

TECNOLOGIA

Computadores em transição para IA

Empresas apostam em unidade de processamento neural (NPU) para acelerar tarefas relacionadas à inteligência artificial. Atualmente, processamento ocorre na nuvem

» PEDRO JOSÉ*

No universo revolucionário da inteligência artificial (IA), o desafio é processar cada vez mais informações, de maneira simultânea. As maiores empresas de tecnologia estão investindo pesado nesse desafio. Na corrida pela inovação, um novo componente está ganhando destaque: a unidade de processamento neural (NPU).

As NPUs são chips especializados projetados para acelerar tarefas relacionadas à IA, especialmente em áreas como aprendizado de máquina (machine learning) e redes neurais artificiais.

Atualmente, a maior parte do processamento de inteligência artificial é realizada na nuvem, em grandes servidores de empresas especializadas. Isso ocorre porque a IA requer uma quantidade significativa de poder de computação para treinar e executar modelos que são realizados em chips chamados de GPU (unidade de processamento gráfico).

Na tentativa de combater o domínio no mercado de chips gráficos, diversas empresas, dentre elas, Microsoft, AMD e Qualcomm, estão apostando nas NPUs. Essa tecnologia tem o foco específico em calcular somente inteligência artificial, diferentemente das placas gráficas, que executam outras funções, como a projeção de imagem na tela do computador — e gastam mais energia.

A principal inovação, anunciada recentemente, é produzir notebooks ultrafinos, capazes de processar redes neurais artificiais nativamente. As NPUs ocupam menos espaço e consomem menos energia, por isso são valorizadas. No entanto, a potência para fazer esses cálculos paralelos é medida em TOPS (Tera Operações por Segundo), esse computadores ultrafinos foram anunciados com o selo da Copilot Plus, pela Microsoft, empresa responsável pelo Windows que é o Sistema Operacional com cerca de 74% dos usuários de PCs, de acordo com o site especializado Statcounter, e terão pelo menos 40 TOPS (Trilhões de Operações por Segundo), enquanto uma placa de vídeo de uso pessoal pode chegar até 1.300 TOPS.

Devido à grande potência que os núcleos especializados como os da Nvidia, que domina o mercado de GPUs em IA, grande parte de companhias que trabalham com redes neurais de supercomputadores que trabalham com demandas de inteligência artificial do mundo inteiro a partir do sistema de nuvem. Nessa situação, qualquer um que usar o sistema, envia um comando que será enviado via internet para esses grandes servidores, que vão calcular e enviar o resultado de volta, porém, há um delay que depende de fatores de internet e conexão, o processamento local

AFP



Conferência Mundial de IA em Shanghai, na China: Microsoft avança na corrida dos processadores

ENTENDENDO O COMPUTADOR POR DENTRO

CPU (Unidade Central de Processamento)
A CPU é o “cérebro” do computador. É responsável pelo funcionamento do sistema operacional a partir de poucos núcleos potentes e complexos, que realizam processos em sequência. No entanto, CPUs têm limitações para processar muitos cálculos ao mesmo tempo.

GPU (Unidade de Processamento Gráfico)
As GPUs são excelentes em processamento paralelo. Originalmente construídas para projetar imagens na tela do computador, elas agora são utilizadas para acelerar tarefas intensivas em computação e IA, como aprendizado profundo e processamento de imagens. Diferentemente das CPUs, as GPUs executam milhares de tarefas simultaneamente, reduzindo o tempo de treinamento de redes neurais. Esse foi o avanço tecnológico que permitiu à Nvidia,

uma fabricante de Placas de Vídeo (GPU), tornar-se uma das empresas mais valiosas do mundo, competindo com Apple e Microsoft.

NPU (Unidade de Processamento Neural)
As NPUs são especializadas em tarefas de inteligência artificial (IA). Elas complementam as funções das CPUs e GPUs. Enquanto as CPUs lidam com uma ampla gama de tarefas e as GPUs se destacam na renderização de gráficos detalhados, as NPUs executam tarefas de IA rapidamente. As NPUs ganharam força recentemente como uma resposta à dominância das placas da Nvidia no mercado de processamento de IA. Elas oferecem processamento local, dispensando o envio de informações para processamento na nuvem e depois o retorno ao usuário.

de processos de IA pode acabar agilizando o resultado.

