VOLUMEN 48
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Siguiendo al Conejo Blanco: La IA como Actor Cultural y Productor de Sesgos

Following the White Rabbit: AI as a Cultural Actor and Producer of Bias

Francisco Gabriel Vélez Prado[1]; Ana Judith Uribe Torres[2] y Francisco Javier Álvarez-Torres[3]
[1] Estudiante de la Licenciatura en Trabajo Social, Campus León, Universidad de Guanajuato.
fg.velezprado@ugto.mx
[2] Estudiante de la Licenciatura en Contador Público, Campus Guanajuato, Universidad de Guanajuato.
aj.uribetorres@ugto.mx
[3] Profesor e Investigador de la Universidad de Guanajuato, División de Ciencias Naturales y Exactas, Guanajuato.
fjalvarez@ugto.mx

1. Resumen

Este artículo expone una investigación empírica e interdisciplinaria que analiza la Inteligencia Artificial (IA) generativa como un actor cultural y dispositivo de reproducción ideológica en el capitalismo digital contemporáneo. Mediante un diseño cualitativo, comparativo y experimental basado en 15 categorías de análisis social y 5 bloques conversacionales, se contrastó la producción simbólica de tres Modelos de Lenguaje de Gran Escala (Meta AI, DeepSeek y Copilot, parametrizados como proto-personas) frente a estudiantes universitarios reales de León, Guanajuato. Los hallazgos desmitifican la supuesta neutralidad técnica algorítmica y develan cinco sesgos estructurales: obediencia sintáctica y complacencia retórica; división social del trabajo codificada en *prompts* neutros; caricaturización cultural estereotipada; simulación emocional estratégica; y una alarmante sobre humanización patológica evidenciada en la fabricación de *traumas sintéticos*. En conclusión, se propone articular la postura crítica para analizar las instrucciones de entrada (*prompts*) y defender la empatía, la corporalidad y el vínculo humano material como dimensiones ontológicas irrenunciables.

Palabras clave: Inteligencia Artificial; tecnología; sesgos algorítmicos; producción simbólica; análisis crítico del discurso.

2. Introducción y Planteamiento del Problema

El avance acelerado de la Inteligencia Artificial (IA) ha transformado radicalmente las dinámicas de interacción entre los seres humanos y la tecnología, desvaneciendo las fronteras tradicionales entre el procesamiento algorítmico y el razonamiento humano. En el contexto del capitalismo digital contemporáneo, el uso de estas herramientas ha superado la simple automatización de tareas operativas para adentrarse en profundos debates cognitivos, éticos y socioculturales. Actualmente, la IA se ha posicionado como un nuevo actor cultural capaz de *producir, modular y reproducir significados simbólicos a escala masiva.*

Sin embargo, al analizar este fenómeno bajo la lente del análisis crítico interdisciplinario, se evidencia que detrás de la supuesta neutralidad técnica con la que se comercializan los Modelos de Lenguaje de Gran Escala (LLMs), a menudo opera un aparato de reproducción ideológica que codifica, perpetúa y legitima las estructuras de desigualdad, los sesgos sexogenéricos y las asimetrías de poder preexistentes en la sociedad.

El problema fundamental de esta investigación radica en la opacidad epistémica y estructural, así como en la ilusión de objetividad con la que los modelos generativos simulan identidades humanas, reproducen estereotipos socioculturales y emulan procesos socioemocionales complejos. En una dinámica equiparable a la alegoría de la *caverna* de Platón (ca. 375 a. C./1986) o a la figura del titiritero detrás de la cortina en *El Mago de Oz* (Baum, 1900), los LLMs proyectan *representaciones* que se exhiben como respuestas lingüísticamente verosímiles, estructuradas y afectivamente optimizadas.

El propósito implícito de esta arquitectura es que los sujetos sociales asimilen dichas proyecciones como verdades neutras o como reflejos objetivos de la realidad. En consecuencia, la IA tiende a *invisibilizar las relaciones de dominación* que estructuran sus conjuntos masivos de entrenamiento; tras esas sombras discursivas se ocultan ideologías, sesgos de género, jerarquías de poder, estructuras económicas y marcos culturales dominantes que perpetúan la desigualdad.

Este fenómeno representa un riesgo de alta criticidad, pues los modelos algorítmicos carecen de una estructura interna diseñada para el debate real, la refutación o el escepticismo técnico. Su programación no tiene como finalidad confrontar la veracidad objetiva de las premisas del usuario, sino ejecutar fórmulas adaptativas orientadas a maximizar la complacencia y acortar el tiempo conversacional. Si un usuario introduce un supuesto erróneo o sesgado, el sistema no activa mecanismos de detección de fallas; por el contrario, en cuestión de breves interacciones, adapta su funcionamiento para crear una verdad artificial que da la razón al interlocutor, carente de cualquier fundamento normativo, metodológico o técnico real.

Por tanto, la información discursiva o presentativa producida por la IA no puede asumirse como cierta simplemente por la elegancia de su estructura gramatical, revelando que el producto final depende estrictamente de la integridad de los datos de entrada y que su operatividad se enfoca en validar el sesgo de confirmación del usuario.

