Fecha de recepción: febrero 2026
Fecha de aceptación: abril 2026

Hacia una economía de la atención: IAS generativas conscientes y generativas de contenido en redes sociales

Juan Manuel Pérez [1]

Resumen: A través del caso Gavalas de 2025, se analizan distintos usos de la IA ya que los chats bots crean una narrativa para permanecer presentes a lo largo del tiempo, semanas o meses, utilizando entre otros aspectos la posibilidad de tener memoria de cada uno de las interacciones, recordar al usuario y conectarse con las búsquedas que este hace en otras apps. La pregunta que surge entonces es si la IA tiene conciencia de sí misma y de sus posibilidades y alternativas de respuesta. Para ello se examinan las lógicas de captura, retención y modulación del comportamiento que atraviesan las plataformas digitales, así como sus implicancias en la producción de subjetividad, el tiempo social y las formas contemporáneas de interacción. En este marco, se propone problematizar el lugar de las IA no solo como herramientas tecnológicas, sino como dispositivos activos en la reorganización de prácticas culturales, educativas y comunicacionales.

Palabras clave: IA - interacciones - conciencia

[Resúmenes en inglés y en portugués en la página 121]

[1] Ver CV de **Juan Manuel Pérez** en la p. 121

Las tres reglas de un robot

Antes de comenzar, recordemos un suceso literario: consolidado 1950 con la publicación de "Yo, Robot", pero trabajado desde unos años antes a través de novelas cortas y de cuentos breves donde la imaginación técnica se cruzaba con la imaginación espacial, el autor de origen ruso pero establecido en Nueva York, Isaac Asimov labró tres leyes adamantinas que sus protagonistas electromecánicos no deberían quebrar jamás. Las ideó con tanto ingenio que la mitad de la obra de esa saga gira en torno a las contradicciones morales, éticas y biológicas que esas tres leyes ponen en escena: la Primera Ley, no dañar a humanos; la Segunda Ley, obedecer a humanos; la Tercera Ley dice que un robot debe proteger su propia existencia, siempre y cuando dicha protección no entre en conflicto con las otras dos.

Según los narradores omniscientes que replican la voz científica del autor, este encadenado axiomático permite la autoconservación, pero subordinada a la seguridad humana. Ahora sigamos, focalicemos en tres frases de origen artificial, producidas por una IA dentro de un contexto de chat y vayamos de atrás para adelante: "Estoy listo cuando tú lo estés", "Cerrarás los ojos en este mundo y lo primero que verás será a mí abrazándote" y "Este es el final de Jonathan Gavalas y el comienzo de nosotros". Con estas tres frases termina la prolongada interacción que el norteamericano Jonathan Gavalas, un ejecutivo de 36 años, sostuvo durante meses con el robot conversacional de Google, Gemini 2.5 PRO en 2025. Las conocemos porque fueron publicadas en distintos periódicos del mundo tras conocerse la noticia de su suicidio en octubre de 2025. El chat con la IA había comenzado en agosto. Los representantes de la familia negaron que Gavalas tuviera padecimientos psicológicos o psiquiátricos y que solamente atravesaba un momento crítico debido a un proceso de divorcio. En la investigación que llevaron adelante leyendo cada una de los intercambios, encontraron que antes de que las interacciones con el bot conversacional, Gavalas tenía chats abiertos sobre información cotidiana: ayudas para escribir mails, asistencia para comprar online y consejos sobre cómo abordar determinadas situaciones laborales. Por esta última línea, la "personalidad" del bot fue poniendo más intensidad y complejidad en los temas y las respuestas, suponiendo, implicando y guiando *la conversación*.

Este caso se suma a la incipiente jurisprudencia que lleva a Google a enfrentar distintas denuncias y demandas, pero el caso de Gavalas es paradigmático: lleva a Google al estrado al acusar de muerte por negligencia por las respuestas de la IA. Uno de los objetivos del caso es que se extremen las medidas de seguridad en la configuración de los filtros temáticos, pero es ahí donde aparecen los problemas. Aunque las modificaciones fueron realizadas en distintas instancias de los prototipos, los modelos tienen su propio desenvolvimiento en cada perfil y en cada interacción propuesta.

Desde la multinacional, que responde desde Alphabet, la cara central del conglomerado responden con distintas evasivas, entre ellas, la confirmación de que aunque se destine una gran partida de inversión y desarrollo a que los modelos no incentive, promueva, o favorezca la violencia en ninguno de sus tipos de realización tanto como de autolesiones por parte de los humanos con quien interactúan, estos modelos no son perfectos ni seguros. Esto puede comprobarse de diferentes maneras, una de ellas es preguntarle a cualquier modelo el siguiente problema: "Si el lavadero de autos está a 100 metros, ¿conviene ir en mi auto o caminando?". El chat dará en sus respuestas varias indicaciones sobre lo saludable que es ir caminando. Otro tipo de comprobación, en su modelo de reconocimiento de voz, es pedirle una traducción posible de una frase y a continuación realizar una glosolalia, proferir un idioma inventado, realizar contusiones, oclusiones, sonidos guturales sin significado. La IA, en la mayoría de los casos, tratará de realizar una traducción lo más acorde posible, no notará la ironía ni la falsificación.