Mudança digital

As NPUs têm potencial para impulsionar grandes mudanças no cotidiano digital. Elas processam uma grande quantidade de dados em paralelo, realizando trilhões de ações paralelas por segundo, explica Euclides Lourenço Chuma, professor e pesquisador do Instituto de Engenharia Eletricistas e Eletrônicos (IEEE). “Contudo, as NPUs têm uso restrito na IA, pois muitas operações não funcionam em paralelo” ressalta.

O professor também explica que NPUs terão “aplicações profissionais úteis”, como trabalho de escritório e questões mais funcionais e simples e que para esse processador

caiba em um computador ultrafino, deve vir integrado em placa-mães, assim como GPUs simples, que tem o foco somente em gerar a imagem da tela. Portanto, “são pouco customizáveis” e devem ser bem escolhidas, porque não há como fazer uma melhoria sem trocar todo o sistema.

“NPUs oferecem uma melhor relação custo-benefício-energia para aplicações de IA. Já as GPUs são mais versáteis, contudo costumam custar ter uma relação custo-benefício-energia pior que as NPUs para aplicações de IA”, conta o professor. Portanto, computadores portáteis que funcionam a base de bateria, seriam beneficiados em relação ao tempo fora da tomada.

Em uma recente declaração, a Nvidia, líder mundial em tecnologia de processamento gráfico, defendeu que suas GPUs RTX são superiores às Unidades de Processamento

Neural para o processamento de inteligência artificial.

Segundo a Nvidia, a empresa domina todo o segmento de PC IA, “já que a RTX 4070 (placa intermediária da fabricante) é suficiente para bater a Radeon (placa gráfica da concorrente AMD mais rápida)” em tarefas que envolvem processamento para inteligência artificial. No último ano, a Nvidia vendeu 3,76 milhões de GPUs para servidores, o que significa 98% de todo o mercado. Nesse mesmo período, a empresa registrou um lucro de US\$12,2 bilhões, montante é 779% maior que valor de 2022. A AMD respondeu dizendo que suas GPUs também conseguem entregar bons resultados em áreas diferentes, como qualidade de vídeo, remoção de ruídos de fundo, IA generativa e edição de imagens e vídeos.

*Estagiário sob a supervisão de Carlos Alexandre de Souza

Brasil S/A
por Antonio Machado



machado@cidadebiz.com.br

Dias de fúria

Dos ataques do presidente Luiz Inácio Lula da Silva aos juros altos, que julga ser uma maquinação do Banco Central independente em conluio com especuladores do mercado financeiro, sobram o espanto da taxa de câmbio chegando a irreais R\$ 5,70 e a certeza de que cabeça quente e impulsividade são uma combinação com consequências desastrosas.

Aparentemente convencido por economistas reunidos pelo ministro Fernando Haddad a manejar os seus ataques contínuos ao BC e permitir o anúncio de algumas promessas para desacelerar o ritmo dos gastos públicos, Lula se aquietou, e o dólar refluíu.

Atribuiu-se a insatisfação presidencial às pesquisas indicando que a opinião média do cidadão segue desfavorável ao governo, apesar de indicadores razoáveis sobre o nível de emprego e de renda, com implicações para os candidatos petistas nas eleições municipais, em outubro. Isso pode ter sido o fato detonador, não a causa.

Intuitivo, o presidente sabe que há muitas esquisitices. O custo do crédito, formado em parte pela taxa Selic definida pelo BC a partir da meta de inflação definida pelo seu próprio governo, é uma das fontes do mal-estar que carcome as entranhas econômicas e sociais, mas não explica tudo e talvez seja a menos relevante.

Importante foi a demora para entender que a orientação de seus assessores políticos a voltar a pautar o noticiário, até então sob controle da oposição e dos chefes do Legislativo, depende muito da pauta. Questões de juros e contas fiscais, ligadas umbilicalmente ao mercado de divisas, não se prestam a proselitismos políticos.

Tais temas são postos em baila quando há propostas para comunicar ou ideias a debater. Não é bem a autonomia do BC uma dessas pautas críticas para a melhoria dos instrumentos da macroeconomia até por ser dependente de orientação superior do Conselho Monetário Nacional (CMN), liderado pelo ministro da Fazenda. Quanto sabem disso? Poucos.

