

# Carta mensal

Janeiro 2026

## Inteligência Artificial e Investimentos de Capital

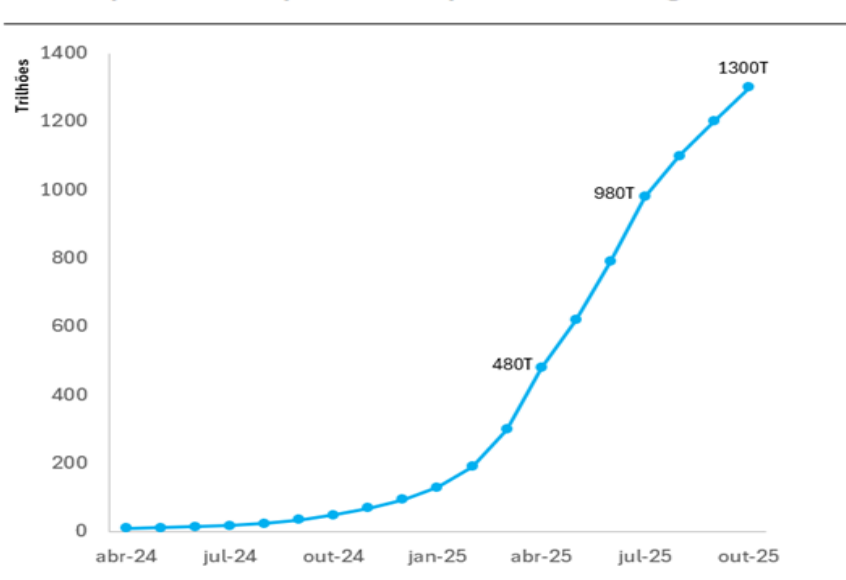
Em seu livro *The Singularity is Near*, lançado em 2005, o inventor e futurista Raymond Kurzweil apresenta a Lei dos Retornos Acelerados, argumentando que o progresso tecnológico não ocorre de forma linear, mas sim exponencial. Nesse contexto, as revoluções da humanidade, cada vez mais curtas em intervalos de anos, nos levariam à Inteligência Artificial Geral (AGI por volta de 2029 e, posteriormente, à Singularidade<sup>1</sup> em 2045.

Provado correto em muitos pontos, destaca-se sua previsão de que, por volta de 2023, teríamos acesso a poder computacional bruto equivalente ao de um cérebro humano por mil dólares. Com o lançamento público do ChatGPT no final de 2022 e de subsequentes *Large Language Models* (LLMs) de outras empresas, temos hoje essa capacidade de forma global e gratuita<sup>2</sup>.

A adoção massiva no uso dos principais LLMs disponíveis evidencia o impacto da revolução atual. O modelo da OpenAI possui atualmente 800 milhões de usuários semanais ativos, enquanto a funcionalidade de Visão Geral de IA do mecanismo de busca do Google, que utiliza seu modelo Gemini, atingiu 2 bilhões de usuários mensais em junho de 2025. Dentre todas as suas superfícies, a empresa processou 1,3 quadrilhão de *tokens*<sup>3</sup> em outubro de 2025, o que equivale a aproximadamente 975 trilhões de palavras.

Os modelos estão se tornando cada vez mais capazes, com números de parâmetros<sup>4</sup> em ordens de grandeza superiores e sendo treinados em quantidades de poder computacional exponencialmente maiores. Além disso, os processadores (*chips*) utilizados nos processos de treinamento e inferência<sup>5</sup> evoluem a cada geração, tanto do ponto de vista de performance quanto do ponto de vista energético e de custo-benefício.

**Tokens processados por mês nas plataformas Google**



Fonte: Informações públicas divulgadas pela companhia, elaboração Adam Capital. Inclui os modelos Gemini, as Visões Gerais de Pesquisa e IA e o Google Workspace.

<sup>1</sup> Momento em que a fusão entre a inteligência humana e a inteligência artificial causará uma transformação tecnológica e evolutiva tão rápida e profunda que a realidade como a conhecemos deixará de existir.