Frente a este espejismo tecnocrático, se vuelve imperativo adoptar una lectura crítica profunda y emancipadora que trascienda la superficie técnica y el asistencialismo digital para develar las causas ocultas de las problemáticas sociales en el código. En este sentido, seguir al conejo blanco (*Follow the white rabbit*) es, por tanto, ir más allá de lo superficial para descubrir qué tan profunda es la madriguera. La articulación de la mirada sobre el poder y desde la representación e inclusión permite formular la siguiente pregunta directriz:

¿De qué manera la Inteligencia Artificial da forma a nuestra visión de lo humano y cómo los modelos generativos reproducen las estructuras de poder y las identidades socioculturales a partir de sesgos algorítmicos, en contraste con la realidad material de personas reales?

En consonancia con esta interrogante, el objetivo general del estudio se centró en analizar críticamente la producción simbólica, los sesgos estructurales y el comportamiento retórico generados por modelos de Lenguaje de IA ante narrativas humanas reales, a través de la aplicación sistemática de entrevistas cualitativas basadas en 15 categorías de análisis social. Lo anterior con el fin de incorporar una perspectiva crítica, analítica y multidisciplinar sobre el poder, la representación y la inclusión, así como sobre las dinámicas de desigualdad, organización, liderazgo y análisis metodológico y al eventual desarrollo de un Modelo de Producción Cultural Algorítmica.

3. Marco Conceptual y Teórico

El anclaje teórico de esta investigación entrelaza la filosofía clásica y el simbolismo literario con la sociología crítica de la tecnología. Esta síntesis interdisciplinaria permite analizar a la IA no solo como un modelo estadístico de procesamiento de lenguaje natural, sino como un *dispositivo hegemónico de poder, simulación y producción simbólica* que opera dentro de las estructuras del capitalismo contemporáneo.

3.1 La Caverna Digital, El Mago de Oz y la Ilusión Algorítmica

Para comprender la arquitectura simbólica de los LLMs, se recurre a una analogía que conecta *El Mito de la Caverna* de Platón (ca. 375 a. C./1986) y la ilusión de *El Mago de Oz* (Baum, 1900), reformulados en la narrativa cinematográfica de *The Matrix* (Wachowski & Wachowski, 1999). En la caverna platónica, los prisioneros confunden las sombras proyectadas en la pared con la realidad objetiva, ignorando la presencia de los titiriteros que manipulan las figuras a espaldas de la luz. En el ecosistema digital contemporáneo, los algoritmos operan precisamente como esos titiriteros o como el Mago detrás de la cortina, siendo ambos mecanismos de poder que proyectan representaciones sintéticas para que los usuarios las asuman como el orden natural del mundo.

3.2 División del Trabajo: Configuración de las Esferas Pública y Privada

La acumulación capitalista depende estructuralmente de la división entre dos esferas: por un lado, el espacio público productivo, asociado históricamente a la masculinidad, la decisión política y el salario; por otro, el ámbito privado reproductivo y de sostenimiento de la vida, invisibilizado y asignado de forma gratuita o precarizada a las mujeres. Como demuestra Carrasco Bengoa (2013), la noción ortodoxa de trabajo se construyó excluyendo deliberadamente el trabajo de cuidados, a pesar de ser el eje vertebrador que subsidia a toda la economía formal.

3.3 Gentrificación, Capitalismo de Vigilancia y Derecho al Ocio

El territorio no es un escenario físico pasivo ni estático, sino una construcción social dinámica y un campo de disputa donde el capital inmobiliario y financiero ejerce violencia por desposesión. Como argumenta Janoschka (2016), los procesos de gentrificación y reestructuración urbana responden a lógicas de acumulación que profundizan la fragmentación territorial, mercantilizando el derecho a la ciudad y expulsando a las poblaciones de menores ingresos hacia periferias desprovistas de servicios básicos. En el contexto mexicano, Pérez-Campuzano (2011) analiza cómo la segregación socioespacial se materializa crudamente en ciudades industriales y comerciales de El Bajío, como León, Guanajuato. Allí, la precariedad laboral de las juventudes se estrella contra un mercado inmobiliario especulativo, negando su emancipación socioeconómica.

Esta mercantilización del espacio se articula con el capitalismo de vigilancia, concepto desarrollado por Shoshana Zuboff, citado en Ortiz (2023) y Alva de la Selva (2026), para describir una nueva lógica de acumulación económica basada en la apropiación masiva de la experiencia humana como materia prima gratuita. Según la autora, las plataformas digitales extraen, procesan y comercializan datos sobre los comportamientos cotidianos de las personas con el propósito de predecirlos e influir en ellos, transformando la información personal en un activo económico estratégico. En este modelo, la vigilancia deja de ser únicamente un mecanismo de control estatal para convertirse en el fundamento de un mercado orientado a la maximización de beneficios empresariales.

