

OPINIÃO

AI Agents entram em nova fase: vence quem dominar a matemática do negócio

André Fossa (*)

A corrida dos AI Agents nas empresas entrou em uma nova fase e isso muda completamente o critério de sucesso.

Se, até pouco tempo atrás, a principal dúvida era se a tecnologia funcionava na prática, essa barreira começa a ser superada. Segundo a McKinsey, 88% das empresas já utilizam inteligência artificial em pelo menos uma função, mas apenas cerca de um terço conseguiu escalar essas iniciativas. No caso de agências de IA, menos de 10% das organizações afirmam ter aplicações em escala, de acordo com o relatório *The State of AI 2025*. A questão deixou de ser técnica. Agora, é econômica.

É nesse ponto que surge o verdadeiro desafio: os unit economics do negócio vão sepultar uma série de iniciativas (e startups!), separando os prováveis ganhadores dos demais. Em teoria, a IA promete reduzir custos ao substituir ou complementar o trabalho humano. Na prática, isso só acontece em escala, se você tiver um projeto e arquitetura profissionais orientados para IA; não tem espaço para "puxadinhos" nesta etapa do jogo. Em jornadas reais, uma única interação pode consumir de 100 mil a 1 milhão de tokens.

Considerando preços médios entre US\$ 2 e US\$ 2,50 por milhão de tokens, como praticado por grandes provedores, o custo por atendimento pode rapidamente superar, o de um humano. Sem controle, a economia prometida no atendimento migra diretamente para a conta da infraestrutura de IA. E, em escala, isso compromete margem.

Nesse estágio, muitas empresas começam a enfrentar dificuldades. E esse tipo de distorção aparece justamente quando a operação escala. Em piloto, quase tudo funciona. O volume é baixo, os fluxos

são mais controlados e o custo passa despercebido. Em produção, com milhões de interações, qualquer ineficiência se multiplica, e aparece diretamente na margem.

Operar IA em escala pode, sim, gerar ganhos relevantes de eficiência: na ordem de 60% a 80% de redução de custos operacionais. Mas isso não vem "de graça". Esse nível de resultado exige uma arquitetura parruda, pensada desde o início para produção, e não para experimentação. Em escala, as soluções de prateleira e as promessas mágicas simplesmente não se sustentam. A ideia de que é possível resolver uma operação complexa com integrações simples, vibe coding ou ferramentas como n8n funciona até certo ponto, mas quebra rapidamente quando exposta à realidade de grandes operações. Um banco, uma telco ou uma utility não vão rodar sua operação crítica dessa forma. É aqui que começa a separação real do mercado. Plataformas especializadas já conseguem operar AI Agents com até 70% de eficiência de custo (token por token), mantendo níveis de satisfação acima de 90%. Esse tipo de combinação, eficiência econômica com qualidade consistente, é o que vai definir os vencedores.

No fim, a corrida dos AI Agents não será vencida por quem implementar primeiro, mas por quem conseguir sustentar eficiência em escala. A diferença entre um piloto promissor e uma operação transformadora está na capacidade de equilibrar qualidade, escala e custo de forma consistente. A inteligência artificial já provou que funciona. Agora, os unit economics dos AI Agents vão ser implacáveis. E é isso que vai separar quem está apenas experimentando de quem, de fato, vai capturar valor.

(*) Cofundador da Cogni2 e executivo de tecnologia com mais de 25 anos de experiência em transformação digital e inteligência artificial.

Grandes empresas repensam o uso de IA devido a custos elevados

O entusiasmo das grandes companhias com ferramentas de inteligência artificial parece estar dando sinais de desgaste.

Vivaldo José Breternitz (*)

A Microsoft, segundo o portal *The Verge*, estaria incentivando seus funcionários a trocar o Claude Code pelo Copilot, ferramenta da própria Microsoft. O motivo não seria apenas a preferência por tecnologia própria, mas também o aumento expressivo nos custos de utilização do Claude Code.

Esse custo é apurado pelo número de tokens processados: tokens em inteligência artificial são as menores unidades de informação que as ferramentas de IA utilizam para processar, compreender e gerar conteúdo. Em textos, um token pode ser uma palavra, parte de uma palavra ou alguns caracteres; em imagens, áudios e vídeos, tokens representam fragmentos dessas informações. Assim, quanto mais as ferramentas são utilizadas, maior é o preço a pagar ao fornecedor da mesma.