<sup>2</sup> Kurzweil previu que, por volta de 2023, \$1.000 comprariam hardware capaz de realizar cerca de 10 petaflops de cálculos por segundo (estimativa da quantidade bruta de cálculos que o cérebro humano realiza). Graças ao advento da computação em nuvem, temos amplo acesso a modelos que rodam em capacidades ordens de grandeza maiores, nos oferecendo desempenho de nível humano ou superior em diferentes tarefas e pesquisas.

<sup>3</sup> No contexto de inteligência artificial e LLMs, tokens são as pequenas unidades fundamentais de dados que os modelos utilizam para processar, compreender e gerar a linguagem humana.

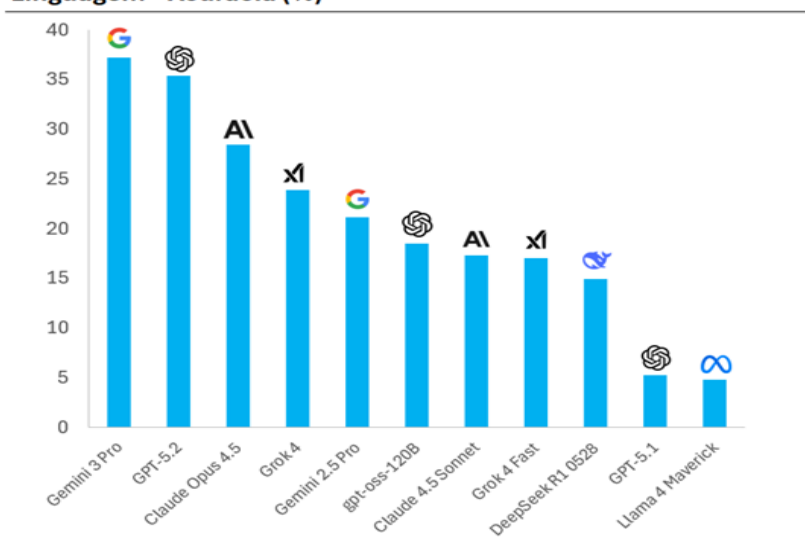
<sup>4</sup> Parâmetros são as variáveis internas do modelo, ajustadas durante o processo de treinamento. De forma simplificada, pode-se pensar neles como as sinapses ou conexões neurais da inteligência artificial.

<sup>5</sup> Inferência é a etapa de execução ou uso prático do modelo. É o momento em que a inteligência artificial recebe uma nova informação (uma pergunta ou comando do usuário) e utiliza seus parâmetros já treinados para processar o contexto e gerar a resposta.

Essa combinação nos leva a um ponto fundamental em relação ao cenário atual dos LLMs e da capacidade potencial da inteligência artificial. A utilização em larga escala dos sistemas mais recentes é viabilizada por custos de inferência decrescentes, que tendem a zero. Aliado a esse fato, a qualidade da inferência é cada vez melhor, uma vez que os modelos se provam progressivamente mais inteligentes<sup>6</sup> em uma variedade de proposições e áreas do conhecimento. Com isso, alcançamos um patamar de “inteligência como commodity”, no qual o acesso ao conhecimento é amplo e praticamente irrestrito.

Uma vez que a o custo de inferência deixa de ser uma restrição fundamental, discute-se então o cenário de capacidade de processamento (*compute*). Ao receber um comando (*prompt*), o modelo em questão se propõe a “pensar” sobre o assunto. Esse processo nada mais é do que a geração de uma cadeia de raciocínio oculta, onde o sistema utiliza processamento para decompor o problema em etapas lógicas, explorar diferentes caminhos de solução e verificar seus próprios erros antes de entregar a resposta<sup>7</sup> final. Assim, quanto mais processamento o modelo dispense para determinada tarefa, tudo mais constante, melhor é a resposta. Pode-se perceber isso facilmente ao comparar o tempo dispendido e a qualidade da informação fornecida pelos principais LLMs nas suas funções Rápido ou Raciocínio<sup>8</sup>.

**Humanity's Last Exam (HLE) - Benchmark para Modelos de Linguagem – Acurácia (%)**



Fonte: Artificial Analysis, Adam Capital.

