesor de ICADE e ICAI
15. **Javier Prades**, Rector, Universidad Eclesiástica San Dámaso
16. **Jesús Avezuela**, Director General de la Fundación Pablo VI
17. **José Luis Calvo**, Director de Inteligencia Artificial en SNGULAR
18. **José Luis Fernández Fernández,** Director de la Cátedra Iberdrola de Ética Económica y Empresarial **ICADE**
19. **José Ramón Amor,** Coordinador del Observatorio de Bioética de la Fundación Pablo VI
20. **Juan Benavides**, Catedrático de comunicación **Universidad Complutense de Madrid**
21. **Julio Martínez s.j**., Rector, Universidad Pontificia Comillas
22. **Moisés Barrio**, Letrado del Consejo de Estado y Director del Diploma de Alta Especialización en Legal Tech y transformación digital (DAELT) de la Escuela de Práctica Jurídica de la Universidad Complutense de Madrid
23. **Paul Dembinski**, Director de Observatoire de la Finance (Ginebra)
24. **Sara Lumbreras**, Subdirectora de resultados de investigación, Profesora titular, Instituto de Investigación Tecnológica , ICAI, Universidad Pontificia Comillas
25. **Richard Benjamins**, Embajador de Data & IA, Telefónica