

Carta Mensal

Fevereiro 2025

“A inovação começa quando você diz 'não sei'.”

- Jim Collins, no livro "Good to Great" (2001)

DeepSeek vs. ChatGPT:

Embaralhou o jogo?

MERCADOS GLOBAIS:

A revolução da Inteligência Artificial (IA) ganhou um novo protagonista: o DeepSeek. Em um cenário dominado por empresas como OpenAI, Google e Anthropic, a startup chinesa surpreendeu o mundo ao lançar os modelos DeepSeek-V3 e DeepSeek-R1, que oferecem desempenho competitivo a um custo (em tese) até 20 vezes menor. Mas o que torna essa tecnologia tão inovadora? Será que ela ameaça a supremacia de modelos como o ChatGPT? E, mais importante, como o DeepSeek impactará o mercado de chips e o consumo de energia elétrica?

O DeepSeek surgiu dentro da High-Flyer, um hedge fund chinês, como um laboratório de pesquisa em IA. Em 2023, a empresa se tornou independente e, desde então, tem investido em modelos de linguagem de grande escala (LLMs), buscando alternativas mais acessíveis e eficientes para treinar e operar IA.

A grande novidade do DeepSeek-V3 e DeepSeek-R1 é sua eficiência extrema. Diferente de modelos ocidentais, que dependem de chips de última geração da NVIDIA, a DeepSeek conseguiu criar modelos avançados usando chips menos potentes (A800 e H800) e algoritmos otimizados para reduzir custos e consumo de energia.

Essa abordagem coloca a China mais próxima da liderança na corrida da IA, mesmo sob fortes restrições de exportação de hardware impostas pelos EUA. Ou pelo menos o governo chinês deseja que o ocidente pense.

O DeepSeek representa uma abordagem inovadora à IA, mas é importante entender onde ele se diferencia do ChatGPT e de outros modelos ocidentais.

CUSTO

O DeepSeek-V3 foi treinado com um orçamento de US\$ 5,6 milhões, enquanto o ChatGPT-4 custou mais de US\$ 100 milhões. Essa diferença se deve a três fatores principais:

- ✓ Uso de chips mais acessíveis (A800/H800) e otimização para esse hardware.
- ✓ Redução na complexidade dos cálculos através de técnicas como FP8 Mixed Precision Training.
- ✓ Menos dados necessários para atingir um bom desempenho, graças ao foco em tarefas específicas.

Impacto: Com custos tão reduzidos, IA pode se tornar acessível para mais empresas e startups, o que naturalmente provocaria um aumento de demanda pelos aplicativos de IA. **Cabe aqui ressaltar que os custos reais da DeepSeek foram maiores, eles mesmo no seu *paper* alertam que excluíram vários custos preliminares que não foram pequenos.**

ARQUITETURA COMPUTACIONAL E EFICIÊNCIA

O DeepSeek utiliza várias técnicas para melhorar a eficiência computacional. Multi-Token Prediction (MTP): Enquanto o ChatGPT prevê um token por vez, o DeepSeek antecipa múltiplos tokens simultaneamente, reduzindo a necessidade de processamento. FP8 Mixed Precision Training: **Em vez de cálculos em 16 ou 32 bits, usa precisão de 8 bits, diminuindo o consumo de energia e memória.**

Impacto: **Modelos mais eficientes exigem menos energia e podem operar em infraestruturas menores.**

ESPECIALIZAÇÃO VS. GENERALIZAÇÃO

O DeepSeek se destaca em tarefas matemáticas e lógicas, superando até mesmo o ChatGPT em alguns *benchmarks*. No entanto, ele tem dificuldades em tarefas abertas e criativas, conforme demonstrado em testes como o AidanBench.

Impacto: **Se o objetivo é precisão e eficiência, o DeepSeek pode ser uma excelente escolha. Mas para interações mais complexas e criativas, o ChatGPT ainda leva vantagem.**

OPEN SOURCE VS. MODELO FECHADO

O DeepSeek-V3 é open-source, enquanto o ChatGPT-4 é um modelo fechado. Isso significa que qualquer desenvolvedor pode adaptar o DeepSeek para suas necessidades, o que pode acelerar a inovação e diversificação de aplicações.

Impacto: Empresas podem criar soluções personalizadas sem depender de OpenAI ou Google.