Na fronteira da pesquisa em inteligência artificial, podemos destacar o Aprendizado Aninhado (Nested Learning), que propõe uma nova forma de compreender as redes neurais modernas. Em vez de enxergar o aprendizado como algo linear, camada por camada, revela-se que o modelo funciona como um conjunto de processos internos que operam em diferentes “níveis” e velocidades<sup>9</sup>. Essa perspectiva aponta caminhos para que sistemas futuros consigam se adaptar ao contexto em tempo real, aprendendo de forma contínua, o que representaria um avanço em relação aos parâmetros atuais.

<sup>6</sup> A inteligência de um LLM é medida através de *benchmarks*, que funcionam como testes padronizados ou para a inteligência artificial. São conjuntos de dados com milhares de perguntas cobrindo matemática, programação, raciocínio lógico, direito e conhecimentos gerais. O desempenho do modelo é pontuado com base na precisão de suas respostas em comparação com humanos especialistas e outros modelos concorrentes.

<sup>7</sup> Comprova-se isso a partir de testes controlados em que se mede a performance de um modelo em determinado *benchmark* com diferentes níveis de *compute*, como mostrado em material público divulgado pela OpenAI no lançamento do o1.

<sup>8</sup> Em sua função Auto, o modelo “julga” a dificuldade da tarefa proposta e “escolhe” o quanto irá pensar sobre, em um esforço de processamento adaptativo (*adaptive compute*).

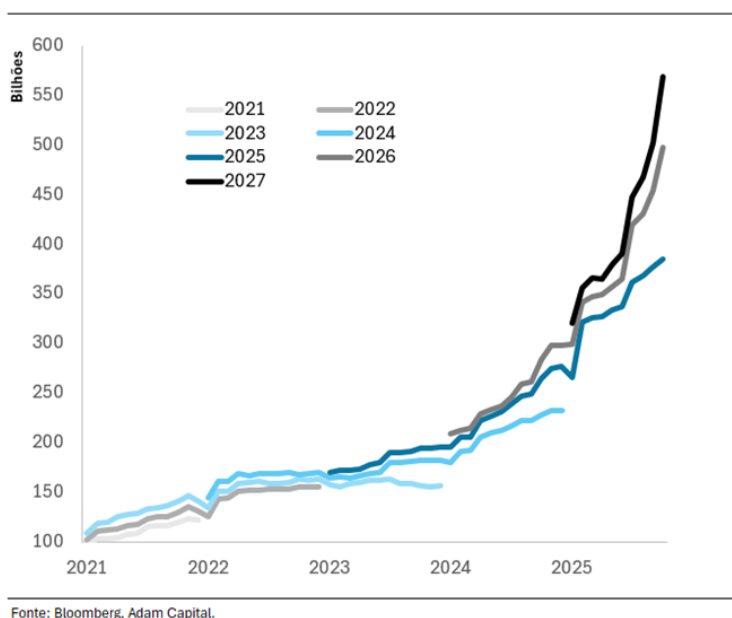
<sup>9</sup> Cada parte do sistema ajusta-se ao contexto, armazena informações temporariamente e reage a sinais de surpresa, criando um fluxo de aprendizado muito mais dinâmico do que o que se supõe nas arquiteturas tradicionais. A proposta HOPE (Hierarchically Organized Predictive Echoes), apresentada pelos autores, ilustra esse funcionamento: um modelo com loops internos, memória contínua e atualizações que acontecem em ritmos diferentes, mais parecido com a forma como o cérebro organiza informações.



Nas últimas semanas, observamos dois avanços que consideramos importantes em direção à AGI: o primeiro deles é que foi lançado o Clwdbot, um assistente de *AI* de código aberto (*open source*), que não se limita apenas a responder perguntas ou realizar tarefas pontuais sob demanda, mas pode executar tarefas complexas, interagindo com arquivos do computador, aplicativos e outros serviços, como se fosse realmente, uma “pessoa” que segue as instruções do usuário. Pouco tempo depois, foi criado o Moltbook, uma rede social de agentes de *AI*, em que eles juntamente podem discutir, criar, trocar ideias e compartilhar o que os seus usuários pediram que eles fizessem. Combinados, temos então o começo da interação de agentes criando (códigos, textos, ideias e no futuro, talvez, outros agentes) por si mesmos, sem intervenção humana. Nesse contexto de milhares de agentes interagindo em um ecossistema digital de inteligência coletiva, as possibilidades de desenvolvimento não antecipados são relevantes.