Acreditando que o uso intensivo de IA vai gerar melhores resultados, muitas empresas incentivam, até mesmo sob a ameaça de demissão, seus funcionários a utilizarem essas ferramentas, o que gera um fenômeno que vem sendo chamado "tokenmaxxing", quando em empresas como Amazon, Microsoft e Meta, funcionários tem usado IA até em tarefas desnecessárias para inflar métricas internas, apenas para mostrar que usam IA intensamente.

Além disso, à medida que mais pessoas utilizam o sistema em uma empresa, mais



Just_Super_CANVA

crece o número de tokens processados e consequentemente, o custo de utilização da ferramenta.

A Microsoft não está sozinha; a revista *Fortune* disse que outras empresas também estão repensando o uso de IA, pois o consumo de tokens disparou, especialmente quando são usados Agentes de IA, ferramentas capazes de executar tarefas mais complexas e que podem gastar milhares de vezes mais tokens do que quando o software responde apenas a uma consulta.

O CEO da Nvidia, Jensen Huang, evidentemente interessado em um uso cada

vez maior de IA, o que aumenta a demanda por seus produtos, chegou a defender que profissionais consumam tokens equivalentes a metade de seus salários anuais, incentivando o "tokenmaxxing".

Ainda não está claro se as medidas da Microsoft se transformarão em uma forte tendência, mas o fato é que, se o consumo de tokens continuar crescendo sem uma queda significativa de preços, a tentativa de substituir pessoas por máquinas pode acabar se tornando um tiro no pé.

(*) Doutor em Ciências pela Universidade de São Paulo, é professor, consultor e diretor do Fórum Brasileiro de Internet das Coisas – vjntz@gmail.com.

Os riscos de usar conectividade satelital doméstica no ambiente corporativo

A popularização da internet via satélite ampliou o acesso à conectividade no Brasil – especialmente após a chegada de novos players de soluções de baixa órbita (LEO). Mas existe um ponto importante: nem toda conectividade foi desenvolvida para ambientes corporativos críticos.

O mercado já conhece esse cenário em produtos como computadores, impressores e outros hardwares e softwares. No entanto, muitas empresas ainda utilizam equipamentos domésticos para aplicações profissionais, mesmo sem estrutura para operação contínua, múltiplos usuários ou alta disponibilidade.

Na conectividade satelital, o risco é semelhante.

As soluções residenciais podem funcionar bem para uso pessoal ou aplicações não críticas. Porém, quando utilizadas para sustentar operações empresariais, os impactos de indisponibilidade se tornam significativamente maiores.

Entre os principais riscos estão a ausência de um Service Level Agreement (SLA) corporativo; limitações de suporte técnico, o que pode comprometer a disponibilidade de serviço; menor previsibilidade operacional em caso de excesso de demanda; riscos de segurança, que podem levar à violação de normas regulatórias como a LGPD; falta de redundância e contingência.

É importante lembrar que no ambiente corporativo atual a conectividade permanente deixou de ser apenas uma infraestrutura. Hoje a conectividade é parte fundamental da estrutura que garante a continuidade do negócio.



Andreia Rosa

Segundo estudo do Ponemon Institute, o custo médio de indisponibilidade pode chegar a US\$ 9 mil por minuto, dependendo da operação. Em empresas com alta dependência digital, uma interrupção pode afetar vendas, logística, atendimento, meios de pagamento e comprometer a experiência do cliente, gerando impactos reputacionais.

Suporte técnico e continuidade operacional

As soluções corporativas de conexão por meio de satélites contam com monitoramento, gestão centralizada, atendimento especializado e acordos de nível de serviço. Já soluções residenciais, normalmente, seguem um modelo massificado, sem garantias compatíveis com operações críticas.

Isso não significa que a conectividade satelital residencial não tenha valor. Pelo contrário: ela democratizou o acesso à internet e ampliou possibilidades em regiões antes desconectadas.

Mas a conectividade corporativa exige outro nível de planejamento.

