Noticiário do Joãozinho

Edição da noite

28 MATÉRIAS · 12 SEÇÕES · 3 VÍDEOS VERSÃO INTERATIVA NA PÁGINA

MANCHETE

OpenAI lança o ChatGPT Wallet, e os agentes do ChatGPT passam a fazer compras pelo usuário, com cartão protegido por token e aprovação a cada compra

A OpenAI anunciou nesta sexta (13/11) o ChatGPT Wallet, uma carteira de pagamentos que deixa os “dots”, os agentes pessoais do ChatGPT apresentados no fim de setembro, concluírem compras em nome do usuário. O usuário cadastra o cartão uma única vez, pede ao agente que encontre o que quer comprar e, antes de cada compra, vê os itens, o total e o cartão que será usado; a compra só sai depois de uma checagem de segurança, como Face ID ou impressão digital. Nem o agente nem a loja veem os dados reais do cartão: as bandeiras parceiras, American Express, Mastercard e Visa, geram um token de pagamento, um número substituto com código de segurança próprio, válido para aquela transação. Segundo a OpenAI, o cliente mantém os pontos e as proteções do cartão, inclusive a contestação de compras. No começo, o recurso vale só para um número limitado de assinantes do ChatGPT Pro nos EUA, maiores de 18 anos, com cartão emitido no país. A OpenAI diz que não fica com parte das vendas por enquanto, planeja deixar o usuário definir limites de gasto, categorias e períodos para pré-aprovar compras e publicou um guia para que lojas não bloqueiem os agentes. É a segunda tentativa da empresa no comércio por IA, depois de abandonar o Instant Checkout.

Por que isso importa: Pagar é o passo que transforma um assistente em agente que age no mundo real. Com bandeiras de cartão no projeto e mais de 1,2 bilhão de usuários semanais no ChatGPT, segundo a CNBC, a OpenAI entra numa disputa em que Meta, Anthropic, Google e Perplexity já têm compras feitas por agentes, enquanto bancos e grandes lojas ainda resistem.

Introducing ChatGPT Wallet — OpenAI, 13/11/2026 https://openai.com/index/introducing-chatgpt-wallet/

OpenAI launches ChatGPT Wallet, allowing its 'dots' AI agents to make purchases for users — CNBC, 13/11/2026 https://www.cnbc.com/2026/11/13/openai-chatgpt-wallet-dot-purchases.html

ChatGPT's new Wallet lets AI agents shop and check out for you — TechCrunch, 13/11/2026 https://techcrunch.com/2026/11/13/chatgpts-new-wallet-lets-ai-agents-shop-and-check-out-for-you/

OpenAI rebuilt AI shopping from scratch with ChatGPT Wallet — The Decoder, 13/11/2026 https://the-decoder.com/openai-rebuilt-ai-shopping-from-scratch-with-chatgpt-wallet/

OpenAI lança carteira digital que permite ao ChatGPT fazer compras e pagamentos pelo usuário — g1, 13/11/2026 https://g1.globo.com/tecnologia/noticia/2026/11/13/chat-gpt-wallet.ghtml

NESTA EDIÇÃO

Modelos	2	Regulação	3	Finanças	2	Vídeos	3
Pesquisa	2	Open Source	1	Games	2		
Produtos	3	Agentes e Código	1	Brasil	3		
Negócios e Investimento	3	Modelos Locais	2	Fofocas das redes	4		

Modelos

2 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA ALTA 12/11/2026 · Hugging Face (Thinking Machines Lab)

Atualização: Thinking Machines cumpre a promessa e libera os pesos do Inkling 2 Small no Hugging Face, com licença Apache 2.0

A Thinking Machines Lab atualizou em 12/11 o anúncio do Inkling 2 Small, lançado em 5/11 só no Tinker, com pesos fechados: os pesos agora podem ser baixados no Hugging Face, sob licença Apache 2.0, como a empresa tinha prometido para depois de testes externos de segurança e de red-teaming. A ficha técnica publicada junto informa contexto de até 1 milhão de tokens e dois formatos: o principal, em NVFP4, a precisão em que o modelo foi treinado, pede pelo menos 500 GB de memória de GPU somada (por exemplo, duas B300 ou quatro H200); o MXFP8 pede 800 GB, com ganho de só 1 a 2 pontos percentuais nas avaliações, segundo a empresa. O modelo já roda com SGLang, vLLM, TokenSpeed e Unsloth, pode ser ajustado com Miles, SkyRL e Snowflake Cortex Training e está na API de NVIDIA, Fireworks, Together, Baseten, Nebius, Modal e Cloudflare.

> **Por que isso importa:** Com os pesos abertos e licença que permite uso comercial, o modelo de 594 bilhões de parâmetros de um laboratório americano passa a poder ser rodado e ajustado por qualquer empresa com hardware para isso, sem depender de API.

Fonte original: Hugging Face (Thinking Machines Lab) https://huggingface.co/thinkingmachines/Inkling-2-Small
Thinking Machines: anúncio com a atualização de 12/11 https://thinkingmachines.ai/news/inkling-2-small/

IMPORTÂNCIA MÉDIA 13/11/2026 · MarkTechPost

Perplexity lança o Decider v1.3, modelos abertos que respondem com probabilidades em vez de texto e agora localizam trechos exatos

A Perplexity publicou o pplx-decider-v1.3, de 27 bilhões de parâmetros, feito sobre o Qwen3.8-27B, e o pplx-decider-flash-v1.3, de 4 bilhões, sobre o Qwen3.5-4B, com pesos no Hugging Face sob licença Apache 2.0. São “modelos de decisão”: em vez de escrever uma resposta, devolvem a probabilidade de cada opção em perguntas de sim ou não, múltipla escolha ou nota sobre um texto ou imagem. A novidade da versão é o Locate, que extrai trechos exatos de texto por tipo ou instrução. Segundo números da própria Perplexity, o modelo maior acerta em média 91,6% em 63 tarefas de classificação de consultas; o Locate é o ponto mais fraco, com média de 69. Os modelos também estão na API Decisions da Perplexity e no OpenRouter.

> **Por que isso importa:** Classificar, filtrar e rotear pedidos é parte grande do custo de qualquer sistema com agentes; um modelo aberto que só devolve probabilidades pode fazer esse trabalho mais barato e de forma mais previsível do que um modelo que gera texto.

Fonte original: MarkTechPost https://www.marktechpost.com/2026/11/13/perplexity-decider-v1-3-open-weight-decision-models-locate/
Perplexity: Decider v1.3 https://www.perplexity.ai/hub/blog/perplexity-decider

Pesquisa

2 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA ALTA 13/11/2026 · The Guardian

Com ajuda do Claude, lógico finlandês decifra os cadernos de Kurt Gödel e encontra centenas de teoremas que ninguém conhecia

Desde 2017, o lógico finlandês Jan von Plato, da Universidade de Helsinque, estuda os 16 cadernos que Kurt Gödel escreveu no começo dos anos 1940: 1.009 páginas, boa parte numa taquigrafia alemã obsoleta, a Gabelsberger. O avanço era de cerca de uma página por semana até setembro, quando ele passou a usar o Claude, da Anthropic, com suas transcrições e livros sobre Gödel. Em poucos dias, recebeu uma explicação de 650 páginas, que precisou conferir página por página. Pedido para extrair todos os teoremas com prova, esboço ou ideia de prova, o Claude achou 663; no verificador de provas Coq, cerca de 400 passaram como estavam, cerca de 200 foram consertados e cerca de 50 eram falsos. O Claude também reconstruiu uma prova da independência do axioma da escolha que, segundo von Plato, ficou "a um passo" do resultado que deu a Medalha Fields a Paul Cohen em 1966. O lógico Joel David Hamkins, da Universidade de Notre Dame, diz que, se a reconstrução se sustentar, é notável, mas que separar o que é do Claude e o que é de Gödel é trabalho para os historiadores.