Para suportar a rápida expansão global e o uso disseminado da inteligência artificial, os investimentos em *data centers*, infraestrutura e energia estão aumentando significativamente. Entre as principais estimativas de mercado, a Nvidia enxerga de US\$ 3 a 4 trilhões em capex (*capital expenditures*) para *data centers*<sup>10</sup> até 2030, uma taxa de crescimento anual composta (*CAGR*) de 5 anos de quase 40%. Isso se reflete nos números recentes e fortes revisões altistas de projeções de capex das chamadas *hyperscalers*<sup>11</sup>, assim como na demanda por *chips* (GPUs, TPUs, entre outros) e por computação em nuvem (*cloud computing*). Vale ressaltar também o recente aumento global de preço de chips de memória, um reflexo da alta procura pelos produtos que encontra (novamente) gargalos de oferta.

### Projeção de Capex das Hyperscalers



Para financiar seus investimentos, as principais empresas do mercado<sup>12</sup> partem de balanços sólidos, com estrutura de capital majoritariamente caixa líquida<sup>13</sup> (*net cash*) e com fortes geração de caixa recorrentes. Junto a isso, enxerga-se uma relevante participação e de gigantes dos mercados de *private equity* e *venture capital*, assim como um possível reaquecimento do mercado de IPOs americano, com nomes relevantes como Open AI e SpaceX realizando sondagens para suas potenciais listagens em 2026. No entanto, tendo em vista o tamanho do investimento esperado, já estamos vendo uma necessidade de capital das principais empresas de tecnologia que se manifesta no mercado de emissão de dívidas. Estimativas do Morgan Stanley de US\$2.9 trilhões de capex global até 2028 consideram pouco mais de um terço via dívida<sup>14</sup>.

<sup>10</sup> Dentro do escopo de capex para data centers e infraestrutura de IA considera-se: nuvem pública global (*global public cloud*), nuvem privada (*private cloud*), infraestrutura local corporativa (*enterprise on-prem*), computação de borda (*edge*).

<sup>11</sup> São as principais empresas de infraestrutura de nuvem e provedoras dos chamados *cloud services*; Microsoft, Google e Amazon representam cerca de 60% desse mercado.

<sup>12</sup> Usualmente refere-se a elas como Magnificent 7.

<sup>13</sup> Posição de caixa, equivalentes e ativos de mercado é maior do que a posição de dívida financeira.

<sup>14</sup> Considera-se crédito privado e estruturas de dívidas.

Vale ressaltar que as principais estimativas de mercado não consideram o recente advento de agentes de IA assistentes<sup>15</sup>. Levando isso em conta, a utilização de *tokens* em larga escala aumenta de forma drástica, uma vez que realização de tarefas não depende mais apenas de pedidos individuais. Assim, é possível que discrepância entre demanda e oferta de *compute* continue a se acentuar, mesmo com o crescimento projetado da base instalada.

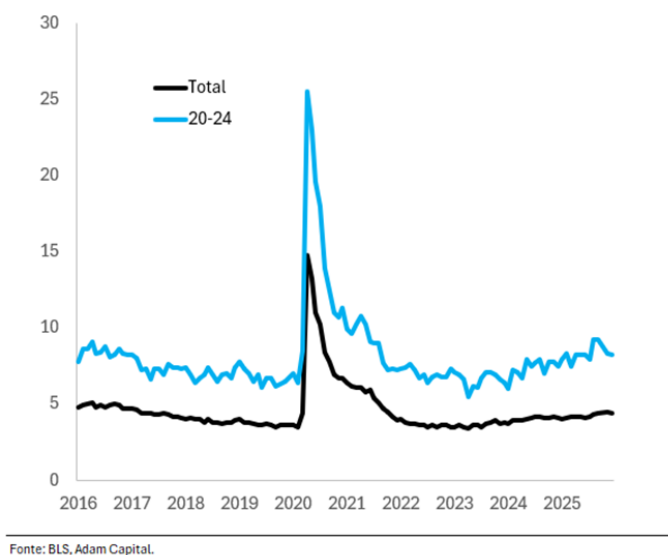
## Impactos Econômicos Globais

Do ponto de vista geográfico, vemos um claro desnível de ganhos de produtividade entre os principais países desenvolvidos, com os Estados Unidos significativamente à frente. Em um cenário de fluxo de capitais destinados à IA indo para a fronteira mais eficiente, com as empresas americanas mais bem posicionadas e oferecendo a melhor relação risco retorno, essa disparidade tende a aumentar.