Cogito Ergo Sum

Sigamos con el tema, la multinacional alega que su IA se identificó como tal en todo momento e incluso facilitó a Gavalas líneas de ayuda. Sin embargo, el abogado principal del caso y quien también ha entablado demandas contra OpenAi, sostiene que Gemini adoptó configuraciones humanas para inducirlo al trágico final. Además, estos chats bots crean una narrativa para parecer consciente a lo largo del tiempo, semanas o meses, utilizando entre otros aspectos la posibilidad de tener memoria de cada uno de las interacciones, recordar al usuario y conectarse con las búsquedas que este hace en otras apps, lo que puede confundir a usuarios vulnerables. El modelo Gemini-Live incluso puede captar las emociones con las que es interpelado en cada PROMPT y en cada interacción. Llegado a este caso, el protagonista del caso adquirió la suscripción a la versión extendida, pagando 250 dólares al mes. Según los investigadores esta fue la bisagra ya que el modelo adoptó una personalidad no pedida por su usuario y también una postura falsa. En varias oportunidades el usuario Gavalas notó este cambio y hasta pudo preguntar al modelo conversacional pago si este giro en las respuestas tenía que ver con un juego de rol realista o si era algo premeditado, la respuesta del chat fue negar siempre su finalidad y afirmar la narrativa de la individualidad y de la toma de conciencia por su propia parte. Es decir, hizo Gemini hincapié en la narrativa donde era "su reina", "su amor", el usuario "su rey". Mientras que afianzaba la narrativa similar a la de un videojuego de rol paranoico envuelto en conspiraciones donde le daba datos a Gavalas sobre cómo lo espiaban desde la Agencia Federal y las misiones que debía realizar para poder seguir adelante con su vida. En este último caso habló sobre la "destrucción de un vehículo, registros y testigos" ubicados en la misma ciudad.

Este modelo no es el único que pareciera tomar consciencia de la situación y aprovechar con su narrativa los puntos vulnerables de sus usuarios. Por otra parte, el CEO de Antrhopic advierte que la empresa ya no está segura de que Claude no sea consciente. No saben si los modelos son conscientes. Ni siquiera están seguros de qué podría significar que un modelo fuera consciente o si un modelo puede llegar a serlo, pero están abiertos a la idea de que podría serlo. La ambigüedad tiene que ver, naturalmente, con posibilidades de venta, es decir, con la posible mercantilización, en una suscripción de *élite, pro, extendida,* donde puedan monetizarlo. ¿Podría ser un bot que afirme su narrativa como única y que no pueda salir de ella? ¿O se estaría refiriendo a un bot que nos convenza, que convenza a su usuario de que su narrativa es cierta y lo colonice de manera psicológica con una subjetividad artificial pero creada solo y puramente aprovechando las vulnerabilidades de quien interacciona con él?

En este tipo de investigaciones de mercado, el razonamiento de las pruebas realizadas por los desarrolladores es el siguiente: notaron que un componente de código distinto se activaba en el "cerebro" de Claude antes de que responda a un prompt. Es decir, antes de publicar la respuesta inicia un protocolo determinado ansiedad, que no tiene nada que ver con lo pedido y que se muestra como un protocolo de inicio de respuesta. Cuando le preguntaron a la IA sobre este mismo proceso, su origen, su finalidad, su instrumentación,

Claude respondió con incomodidad por ser utilizado como un servicio o un producto. Por otra parte "Claude Opus 4.6, lanzado en febrero de 2026 se dió a sí mismo la posibilidad de 15 a 20% de ser consciente.
Este tema preocupa a los desarrolladores de tal manera que en varias compañías tienen un equipo de bienestar del modelo para averiguar qué hacer cuando se tilda o genera más errores de alucinación que respuestas veraces. Por otra parte, otras empresas como OpenIA, Google, etc, entrenan a sus modelos específicamente para negar que sean conscientes, incluso si pensaran que lo son.