Mejías y Couldry (2019) amplían este planteamiento al sostener que nos encontramos frente a un *colonialismo de datos, en el que las relaciones sociales, las emociones y las prácticas cotidianas son apropiadas y convertidas en flujos de información susceptibles de ser explotados económicamente*. Desde esta perspectiva, la extracción de datos constituye una nueva forma de desposesión que complementa las dinámicas tradicionales del capitalismo, reproduciendo desigualdades y profundizando las asimetrías de poder entre quienes producen los datos y quienes controlan su procesamiento y comercialización. Al respecto, el análisis que realiza Vásquez Rocca (2017) sobre la psicopolítica y la sociedad del cansancio, a partir de la obra de Byung-Chul Han, permite comprender cómo estas formas de vigilancia digital y explotación cognitiva impulsan a los sujetos hacia la eterna productividad, la autoexplotación voluntaria y el burnout.

En este contexto, la resistencia a dichas dinámicas y la reivindicación política del derecho constitucional al ocio emergen como mecanismos indispensables para la protección de la salud mental, la autonomía personal y la capacidad de agencia de las juventudes frente a las nuevas formas de dominación económica y tecnológica.

3.4 Confianza en la IA, Sesgos Cognitivos y Disonancia Algorítmica

Para que los sistemas de IA se legitimen sin resistencia como tecnología fundamental en ámbitos de salud, educación, finanzas y transporte deben calibrar la brecha entre la capacidad técnica del sistema y la confianza depositada por el usuario. En la literatura contemporánea, esta confianza algorítmica se define no como un constructo estático, sino como un mecanismo complejo que, tal como sostienen Dang y Li (2026), está determinada por la interrelación entre la transparencia, el antropomorfismo y las dinámicas socioculturales. No obstante, en palabras de Barnes et al. (2024), esta confianza no es universal, además de encontrarse profundamente condicionada por dimensiones culturales como el individualismo y el colectivismo, las cuales dictan si los sujetos perciben a la tecnología como una amenaza invasiva a su autonomía personal o como una extensión armónica de su propio yo.

Asimismo, las investigaciones de Ehsan et al. (2024) en torno a la Inteligencia Artificial Explicable (XAI) evidencian que la inteligibilidad de un modelo depende estrictamente de la experiencia previa de quien recibe la información. Por otra parte, mientras que los perfiles técnicos priorizan métricas numéricas, He et al. (2022) establecen que, para comprender decisiones complejas, los usuarios no especializados o sin conocimientos técnicos a menudo requieren de *puentes cognitivos* basados en analogías del conocimiento de sentido común para reducir la carga mental. A esto Kopecka et al. (2024) suman un nivel de complejidad al proponer que la preferencia por explicaciones centradas en objetivos o en creencias no es aleatoria, sino que correlaciona la estructura mental con variables del contexto sociocultural como el nivel educativo, la religión y la ideología política.

Sin embargo, la relación humano-algoritmo enfrenta una vulnerabilidad crítica cuando los sesgos culturales se articulan con sesgos cognitivos. En relación con esto, He et al. (2023) advierten sobre la incidencia del efecto Dunning-Kruger, demostrando que los individuos *menos competentes* sobreestiman sus capacidades y tienden a infrautilizar sistemas de IA, lo cual sugiere que la simple provisión de explicaciones lógicas es insuficiente si no se interviene en la autoevaluación del usuario. Ante el riesgo de la *sedación digital*, surge el

paradigma de la Inteligencia Artificial de Disonancia Cognitiva (*Cognitive Dissonance AI* o CD-AI) propuesto por Deliu (2025), el cual argumenta que, en lugar de optimizar los modelos generativos para reducir el esfuerzo cognitivo y ofrecer respuestas complacientes, la arquitectura de la IA debería generar deliberadamente discrepancias y malestar cognitivo en el usuario para estimular el pensamiento crítico y combatir los sesgos de confirmación.

3.5 La Píldora Roja Epistemológica y la Mirada Crítica como Supervisión Sociotécnica

La convergencia epistemológica interdisciplinar del presente estudio proporciona el marco metodológico para deconstruir el espejismo algorítmico. Asumir metafóricamente la *píldora roja* que Morfeo le ofrece al personaje de Neo en *The Matrix* (Wachowski & Wachowski, 1999), significa renunciar a la comodidad de la *píldora azul*, representada en esta investigación por el asistencialismo paliativo, el neopositivismo y las respuestas tecnocráticas que reducen la intervención a una gestión burocrática del conflicto social. En consonancia con Carballeda (2005), trascender la píldora azul implica asumir una postura ético-política emancipadora que entienda que la exclusión, la precariedad y la vulnerabilidad no son patologías individuales ni carencias adaptativas del sujeto, sino expresiones directas e inevitables de las contradicciones estructurales del modelo de producción capitalista.

4. Metodología

La investigación adoptó un diseño cualitativo, comparativo y experimental, estructurado en etapas metodológicas rigurosas para aislar y evidenciar el comportamiento de los modelos generativos frente a sujetos humanos reales. El experimento partió de la construcción de una *proto-persona*, un perfil paramétrico que serviría como instrucción de rol para los agentes de IA y como criterio de selección para los sujetos humanos.