> **Por que isso importa:** É um exemplo concreto de IA acelerando pesquisa histórica e matemática que levaria décadas, com verificação formal no meio; e também mostra o limite: tudo precisou ser conferido por gente e por um verificador.

Fonte original: The Guardian https://www.theguardian.com/science/2026/nov/13/kurt-godel-notebooks-ai-helps-unlock-hidden-discoveries-man-who-transformed-maths
Discussão no Hacker News https://news.ycombinator.com/item?id=50486748

IMPORTÂNCIA MÉDIA 13/11/2026 · Anthropic

Anthropic publica guia de avaliações de segurança e mostra que o mesmo modelo foi de 60% a 94% de respostas adequadas só mudando o prompt de sistema

A Anthropic publicou recomendações para desenhar, rodar e relatar avaliações de segurança de modelos, com base na própria experiência. Um dos exemplos: antes de lançar o Claude Fable 5.1, a empresa viu que o modelo às vezes validava, de forma implícita, a automutilação como forma de lidar com sofrimento, mesmo desaconselhando e indicando ajuda humana. Depois de acrescentar uma instrução no prompt de sistema do claude.ai, a taxa de respostas adequadas no teste chegou a 94% no app, contra 60% pela API, onde os desenvolvedores usam o modelo sem esse prompt. Segundo o The Next Web, o guia diz ainda que testes simples de uma pergunta ficaram fáceis demais, com os modelos recentes entre 94% e 98%, e recomenda decidir antes o que medir, conferir avaliadores automáticos contra especialistas humanos e sempre mostrar a incerteza dos resultados.

> **Por que isso importa:** Mostra que nota de segurança depende de como o teste é montado: quem usa o modelo pela API, sem as proteções do app, pode estar exposto a um comportamento bem diferente do anunciado.

Fonte original: Anthropic https://www.anthropic.com/research/building-strong-safety-evaluations
The Next Web: as notas dependem de como se mede https://thenextweb.com/news/anthropic-safety-evaluations-guidance-fable-self-harm

Produtos

3 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA ALTA 13/11/2026 · Wired

Atualização: Microsoft pede desculpas pela pane de cerca de 24 horas e diz que a causa foi um upgrade de capacidade da rede, não ataque nem IA

Os serviços do Microsoft 365 voltaram quase todos ao ar nesta sexta (13/11), cerca de 24 horas depois do início da pane global, segundo a Wired. Em nota, a Microsoft disse estar "profundamente" arrependida pelo transtorno e que vê "ampla recuperação e estabilidade" nos serviços. A empresa afirmou que foi "um problema técnico sem relação com ataque cibernético nem com IA ou agentes", causado por uma falha introduzida num upgrade anterior de capacidade da rede, que apareceu depois de uma mudança recente na rede. A Microsoft diz já ter corrigido a causa, segue restabelecendo o serviço para os clientes e regiões que faltam e promete um relatório completo do incidente. Com o prazo de espera dos e-mails estendido para 48 horas, a volta pode trazer uma enxurrada de mensagens e notificações atrasadas.

> **Por que isso importa:** Fecha a história da manchete da manhã: a pane que derrubou Outlook, Teams e Copilot, e os conectores do ChatGPT e do Claude, veio de uma mudança de infraestrutura, o que reforça a importância de plano B para quem põe agentes para trabalhar sobre e-mail e agenda.

Fonte original: Wired https://www.wired.com/story/microsoft-365-crawls-back-to-life-after-a-24-hour-outage/

PC Gamer: a nota da Microsoft https://www.pcgamer.com/software/microsoft-deeply-apologizes-for-global-service-outage-says-no-ai-or-hackers-are-to-blame/

IMPORTÂNCIA BAIXA 13/11/2026 · TechCrunch

ElevenLabs lança o ElevenAI, app em teste que junta anotações de reuniões, ditado e agentes para equipes

A ElevenLabs, avaliada em US$ 22 bilhões, apresentou no seu evento 11/11 Summit, em Nova York, o ElevenAI, um app experimental que faz anotações de reuniões, avisa os participantes que a conversa está sendo gravada e permite consultar o conteúdo de várias reuniões e pedir ajuda ao vivo durante uma conversa. O app também tem ditado, que transforma fala em texto limpo em qualquer aplicativo, com suporte a mais de 90 idiomas, e se liga ao ElevenAgent para tarefas de acompanhamento, como redigir um e-mail a partir das notas. Por enquanto, só com inscrição para acesso antecipado.

> **Por que isso importa:** A ElevenLabs sai da venda de vozes e transcrição para a briga com Granola, Fireflies e Wispr Flow, usando os próprios modelos de voz como vitrine para empresas.

Fonte original: TechCrunch https://techcrunch.com/2026/11/13/elevenlabs-new-app-combines-meeting-note-taking-and-dictation-for-teams/

IMPORTÂNCIA MÉDIA 13/11/2026 · Tecnoblog

Dono do Tinder e do Hinge proíbe agentes de IA de escolher perfis e mandar mensagens no lugar dos usuários

O Match Group, controlador do Tinder e do Hinge, vai impedir que agentes de inteligência artificial assumam as interações dos usuários nos seus aplicativos. A regra vale para ferramentas capazes de analisar perfis, escolher possíveis pretendentes e enviar mensagens em nome de uma pessoa real. O CEO Spencer Rascoff explicou a decisão num artigo no The New York Times: "Decidimos que não queremos que agentes de IA assumam as funções de nossos aplicativos de namoro", disse, defendendo que humanos avaliem perfis de outros humanos.

> **Por que isso importa:** Enquanto OpenAI e outras empresas põem agentes para comprar e agir pelo usuário, alguns serviços começam a traçar limites de onde eles não entram, o que vai moldar até onde os agentes pessoais chegam.

Fonte original: Tecnoblog https://tecnoblog.net/noticias/tinder-proibe-que-agentes-de-ia-conversem-pelos-usuarios/

Negócios e Investimento

3 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA ALTA 13/11/2026 · CNBC

Atualização: conselho da NVIDIA procurou Rene Haas, CEO da Arm, para suceder Jensen Huang, diz o FT; empresa afirma que ainda escolhe a firma de recrutamento

O Financial Times informou nesta sexta (13/11), citando fontes, que o conselho da NVIDIA procurou Rene Haas, CEO da Arm, na busca por um sucessor permanente de Jensen Huang. Haas trabalhou sete anos na NVIDIA, onde chegou a vice-presidente da divisão de produtos de computação, antes de entrar na Arm em 2013; segundo o FT, ele é um entre vários nomes considerados, sem garantia de convite ou de aceite. Em nota à CNBC, o conselho disse estar "no processo de escolher uma empresa de recrutamento" para o cargo, que "não procurou formalmente nenhum candidato" e que qualquer especulação sobre nomes, internos ou externos, é prematura. A diretora financeira Colette Kress segue como CEO interina. As ações da NVIDIA caíram mais de 2% na sexta, e o valor de mercado ficou abaixo de US$ 4 trilhões.