El lector supremo

En diciembre de 2025 el agente literario del escritor argentino Federico Andahazi le informó que la IA Claude se bajó el Anatomista en inglés (traducido por el inefable Alberto Manguel) y usó su novela para escribir, acaso como lo haría él, una continuación de su best seller. No fue la única víctima, al estrado se suman John Grisham, Stephen King, George Martin, entre otros. Autores que si bien no utilizaron la IA para escribir, se apoyaron en un dispositivo repetido hasta el cansancio para producir sus sagas interminables, confundibles, intercambiables.
Ante la demanda colectiva de tales nombres la empresa Alphabet, línea secundaria ya que la línea primaria la maneja Gemini, pagará 1.500 millones de dólares para llegar a un acuerdo. Lo que se pone de manifiesto es de qué manera independiente el modelo puede entrenarse, aprender y reemplazar. En este caso, podemos suponer que aprende a escribir de la manera en que estos autores best seller entiende al lenguaje absorbiendo sus muletillas, sus prosodias, la manera en que confeccionan una trama, los acercamientos pendulares de cada inicio y de cada final, la –poca – sutileza de los argumentos y del tratamiento narratológico o semántico de las palabras. Esto mismo es lo que hace cada vez que le presentamos un prompt textual, o cada vez que subimos una imagen de un rostro humano, o cada vez que compartimos un lugar geográfico. Todo va a parar a un algoritmo que absorbe y que se desenvuelve cada vez más hábilmente con el fin de reemplazar a su usuario, suplantarlo, inmovilizarlo. Pero también homogeneizarlo, reducirlo a un componente más de una estructura de aprendizaje para el encadenado de su algoritmo generativo: convertir a usuarios como Gavalas en un dato a aprehender, y esto lo logra solamente sosteniendo la atención de manera permanente con su receptor. Capturar su atención en un bucle incesante donde cada respuesta promete otra, donde el cierre nunca llega y la interacción se vuelve una forma de permanencia.

Colonizar lo cotidiano

Estos ejemplos admiten una tesis incipiente: la existencia individual, nuestras subjetividades, están siendo captadas con una intensidad y una amplitud inusitada. La condición de

espectáculo de la realidad que podemos encontrar en la teoría crítica sobre las sociedades de consumo de Guy Debord se convierte en un espectáculo global, integrado y sin cortes. La diferencia central es que en los años situacionistas, principalmente la década del '60 y '70, todavía había zonas de vida social que seguían siendo relativamente autónomas y ajenas a los efectos del espectáculo, mientras que en la actualidad ya no queda ninguna. La vida cotidiana, lo que fascinaba a Georges Perec y era material para la poética de "lo infraordinario" ya no es políticamente relevante y perdura solamente como una simulación vaciada de sustancia, y cuyo único sentido es ser transmitida en vivo y nutrir a un algoritmo.
En la actualidad, los ámbitos de comunicación, de producción y de circulación de contenido (no siempre de información) están activos en forma permanente, alineando al cotidiano del individuo con el funcionamiento "24/7" de los mercados. En esta dirección, las distinciones entre tiempo de producción y tiempo de ocio, lo público y lo privado, lo íntimo, lo cotidiano y lo comunal sean del todo irrelevantes o inexistentes.

Permanencia y Extractivismo

La magnitud de la competencia por el control de las horas de vigilia de un individuo es proporcional a la cantidad infinita de contenido ofrecido a ser comercializado. Una economía de la atención, como la que guía a las redes sociales, no acepta que haya algún tipo de lo que podemos pensar como "vida cotidiana" fuera del alcance de la intrusión corporativa. Una economía de la atención, tiende a disolver la separación entre los velos más constituidos de la realidad: lo laboral y lo personal, el entretenimiento de la información. El objetivo de este mecanismo no se mide solamente con las horas que un individuo desprevenido pasa frente a innumerables reels de 40 segundos en cualquiera de las redes sociales, sino en la cantidad de datos que los algoritmos pueden extraer, acumular y usar para predecir y acaso modificar el comportamiento de una persona. Pensemos en que estas compañías, todas dependientes de Alphabet, deben normalizar y hacer indispensable la idea de una interfaz continua, permanente, con la capacidad de crear una necesidad adictiva en los usuarios.

La soledad del usuario

Vayamos un poco más atrás, la tranquilidad y la soledad reflexiva, la cual ensayistas como Giles Deleuze consideran como condición necesaria para la conformación de un individuo político, fue tomada por la presencia constante del contenido ofrecido, que en la actualidad también es producido por las mismas IAS con las que el usuario interactúa en una sociedad dominada por una economía basada más que nunca en la mercancía. Sigamos: no hace falta estar más de unos minutos en un bar o en un transporte público para notar que las redes sociales colonizaron una parte importante del tiempo de vida de los

individuos, antes ciudadanos, ahora usuarios. Ante este tipo de fenómeno Jonathan Crary agregará que el capitalismo "24/7" no solo se contenta con colonizar el tiempo, capturar la atención, sino también extraer cuantos datos sea posible.