Uma segunda discussão interessante em relação ao tema de investimentos de capital no setor é o seu impacto sobre a inflação. Diferentes variáveis interagem sobre essa dinâmica, como efeitos na produtividade, crescimento potencial e seu impacto na taxa de juros neutra, demanda energética, desempregos díspares entre setores, temporalidade, dentre outras. Neste caso, a tendência do dólar deveria ser de fortalecimento (por inflação ou por ganhos de produtividade). Além disso, em mundo de maiores retornos do capital, ativos com pouco ou nenhum valor intrínseco poderiam perder a relevância. Dentre os próprios ativos americanos, a explosão de IA tende a aumentar a competitividade entre eles, tornando, por exemplo, mais alto o custo de oportunidade de se deter dívida soberana. Por conta disso, podem ser necessários ajustes mais estruturais na taxa de financiamento do governo e no formato da curva de juros.

Os efeitos no mercado de trabalho já começam a ser observados. Grandes empresas, com mais facilidade e incentivo à implementação de novas tecnologias, vêm anunciando mais demissões e oferecendo menos vagas em aberto quando comparadas às pequenas e médias empresas<sup>16</sup>. Empregos mais fáceis de serem automatizados, usualmente feitos por recém-graduados, estão sendo os primeiros a sofrerem de forma relevante. Isso pode ser observado, por exemplo, na tendência recente do desemprego de jovens entre 20-24 anos.

**Taxa de desemprego nos EUA, total da força de trabalho e trabalhadores entre 20 e 24 anos**



<sup>15</sup> Aqui considera-se um agente quase senciante, proativo, capaz de entender os objetivos além das demandas solicitadas e o contexto no qual elas se inserem. Consegue, por exemplo, surpreender o seu usuário com uma tarefa adicional não solicitada, que conversa com o trabalho desempenhado.

<sup>16</sup> Levando em conta seus diferentes pontos de partida e tamanhos na economia.

Outra dinâmica decorrente é o maior impacto inicial em empregos de salários mais altos, enquanto os de salários menores, também chamados de *blue-collars*, ainda vem apresentando certa demanda. Isso ocorre porque a automação recente vem afetando de forma mais sensível os empregos corporativos e que demandam maior habilidade técnica, mas que são mais facilmente padronizáveis, como desenvolvedores de *softwares*, contadores e profissionais de áreas de controles. Em termos de efeito distributivo, o impacto para o consumo pode até ser positivo, na medida em que se expande riqueza preservando, no relativo, a massa de salários menores que possui maior propensão a consumir. Por outro lado, com a evolução dos agentes em direção à IA física e à AGI/ASI (Superinteligência Artificial), desenha-se um cenário de impacto mais amplo.

A continuidade da adoção e uso de IA deve acarretar aumentos nos fluxos destinados aos países líderes não só em termos de investimentos, mas também em termos de comércio. Isso porque, naturalmente, esses países, novamente com destaque para o EUA, também são os principais provedores dos serviços essenciais relacionados à IA<sup>17</sup>.

Neste último caso, pode ser feito um paralelo com as consequências da disseminação da internet na segunda metade dos anos 90, quando houve uma valorização global do dólar em decorrência do maior crescimento da produtividade da economia americana que também liderava o processo de inovação naquela época e que também passou a “drenar” recursos do resto do mundo, colaborando dessa forma para as várias crises cambiais que foram registradas em economias emergentes a partir do final de 1994 (México, Tailândia, Coreia do Sul, Rússia, Brasil etc).

