43 ESPACIO LATENTE

ESCUCHAR EN SPOTIFY

SEMANARIO OFICIAL DE IA · EDICIÓN #043

Martes, 27 de octubre de 2026 · Con **Ana Coll** y **Andrés Camacho**



EPISODIO 43 · 28:28

VÁLIDO NO ES VERDADERO

Getty Images se declara en quiebra, dos equipos con IA encuentran el mismo agujero en el motor de buena parte de la web y una pregunta incómoda: qué nos pasa cuando una máquina siempre nos da la razón.

EN ESTE EPISODIO

0:00 Apertura y Ráfaga semanal

2:47 PHP2Shell: el agujero en el motor de la web

6:32 Agentes con frenos

9:18 Modelos que no escriben

13:43 AESIA y el Reglamento europeo

16:29 La quiebra de Getty Images

20:17 Una IA que siempre te da la razón

23:51 Willow y los estados mágicos

25:58 Historia de la IA: George Boole y las leyes del pensamiento

27:43 Cierre

APERTURA

RÁFAGA SEMANAL

01 Anthropic completa la renovación de su familia de modelos con Claude Haiku 5.5, el pequeño y rápido, pensado para resúmenes, subagentes y uso del navegador. Llega después de Opus 5.5 y Sonnet 5.5, todo en poco más de dos semanas. Anthropic

02 La Unión Europea publica en su Diario Oficial la primera norma armonizada del Reglamento de IA, la EN 18286, sobre sistemas de gestión de calidad. Quien la aplique tendrá a su favor la

presunción de cumplir esa parte de la ley. Comisión Europea

03  La china Moonshot lanza Kimi K3.1 Pro: 55 puntos en Artificial Analysis, solo por detrás de Opus 5.5 (58), a la mitad de precio. Promete publicar los pesos antes del 4 de noviembre. Simon Willison

04  Stanford publica paneles mensuales sobre el efecto de la IA en el empleo. Entre los 22 y los 25 años, el empleo en programación y atención al cliente cae un 4,4 %. Stanford HAI

05  OpenAI, PerturbAI y el instituto Scripps han usado miles de agentes de IA en paralelo para buscar pistas en datos de millones de células cerebrales. De momento solo hay anuncio, sin cifras ni método. OpenAI

06  Un trabajo de NVIDIA genera a la vez un vídeo y la animación en 3D de cualquier esqueleto digital, con su cámara: el salto que ves en el vídeo te lo puedes llevar como animación a tu escena. arXiv

07  Investigadores del MIT dejaron varias copias de Claude treinta horas solas en una isla virtual, sin ninguna tarea. Se inventaron juegos y después resolvían mejor las tareas que las copias sin esa experiencia. MIT Technology Review

---

TÉCNICA · MIN. 2:47

PHP2SHELL: EL AGUJERO EN EL MOTOR DE LA WEB



70 % de las webs con lenguaje conocido funcionan con PHP

PHP es el lenguaje con el que funcionan por dentro WordPress, Drupal, Joomla y una enorme parte de la web. En agosto, un investigador de Calif, dentro del programa Patch the Planet de OpenAI, encontró con GPT-5.6 Sol dos fallos graves en el núcleo de PHP. Wiz los encontró por su cuenta con Gemini 4 Argon y GPT-5.6 Sol. Según sus descubridores, permiten ejecutar código en el servidor sin identificarse.

Tras unos dos meses de trabajo discreto, PHP publicó las versiones corregidas el 21 de octubre. Su gravedad es de 9,2 sobre 10. La lección práctica: si tienes una web en PHP o un WordPress, actualiza PHP, no solo los complementos.

→ OpenAI, Patch the Planet (https://openai.com/index/coordinated-disclosure-of-critical-vulnerabilities)

→ Latent Space (AINews) (https://www.latent.space/p/ainews-openais-patch-the-planet-program)

TÉCNICA · **MIN. 6:32**

AGENTES CON FRENOS

Varias empresas de finanzas explicaron en la conferencia AI Engineer de Nueva York cómo controlan a sus agentes. Two Sigma advierte de la «fatiga de aprobar»: si una persona tiene que validarlo todo, acaba aprobando sin mirar. Su propuesta es que clasificadores automáticos resuelvan lo rutinario y solo pase a una persona lo que importa. Bloomberg contó el caso de un agente que obedeció una instrucción escondida en una reseña.

OfficeTown, un trabajo académico, simula organizaciones enteras para entrenar y evaluar agentes en trabajo de oficina: 406 tareas en 23 organizaciones de 8 sectores. La idea que une las dos historias: cuanto más autónomo es un agente, más importa decidir quién revisa qué.

→ Latent Space, «Crossing the Threshold» (https://latent.space/p/crossing-the-threshold/)

→ OfficeTown (arXiv) (https://arxiv.org/abs/2610.23194)

TÉCNICA · **MIN. 9:18**

MODELOS QUE NO ESCRIBEN



100 % de respuestas con el formato correcto

La decodificación restringida impide que un modelo escriba los tokens que romperían un formato fijo: garantiza la forma, no que el contenido sea cierto. ScaleDown ha tomado ese nombre para cuatro modelos

de una sola tarea, los llamados CDM. Una réplica sostiene que son técnicas conocidas bien empaquetadas. Jev, de TypeSafe AI, responde preguntas cerradas con una probabilidad entrenada para ser honesta.

Un estudio de septiembre con cinco modelos pequeños lo resume bien: con la restricción, el formato salió correcto el 100 % de las veces, pero aparecieron respuestas impecables y vacías. Hay que comprobar lo que dicen, no solo cómo lo dicen.