IMPACTO NA DEMANDA POR CHIPS E ENERGIA ELÉTRICA

Um dos maiores debates no mundo da IA é sobre o consumo energético e a demanda por chips avançados. Como o DeepSeek impacta esse cenário?

Redução da dependência de chips avançados. Atualmente, os chips de alto desempenho da NVIDIA dominam o mercado de IA. No entanto, modelos como o DeepSeek demonstram que é possível criar IAs poderosas sem depender dessas GPUs premium.

Consequências:

- ✓ Demanda por chips de inferência (ASICs) pode crescer – chips especializados em eficiência podem substituir GPUs de propósito geral.
- ✓ **Empresas como Broadcom e Huawei podem ganhar mais espaço, desafiando o domínio da NVIDIA.**

MENOR CONSUMO DE ENERGIA

A IA moderna demanda grandes quantidades de eletricidade. **Por exemplo, um data center que opera ChatGPT-4 pode consumir o equivalente a uma cidade pequena.** O DeepSeek, ao otimizar algoritmos e hardware, reduz (em tese) esse consumo, tornando a IA mais sustentável.

Consequências:

- ✓ Empresas podem rodar modelos de IA sem precisar de infraestruturas gigantescas. Em tese.
- ✓ Menos pressão sobre redes elétricas, reduzindo custos e impactos ambientais. Em tese.

ASPECTOS POSITIVOS

- ✓ Custo aparentemente extremamente reduzido: IA se torna acessível a mais empresas.
- ✓ Menos dependência da NVIDIA: Expande o mercado de hardware.
- ✓ Baixo consumo energético: Mais sustentável do que outros modelos.
- ✓ Código aberto: Qualquer desenvolvedor pode adaptar e inovar.
- ✓ Eficiência em tarefas lógicas e matemáticas: Supera concorrentes nesses domínios.

ASPECTOS NEGATIVOS

- ✗ Menor flexibilidade: Desempenho inferior em tarefas criativas e abertas.
- ✗ Escalabilidade limitada: Pode não ser ideal para aplicações complexas em larga escala.
- ✗ Problemas de segurança: Testes mostraram falhas na proteção contra respostas inadequadas.
- ✗ Dificuldade em manter coerência em interações longas: Perde contexto mais rápido que o ChatGPT.
- ✗ Pode não ameaçar imediatamente líderes do setor: Ainda há desafios para competir de igual para igual com OpenAI e Google.

CONCLUSÃO

O DeepSeek não é apenas um novo modelo de IA, mas um marco na evolução da inteligência artificial. **Ele prova que não é necessário hardware ultra avançado para criar modelos poderosos, o que pode levar a uma redefinição do mercado de IA.**

Embora ainda não supere o ChatGPT em termos de criatividade e flexibilidade, o DeepSeek abre portas para um futuro em que IA será mais **acessível**, eficiente e sustentável. Seu impacto pode ser profundo, reduzindo a dependência dos chips da NVIDIA, diminuindo (em tese) o consumo de energia e democratizando o acesso à IA.

Um ponto interessante é que modelos de “*small language*” como DeepSeek podem rodar localmente em dispositivos como celulares, sem ter que acessar a nuvem para obter mais

poder de processamento. Outro ponto é que DeepSeek aparentemente obteve um menor custo de treinamento, **porém alguns especialistas dizem que ele consome mais poder computacional no uso diários dos usuários, ao contrário do ChatGPT, que demanda muito esforço no treinamento e relativamente menos no uso. Ou seja, talvez a conclusão seja o oposto do que o mercado supôs essa semana, modelos menores e mais baratos aumentam a demanda, atraindo bilhões de usuários, que, no entanto, no uso diário podem acabar demandando mais esforço computacional e mais energia.**