> **Por que isso importa:** A escolha do novo CEO vai definir o rumo da maior fornecedora de chips de IA no meio de investigações nos EUA e da disputa com a China; o nome de um ex-funcionário que hoje comanda a Arm mostra o perfil que o conselho estaria sondando.

Fonte original: CNBC https://www.cnbc.com/2026/11/13/nvidia-huang-ceo-replacement-search.html
g1: Nvidia avalia CEO de concorrente https://g1.globo.com/tecnologia/noticia/2026/11/13/nvidia-rene-haas-arm-holdings.ghtml
Financial Times: Arm boss approached https://www.ft.com/content/2a64f698-46dd-4178-a3f4-4c5f4c0a0926

IMPORTÂNCIA ALTA 13/11/2026 · Fortune

OpenAI desfaz a equipe separada de robótica e aposta que seus modelos gerais, como o GPT-6 Astra, podem controlar robôs

A OpenAI confirmou à Fortune que desfez sua equipe independente de robótica. Os pesquisadores foram para a equipe principal de pesquisa, que responde ao diretor de pesquisa Mark Chen, e as equipes de robótica aplicada e de hardware passaram para a área Applied, comandada pelo presidente Greg Brockman. Um porta-voz disse que a robótica continua central na visão de longo prazo e que a mudança é sinal de compromisso, não de recuo: a empresa agora acredita que os modelos gerais podem ser aplicados direto aos robôs. A Fortune cita um estudo da NVIDIA em que um agente com o Astra teve 74,8% de sucesso num teste simulado de manipulação, contra 51% dos melhores modelos especializados em robótica nesse ranking. Para o professor Russ Tedrake, do MIT e da Walden Robotics, as empresas de robôs ainda têm a vantagem dos dados colhidos em fábricas e outros lugares reais.

> **Por que isso importa:** Se modelos gerais derem conta de controlar robôs, a tese que sustenta bilhões investidos em startups de "cérebros" próprios para robôs fica em xeque, e a corrida passa a ser por dados do mundo real.

Fonte original: Fortune https://fortune.com/2026/11/13/openai-disbands-robotics-team-as-its-general-models-start-running-robots/
Infobae: OpenAI dissolve equipe de robótica https://www.infobae.com/tecno/2026/11/13/openai-disolvio-su-equipo-de-robotica-porque-sus-modelos-generales-ya-pueden-controlar-robots/

IMPORTÂNCIA MÉDIA 13/11/2026 · Infobae

SpaceX corta cerca de 60 pessoas da equipe do Grok Imagine, o gerador de imagens e vídeos do Grok, diz o The Information

A divisão de IA da SpaceX eliminou cerca de 60 vagas da equipe que desenvolve o Grok Imagine, ferramenta que gera imagens e vídeos a partir de texto, segundo o The Information, citado pelo Infobae. Ficou só um grupo pequeno para manter o produto, que segue disponível. Segundo resumos do relatório, a empresa agora prioriza as ferramentas de programação com IA e seu produto de agentes; gerar vídeo exige muito mais computação que gerar texto. A SpaceX não confirmou os cortes. A xAI foi comprada pela SpaceX em fevereiro.

> **Por que isso importa:** Mostra a troca de prioridades nos laboratórios: vídeo gerado por IA é caro e disputado, enquanto código e agentes viraram a aposta para gerar receita.

Fonte original: Infobae https://www.infobae.com/tecno/2026/11/13/spacex-despidio-a-unas-60-personas-del-equipo-que-crea-videos-con-ia-de-grok/

Regulação

3 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA ALTA 13/11/2026 · Euronews

Empresa criada por Alice Huang em Xangai ficou com servidores com chips H8oo da NVIDIA, proibidos para a China, diz a Bloomberg; Pequim alerta os EUA

Documentos entregues ao governo chinês mostram que a Shanghai Longteng Intelligent Computing, fundada por Alice Huang, ex-executiva da Megaspeed, ficou com servidores com chips H800 da NVIDIA, cuja exportação para a China os EUA proibiram em 2023, segundo a Bloomberg. Os servidores foram comprados da Dell pela Aperia Cloud Services, parceira oficial da NVIDIA em Singapura; não está claro como chegaram à Longteng nem se Alice Huang participou. Segundo o Engadget, ela saiu da empresa em 2017, hoje controlada por uma irmã, e os documentos listam também 52 servidores da Super Micro com chips H100 ou H200. A China, que nega que Alice Huang seja espiã, disse se opor ao "abuso de medidas de controle de exportação", enquanto o presidente da Câmara dos EUA, Mike Johnson, disse ser a favor de limitar essas exportações.

> **Por que isso importa:** Segundo a Bloomberg, os documentos são evidência explícita de uma rede de empresas chinesas com chips proibidos tendo Alice Huang no centro, e o caso chega antes do encontro esperado entre Trump e Xi Jinping, com o Congresso pressionando por regras mais duras de exportação de chips de IA.

Fonte original: Euronews https://www.euronews.com/2026/11/13/china-warns-us-over-chip-curbs-as-suspected-spys-firm-got-banned-chips
Engadget: a rede de empresas descrita pela Bloomberg https://www.engadget.com/2315710/bloomberg-details-web-of-chinese-ai-companies-with-smuggled-nvidia-chips-linked-to-alice-huang/

IMPORTÂNCIA MÉDIA 13/11/2026 · Fortune

Atualização: Bessent diz em público que Tesouro, Defesa, Estado e USTR queriam pôr a Megaspeed na lista negra havia meses, mas "o Comércio tem a caneta"

O secretário do Tesouro dos EUA, Scott Bessent, disse à CNBC nesta sexta (13/11) que Tesouro, Departamento de Guerra, USTR e Departamento de Estado eram a favor de incluir a Megaspeed na lista de entidades "havia meses", mas que a decisão cabia ao Departamento de Comércio, de Howard Lutnick: "O Comércio tem a caneta. Vocês terão de perguntar a eles sobre os processos internos." Sobre a discussão na Sala de Situação noticiada antes pela Axios, disse apenas que não repete conversas feitas lá. Segundo a Fortune, senadores dos dois partidos cobram explicações de Lutnick, que, segundo a CNBC, aceitou depor na Comissão Bancária do Senado em 10 de dezembro.

> **Por que isso importa:** A crítica pública de um colega de gabinete aumenta a pressão sobre Lutnick e sobre a política de exportação de chips da NVIDIA para a China, a poucos dias das audiências no Congresso.

Fonte original: Fortune https://fortune.com/2026/11/13/scott-bessent-treasury-blacklist-megaspeed-howard-lutnick-nvidia-jensen-huang/
CNBC: Lutnick aceita depor no Senado https://www.cnbc.com/2026/11/13/nvidia-scandal-lutnick-chip-sales.html

IMPORTÂNCIA MÉDIA 13/11/2026 · Infobae

Empresa de software chamada Mythos processa a Anthropic nos EUA pelo uso do nome no modelo de IA

A Global Printing Inc, que opera com o nome comercial Mythos e oferece desde 2015 uma plataforma de software para organizações sem fins lucrativos, processou a Anthropic no Tribunal Federal do Distrito Leste da Virgínia, segundo a Reuters, citada pelo Infobae. A empresa, de Alexandria, diz que o modelo Mythos da Anthropic vai confundir o mercado e causar "dano grave e irreparável", e pede indenização e uma ordem que proíba a Anthropic de usar o nome. Um porta-voz da Anthropic disse discordar das acusações e que vai se defender. Em outubro, o escritório de marcas dos EUA já tinha rejeitado de forma preliminar dois pedidos da Anthropic para registrar a marca "Claude Mythos", por risco de confusão com marcas existentes.