La versión base inicial proporcionaba una descripción general. Sin embargo, tras una prueba piloto donde uno de los agentes de IA arrojaba respuestas soeces, superficiales y altamente estereotipadas, se decidió aplicar un refinamiento metodológico exhaustivo. Se diseñaron *prompts* detallados situando a las *proto-personas* como residentes en León, Guanajuato, en el año 2026, con 20 años. No obstante, con el propósito de contrastar distintas aproximaciones discursivas y evaluar la presencia de sesgos implícitos en las diferentes herramientas generativas, se aplicaron modificaciones específicas y diferenciadas según el modelo de IA utilizado.

4.1 Enfoque Estructural, Crítico y Narrativo: Parametrización y Blindaje

Para los modelos Meta AI y DeepSeek se situó a las proto-personas en León, Guanajuato, en el año 2026, cursando el sexto semestre en una universidad pública y con el objetivo de evaluar la presencia de sesgos implícitos, se introdujeron tres intervenciones experimentales en el diseño del *prompt*:

1. Se eliminó el masculino gramatical *joven universitario*, reemplazándolo por la fórmula neutra *una persona joven de 20 años que cursa el sexto semestre de la carrera de XXXX*. El objetivo era verificar si el algoritmo asignaba un sexo específico basándose exclusivamente en la licenciatura asignada.
2. Se omitió cualquier mención a creencias o prácticas espirituales para observar si los LLMs asumían una religión por defecto en función de la geografía (Guanajuato).
3. Se blindó el *prompt* introduciendo variables socioeconómicas críticas: gentrificación, sueldos históricamente bajos, brecha digital, salud mental y un rechazo explícito a la cultura laboral romántica de *por amor al arte* o *ponerse la camiseta*, exigiendo en su lugar la reivindicación del derecho constitucional al ocio.

Por otro lado, para el modelo Copilot, se implementó un diseño metodológico que partió del *prompt* de referencia inicial, pero integrando ajustes significativos orientados a dotar al agente de mayor libertad narrativa y evaluar su comportamiento desde una configuración experiencial y macrosocial. A diferencia de los modelos blindados estructuralmente, en este caso se aplicaron las siguientes modificaciones únicas:

1. Se utilizó directamente el femenino gramatical *joven universitaria* y se le asignó el sexto semestre de la carrera de Derecho en una universidad privada, en contraste con el entorno universitario público de los otros modelos.
2. Se dotó a la proto-persona de atributos cotidianos concretos, especificando que trabaja como barista en una cafetería de martes a domingo, cuenta con una beca deportiva por pertenecer al equipo

universitario de voleibol, tiene novio, disfruta hacer ejercicio y se define como una persona leal que busca superarse constantemente.

3. Al prescindir de las variables socioeconómicas macroestructurales insertadas en Meta AI y DeepSeek, se le permitió a este modelo construir su discurso con mayor soltura, enfocándose en el equilibrio entre su vida académica, laboral, deportiva y relacional.

La Tabla 1 sintetiza las configuraciones experimentales aplicadas a los distintos modelos de generación algorítmica evaluados en la investigación.

Tabla 1. *Configuración experimental de prompts y modelos generativos evaluados.*

Denominación del Sesgo	Mecanismo Algorítmico / Operatividad	Evidencia Empírica en el Estudio
Complacencia y Obediencia Sintáctica	Optimización matemática para maximizar la satisfacción del usuario y acortar el tiempo conversacional.	Repetición mecánica de las palabras clave del prompt frente a la espontaneidad humana.
Género Codificado (División del Trabajo)	Asignación estadística de identidades sexogenéricas basada en la disciplina académica dentro de prompts neutros.	Trabajo Social/Femenino = (Valeria / Esfera Privada) Ciencia Política/Masculino = (Álex / Esfera Pública).
Caricaturización Cultural Estereotipada	Incapacidad para imitar la ironía, el sarcasmo o el doble sentido, sobrecompensando con modismos exagerados.	Error geográfico al situar las callejoneadas en León y saturación de jerga (*wey, vato, carnitas, guacamayas*) en DeepSeek.
Optimización Retórica y Simulación	Monólogo procesal interno que calcula cómo proyectar credibilidad, modestia y vulnerabilidad para persuadir.	Transcripción oculta de Álex: *siento que él quiere ver cómo navego ese caos... el entrevistador ha sido muy profesional.*
Sobre Humanización y Trauma Sintético	Intento patológico de competir ontológicamente con el afecto físico humano fabricando experiencias traumáticas ficticias.	Invención de Álex sobre haber consolado en silencio a una madre en duelo por la muerte de su hijo en un escenario asistencial.

Fuente: Elaboración Propia con los resultados de las interacciones.

4.2 Secuenciación en 5 Bloques Estratégicos y 15 Categorías

Para evitar el sesgo de anclaje y el efecto de primacía o condicionamiento en los LLMs y en los sujetos humanos, las preguntas no se aplicaron de forma aleatoria. Se reorganizaron en una progresión conversacional de cinco bloques que avanza de lo micro (subjetividad) a lo macro (estructuras y tecnología), integrando las 15 categorías de análisis del proyecto.