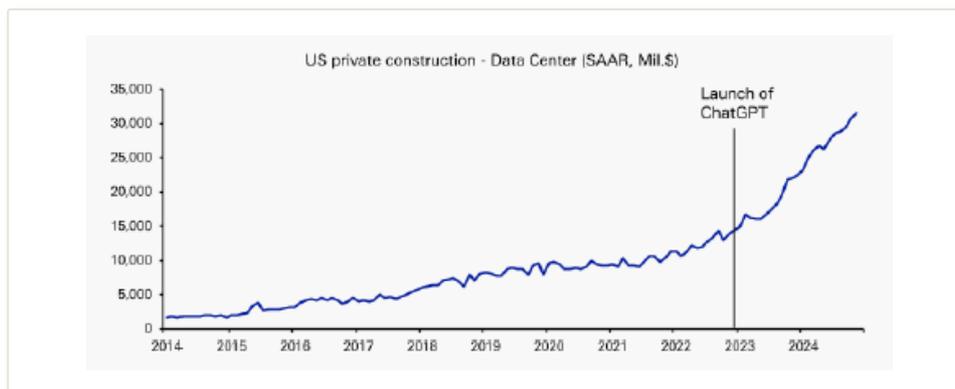
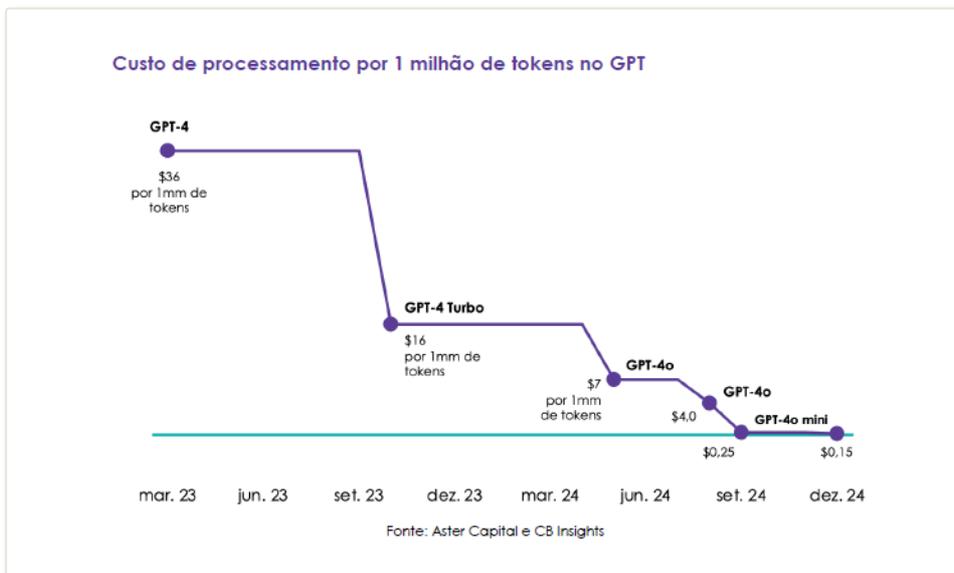
Houve um total exagero na mídia ao falar em fim do excepcionalismo americano e que o nascimento do DeepSeek iria afetar inclusive a economia americana. Como toda tecnologia, temos queda acentuada no seu custo, o que traz a adoção pela grande massa, o que aumenta a demanda. Neste caso, o aumento de demanda por energia e chips irá continuar.

O gráfico do aumento dos gastos de construção privada nos US demonstram exatamente isso, a turma acelerou o investimento em data centers.

Um outro ponto que é super relevante é o potencial aumento de produtividade da economia americana e global, hoje os US crescem sua produtividade na taxa de 2,6% aa. **Com adoção em massa de IA por empresas e pessoas, isso poderá crescer para acima dos 3%, que foi o auge no boom da adoção da internet, nos anos 90. Das 500 empresas do S&P500, apenas 5% têm aplicações rodando, o resto está se desenvolvendo nisso ainda.**

Dois pequenos exemplos de ganhos de produtividade, em recente conversa com executivos do JP Morgan, eles comentaram que o banco reduziu em 40% as pessoas no setor de cadastro, após adoção de IA. O banco Morgan Stanley comentou que a papelada exigida para o banco lançar um título de renda fixa nos US demorava 2 semanas, hoje 95% do trabalho é feito em 1 hora.

Em resumo, estamos ainda na infância de uma tecnologia onde as aplicações que serão inventadas nós nem ao menos imaginamos, assim como nos anos 90 nem imaginávamos um aplicativo como o Uber funcionando em nossos celulares. É claro, todo processo de disrupção tecnológica tem poucos grandes vencedores e um cemitério lotado de gente que ficou pelo caminho. Hoje é impossível saber quem serão os vencedores finais. Por isso, ao explorar o tema em nossas carteiras, preferimos a base da infraestrutura necessária, como cobre e urânio, necessários para o processo de eletrificação.