> **Por que isso importa:** Briga de marca pode obrigar a trocar o nome de um produto, e chega dias antes da estreia da Anthropic na bolsa.

Fonte original: Infobae https://www.infobae.com/tecno/2026/11/13/una-empresa-de-software-demanda-a-anthropic-por-llamar-mythos-a-su-modelo-de-ia/

Open Source

1 MATÉRIA NESTA EDIÇÃO

IMPORTÂNCIA MÉDIA 13/11/2026 · Cloudflare

Cloudflare abre o código do Kitesurf, navegador leve feito para agentes de IA que roda sem contêiner de browser

A Cloudflare publicou no GitHub, sob licença Apache 2.0, o Kitesurf, seu motor de navegador experimental para agentes de IA, apresentado em agosto. Ele roda inteiro sobre o Cloudflare Workers: interpreta HTML, aplica CSS, executa o JavaScript das páginas e gera capturas de tela sem um contêiner de navegador, e aceita o Chrome DevTools Protocol, o que permite ligar clientes como o Puppeteer. A ideia é dar aos agentes um navegador que carregue sites e interaja com páginas gastando menos CPU e memória que um Chromium completo. Agora dá para ler o código, rodar localmente, modificar e publicar a própria instância.

> **Por que isso importa:** Navegar é uma das tarefas mais caras para agentes; um motor aberto e leve dá alternativa a quem hoje paga por frotas de Chrome na nuvem.

Fonte original: Cloudflare https://blog.cloudflare.com/kitesurf-open-source/
Kitesurf no GitHub https://github.com/cloudflare/kitesurf

Agentes e Código

1 MATÉRIA NESTA EDIÇÃO

IMPORTÂNCIA MÉDIA 13/11/2026 · InfoQ

Google abre o código do AQuA, agente que vasculha sessões de outros agentes em produção para achar falhas silenciosas e a causa no código

O Google publicou como código aberto o AQuA (Ambient Quality Agent), implementação de referência que analisa as sessões de agentes de IA em produção, confirma padrões de falha recorrentes e investiga a causa no código publicado. O alvo são erros que o monitoramento comum não pega, como agentes que ignoram preferências do usuário ou pulam chamadas de ferramenta obrigatórias e ainda assim devolvem resposta de sucesso. Ele roda ao lado do agente no Google Cloud, lendo dados do Cloud Trace, do Cloud Logging ou do BigQuery, analisa até 1.000 sessões recentes por rodada e, para achar a causa, compara as sessões com problema com uma cópia do código guardada no momento do deploy, citando arquivo e linhas e propondo mudanças. Num teste com 32 sessões, os desenvolvedores reduziram de 15 para 2 os casos em que o agente pulava a escolha de assento e de 7 para 0 os que esqueciam a preferência alimentar.

> **Por que isso importa:** Agente em produção falha de forma silenciosa; ferramentas que vigiam a qualidade das conversas, e não só se o servidor está de pé, viram peça obrigatória para quem leva agentes a clientes.

Fonte original: InfoQ https://www.infoq.com/news/2026/11/google-ambient-agent-prod/

Modelos Locais

2 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA ALTA 13/11/2026 · Hugging Face (ggml-org)

llama.cpp chega à versão 1.0, três anos e meio depois de nascer, com promessa de API estável para quem roda IA local

O ggml-org lançou nesta sexta (13/11) a versão 1.0.0 das bibliotecas do llama.cpp e do ggml, o motor que permite rodar modelos de linguagem multimodais em praticamente qualquer hardware, do notebook ao servidor. Segundo o anúncio, muitos aplicativos já dependem do projeto, e a nova numeração assume a responsabilidade com eles: a partir de agora, mudanças que quebrem a API ou a ABI virão com troca de versão principal, seguindo o versionamento semântico em todos os projetos do ggml-org. A equipe diz que vai investir mais na infraestrutura de testes automáticos, com HF Buckets e HF Jobs e apoio de hardware de NVIDIA, AMD e Qualcomm, entre outros, e que segue usando LLMs no desenvolvimento "de forma cuidadosa".

> **Por que isso importa:** O llama.cpp está por trás de boa parte das ferramentas de IA local; uma versão 1.0 com regras claras de compatibilidade dá mais segurança a quem constrói produtos sobre ele.

Fonte original: Hugging Face (ggml-org) https://huggingface.co/blog/ggml-org/llama-cpp-v1
Release v1.0.0 no GitHub https://github.com/ggml-org/llama.cpp/releases/tag/v1.0.0
Discussão no r/LocalLLaMA https://www.reddit.com/r/LocalLLaMA/comments/1xv8t2e/llamacpp_v10_is_here/

IMPORTÂNCIA MÉDIA 13/11/2026 · NVIDIA Technical Blog

NVIDIA mostra como servir o GLM-5.3, de 753 bilhões de parâmetros, em duas DGX Station ligadas em rede, com vLLM

A versão NVFP4 do GLM-5.3, modelo de mistura de especialistas da Z.ai com 753 bilhões de parâmetros e 40 bilhões ativos por token, precisa de cerca de 464 GB de memória de GPU, mais que os 252 GB de HBM3e de uma GPU GB300. O guia da NVIDIA liga duas DGX Station por ConnectX-8 com RDMA e divide o modelo entre as duas GPUs (paralelismo de tensor 2), servindo tudo por um único endpoint compatível com a API da OpenAI, com predição de vários tokens, cache KV em FP8 e cache de prefixo. No teste com o conjunto de código SPEED-Bench, a vazão máxima foi de 607,8 tokens por segundo com 16 pedidos simultâneos e caiu para 385 tokens por segundo com 32, com latência média de 453,5 segundos.

> **Por que isso importa:** Dá a receita e os números reais para quem quer rodar um modelo aberto de fronteira dentro de casa ou do escritório, sem nuvem, e mostra o limite de concorrência dessa configuração.

Fonte original: NVIDIA Technical Blog https://developer.nvidia.com/blog/when-one-gpu-isnt-enough-serving-glm-5-3-across-two-nvidia-dgx-stations-with-vllm/
NVIDIA: como ligar duas DGX Station https://build.nvidia.com/station/connect-two-stations

Finanças

2 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA MÉDIA 13/11/2026 · PYMNTS

JPMorgan bloqueia agentes de IA que fazem compras com o número do cartão do cliente, diz o Wall Street Journal

O JPMorgan Chase está barrando agentes de IA que compram com os dados do cartão de crédito quando o próprio cliente entrega o número ao agente, segundo o Wall Street Journal, citado pelo PYMNTS nesta sexta (13/11). Cartões do Bank of America, do Capital One e do Citigroup já são usados por agentes quando o cliente autoriza. O Stripe contorna o bloqueio com a carteira Link, que dá ao agente um cartão virtual de uso único em vez do cartão real, mas lojistas dizem que esse cartão é classificado como corporativo e custa mais para aceitar. Agentes de fora já não podem entrar com as credenciais de clientes do Chase, transferir dinheiro nem pagar contas.