A omissão do CMN é um indicativo da incompreensão sobre o seu papel na governança da macroeconomia. Tanto quanto do presidente do BC falar sobre política fiscal, um tema fora de sua alçada, ou a timidez de seus pares na autarquia, todos com iguais direitos e mandato, não submetidos a regras de hierarquia durante, por exemplo, as votações do Comitê de Política Monetária (Copom)

Nem feia nem fulgurante

A realidade não é tão feia quanto os movimentos mambembes da taxa cambial nos últimos dias quiseram sugerir, supondo que os voleios dos traders do mercado financeiro a cada bronca de Lula esvançam, nem fulgurante quanto ele quer fazer crer. O problema para valer é sério, mas o juro de dois dígitos é mais consequência que causa.

O juro à vontade do carregador do estoque dos papéis de dívida do Tesouro Nacional mudará de mãos quando seu total deixar de crescer no piloto automático. Como proporção do Produto Interno Bruto (PIB), a dívida bruta hoje é de 76,8%, 4,6 pontos percentuais acima do nível de um ano atrás.

Que Lula fosse um presidente mãos de tesoura e pouco poderia ter feito para segurar esse ritmo. Contra a vontade de Jair Bolsonaro, foi o que o ministro Paulo Guedes fez por um tempo. Mas, o fez em cima da coxa, trancando aumentos salariais do funcionalismo, retendo concursos públicos e atrasando administrativamente os pedidos de aposentadoria e pensões. Não houve nada estrutural, tal como também não houve no governo Temer, já que o teto de gastos era o que diz o nome, um teto, não uma reforma do sistema público.

O estilo fiscal de Lula, portanto, não difere muito das gestões anteriores, o que inclui seus dois primeiros governos e de Dilma. O que o atrapalha é a boca, ao querer renomear categorias de gasto com saúde e educação como investimento, cujo conceito significa aumento da entrega de bens e serviços públicos e privados, e dizer que não fará ajuste do orçamento cortando despesa social. Não deve cortar mesmo, o que não significa que não haja o que aprimorar.

Copom e Tesouro em sintonia

Se dividirmos o tempo até a próxima eleição geral em 2026 em três blocos, sendo o curtíssimo prazo este ano, curto prazo ano que vem e médio prazo até a virada para 2027, as coisas ficam mais claras.

O mandato do presidente do BC, Roberto Campos Neto — que Lula pôs na categoria de desafio depois que foi votar vestindo camiseta da Seleção Brasileira, um símbolo nacional apropriado por Bolsonaro —, acaba no fim do ano. O quanto antes indicar seu substituto ao escrutínio do Senado e ele se impuser aos críticos, mais autonomia em relação ao mercado ambos terão. É o que há para fazer no curtíssimo prazo.

Mais que isso, apenas se Haddad for autorizado a contingenciar um volume gordo de despesas da lei orçamentária atual, algo da ordem de R\$ 40 bilhões. Não vai, mas vale a intenção, até para não se desperdiçar a imagem de bom moço cultivada por parte da imprensa.

Em paralelo, o CMN poderia ser ampliado para Conselho Monetário Nacional e Fiscal (CMFN) pondo o secretário do Tesouro ao lado dos ministros da Fazenda e do Planejamento e do presidente do BC. As duas políticas são complementares uma da outra e não adversárias.

Para inovar com substância, poderia também ser criado um conselho consultivo de economistas notáveis em ambas as áreas, dois a três, para não virar plenária de partido, e uma primeira missão: passar a limpo os modelos analíticos e prospectivos usados pelo Copom.

Só se queixar não resolve

Ah! Vale também a área de controle dos projetos de infraestrutura atentar para a exaustão de balanço dos candidatos a participar dos leilões agendados de concessões de rodovias, saneamento e outros.

Os últimos 30 anos de juros colossais exauriram as finanças dos investidores potenciais para algo mais do que exige o Tesouro para emitir e refinanciar seus títulos, um sugadouro da tesouraria de bancos e empresas. É duro, além disso, disputar o *funding* de quem recebe IPCA+6% praticamente sem riscos. Depois da crise da Operação Lava-Jato, as grandes da construção se foram e os bancos se fecharam.

Restaram empresas boas, sem dívida nem passivo reputacional e com bom acervo técnico, mas pequenas para o que se impõe. Há soluções para além da discussão sobre juros, cuja saída é mais estrutural e possível a médio prazo. O governo precisa aconselhar-se com quem lhe possa suprir a carência de estratégias financeiras. Com jeito e pragmatismo, muito mais pode ser feito. O que não dá é fazer da queixa um meio de vida sem nem ao menos tentar ser diferente.