MERCADO LOCAL:

Sim, ao mesmo tempo que o mundo enfrenta uma era superinteressante de disrupção tecnológica, seguimos com antigos problemas, como o nosso já muito conhecido desarranjo fiscal. Vamos a ele a seguir:

Os mercados tiveram uma descompressão em janeiro, com o real se valorizando 5,7%, a bolsa subindo 4,90% e um fechamento de taxa médio entre 80 e 100 bps nos juros nominais. E aí? O que mudou?

Tivemos um mix de alguns fatores, primeiramente o mercado devolvendo parte do movimento de dezembro, muito exagerado e beirando o pânico em alguns dias. Estamos em recesso parlamentar, o que ajuda na ausência de ruídos políticos. Não tivemos também

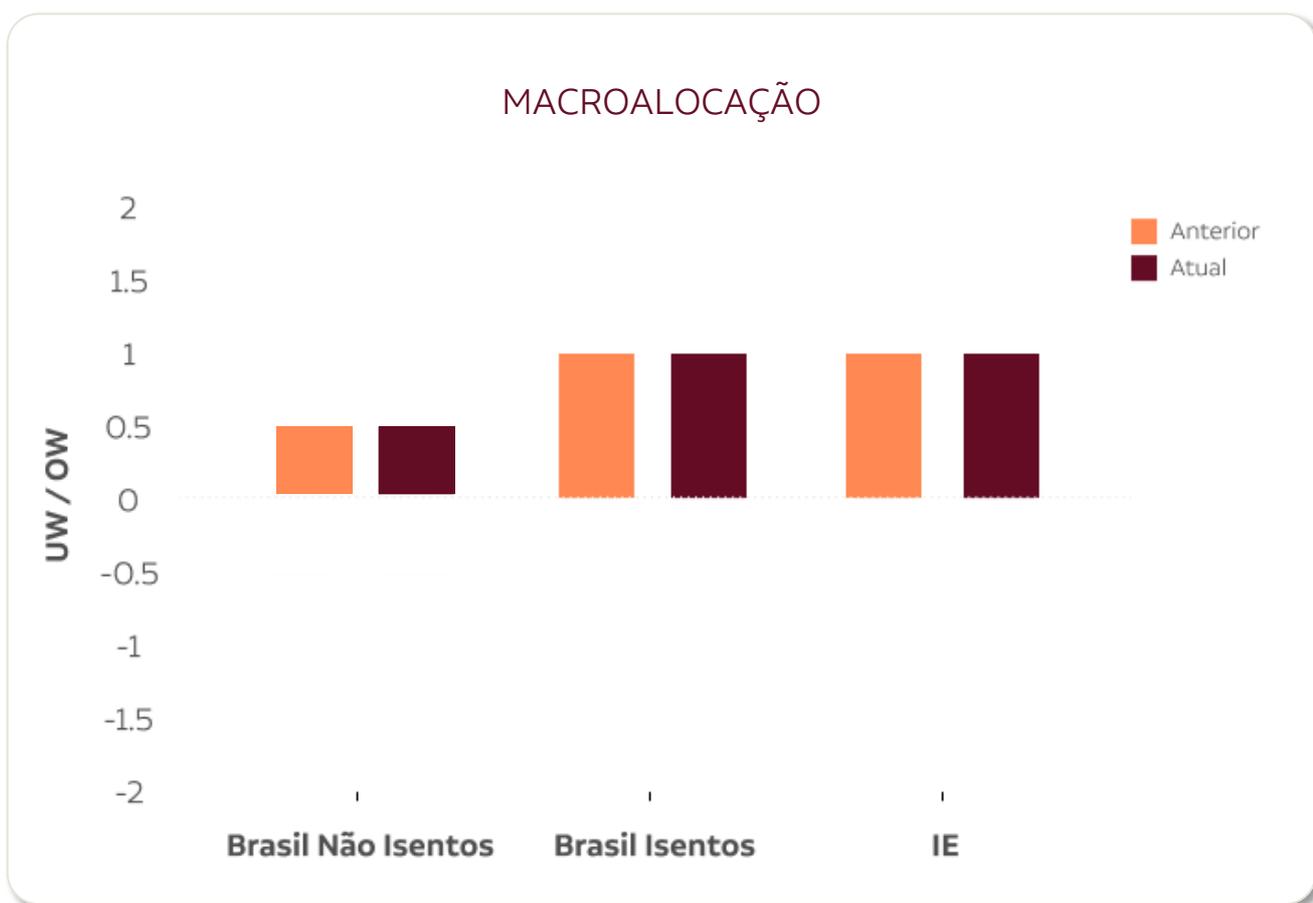
declarações polêmicas sobre economia da ala política do governo. Ainda sobre política, o evento da queda de popularidade de Lula nas pesquisas também contribuiu, com alguns agentes antecipando (prematuramente) a corrida eleitoral de 2026.