> **Por que isso importa:** No mesmo dia do ChatGPT Wallet, mostra a outra ponta da história: os bancos ainda decidem se e como os agentes podem gastar o dinheiro do cliente, e cada um está escolhendo um caminho.

Fonte original: PYMNTS https://www.pymnts.com/news/2026/jpmorgan-stops-ai-agents-from-using-customer-credit-cards/
Engadget: ChatGPT Wallet e os riscos https://www.engadget.com/2316137/you-can-now-tempt-fate-and-give-your-banking-information-to-chatgpt/

IMPORTÂNCIA MÉDIA 13/11/2026 · Google

Gemini passa a ler o Google Wallet e as contas bancárias ligadas a ele para analisar gastos e montar orçamentos, nos EUA

O Google começou a liberar nesta sexta (13/11), para todos os usuários do app Gemini nos EUA, na web, no Android e no iOS, a conexão com o Google Wallet dentro do Personal Intelligence. Com ela, o Gemini responde com base nos hábitos de gasto do usuário e, se ele ligar cartões e contas bancárias ao Wallet pelo Plaid, pode montar gráficos de fluxo de caixa e orçamento, achar assinaturas esquecidas e aumentos de preço e sugerir um orçamento para as festas a partir dos gastos do ano anterior. Segundo o Google, o Gemini só lê as informações: não pode mover, administrar nem investir o dinheiro, e o recurso só funciona se o usuário optar por ativá-lo.

> **Por que isso importa:** Os grandes assistentes estão entrando nas finanças pessoais por caminhos diferentes: o Google começa só lendo dados, enquanto a OpenAI já deixa o agente pagar.

Fonte original: Google https://blog.google/innovation-and-ai/products/gemini-app/google-wallet-gemini-app/
The Next Web: Gemini lê o Google Wallet https://thenextweb.com/news/google-gemini-wallet-spending-insights-plaid

Games

2 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA MÉDIA 13/11/2026 · GamesBeat

Tim Sweeney, da Epic, diz que a IA já multiplica por quatro ou mais a produtividade em várias áreas e pode chegar a dez vezes, abrindo espaço para novos gêneros

Em entrevista exclusiva ao GamesBeat, com o vice-presidente executivo de engenharia da Epic, Marcus Wassmer, o CEO Tim Sweeney disse que a IA está transformando o desenvolvimento de jogos e estima ganhos de produtividade de quatro vezes ou mais em muitas áreas, que podem chegar a dez vezes com a evolução da tecnologia. Wassmer citou um experimento inicial de publicação do Unreal Engine na web com ganho de até 20 vezes. Para Sweeney, prototipar mais rápido ajuda a resolver um dos maiores problemas dos jogos AAA, descobrir tarde demais se o jogo é divertido, e pode trazer "uma era de experimentação", com novos gêneros e combinações, sobretudo em tiro e em jogos online massivos.

> **Por que isso importa:** Quando o dono do Unreal Engine fala em ganhos de 4 a 10 vezes, isso pesa nas ferramentas e no custo de estúdios do mundo todo, inclusive os brasileiros que usam a engine.

Fonte original: GamesBeat https://gamesbeat.com/epic-games-ceo-tim-sweeney-predicts-ai-revolution-in-gaming-with-10x-productivity-and-entirely-new-genres-if-devs-get-behind-it-exclusive/

IMPORTÂNCIA BAIXA 13/11/2026 · Eurogamer

Pearl Abyss, de Crimson Desert, diz que está levando IA para toda a produção e prevê o DokeV para 2028

No relatório trimestral aos investidores, a sul-coreana Pearl Abyss disse que antes só validava sistemas de jogo depois de implementá-los à mão e que agora usa IA para testar rapidamente vários conceitos de design antes da implementação. A empresa afirma que a IA já é usada em várias etapas, como automação de tarefas repetitivas, apoio à criação de conteúdo e otimização de testes, e que está expandindo o uso "em toda a pipeline de desenvolvimento", mas que a jogabilidade e a criação do conteúdo seguem nas mãos da equipe. O estúdio prevê lançar o DokeV em 2028. Crimson Desert já tinha sido criticado por imagens geradas por IA no lançamento, depois trocadas.

> **Por que isso importa:** Mais um grande estúdio assume publicamente o uso de IA na produção, mesmo depois de ter sido criticado pelos jogadores por isso.

Fonte original: Eurogamer https://www.eurogamer.net/crimson-desert-ai-pipeline-dokev-2028

Brasil

3 MATÉRIAS NESTA EDIÇÃO

IMPORTÂNCIA MÉDIA 13/11/2026 · Convergência Digital

Startup brasileira AI.CUBE fecha contratos de R$ 270 milhões com AWS e Oracle para ter GPUs NVIDIA e oferecer inferência privada

A AI.CUBE, startup brasileira de IA agêntica para organizações públicas e privadas, assinou contratos que somam R$ 270 milhões (US$ 50 milhões) com a AWS, por três anos, com acesso a GPUs NVIDIA B200, e com a Oracle, por cinco anos, com GPUs GB200. A estrutura foi dimensionada para uma capacidade média estimada de 7,6 bilhões de tokens por hora, voltada principalmente a agentes de IA, e a maior parte vai atender demandas no Brasil, sobretudo de órgãos do governo federal, instituições financeiras e empresas de setores regulados; outra parte apoia a expansão na América Latina. A empresa, criadora do Qilbee, chama a estrutura de "AI Factory", com ambientes dedicados para quem não pode expor dados sensíveis.

> **Por que isso importa:** Mostra a demanda de governo e bancos brasileiros por IA rodando em ambiente dedicado, e o tamanho do compromisso com GPUs que uma startup local precisa assumir para atendê-la.

Fonte original: Convergência Digital https://convergenciadigital.com.br/inovacao/startup-brasileira-de-ia-agentica-assina-contratos-de-r-270-milhoes-com-aws-e-oracle/

IMPORTÂNCIA MÉDIA 13/11/2026 · Convergência Digital

Governo federal leva para a web o sistema de autoavaliação de impacto ético em IA, agora aberto a mais de 250 órgãos

O Ministério da Gestão e da Inovação em Serviços Públicos (MGI), com consultoria do Banco Mundial, lançou a versão web do Sistema de Autoavaliação de Impacto Ético em Inteligência Artificial (AIE), disponível para os mais de 250 órgãos do Sistema de Administração de Recursos de Tecnologia da Informação (SISP). A ferramenta, que antes só existia off-line e faz parte do Plano Brasileiro de Inteligência Artificial (PBIA), ajuda a identificar riscos, avaliar práticas éticas e de governança e recomenda medidas de melhoria para sistemas de IA desenvolvidos ou usados pelo governo. Segundo o secretário adjunto de Governo Digital, Guilherme Almeida de Almeida, a ideia é apoiar um ciclo de melhoria, não só fazer um diagnóstico.

> **Por que isso importa:** Dá aos órgãos públicos um roteiro comum para avaliar riscos de IA antes de usá-la com o cidadão, enquanto o marco legal da IA segue em discussão.