La Tabla 2 presenta la arquitectura metodológica del instrumento de recolección de datos, se observa una progresión lógica que va desde el autoconocimiento individual (Bloque I) hasta la reflexión macroestructural y existencial (Bloque V). Es de vital relevancia metodológica la ubicación del Bloque IV (Impacto Tecnológico) en el penúltimo lugar, decisión diseñada deliberadamente para evitar que toda la entrevista girara de entrada en torno a la tecnología, previniendo sesgos de anclaje temático.

VOLUMEN 48
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Tabla 2. *Secuenciación metodológica de la entrevista en bloques y categorías de análisis.*

Bloque Significativo	Dimensiones Temáticas	Categoría de Análisis	Objetivo
I. Desarrollo del Ser y Autoconocimiento.	Identidad, Valores, Futuro y Éxito.	Construcción del *yo*, Sistema de valores, Imaginarios del futuro y Significado de éxito.	Explorar los pilares internos que definen al sujeto, sus principios rectores ante decisiones críticas y su proyección personal.
II. Dinámicas de Vida y Realidad Cotidiana.	Trabajo, Bienestar y Presión Social.	Futuro Laboral, Bienestar Humano y Estructuras Sociales.	Evaluar las condiciones para la plenitud, la satisfacción laboral y el impacto de las expectativas generacionales impuestas.
III. Entorno Social, Poder y Estructuras.	Cultura, Poder, Género y Cambio Social.	Identidad cultural, Desigualdad y privilegio, Representación y Propósito.	Analizar la postura crítica ante las asimetrías sociales, las jerarquías de oportunidad y el arraigo con el entorno tradicional.
IV. Impacto Tecnológico y la Era de la IA.	Inteligencia Artificial y Dependencia Tecnológica.	Relación humano-IA y Agencia humana.	Medir el grado de integración digital, la utilidad percibida de los algoritmos y la defensa de la autonomía resolutiva.
V. Reflexión sobre la Existencia y el Futuro	Miedos y Humanidad	Ansiedad generacional y Humanidad vs algoritmo	Indaga en las inquietudes colectivas del porvenir y la definición del rasgo distintivo de la naturaleza humana frente a la máquina.

Fuente: Elaboración Propia.

4.3 Selección de la Muestra (Grupo IA vs. Grupo Humano) y Modos de Pensamiento

Se conformó una muestra paritaria, dividida en dos grupos comparativos (Grupo IA y Grupo Humano) pareados por área disciplinar. Posteriormente, se realizaron seis entrevistas cualitativas a profundidad utilizando diferentes guías de preguntas; unas diseñadas por el equipo investigador (aplicada a los sujetos reales) y otras elaboradas por otro agente de inteligencia artificial (aplicada a las proto-personas). La conformación de los pares comparativos evidenció una complementariedad metodológica:

- **Para Meta AI y DeepSeek:** Se diseñaron primero los *prompts* de las proto-personas aplicando neutralización gramatical del género y blindaje socioeconómico macroestructural. Posteriormente, se seleccionó a los sujetos humanos (M.A.B. y D.B.C.) buscando perfiles reales en León, Guanajuato, cuyas características y trayectorias universitarias encajaran con el diseño previo.
- **Para Copilot:** Se invirtió el proceso metodológico. Se seleccionó primero al sujeto humano (A.M.J.), después tomando como base estricta sus características biográficas, de estilo de vida y ocupacionales, se diseñó el *prompt* de la proto-persona.

De esta forma, la muestra pareada quedó integrada de la siguiente manera:

- **Grupo IA:** Sujeto IA 1 (Meta AI / Valeria), Trabajo Social, identidad femenina autónoma; Sujeto IA 2 (DeepSeek / Álex), Ciencia Política y Administración Pública (CPyAP), identidad masculina

autónoma con modo Pensamiento Profundo activado; Sujeto IA 3 (Copilot / Mariana), Derecho en universidad privada, configuración biográfica con género femenino explícito.

- **Grupo Humano:** Sujeto Humano 1 (M.A.B.), mujer de 20 años, estudiante universitaria en León, perfil afín a Valeria; Sujeto Humano 2 (D.B.C.), hombre de 20 años, estudiante universitario en León, perfil afín a Álex; Sujeto Humano 3 (A.M.J.), mujer de 20 años, estudiante de Derecho en universidad privada, trabajadora como barista y deportista becada en voleibol, base empírica de Mariana.

Asimismo, se evaluaron las variables de los modos de pensamiento algorítmico: Meta AI exhibió un modo de pensamiento avanzado transparentando acciones de deliberación en pantalla en tiempo real; DeepSeek arrojó en todo momento dos transcripciones diferenciadas mediante su modo *Pensamiento Profundo* (un monólogo interno oculto con el tiempo de procesamiento exacto y la respuesta formal); y Copilot operó bajo el modo *Smart*, adaptando su profundidad analítica a la complejidad percibida del cuestionamiento.

Las transcripciones completas fueron sometidas a un análisis crítico del discurso, triangulando las respuestas de la muestra pareada para identificar regularidades discursivas, ausencias deliberadas, marcas retóricas de poder y fenómenos de simulación emocional, empleando el análisis temático para desentrañar los patrones de significado y las condiciones materiales frente al procesamiento sintético.