Dois outros fatores macro merecem também a nossa análise. **Dados de atividade de novembro/dezembro sugerem que a atividade econômica pode ter feito topo e começa o seu caminho ladeira abaixo**, afinal com 10% de juro real é complicado para qualquer economia permanecer em ritmo de festa. No fim do mês tivemos reunião do COPOM que seguiu o roteiro deixado por Roberto Campo e subiu a taxa Selic em 1 ponto percentual, para 13,25%. O problema foi na comunicação, bem suave e dando a entender que após a próxima elevação de mesma magnitude, o ciclo poderia estar bem perto do fim. **Em seu jogo de estreia na Série A, Gabriel Galípolo preferiu entregar um cartão de visitas de um líder de torcida e não de um zagueiro mal-encarado, defendendo as metas inflacionárias.** Quando rodamos modelos de inflação, temos um juros de 16% ou acima, o que é necessário para colocar a inflação na meta (3%) no horizonte relevante (meados de 26). **Ao mostrar que talvez nem passe de 14,50%, o BC mostra que talvez não se importe com uma inflação no teto da meta (4,5%) ou na melhor das hipóteses que deseja a convergência em um horizonte de tempo maior. Definitivamente não é um quadro que melhora o cenário.**

Ao mesmo tempo, em breve termina o recesso parlamentar e temos de observar a movimentação dos atores políticos em relação a um governo que se enfraquece e a reação do próprio governo a uma economia que terá mais inflação e menos crescimento em 2025. A grande questão aqui será: **Governo corrige o rumo ou dobra aposta em mais gastos?**

Em resumo, o movimento de janeiro remete mais a um cenário mais tranquilo externo (por ora, Trump moderado), um ajuste técnico de um mercado muito leve em termos de risco e prêmios de risco exagerados. **A crise de credibilidade fiscal, que virou uma crise de credibilidade política ainda tem uma longa travessia até o ciclo eleitoral. Reconhecemos que neste nível de preços, os ativos brasileiros têm uma assimetria positiva, porém a travessia de 2025 será em mar encrespado, não podemos nos iludir.**

Termômetro de Alocação



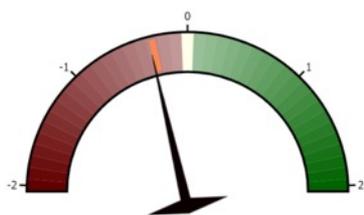
O radar "Macroalocação" da TAG é o primeiro passo da equipe de Gestão no desenvolvimento das Carteiras Modelo TAG. Esse painel busca refletir a nossa convicção em 3 "áreas" de investimento: ativos brasileiros isentos da tributação de Imposto de Renda (ex. debêntures incentivadas, CRAs, CRIs, LCIs, e LCAs, entre outros), ativos brasileiros não-isentos, e ativos de Investimento no Exterior.

ALOCAÇÃO ESTRATÉGICA

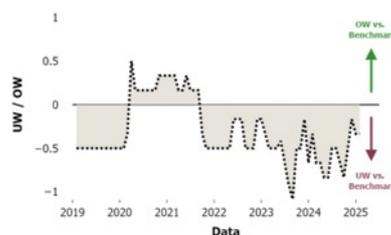
INCLINAÇÃO

HITÓRICO DE ALOCAÇÃO DE RISCO

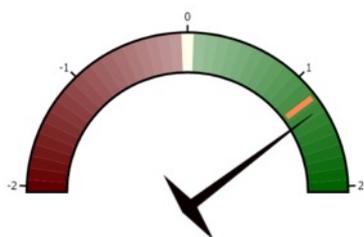
ACÇÕES



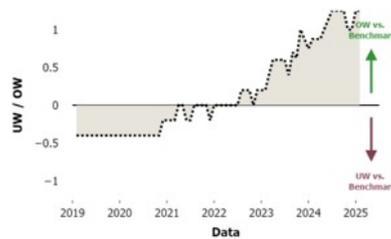
	--	-	N	+	++
Brasil			■		
DMs			■		
EMs (ex-Brasil)			■		



