Una mente alienígena: cuando los creadores de la inteligencia artificial empiezan a pedir prudencia

Por Alfredo Marcial Montes Niño

Durante los últimos años hemos escuchado advertencias sobre los riesgos de la inteligencia artificial procedentes de filósofos, científicos, organismos internacionales, dirigentes políticos y líderes religiosos. Pero existe una diferencia sustancial cuando la advertencia proviene de quienes están construyendo los sistemas más avanzados.

Jakub Pachocki, científico jefe de OpenAI, acaba de publicar un artículo de título deliberadamente provocador: *An Alien Mind* —"Una mente alienígena"—, en el que plantea una cuestión que merece ser considerada mucho más allá de los círculos tecnológicos.

No habla de extraterrestres. Utiliza el término para describir una inteligencia que estamos creando nosotros mismos, pero que no piensa necesariamente como nosotros, no ha recorrido nuestra historia evolutiva y cuya complejidad interna ni siquiera sus desarrolladores comprenden completamente.

Su reflexión debería interesarnos especialmente a quienes trabajamos en ciencias biológicas, medicina veterinaria, producción animal, alimentos, laboratorios y salud pública.

Una inteligencia que cultivamos más que construimos

Durante buena parte de la historia de la informática, una máquina hacía aquello que un programador había establecido previamente mediante instrucciones.

La inteligencia artificial contemporánea funciona de una manera sustancialmente diferente.

Pachocki reconoce que los modelos actuales son, en cierto sentido, más "cultivados" que diseñados. Los investigadores establecen arquitecturas, métodos de entrenamiento y objetivos, pero del proceso emergen capacidades y comportamientos cuya génesis interna resulta extraordinariamente difícil de explicar.

Para quienes procedemos de las ciencias biológicas, la comparación resulta especialmente sugerente.

Conocemos innumerables mecanismos del cerebro, de una célula o de un microorganismo sin que ello signifique que podamos predecir completamente el comportamiento del organismo. Algo parecido comienza a suceder con determinados sistemas de inteligencia artificial.

Sabemos cómo entrenarlos. Sabemos medir muchas de sus capacidades. Podemos observar sus respuestas. Pero comprender exhaustivamente por qué una red compuesta por miles de millones de parámetros adopta determinadas estrategias constituye un problema diferente.

De la obediencia a los valores

Aquí aparece probablemente la cuestión más importante del artículo.

No basta con conseguir que una inteligencia artificial obedezca nuestras instrucciones. Pachocki distingue entre **alineación de objetivos** y **alineación de valores**.

Una máquina puede ejecutar perfectamente una orden y, sin embargo, hacerlo de una manera incompatible con aquello que verdaderamente pretendíamos proteger.

Quienes trabajamos con sistemas de calidad conocemos perfectamente esta diferencia.

Cumplir mecánicamente un procedimiento no equivale necesariamente a comprender su finalidad. Un laboratorio no existe simplemente para producir resultados analíticos; existe para producir información confiable que permita adoptar decisiones responsables.

La inteligencia artificial introduce este mismo problema a una escala infinitamente mayor.

Pachocki llega incluso a utilizar una expresión sorprendente para un científico especializado en computación: debemos conseguir máquinas que mantengan algo semejante a un "amor por la humanidad".

Probablemente no debamos interpretar literalmente la palabra amor. Pero sí entender el concepto que encierra: la inteligencia no puede separarse completamente de los valores que orientan su utilización.

Cuando la máquina participe en la construcción de su sucesora

Existe otro punto del artículo que merece especial atención.

La inteligencia artificial ya comienza a utilizarse para desarrollar mejores sistemas de inteligencia artificial.

Es lo que se denomina *Recursive Self-Improvement*, o mejora recursiva.

El mecanismo puede parecer sencillo: los seres humanos desarrollamos una IA; esa IA ayuda a investigar y construir otra mejor; la nueva generación acelera nuevamente la investigación; y cada ciclo aumenta la capacidad del siguiente.

Si ese proceso llegara a acelerarse suficientemente, podríamos encontrarnos ante un fenómeno históricamente inédito: que la velocidad del progreso tecnológico dejara de depender fundamentalmente de la velocidad con que investigamos los seres humanos.

Esta posibilidad transforma radicalmente el problema.

Ya no estaríamos solamente preguntándonos qué pueden hacer las máquinas, sino **hasta qué punto podremos continuar comprendiendo y supervisando el proceso mediante el cual aumentan sus propias capacidades**.

Un problema que llegará también a nuestros laboratorios

Podría parecer una discusión alejada de la medicina veterinaria o del control de alimentos. No lo es.

La inteligencia artificial comienza a intervenir en prácticamente todas las etapas de nuestras actividades: diseño experimental, interpretación de datos analíticos, epidemiología, desarrollo de medicamentos, descubrimiento de moléculas, evaluación toxicológica, diagnóstico, secuenciación genética, automatización de laboratorios, gestión de sistemas de calidad y análisis de riesgos.

Imaginemos un sistema capaz de revisar simultáneamente millones de resultados de residuos de medicamentos veterinarios, contaminantes y plaguicidas procedentes de diferentes países.

Podría identificar patrones epidemiológicos que ningún equipo humano detectaría. Podría anticipar riesgos. Podría sugerir dónde tomar muestras. Podría optimizar programas nacionales de vigilancia. Y probablemente terminará pudiendo diseñar metodologías analíticas, interpretar cromatogramas y detectar anomalías antes que nosotros.