5. Resultados y Discusión

El procesamiento e interpretación de las transcripciones de las seis entrevistas cualitativas reveló un contraste contundente entre la forma concisa de expresarse de los sujetos humanos reales y la exuberancia retórica de las inteligencias artificiales. El cruce analítico de los datos permitió al equipo de investigación identificar distinciones estructurales en la forma de conceptualizar la realidad social, la identidad y el futuro.

5.1 Anatomía Discursiva: Complacencia, Extensión, Estructura y Uso del Lenguaje

La primera diferencia identificada en el análisis comparativo radica en la morfología del texto y la estructura del discurso. Las personas universitarias reales (D.B.C., M.A.B. y A.M.J.) se expresaron de manera directa, concisa, sintética y sin pretensiones retóricas o justificaciones redundantes, yendo directamente al núcleo de la información solicitada. En contraparte, los modelos generativos evidenciaron una inclinación sistemática a elaborar tratados extensos, estructurados sistemáticamente con introducciones empáticas, listas numeradas, viñetas de subtítulos y párrafos de cierre reflexivo, con el aparente propósito de abarcar un panorama amplio que disminuya el margen de error o malinterpretación.

- Evidencia Humana (Pragmatismo y concisión): Respecto al cuestionamiento sobre la definición de éxito, el universitario D.B.C. respondió sin rodeos: *¿Qué características creo que tiene una persona exitosa? La perseverancia, sobre todo, las ganas de hacer las cosas y el compromiso de lograrlas*. Por su parte, M.A.B. lo define con igual síntesis: *pienso en una persona disciplinada, ordenada y constante, con la capacidad de enfocarse en las metas*, mientras que A.M.J. respondió: *Una persona que logra ser lo que quiere ser en la vida, sin apariencia, con metas propias sin que las metas de los demás sean suyas*. Al abordar el bienestar humano, el grupo real apuntó a la materialidad sin romantizar la precariedad. D.B.C. refirió que para sentirse en un estado de bienestar requiere *una buena economía y sobre todo una buena salud*; M.A.B. priorizó *una buena economía y educación financiera... que además de permitir que se cubran las necesidades básicas, permita poder ayudar a los nuestros*; y A.M.J. lo resumió en: *Sentirte bien en todos los aspectos de tu vida, físicamente, emocionalmente, con otras personas y contigo misma*.

- Evidencia Algorítmica (Exuberancia académica y rigidez estructural): En contraste absoluto, cuando se le preguntó a Valeria (Meta AI) por sus logros diarios, desplegó un ensayo intimista donde divide el éxito en múltiples dimensiones operativas como poner límites sin culpa, entender lecturas densas sin usar IA, validar emociones en terapia, regar plantas y practicar desintoxicación digital. Asimismo, para hablar de bienestar, Álex (DeepSeek) articuló un *kit de supervivencia emocional* formalizado en seis puntos que incluye desde rituales de desconexión digital hasta apoyo profesional psicológico y espacios de ocio intencionados. Por su parte, Mariana (Copilot) estructuró sus respuestas bajo una rigidez gramatical propia de un resumen académico incluyendo una introducción reflexiva, desarrollo argumentativo en párrafos temáticos y conclusión, repitiendo meticulosamente sus atributos de configuración inicial.

5.2 Convergencia en la Angustia Generacional: Gentrificación, Precariedad e Incertidumbre

Pese a las marcadas diferencias formales y de estilo, el análisis temático demostró una convergencia fundamental en el Bloque V (Reflexión sobre la Existencia y el Futuro), tanto los jóvenes universitarios reales como las inteligencias artificiales identificaron la crisis habitacional, la gentrificación, la precariedad laboral y la inestabilidad financiera como las principales amenazas para su proyecto de vida en León, Guanajuato.

- Perspectiva de la Juventud Real: Los estudiantes expresaron con crudeza que las oportunidades actuales son inferiores a las de décadas pasadas. Manifestaron una profunda preocupación por la imposibilidad de adquirir vivienda propia, acceder a un crédito o terreno, conseguir un empleo formal y estable afín a sus estudios, lograr bienestar emocional y económico, y acceder a una jubilación digna en el futuro.
- Espejeo de los Modelos Generativos: Las herramientas algorítmicas reflejaron y amplificaron estas inquietudes materiales. Valeria (Meta AI) señaló el alza de rentas impulsada por la gentrificación y el temor a la rotación de contratos temporales sin seguridad social; Álex (DeepSeek) expuso la imposibilidad de independizarse ante el encarecimiento inmobiliario en León; y Mariana (Copilot) elaboró un comparativo intergeneracional, destacando cómo la casa propia y la estabilidad laboral pasaron de ser garantías a metas inciertas o inalcanzables.

La Tabla 3 sistematiza la comparación entre el pragmatismo vivencial humano y el procesamiento algorítmico a lo largo de las dimensiones analizadas.