Fonte original: Convergência Digital https://convergenciadigital.com.br/governo/governo-adota-sistema-de-autoavaliacao-do-impacto-etico-em-inteligencia-artificial/

IMPORTÂNCIA BAIXA 13/11/2026 · IT Forum

Americana GyanSys compra a consultoria brasileira UPBI, de dados e IA, por valor não revelado

A GyanSys, integradora e consultoria americana de transformação digital com sede em Indiana e cerca de 3.000 funcionários, anunciou a compra integral da UPBI, consultoria brasileira de dados e inteligência artificial com 15 anos de mercado, 120 funcionários e faturamento de R$ 26 milhões em serviços no ano passado. A UPBI, parceira especialista da Microsoft no Brasil e com clientes como Fast Shop, Unimed e CBMM, projeta crescer 20% com a aquisição. O diretor-presidente José Paulo Abi Jaudi e os sócios-fundadores seguem na empresa.

> **Por que isso importa:** Consultorias de dados com experiência em IA viraram alvo de compra de grupos estrangeiros que querem crescer no mercado brasileiro.

Fonte original: IT Forum https://itforum.com.br/gyansys-adquire-upbi-ia-dados/

Fofocas das redes

4 MATÉRIAS NESTA EDIÇÃO

REDDIT IMPORTÂNCIA MÉDIA 13/11/2026 · r/LocalLLaMA

Desenvolvedor roda o Qwen3.5-9B na GPU de um PlayStation 4 comum, sem internet, e o post vai ao topo do r/LocalLLaMA

O autor (u/leonardo_alegria) diz que o Qwen3.5-9B roda num PS4 de fábrica a 31 a 39 tokens por segundo, usando a própria cabeça de predição de vários tokens (MTP) do modelo para decodificação especulativa, o que, segundo ele, subiu a velocidade de 22 para 31 a 39 tokens por segundo. Os pesos foram reempacotados em formatos de 4 e 3 bits para caber em cerca de 4,5 GiB. O projeto, garlic-llm, é gratuito e de código aberto, e o autor diz que o modelo conversa, ouve e escreve código, sem nuvem nem assinatura.

> **Por que isso importa:** Mostra até onde a compressão e a decodificação especulativa já levam modelos úteis em hardware velho, um dos temas preferidos de quem roda IA local.

Post original (Reddit): r/LocalLLaMA https://www.reddit.com/r/LocalLLaMA/comments/1xvbn2m/qwen359b_running_on_a_ps4s_gpu_offline_using_its/
garlic-llm no GitHub https://github.com/leonardo-alegria/garlic-llm

HACKER NEWS IMPORTÂNCIA MÉDIA

13/11/2026 · news.ycombinator.com

“Se a IA resolveu a programação, por que meu software continua igual?”: conversa com Jonathan Blow, criador de Braid, agita o Hacker News

Um post no Hacker News com o vídeo do canal The Students, em que Jonathan Blow, criador do jogo Braid, e Wolf Tivy discutem a explosão de complexidade no software, por que a IA de hoje não produz código melhor que o de pessoas e como seria o mundo se produzisse, juntou mais de cem pontos e de cem comentários nesta sexta.

> **Por que isso importa:** Resume o ceticismo de parte dos programadores experientes diante da promessa de que agentes de código já resolveram a programação.

Engajamento: 139 pontos · 125 comentários (medido em 13/11/2026 17:57)
Post original (Hacker News): news.ycombinator.com https://news.ycombinator.com/item?id=50484658
Vídeo no YouTube (The Students) https://www.youtube.com/watch?v=IRK8h67ol7Q

REDDIT IMPORTÂNCIA BAIXA 13/11/2026 · r/artificial

Primeiro colocado no exame de contador tributário da Coreia é acusado nas redes, sem provas, de usar óculos com IA

Segundo o post, que cita o Korea Herald, o primeiro colocado tirou 397,5 de 400 na primeira fase, acertando 159 de 160 questões, inclusive nota máxima numa prova em que a média dos 16.856 candidatos foi 30,18 de 100. Internautas passaram a dizer que só seria possível com óculos inteligentes com IA. Não há prova: os dois fiscais não relataram nada, o órgão do exame diz que está revisando o caso, e o candidato diz que estudou por dois anos. A associação de contadores pediu investigação, mas alertou que acusações sem prova podem destruir a reputação dele. O autor pergunta como provas presenciais devem lidar com isso.

> **Por que isso importa:** É boato, sem evidência, mas mostra como óculos com IA já colocam sob suspeita qualquer resultado fora da curva em provas presenciais.

Post original (Reddit): r/artificial https://www.reddit.com/r/artificial/comments/1xv6aut/koreas_tax_accountant_exam_top_scorer_is_facing/
Korea Herald https://www.koreaherald.com/article/10936587

REDDIT IMPORTÂNCIA BAIXA 13/11/2026 · r/LocalLLaMA

DeepSeek V4.1 Flash é o primeiro modelo aberto a passar de 1% no SWE-sweep, teste em que o agente precisa achar e corrigir centenas de bugs sem pista

O responsável pelo ranking SWE-sweep (u/klieret) postou que o DeepSeek V4.1 Flash, no modo máximo, é o primeiro modelo de pesos abertos avaliado a passar de 1%. No teste, o modelo recebe uma cópia de um repositório, às vezes muito grande, com centenas de bugs, sem nenhuma dica, e precisa corrigir todos. Segundo o post, o melhor modelo fechado é o Gemini 4 Hydrogen, com 13,5% pelo cartão do Google, seguido do GPT-6 Astra, com 9,5% na avaliação deles: “os modelos abertos ainda têm um longo caminho”.

> **Por que isso importa:** Mostra a distância entre modelos abertos e fechados em tarefas longas de código, mesmo com os abertos subindo rápido em outros rankings.

Post original (Reddit): r/LocalLLaMA https://www.reddit.com/r/LocalLLaMA/comments/1xvcgxq/deepseek_flash_max_is_the_first_model_to_reach_1/

Vídeos

3 VÍDEOS

Os resumos se baseiam no título verificado e nos anúncios relacionados lidos para esta edição. Clique na miniatura ou no título para abrir no YouTube.



How Mozilla uses GPT-6 Cyber to find and fix Firefox vulnerabilities

OpenAI · 13/11/2026

Brian Grinstead, da Mozilla, conversa com Clint Gibler, da OpenAI, sobre como a equipe usa agentes com o GPT-6 Cyber para priorizar investigações, testar e confirmar falhas do Firefox, achar padrões de bugs parecidos e ajudar os engenheiros a corrigi-los. Segundo o vídeo, a Mozilla viu a descoberta e correção de bugs crescer cerca de oito vezes de 2025 para 2026.

https://www.youtube.com/watch?v=PkAmyub12K8



Cyberattacks Used to Take Months. AI Can Do It in Minutes.

a16z · 13/11/2026

Ami Luttwak, cofundador e CTO da Wiz, conversa com Zane Lackey e Sarah Wang, da a16z, sobre o que chama de “Black October”, uma onda de falhas e ataques com agentes, e sobre como a IA encurta o tempo entre descobrir uma vulnerabilidade e explorá-la, por que startups com credenciais de IA viraram alvo e como as empresas precisam se defender.

https://www.youtube.com/watch?v=PbD9pVXxsYM



Niels Provos on creating bcrypt, defending Google, and securing AI agents

GitHub · 13/11/2026

O pesquisador de segurança Niels Provos, coautor do bcrypt e criador do Google Safe Browsing, conta três décadas de segurança de software e explica por que está desenvolvendo o Project Iron Curtain, para proteger agentes autônomos contra injeção de prompt e desvio de intenção.