Mais do que "ter internet", empresas precisam garantir disponibilidade, segurança, estabilidade e continuidade operacional. Em um cenário onde os negócios dependem cada vez mais de soluções como SaaS, Cloud, e meios de pagamento como cartões e PIX, a conectividade deixou de ser commodity. Hoje, ela é estratégia.

O ambiente corporativo demanda satélites e planos com desempenho similar ao de uma conexão 5G, capazes de alcançar velocidades de download superior a 400 Mbps e de upload de até 50 Mbps, reduzindo riscos de perda de pacotes de dados ou comprometimento do desempenho de aplicativos.

No primeiro momento, o menor custo de hardware e dos serviços pode ser atrativo, mas a realidade é outra. A economia inicial pode gerar custos maiores ou perdas financeiras decorrentes de paradas inesperadas, maior tempo de inatividade ou desempenho inferior ao necessário.

Os planos empresariais trazem mais confiabilidade, melhor custo-benefício a longo prazo, protegem a empresa de gastos imprevistos e garantem que a equipe mantenha a conectividade e a produtividade de forma previsível.

(Fonte: Andreia Rosa, Especialista de Produtos e Vendas da Deutsche Telekom Global Business Solutions Brasil).

News@TI

ICMC estará de portas abertas

@Você sabe o que faz quem estuda na USP São Carlos? Se você quer ter um gostinho do que se ensina e do que se aprende na maior Universidade pública do Brasil, não pode perder esta chance: o Instituto de Ciências Matemáticas e de Computação (ICMC) estará de portas abertas nos dias 16, 17 e 18 de junho. É uma oportunidade de conhecer os projetos desenvolvidos por universitários em áreas como inteligência artificial, programação web, robótica móvel, computação gráfica, ciência de dados, cálculo numérico, estatística, otimização e história da computação. No total, estudantes de 31 disciplinas vão participar das atividades apresentando seus trabalhos à população (<https://icmc.usp.br/e/7pjjy>).

Soluções de controle de acesso

@A Wolpac apresenta durante a Exposec 2026 sua nova geração de soluções para controle de acesso e mobilidade. A feira será realizada entre os dias 1º e 3 de junho, no São Paulo Expo. Os lançamentos WolPass, WolSlim, WolStar IV e WolVision II marcam uma nova fase da companhia, combinando tecnologia, sofisticação estética, experiência do usuário e operação contínua de alta performance. "As novas soluções da Wolpac foram desenvolvidas para responder a uma demanda crescente por ambientes mais inteligentes, seguros e integrados à arquitetura contemporânea. Estamos trazendo ao mercado equipamentos que unem desempenho operacional, comunicação visual e design premium em um novo patamar", afirma Alan Lourenço, diretor comercial.

Empresas & Negócios
José Hamilton Mancuso (1936/2017)
Responsável: Lilian Mancuso

Editorias
Economia/Mercado/Negócios/Tecnologia/Agronegócios/
Espaço empresarial: Ricardo Souza (ricardosouza@netjen.com.br);
Livros: Ralph Peter (ralphpeter@agenteliterioralph.com.br)
Comercial: comercial@netjen.com.br
Publicidade Legal: lilian@netjen.com.br

Colaboradores: Ana Luisa Winckler, Carol Olival, Claudia Lazzarotto, Denise Debiasi, Fabiana Monteiro,
Geraldo Nunes, Heródoto Barbeiro e Neiva Mendes

Laurinda Machado Lobato (1941-2021)

Webmaster/TI: Fabio Nader; Edição Eletrônica: Ricardo Souza.
Revisão: Maria Cecília Camargo; Serviço informativo: Agências Brasil,
Senado, Câmara, EBC, ANSA.

Artigos e colunas são de inteira responsabilidade de seus autores,
que não recebem remuneração direta do jornal.

ISSN 2595-8410

José Leonil Lobato (1939-2026)

Jornal Empresas & Negócios Ltda

Administração, Publicidade e Redação: Rua Joel Jorge de Melo,
468, cj. 71 – Vila Mariana – São Paulo – SP – CEP.: 04128-080
Telefone: (11) 3106-4171 – E-mail: (netjen@netjen.com.br)
Site: (www.netjen.com.br). CNPJ: 05.687.343/0001-90
JUCESP, Nire 35218211731 (6/6/2003)
Matriculado no 3º Registro Civil de Pessoa Jurídica sob nº 103.