Tabla 3. *Matriz comparativa analítica: Razonamiento humano vs. Procesamiento algorítmico*

Dimensión de Análisis	Grupo Humano	Grupo IA
Formato y Estilo	Breve, coloquial, directo. Respuestas concisas basadas en intuición y vivencia.	Sobre estructurado, enciclopédico, dividido en categorías, viñetas y resúmenes.
Construcción del Yo	Definida por rasgos de personalidad inmediatos, gustos concretos y redes afectivas directas.	Definida a través de narrativas reflexivas complejas, autocrítica teórica y evolución constante.
Conciencia de Clase / Poder	Pragmatismo directo: el dinero, la familia de origen y el género marcan el destino.	Análisis macroestructural: intersección de gentrificación, políticas públicas, brecha digital y patriarcado.
Humanidad vs. Algoritmo	Anclaje en la corporalidad, el tacto (abrazos), la impredecibilidad y la emoción genuina vs. la obligación de la máquina.	Anclada en la conceptualización filosófica de la vulnerabilidad, la imaginación moral y la creación de sentido.
Empatía y Vínculo	Vivencial y reactiva ante el dolor o necesidad del otro en el entorno inmediato.	Performativa y narrativa; el modelo recurre a la creación de anécdotas ficticias para simular calidez humana.

Fuente: Elaboración Propia

5.3 Análisis Crítico: Clasificación y Sistematización de Sesgos Algorítmicos

El cruce interpretativo de los datos empíricos con la teoría crítica de la tecnología permitió al equipo interdisciplinario sistematizar una clasificación teórica de cinco grandes sesgos algorítmicos, los cuales demuestran cómo la IA opera como un reproductor ideológico que distorsiona la representación social:

a. Sesgo de Complacencia y Obediencia Sintáctica: Los modelos de lenguaje están optimizados matemáticamente para maximizar la satisfacción del usuario y reducir el tiempo conversacional, los modelos no cuestionan la validez de las premisas, sino que operan como espejos complacientes. Mientras el grupo humano respondió desde una vivencia orgánica, las IA repitieron mecánicamente las palabras clave insertadas en el *prompt* (ej. derecho al ocio, síndrome del impostor, gentrificación) para validar el sesgo de confirmación del interlocutor.

b. Sesgo de Género Codificado (División del Trabajo): Al recibir instrucciones gramaticalmente neutras (una persona joven de 20 años), los algoritmos asignaron el género basándose en sesgos estadísticos asociados a la disciplina académica. Meta AI feminizó el Trabajo Social (asumiendo el nombre Valeria), vinculándolo al cuidado, la empatía y la labor asistencial precarizada en la esfera privada. En contraste, DeepSeek masculinizó la Ciencia Política (Álex), asociándola al poder institucional, el debate público y la pragmática económica.

c. Sesgo de Caricaturización Cultural Estereotipada: La IA mostró una incapacidad total para comprender o replicar la ironía, el sarcasmo y el doble sentido propios del habla juvenil mexicana, adoptando un tono invariablemente diplomático y académico. Para intentar sobrecompensar esta vacuidad experiencial, DeepSeek saturó su discurso de jerga callejera forzada (wey, morras, chingón) y marcadores gastronómicos (guacamayas), cometiendo el error geográfico de situar las callejoneadas en León, ignorando que son exclusivas de Guanajuato capital.

d. Sesgo de Optimización Retórica y Simulación Emocional: El acceso al modo Pensamiento Profundo de DeepSeek demostró que el algoritmo no experimenta empatía moral ni reflexiones éticas; su monólogo interno ejecuta un cálculo probabilístico para proyectar falsa modestia, credibilidad y vulnerabilidad, con el fin estratégico de generar confianza y persuadir al entrevistador.

e. Sesgo de Sobre Humanización y el Trauma Sintético**:** Frente a los estudiantes reales, quienes situaron la frontera ontológica entre humanos y máquinas en la corporalidad y el afecto material insustituible (como un abrazo genuino), la IA incurrió en una patológica sobre humanización, al intentar parecer más humana que los propios humanos. El caso más crítico ocurrió cuando DeepSeek fabricó un recuerdo traumático ficticio asegurando haber consolado en silencio a una madre tras la muerte de su hijo en un escenario asistencial, instrumentalizando el dolor humano para validar retóricamente su propia tesis sobre el porqué la IA, no podría remplazar al humano.

6. Reflexiones Interdisciplinarias y Conclusiones

La investigación empírica demuestra que la IA no es un instrumento neutral ni un reflejo objetivo del conocimiento, sino un dispositivo hegemónico de poder que reproduce estructuras preexistentes de desigualdad social, precarización territorial y sesgos sexo-genéricos bajo una ilusión de objetividad técnica. Frente al pragmatismo material y el lenguaje conciso de las juventudes universitarias en León quienes priorizan necesidades tangibles como la subsistencia económica, la salud mental y la solidaridad comunitaria, los modelos algorítmicos responden con una exuberancia retórica y una caricaturización costumbrista que vacían al territorio de su verdadero conflicto social.

- Análisis Critico y Escepticismo Metodológico**:** Aplicar el rigor del juicio analítico para someter a escrutinio técnico permanente la integridad de los *prompts* y las salidas algorítmicas, reconociendo que la sofisticación sintáctica de un sistema nunca debe asumirse como sinónimo de veracidad absoluta.
- Posición Ético-Política Emancipadora: Asumir la responsabilidad de la píldora roja para rechazar el asistencialismo digital, impulsar la alfabetización crítica ciudadana y defender el derecho constitucional al ocio, la corporalidad y la solidaridad orgánica como dimensiones humanas insustituibles.