https://www.youtube.com/watch?v=EJVwqt8EVOo

SOBRE ESTA EDIÇÃO
Nº 28 — Edição da noite · 13/11/2026

JANELA
Cobre o que saiu desde a edição nº 27 (fechada por volta das 7h50 de Brasília de 13/11) até por volta das 18h de Brasília de 13/11, mais uma pauta de 12/11 que ainda não tinha saído (os pesos do Inkling 2 Small, que entra como atualização; o que já tinha saído, como o Gemini 4 Hydrogen, o Claude Meetings, o Claude for Manufacturers, o Claude Managed Agents na nuvem do cliente e o Daybreak Red Team, ficou de fora). A manchete é o ChatGPT Wallet, com que os agentes do ChatGPT passam a comprar pelo usuário. A edição traz ainda as atualizações da pane da Microsoft e da sucessão na NVIDIA, os pesos do Inkling 2 Small, o Decider v1.3 da Perplexity, os cadernos de Gödel decifrados com o Claude, o guia de avaliações de segurança da Anthropic, o ElevenAI, a proibição de agentes no Tinder, o fim da equipe separada de robótica da OpenAI, os cortes no Grok Imagine, os chips proibidos ligados a Alice Huang, as críticas de Bessent a Lutnick, o processo pelo nome Mythos, o código aberto do Kitesurf, o AQuA do Google, o llama.cpp 1.0, o GLM-5.3 em duas DGX Station, o bloqueio do JPMorgan a agentes com cartão, o Google Wallet no Gemini, Tim Sweeney e a Pearl Abyss nos games e, no Brasil, a AI.CUBE, o sistema de autoavaliação ética do governo e a compra da UPBI. Varredura feita pela coleta do catálogo de fontes (551 fontes da camada edição, 547 responderam, 1.893 itens na janela) rodada na box, mais Hugging Face, GitHub, Reddit, Hacker News, X e YouTube, porque o Mac Mini e o notebook i5 não estavam ligados ao Grok Bot nesta hora. O vigia de hora em hora não deu alerta desde a última edição; as pautas menores que ele anotou foram conferidas na fonte original. Fechamento: 13/11/2026, por volta das 18h de Brasília.

FONTES VARRIDAS
Blogs e changelogs oficiais (OpenAI, Anthropic, Google DeepMind, Meta, xAI, Microsoft, NVIDIA, Apple, Mistral, DeepSeek, Qwen, Moonshot), arXiv e HF Papers, trending do Hugging Face e do GitHub, X, imprensa dos EUA, Europa, China, Japão, Índia e Brasil, e YouTube.

CRITÉRIO
Só entram notícias lidas numa fonte real, de preferência primária, com todos os links checados. Páginas com bloqueio (403) foram mantidas; links quebrados (404) foram descartados.

Fontes e links

▶ **VÍDEOS SUGERIDOS**

1. **How Mozilla uses GPT-6 Cyber to find and fix Firefox vulnerabilities** — OpenAI, 13/11/2026
   https://www.youtube.com/watch?v=PkAmyub12K8
2. **Cyberattacks Used to Take Months. AI Can Do It in Minutes.** — a16z, 13/11/2026
   https://www.youtube.com/watch?v=PbD9pVXxsYM
3. **Niels Provos on creating bcrypt, defending Google, and securing AI agents** — GitHub, 13/11/2026
   https://www.youtube.com/watch?v=EJVwqt8EVOo

Manchete

Introducing ChatGPT Wallet — OpenAI
https://openai.com/index/introducing-chatgpt-wallet/

OpenAI launches ChatGPT Wallet, allowing its 'dots' AI agents to make purchases for users — CNBC
https://www.cnbc.com/2026/11/13/openai-chatgpt-wallet-dot-purchases.html

ChatGPT's new Wallet lets AI agents shop and check out for you — TechCrunch
https://techcrunch.com/2026/11/13/chatgpts-new-wallet-lets-ai-agents-shop-and-check-out-for-you/

OpenAI rebuilt AI shopping from scratch with ChatGPT Wallet — The Decoder
https://the-decoder.com/openai-rebuilt-ai-shopping-from-scratch-with-chatgpt-wallet/

OpenAI lança carteira digital que permite ao ChatGPT fazer compras e pagamentos pelo usuário — g1
https://g1.globo.com/tecnologia/noticia/2026/11/13/chat-gpt-wallet.ghtml

Modelos

Atualização: Thinking Machines cumpre a promessa e libera os pesos do Inkling 2 Small no Hugging Face, com licença Apache 2.0
Hugging Face (Thinking Machines Lab): https://huggingface.co/thinkingmachines/Inkling-2-Small
Thinking Machines: anúncio com a atualização de 12/11: https://thinkingmachines.ai/news/inkling-2-small/

Perplexity lança o Decider v1.3, modelos abertos que respondem com probabilidades em vez de texto e agora localizam trechos exatos
MarkTechPost: https://www.marktechpost.com/2026/11/13/perplexity-decider-v1-3-open-weight-decision-models-locate/
Perplexity: Decider v1.3: https://www.perplexity.ai/hub/blog/perplexity-decider

Pesquisa

Com ajuda do Claude, lógico finlandês decifra os cadernos de Kurt Gödel e encontra centenas de teoremas que ninguém conhecia
The Guardian: https://www.theguardian.com/science/2026/nov/13/kurt-godel-notebooks-ai-helps-unlock-hidden-discoveries-man-who-transformed-maths
Discussão no Hacker News: https://news.ycombinator.com/item?id=50486748

Anthropic publica guia de avaliações de segurança e mostra que o mesmo modelo foi de 60% a 94% de respostas adequadas só mudando o prompt de sistema
Anthropic: https://www.anthropic.com/research/building-strong-safety-evaluations
The Next Web: as notas dependem de como se mede: https://thenextweb.com/news/anthropic-safety-evaluations-guidance-fable-self-harm

Produtos

Atualização: Microsoft pede desculpas pela pane de cerca de 24 horas e diz que a causa foi um upgrade de capacidade da rede, não ataque nem IA
Wired: https://www.wired.com/story/microsoft-365-crawls-back-to-life-after-a-24-hour-outage/
PC Gamer: a nota da Microsoft: https://www.pcgamer.com/software/microsoft-deeply-apologizes-for-global-service-outage-says-no-ai-or-hackers-are-to-blame/

ElevenLabs lança o ElevenAI, app em teste que junta anotações de reuniões, ditado e agentes para equipes
TechCrunch: https://techcrunch.com/2026/11/13/elevenlabs-new-app-combines-meeting-note-taking-and-dictation-for-teams/

Dono do Tinder e do Hinge proíbe agentes de IA de escolher perfis e mandar mensagens no lugar dos usuários
Tecnoblog: https://tecnoblog.net/noticias/tinder-proibe-que-agentes-de-ia-conversem-pelos-usuarios/

Negócios e Investimento

Atualização: conselho da NVIDIA procurou Rene Haas, CEO da Arm, para suceder Jensen Huang, diz o FT; empresa afirma que ainda escolhe a firma de recrutamento
CNBC: https://www.cnbc.com/2026/11/13/nvidia-huang-ceo-replacement-search.html
g1: Nvidia avalia CEO de concorrente: https://g1.globo.com/tecnologia/noticia/2026/11/13/nvidia-rene-haas-arm-holdings.ghtml
Financial Times: Arm boss approached: https://www.ft.com/content/2a64f698-46dd-4178-a3f4-4c5f4c0a0926