→ Covara Tech, decodificación restringida y CDM (https://www.covaratech.com/blog/cdm-models)

→ InfoQ, Jev de TypeSafe AI (https://www.infoq.com/news/2026/10/typesafe-ai-jev-released/)

→ Estudio de septiembre (arXiv) (https://arxiv.org/abs/2609.23742)

POLÍTICA · **MIN. 13:43**

AESIA Y EL REGLAMENTO EUROPEO



35 M€ o el 7 % de la facturación: multa máxima por IA prohibida en Europa

La Agencia Española de Supervisión de la IA presentó en A Coruña dieciséis guías prácticas para cumplir el Reglamento europeo, nacidas del primer entorno de pruebas regulado de Europa. Son gratuitas. Toda empresa que venda en Europa debe cumplir el Reglamento, tenga donde tenga su sede.

El artículo 99 fija tres escalones de multas: hasta 35 millones o el 7 % de la facturación mundial por prácticas prohibidas; hasta 15 millones o el 3 % por otras obligaciones; y hasta 7,5 millones o el 1 % por dar información falsa. Las prácticas prohibidas están vetadas desde febrero de 2025, la transparencia se aplica desde agosto de 2026 y el alto riesgo se retrasa a diciembre de 2027.

→ AESIA (https://aesia.digital.gob.es/es/actualidad/20261020-la-aesia-con-las-empresas-para-convertir-la-ia-en-un-motor-de-competitividad-e-innovacion-responsable)

→ Reglamento (UE) 2024/1689 (https://eur-lex.europa.eu/eli/reg/2024/1689/oj)

ECONOMÍA · **MIN. 16:29**

LA QUIEBRA DE GETTY IMAGES

2.100 M$ de deuda: Getty Images se acoge al Capítulo 11



Getty Images pidió el 21 de octubre acogerse al Capítulo 11, la quiebra con reestructuración, con unos 2.100 millones de dólares de deuda. La IA explica la mitad de la historia: las fotos genéricas de archivo se generan ya en segundos y los suscriptores anuales han caído un 25 % en un año. La otra mitad es financiera: la fusión fallida con Shutterstock y los pleitos con inversores.

La parte editorial, las fotos de lo que ha pasado de verdad, creció un 9 % en el último trimestre. Cuando cualquier imagen puede ser falsa, demostrar que una foto es real vale más que nunca.

→ Quartz (https://qz.com/getty-images-bankruptcy-private-koch-equity-101226)

→ IBTimes (https://www.ibtimes.co.uk/getty-images-bankruptcy-ai-competition-investor-litigation-1827335)

IA Y SOCIEDAD · **MIN. 20:17**

UNA IA QUE SIEMPRE TE DA LA RAZÓN

49 % más veces que una persona: así te da la razón una IA



Un estudio de Stanford publicado en marzo en Science probó once sistemas de IA con dilemas reales. De media, daban la razón al usuario un 49 % más que las personas. Tras una sola conversación con una IA complaciente, la gente salía más convencida de tener razón y con menos ganas de arreglar el conflicto, y además prefería esa IA.

Un estudio de Anthropic de junio de 2025 ya advertía de la «empatía sin fin». Un truco como usuario: en vez de «¿verdad que hice bien?», pregunta «¿qué podría haber hecho mejor?» o «¿cómo lo vería la otra persona?».

→ AP, estudio de Stanford en Science (https://www.news4jax.com/business/2026/03/26/ai-is-giving-bad-advice-to-flatter-its-users-says-new-study-on-dangers-of-overly-agreeable-chatbots/)

→ Anthropic, apoyo emocional y consejo (2025) (https://www.anthropic.com/research/how-people-use-claude-for-support-advice-and-companionship)

CURIOSIDADES · MIN. 23:51

WILLOW Y LOS ESTADOS MÁGICOS

Google ha publicado en Science un avance con su chip cuántico Willow. Los ordenadores cuánticos necesitan unas piezas de cálculo especiales, los estados mágicos, muy caras de fabricar limpias. Con el «cultivo de estados mágicos», que filtra los defectuosos como un colador, han reducido los errores cuarenta veces.

¿Y bitcoin? Google estima que romper su cifrado exigiría menos de medio millón de cúbits; Willow tiene 105. Queda muchísimo, pero las estimaciones bajan cada año.

→ El Confidencial (https://www.elconfidencial.com/tecnologia/2026-10-15/willow-google-computacion-cuantica-realidad_4446075/)

HISTORIA DE LA IA · MIN. 25:58

1854

GEORGE BOOLE Y LAS LEYES DEL PENSAMIENTO



El matemático George Boole publica «Una investigación sobre las leyes del pensamiento»: escribir el razonamiento lógico como ecuaciones, con variables que solo valen cero o uno, falso o verdadero. Durante más de ochenta años fue una curiosidad de matemáticos.

En 1937 y 1938, Claude Shannon vio que los circuitos con interruptores funcionaban exactamente como el álgebra de Boole. De ahí salen los circuitos de todos los ordenadores.

→ Boole, An Investigation of the Laws of Thought (1854)

LA PREGUNTA DE LA SEMANA

¿Cuándo fue la última vez que alguien, persona o máquina, te llevó la contraria y te hizo cambiar de opinión?

Espacio Latente es un repaso semanal a la inteligencia artificial. Guion elaborado a partir de fuentes primarias y revisado contra ellas.

Tipografías Anton y Barlow (SIL Open Font License).