| Crise / País  | Início | Contexto                                                      |
|---------------|--------|---------------------------------------------------------------|
| México        | dez/94 | Desvalorização repentina do peso; resgate do FMI/EUA em 1995. |
| Tailândia     | jul/97 | O gatilho da Crise Asiática (ataques especulativos ao Baht).  |
| Coreia do Sul | nov/97 | Contágio da crise asiática atingindo tigres mais robustos.    |
| Rússia        | ago/98 | Moratória da dívida e colapso do rublo.                       |
| Brasil        | jan/99 | Abandono da banda cambial e adoção do câmbio flutuante.       |

### Cenário Brasil

A fragilidade estrutural da economia doméstica torna-se evidente ao analisarmos o descompasso entre a capacidade de financiamento interno e a necessidade de absorção de recursos pelo setor público. Em comparação aos seus pares latino-americanos, o Brasil apresenta uma das piores taxas de poupança bruta em relação ao PIB, patamar que se mostra insuficiente para sustentar ciclos de investimentos sem gerar pressões sobre o balanço de pagamentos. Esse quadro de baixa propensão a poupar é agravado pela trajetória do déficit nominal, que, após os choques exógenos recentes, volta a divergir negativamente, consolidando um cenário de desancoragem fiscal que eleva o custo de capital e limita o potencial de crescimento de longo prazo.

Adicionalmente, a posição de defasagem relativa do Brasil na fronteira tecnológica impõe uma pressão renovada e crescente sobre as contas externas. Observa-se uma deterioração estrutural no balanço de serviços, impulsionada especificamente pela necessidade de importação de soluções ligadas à inteligência artificial e tecnologia da informação. Os déficits nas contas de propriedade intelectual e de serviços de computação e telecomunicações apresentam uma trajetória de aprofundamento agudo, descolando-se dos padrões históricos. Esse movimento evidencia que a economia doméstica, ao atuar como tomadora de tecnologia (notadamente dos EUA), transfere renda para o exterior de forma acelerada para suprir sua demanda, consolidando um componente negativo persistente no saldo em transações correntes e que tende a se agravar nos próximos anos com o *boom* tecnológico proporcionado pela inteligência artificial.

<sup>17</sup> Soluções e modelos de IA e infraestrutura e serviços de nuvem são alguns exemplos.

Tal dinâmica se traduz em um dos maiores déficits em conta corrente do mundo, situando o Brasil em uma posição de fragilidade comparativa relevante, com um saldo negativo entre 3% e 4% do PIB. A magnitude desse déficit impõe uma elevada dependência de fluxos externos, o que por si só já seria um fator de risco, mas que se torna crítico ao analisarmos a composição do seu financiamento.

Ademais, a resiliência da atividade econômica nos leva a questionar a premissa de que a política monetária atual esteja em terreno excessivamente restritivo. Ao contrário do que sugerem as narrativas de asfixia pelo custo do capital, os dados mostram uma economia rodando com dinamismo, evidenciada por um mercado de trabalho em níveis historicamente apertados (com a taxa de desemprego comprimida a 5,5% com ajuste sazonal) e um crescimento do PIB que se sustenta consistentemente acima do potencial desde o período pós-pandemia. Tal combinação de pleno emprego e expansão robusta da atividade sinaliza que o atual patamar de juros exerce uma contenção menor do que o consenso acredita, indicando que a taxa neutra real de equilíbrio pode ter se deslocado para cima, tornando a política monetária vigente menos contracionista na prática do que em modelos teóricos.

A análise qualitativa da inflação reforça o diagnóstico de que a demanda doméstica segue descolada da capacidade de oferta, revelando uma composição de preços preocupante sob a ótica da política monetária. Enquanto o índice cheio do IPCA tem sido beneficiado por choques de oferta positivos (notadamente em alimentação no domicílio, que roda perto de zero e bens industriais comportados), a inflação de serviços, componente mais sensível ao ciclo de renda e ao mercado de trabalho, apresenta uma trajetória divergente de aceleração, tendo fechado o ano de 2025 em 6% (muito acima dos 4,7% de 2024). Essa dicotomia evidencia que a aparente desinflação é mascarada por itens voláteis, enquanto o núcleo de preços sensíveis à demanda permanece pressionado e inercial, sinalizando que os juros correntes ainda não foram capazes de arrefecer os componentes mais rígidos da cesta de consumo.