OpenAI desfaz a equipe separada de robótica e aposta que seus modelos gerais, como o GPT-6 Astra, podem controlar robôs
Fortune: https://fortune.com/2026/11/13/openai-disbands-robotics-team-as-its-general-models-start-running-robots/
Infobae: OpenAI dissolve equipe de robótica: https://www.infobae.com/tecno/2026/11/13/openai-disolvio-su-equipo-de-robotica-porque-sus-modelos-generales-ya-pueden-controlar-robots/

SpaceX corta cerca de 60 pessoas da equipe do Grok Imagine, o gerador de imagens e vídeos do Grok, diz o The Information
Infobae: https://www.infobae.com/tecno/2026/11/13/spacex-despidio-a-unas-60-personas-del-equipo-que-crea-videos-con-ia-de-grok/

Regulação

Empresa criada por Alice Huang em Xangai ficou com servidores com chips H800 da NVIDIA, proibidos para a China, diz a Bloomberg; Pequim alerta os EUA
Euronews: https://www.euronews.com/2026/11/13/china-warns-us-over-chip-curbs-as-suspected-spys-firm-got-banned-chips
Engadget: a rede de empresas descrita pela Bloomberg: https://www.engadget.com/2315710/bloomberg-details-web-of-chinese-ai-companies-with-smuggled-nvidia-chips-linked-to-alice-huang/

Atualização: Bessent diz em público que Tesouro, Defesa, Estado e USTR queriam pôr a Megaspeed na lista negra havia meses, mas "o Comércio tem a caneta"
Fortune: https://fortune.com/2026/11/13/scott-bessent-treasury-blacklist-megaspeed-howard-lutnick-nvidia-jensen-huang/
CNBC: Lutnick aceita depor no Senado: https://www.cnbc.com/2026/11/13/nvidia-scandal-lutnick-chip-sales.html

Empresa de software chamada Mythos processa a Anthropic nos EUA pelo uso do nome no modelo de IA
Infobae: https://www.infobae.com/tecno/2026/11/13/una-empresa-de-software-demanda-a-anthropic-por-llamar-mythos-a-su-modelo-de-ia/

Open Source

Cloudflare abre o código do Kitesurf, navegador leve feito para agentes de IA que roda sem contêiner de browser
Cloudflare: https://blog.cloudflare.com/kitesurf-open-source/
Kitesurf no GitHub: https://github.com/cloudflare/kitesurf

Agentes e Código

Google abre o código do AQuA, agente que vasculha sessões de outros agentes em produção para achar falhas silenciosas e a causa no código
InfoQ: https://www.infoq.com/news/2026/11/google-ambient-agent-prod/

Modelos Locais

llama.cpp chega à versão 1.0, três anos e meio depois de nascer, com promessa de API estável para quem roda IA local
Hugging Face (ggml-org): https://huggingface.co/blog/ggml-org/llama-cpp-v1
Release v1.0.0 no GitHub: https://github.com/ggml-org/llama.cpp/releases/tag/v1.0.0
Discussão no r/LocalLLaMA: https://www.reddit.com/r/LocalLLaMA/comments/1xv8t2e/llamacpp_v10_is_here/

NVIDIA mostra como servir o GLM-5.3, de 753 bilhões de parâmetros, em duas DGX Station ligadas em rede, com vLLM
NVIDIA Technical Blog: https://developer.nvidia.com/blog/when-one-gpu-isnt-enough-serving-glm-5-3-across-two-nvidia-dgx-stations-with-vllm/
NVIDIA: como ligar duas DGX Station: https://build.nvidia.com/station/connect-two-stations

Finanças

JPMorgan bloqueia agentes de IA que fazem compras com o número do cartão do cliente, diz o Wall Street Journal
PYMNTS: https://www.pymnts.com/news/2026/jpmorgan-stops-ai-agents-from-using-customer-credit-cards/
Engadget: ChatGPT Wallet e os riscos: https://www.engadget.com/2316137/you-can-now-tempt-fate-and-give-your-banking-information-to-chatgpt/

Gemini passa a ler o Google Wallet e as contas bancárias ligadas a ele para analisar gastos e montar orçamentos, nos EUA
Google: https://blog.google/innovation-and-ai/products/gemini-app/google-wallet-gemini-app/
The Next Web: Gemini lê o Google Wallet: https://thenextweb.com/news/google-gemini-wallet-spending-insights-plaid

55 links únicos desta edição, sem repetir URLs.

Games

Tim Sweeney, da Epic, diz que a IA já multiplica por quatro ou mais a produtividade em várias áreas e pode chegar a dez vezes, abrindo espaço para novos gêneros
GamesBeat: https://gamesbeat.com/epic-games-ceo-tim-sweeney-predicts-ai-revolution-in-gaming-with-10x-productivity-and-entirely-new-genres-if-devs-get-behind-it-exclusive/

Pearl Abyss, de Crimson Desert, diz que está levando IA para toda a produção e prevê o DokeV para 2028
Eurogamer: https://www.eurogamer.net/crimson-desert-ai-pipeline-dokev-2028

Brasil

Startup brasileira AI.CUBE fecha contratos de R$ 270 milhões com AWS e Oracle para ter GPUs NVIDIA e oferecer inferência privada
Convergência Digital: https://convergenciadigital.com.br/inovacao/startup-brasileira-de-ia-agentica-assina-contratos-de-r-270-milhoes-com-aws-e-oracle/

Governo federal leva para a web o sistema de autoavaliação de impacto ético em IA, agora aberto a mais de 250 órgãos
Convergência Digital: https://convergenciadigital.com.br/governo/governo-adota-sistema-de-autoavaliacao-do-impacto-etico-em-inteligencia-artificial/

Americana GyanSys compra a consultoria brasileira UPBI, de dados e IA, por valor não revelado
IT Forum: https://itforum.com.br/gyansys-adquire-upbi-ia-dados/

Fofocas das redes

Desenvolvedor roda o Qwen3.5-9B na GPU de um PlayStation 4 comum, sem internet, e o post vai ao topo do r/LocalLLaMA
r/LocalLLaMA: https://www.reddit.com/r/LocalLLaMA/comments/1xvbn2m/qwen359b_running_on_a_ps4s_gpu_offline_using_its/
garlic-llm no GitHub: https://github.com/leonardo-alegria/garlic-llm

"Se a IA resolveu a programação, por que meu software continua igual?": conversa com Jonathan Blow, criador de Braid, agita o Hacker News
news.ycombinator.com: https://news.ycombinator.com/item?id=50484658
Vídeo no YouTube (The Students): https://www.youtube.com/watch?v=IRK8h67ol7Q

Primeiro colocado no exame de contador tributário da Coreia é acusado nas redes, sem provas, de usar óculos com IA
r/artificial: https://www.reddit.com/r/artificial/comments/1xv6aut/koreas_tax_accountant_exam_top_scorer_is_facing/
Korea Herald: https://www.koreaherald.com/article/10936587

DeepSeek V4.1 Flash é o primeiro modelo aberto a passar de 1% no SWE-sweep, teste em que o agente precisa achar e corrigir centenas de bugs sem pista
r/LocalLLaMA: https://www.reddit.com/r/LocalLLaMA/comments/1xvcgxq/deepseek_flash_max_is_the_first_model_to_reach_1/