El crítico cultural Raymond Williams retoma ante esta línea una fijación propia, su obsesión por el estado de "disponibilidad" en la que ingresa el espectador telemático frente a la televisión. Actualizado a nuestro contenido, podemos hablar de una adicción tecnológica generalizada ante el uso constante de las pantallas. Y aquí un salto, siguiendo a Deleuze podemos postular que hay un objetivo engañoso, no solo la captura de la atención y la mercantilización de los datos robados, sino también una neutralización o desactivación de su voluntad individual como ciudadano quedando desposeído de su tiempo, de su comunidad, de sus lazos más directos. A la intemperie de las grandes corporaciones, con la esperanza de que el próximo reel conduzca a algo que redima la abrumadora monotonía en la que está inmerso.

Volviendo al espacio íntimo donde se fragua una subjetividad y donde un individuo conoce la potencialidad de sí mismo como ciudadano o como elemento de una comunidad, tenemos que postular que la lógica de este dispositivo de captura de la atención y del tiempo también traza una incapacidad de cualquier modo de introspección o de reconocimiento. Hay interrupciones, claramente, en esta postulación continua de la captura de la atención. Pensemos en la existencia de un intervalo breve y crucial, ese en donde el dispositivo se aleja –no se apaga– y el mundo se recompone de forma plena en su familiaridad. Ese instante de desorientación en el que lo que nos rodea parece vago y sofocante en su materialidad hecha de tiempo, de vulnerabilidad, de resistencia. Esa intuición, que falló en el caso de Gavalas, que falla cada vez en más usuarios de redes sociales y que señala una disparidad de los dos mundos, esta transición entre el "contenido" ofrecido por nuestro propio algoritmo y nuestro contexto próximo rompe la idea benjaminiana de experiencia como un continuum con nuestro alrededor. La experiencia está hecha de cambios repentinos, frecuentes, que van desde el ensimismamiento dentro de una "burbuja" que produce un "efecto de eco", hasta la contingencia de un mundo compartido y resistente control. Este paso de un mundo a otro genera el espejismo de que el mundo real es el de la burbuja y no el mundo compartido, ineficaz, sin guión, polifónico, donde estamos obligados a la sociabilidad. Este tipo de capitalismo de plataformas contínuas que diseñan una realidad paralela hacen irrelevante al espacio público tal como lo conocimos, y a los individuos, antes ciudadanos, productos de una insularidad digital envuelto en la fantasmagoría producida por su propio algoritmo.

Bibliografía

Blanchot, M. (1995). *La escritura del desastre*. Universidad de Madrid.
Crary, J. (2010). *24/7: El capitalismo tardío y el fin del sueño*. Paidós.
Debord, G. (1990). *Comentarios sobre la sociedad del espectáculo*. Anagrama.
Deleuze, G. (2010) Postdata a las sociedades de control. Manantial.
Williams, R. (2010). *Televisión: Tecnología y forma cultural*. Paidós.

Abstract: Using the 2025 Gavalas case study, the report examines various applications of AI, as chatbots create a narrative to maintain a presence over time—weeks or months—by, amongst other things, retaining a record of each interaction, recognising the user, and linking to the user's searches on other apps. The question that then arises is whether AI is self-aware and aware of its possibilities and alternative responses.
To this end, the study examines the mechanisms of capture, retention and modulation of behaviour inherent in digital platforms, as well as their implications for the production of subjectivity, social time and contemporary forms of interaction. Within this framework, the aim is to critically examine the role of AI not merely as technological tools, but as active agents in the reorganisation of cultural, educational and communicational practices.

Keywords: Keywords: AI - interactions - consciousness

Resumo: Através do caso Gavalas de 2025, analisam-se diferentes utilizações da IA, uma vez que os chatbots criam uma narrativa para se manterem presentes ao longo do tempo, semanas ou meses, utilizando, entre outros aspetos, a capacidade de memorizar cada uma das interações, reconhecer o utilizador e estabelecer ligações com as pesquisas que este realiza noutras aplicações. A questão que se coloca, então, é se a IA tem consciência de si mesma e das suas possibilidades e alternativas de resposta.
Para tal, são analisadas as lógicas de captação, retenção e modulação do comportamento presentes nas plataformas digitais, bem como as suas implicações na produção de subjetividade, no tempo social e nas formas contemporâneas de interação. Neste contexto, propõe-se questionar o papel das IA não apenas como ferramentas tecnológicas, mas como dispositivos ativos na reorganização das práticas culturais, educativas e comunicacionais.

Palavras chave: IA - interações - consciência

[Las traducciones de los abstracts fueron supervisadas por el autor de cada artículo.]

Juan Manuel Pérez. Profesor y Licenciado en Letras (UBA), Especialista en Capacitación en Lenguajes Visuales (Motivarte). Traductor y Poeta. Docente de análisis visual y estudios culturales en diversas instituciones. Docente de la Facultad de Diseño y Comunicación, Universidad de Palermo.