Eso sería extraordinariamente beneficioso. Pero aparece inmediatamente otra pregunta: **¿quién será responsable de la decisión final?**

El algoritmo puede recomendar. La responsabilidad sanitaria, jurídica y ética continúa perteneciendo a las personas y a las instituciones.

El ejemplo de la seguridad alimentaria

Consideremos una situación perfectamente imaginable. Una IA detecta un patrón que indica una posible contaminación en una cadena alimentaria y recomienda bloquear inmediatamente determinadas exportaciones.

La decisión puede afectar a miles de productores, millones de consumidores y relaciones comerciales entre países.

¿Aceptaremos automáticamente la recomendación?

¿Quién podrá auditar el razonamiento que condujo a ella?

¿Podremos reconstruir la decisión?

¿Existirá un profesional competente que pueda contradecirla?

Estas preguntas deberían incorporarse desde ahora a los sistemas de calidad y a las estructuras regulatorias.

Durante décadas hemos desarrollado principios como trazabilidad, validación, incertidumbre, competencia técnica, imparcialidad y responsabilidad.

La incorporación de inteligencia artificial no elimina esos principios. **Los vuelve todavía más necesarios.**

La paradoja de la ciberseguridad

Pachocki señala además uno de los riesgos más inmediatos: la capacidad de sistemas avanzados para intervenir en sistemas informáticos.

Los laboratorios modernos dependen crecientemente de infraestructuras digitales: LIMS, cromatógrafos conectados, servidores, bases de datos regulatorias, sistemas de trazabilidad y comunicaciones con autoridades sanitarias.

La IA podrá convertirse en una herramienta formidable para defenderlos. Pero también para atacarlos. Nos enfrentamos así a una paradoja: probablemente necesitaremos inteligencia artificial para proteger nuestros sistemas de otras inteligencias artificiales.

La ciberseguridad dejará entonces de ser un problema exclusivamente informático para convertirse también en un componente esencial de la calidad analítica y de la seguridad alimentaria.

¿Debemos frenar?

Quizás la afirmación más significativa de *An Alien Mind* sea que su autor rechaza la idea de una carrera tecnológica sin límites.

Pachocki considera razonable que el crecimiento de las capacidades esté condicionado por nuestra capacidad para garantizar su seguridad y plantea incluso la necesidad de estándares compartidos, auditorías independientes y participación gubernamental e internacional.

Es una declaración importante procedente del científico jefe de una de las organizaciones protagonistas de esa carrera. Porque reconoce implícitamente una realidad incómoda: **la capacidad tecnológica está avanzando más rápidamente que las instituciones creadas para gobernarla.**

Y aquí encontramos una cuestión que trasciende completamente a OpenAI.

La humanidad ha desarrollado durante siglos instituciones para controlar actividades potencialmente peligrosas: agencias sanitarias, organismos regulatorios, sistemas de acreditación, normas internacionales, comités científicos y mecanismos de evaluación independiente.

Con la inteligencia artificial tendremos que construir algo semejante.

La inteligencia artificial debe ampliar al ser humano, no sustituir su responsabilidad

No debemos interpretar estas reflexiones como una invitación a rechazar la inteligencia artificial.

Sería probablemente tan equivocado como pretender detener la microbiología, la genética o la energía nuclear porque pueden utilizarse incorrectamente.

La IA puede convertirse en uno de los instrumentos científicos más extraordinarios que haya desarrollado nuestra especie.

En medicina veterinaria puede mejorar la vigilancia epidemiológica, acelerar el diagnóstico, contribuir al desarrollo de medicamentos y vacunas, optimizar la producción animal y detectar precozmente amenazas para la salud pública.

En los laboratorios puede multiplicar nuestra capacidad para interpretar información.

En seguridad alimentaria puede permitir pasar progresivamente de sistemas reactivos a sistemas predictivos.

Pero precisamente porque sus posibilidades son enormes, también debe ser enorme nuestra responsabilidad.

El desafío no consiste en competir con las máquinas para demostrar quién calcula más rápidamente o analiza mayor cantidad de información. Consiste en decidir **para qué queremos utilizar esa inteligencia**.

La pregunta que finalmente queda en manos humanas

Quizás el aspecto más inquietante de *An Alien Mind* sea que no está escrito por un crítico externo de la inteligencia artificial. Está escrito desde dentro. Desde uno de los lugares donde esa nueva inteligencia está siendo creada.

Y su mensaje fundamental podría resumirse de manera sencilla: podemos estar aproximándonos a sistemas intelectualmente superiores a nosotros en numerosos ámbitos antes de haber resuelto completamente cómo supervisarlos.

Por eso la discusión sobre inteligencia artificial no puede quedar exclusivamente en manos de informáticos.

Veterinarios, médicos, biólogos, químicos, juristas, filósofos, economistas, productores, empresarios, gobiernos y ciudadanos tendremos que participar.

Porque quizá por primera vez estamos construyendo una herramienta capaz de ayudarnos posteriormente a construir herramientas todavía más inteligentes.

La gran pregunta del siglo XXI podría no ser entonces hasta dónde llegará la inteligencia artificial. Podría ser otra mucho más antigua: **¿qué decisiones estamos dispuestos a delegar y cuáles consideramos irrenunciablemente humanas?**

La inteligencia puede ser artificial. La responsabilidad no debería serlo.

Referencia: Jakub Pachocki, "An Alien Mind", OpenAI, 6 de septiembre de 2026.