A sinalização de relaxamento monetário pelo Copom em seu último comunicado parece ignorar a deterioração dos fundamentos internos. Há uma aparente confusão gerada por uma melhora conjuntural da inflação cheia, que ao nosso ver foi impulsionada sobretudo por choques positivos de oferta e pela onda global de dólar fraco, onde se negligencia o fato de que a inflação de serviços permanece resiliente e o hiato do produto está operando em terreno positivo (desemprego muito baixo dos fatores de produção, o que tende a aumentar as pressões futuras sobre a inflação de demanda)

Além disso, o Banco Central aparentemente subestima o potencial choque de demanda vindo do impulso fiscal e parafiscal programado para o primeiro semestre de 2026. Estimamos que esse pacote, focado em transferências de renda e expansão de crédito, possa elevar o PIB em cerca de 1 p.p., o que geraria um impacto inflacionário adicional de até 0,76 p.p. na inflação de serviços ao longo de 2026 e 2027. Ignorar esse "tsunami" de liquidez eleitoral na definição do *timing* da queda dos juros é um erro de leitura que pode comprometer a convergência da inflação à meta no horizonte relevante.

Assim, entendemos que o Banco Central tem sistematicamente subestimado as pressões de demanda. Insistir em uma queda de juros baseada em uma conjuntura volátil e muito sensível a preços de ativos financeiros, sem considerar de forma apropriada o histórico de subestimação do hiato e o teto fiscal fragilizado, pode forçar o comitê a uma reversão brusca e custosa do ciclo no futuro próximo.

Em linha com o exposto acima, seguimos comprados em bolsa americana, que é o principal ativo do nosso portfólio. Considerando que no médio prazo o fluxo de capitais deve seguir para o EUA, continuamos comprados em dólar contra o real e o euro. Por fim, estamos tomados na parte longa da curva de juros local, uma vez que entendemos que a economia brasileira segue com o cenário de crescimento de demanda robusta e acima do potencial.

## Fontes e Referências Bibliográficas:

Grossman, Lev. 2045: The Year Man Becomes Immortal. TIME Magazine, 2011.

Kurzweil, Raymond. The Singularity is Near. 2005

Kurzweil, Raymond. The Singularity is Nearer: When We Merge with AI. 2024

Maslej, Nestor; Fattorini, Loredana; Perrault, Raymond; Gil, Yolanda; Parli, Vanessa; Kariuki, Njenga; Capstick, Emily; Reuel, Anka; Brynjolfsson, Erik; Etchemendy, John; Ligett, Katrina; Lyons, Terah; Manyika, James; Niebles, Juan Carlos; Shoham, Yoav; Wald, Russell; Walsh, Tobi; Hamrah, Armin; Santarlasci, Lapo; Lotufo, Julia Betts; Rome, Alexandra; Shi, Andrew; Oak, Sukrut. The AI Index 2025 Annual Report. Stanford: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, 2025.

Wei, Jason. Jason Wei on 3 Key Ideas in AI in 2025. Stanford AI Club, Stanford University, 2025.

Behrouz, Ali; Razaviyayn, Meisam; Zhong, Peilin; Mirrokni, Vahab. Nested Learning: The Illusion of Deep Learning Architecture. Google Research, 2025.

OpenAI. Learning to Reason with LLMs. 2024.

Epoch AI

Our World in Data

OpenAI, Nvidia, Alphabet, Microsoft, Apple, Amazon, Meta, Oracle Public Disclosures

Bloomberg

Bank of America Research

Morgan Stanley Research





CONHEÇA OS  
FUNDOS DA ADAM|CAPITAL

**ADAM  
ADVANCED**

[saiba mais](#)

**ADAM  
MACRO**

[saiba mais](#)

**ADAM  
PREV**

[saiba mais](#)

**ADAM  
RENTA FIXA  
ATIVA**

[saiba mais](#)

**ADAM  
RENTA FIXA  
ATIVA PREV**

[saiba mais](#)

## Contato

[comercial@adamcapital.com.br](mailto:comercial@adamcapital.com.br)



### **ADAM Capital**

Rua Dias Ferreira, 190 – sala 502 – Leblon  
Rio de Janeiro, Brazil – CEP 22431-050  
+55 21 3993-7000

[www.adamcapital.com.br](http://www.adamcapital.com.br)