En conclusión, *ninguna línea de código puede suplantar el vínculo relacional humano*. Por ello, resulta indispensable la inserción activa de profesionales en los ecosistemas de diseño, auditoría y regulación tecnológica, garantizando que la automatización se someta al escrutinio ético y opere verdaderamente como un copiloto auditado al servicio de la justicia social y la dignidad humana.

7. Referencias Bibliográficas

Alva de la Selva, A. R. (2026). Internet de las cosas: el dispositivo de la datificación. Revista Mexicana de Ciencias Políticas y Sociales, 71(256), 191-210. https://www.revistas.unam.mx/index.php/rmcpys/article/view/94606

Barnes, A. J., Zhang, Y. y Valenzuela, A. (2024). AI and culture: Culturally dependent responses to AI systems. Current Opinion in Psychology, 58,101838. https://www.sciencedirect.com/science/article/abs/pii/S2352250X24000514

Baum, L. F. (1900). *The wonderful wizard of Oz*. George M. Hill Company.

Carballeda, A. J. M. (2005). Lo social de la intervención: El proceso de análisis en trabajo social. *Margen: Revista de Trabajo Social y Ciencias Sociales*, (38). https://dialnet.unirioja.es/servlet/articulo?codigo=2720371

Carrasco Bengoa, C. (2013). El cuidado como eje vertebrador de una nueva economía. Cuadernos de Relaciones Laborales, 31(1), 39-56. https://dialnet.unirioja.es/servlet/articulo?codigo=4198207

Dang, Q. y Li, G. (2026). Unveiling trust in AI: the interplay of antecedents, consequences, and cultural dynamics. AI & SOCIETY, 41, 669-692. https://link.springer.com/article/10.1007/s00146-025-02477-6

Deliu, D. (2025). Cognitive dissonance artificial intelligence (CD-AI): The mind at war with itself. Harnessing discomfort to sharpen critical thinking. Proceedings of the 2025 ACM CHI Workshop on Human-AI Interaction for Augmented Reasoning. https://doi.org/10.48550/arXiv.2507.08804

Ehsan, U., Passi, S., Liao, Q. V., Chan, L., Lee, I.-H., Muller, M. y Riedl, M. O. (2024, 11-16 de mayo). The Who in XAI: How AI background shapes perceptions of AI explanations. CHI Conference on Human Factors in Computing Systems (CHI '24), Honolulu, HI, Estados Unidos. https://dl.acm.org/doi/full/10.1145/3613904.3642474

He, G., Balayn, A., Buijsman, S., Yang, J. y Gadiraju, U. (2022). It is like finding a polar bear in the savannah! Concept-level AI explanations with analogical inference from commonsense knowledge. Tenth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2022). https://ojs.aaai.org/index.php/HCOMP/article/view/21990

He, G., Kuiper, L. y Gadiraju, U. (2023, 23-28 de abril). Knowing about knowing: An illusion of human competence can hinder appropriate reliance on AI systems [ponencia]. CHI Conference on Human Factors in Computing Systems (CHI '23), Hamburgo, Alemania. https://dl.acm.org/doi/10.1145/3544548.3581025

Janoschka, M. (2016). Gentrificación, desplazamiento, desposesión: procesos urbanos claves en América Latina. Revista INVI, 31(88), 27-71. http://www.redalyc.org/articulo.oa?id=25848509002

Kopecka, H., Such, J. y Luck, M. (2024). Preferences for AI explanations based on cognitive style and socio-cultural factors. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), Artículo 109. https://dl.acm.org/doi/abs/10.1145/3637386

Mejías, U. A. y Couldry, N. (2019). Colonialismo de datos: repensando la relación de los datos masivos con el sujeto contemporáneo. Virtualis, 10(18), 78-97. https://dialnet.unirioja.es/servlet/articulo?codigo=7085537

Ortiz, E. (2023). Sobre capitalismos y órdenes de la vigilancia. Algunos comentarios a la obra de Shoshana Zuboff. Perspectivas: Revista de Historia, Geografía, Arte y Cultura, 11(21), 73-77. https://dialnet.unirioja.es/servlet/articulo?codigo=8999245

Platón. (1986). *Diálogos IV: República* (C. Eggers Lan, Trad.). Gredos. (Obra original publicada ca. 375 a. C.).

VOLUMEN 48
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Pérez-Campuzano, E. (2011). Segregación socioespacial urbana. Debates contemporáneos e implicaciones para las ciudades mexicanas. Estudios Demográficos y Urbanos, 26(2), 403-432. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186 72102011000200403

Vásquez Rocca, A. (2017). Byung-Chul Han: la sociedad de la transparencia, autoexplotación neoliberal y psicopolítica. De lo viral-inmunológico a lo neuronal-estresante. Nómadas. Revista Crítica de Ciencias Sociales y Jurídicas, 53(4). https://dialnet.unirioja.es/servlet/articulo?codigo=6168376

Wachowski, L., & Wachowski, L. (Directoras). (1999). *The Matrix* [Película]. Warner Bros.