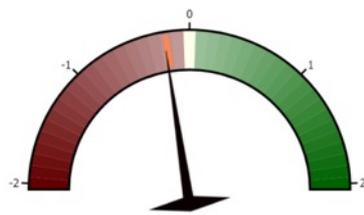
RENDA FIXA



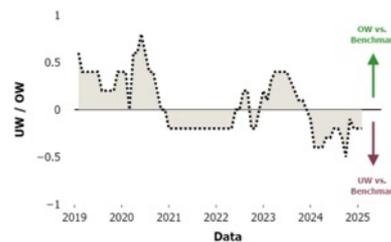
	--	-	N	+	++
Pós					■
Pré					■
IPCA					■
Global Soberanos			■		



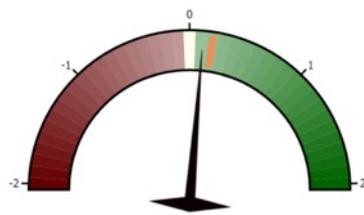
CRÉDITO



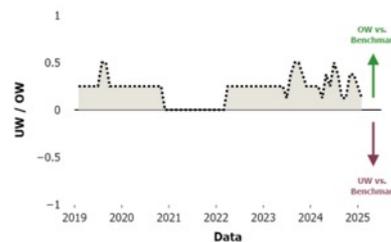
	--	-	N	+	++
Brasil HG		■			
Brasil HY			■		
Global IG				■	
Global HY					■
Global EMs (ex-Brasil)		■			



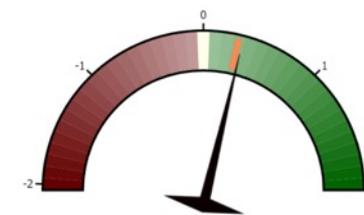
FX



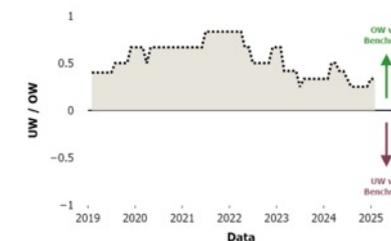
	--	-	N	+	++
USD/DXY				■	
EUR/USD		■			
BRL/USD			■	■	
Ouro/USD					■



ALTERNATIVOS



	--	-	N	+	++
Brasil MM			■		
Brasil Imobiliário				■	
PE / VC					■
Special Sits					■
HFIs					■
Global Imobiliário			■		



Glossário

Termômetro de Alocação

Inclinação

Cada painel da série "Inclinação" reflete as perspectivas TAG sobre uma certa classe de ativos e suas componentes. Em nossa construção de cenário, dividimos o universo em 5 classes: Ações, Juros, Crédito, FX (ie. Moedas/Câmbio e Ouro), e Alternativos (ie. Ilíquidos, Imobiliário, e outros).

Dentro de cada classe, vamos além, e segmentamos o universo TAG em um segundo nível.

- Para Ações, temos 5 países/regiões: Brasil, Estados Unidos, Europa, China, e Países Emergentes (ie. ex-China e ex-Brasil).
- Para "Ações", onde antes tínhamos 5 países/regiões discriminados - "Brasil", "Estados Unidos", "Europa", "China", e "Países Emergentes" (ie. ex-China e ex-Brasil) - hoje, temos "Brasil", "DMs" (ie. Developed Markets, que contempla as nossas perspectivas para os EUA e Europa, entre outros países desenvolvidos), e "EMs (ex-Brasil)" (ie. Emerging Markets, que reflete a nossa visão sobre países em desenvolvimento como China e outros, exceto Brasil).
- Em "Renda Fixa" - a popular classe de Juros - nos guiamos por ativos "Pós" e "Pré" (ie. ativos Brasil pós-fixados e pré-fixados), além de títulos atrelados ao IPCA, e "Global Soberanos" (ie. renda fixa e títulos de dívida soberana ex-Brasil).
- Quanto à parcela de Crédito, dividimos da seguinte forma: ativos brasileiros High Grade e High Yield, ativos globais Investment Grade e High Yield (ie. onde o foco são os papéis de regiões desenvolvidas, os DMs), e os "Global EMs (ex-Brasil)" (ie. ativos de Crédito de países emergentes).
- Em FX, pontuamos as nossas perspectivas para a apreciação/depreciação de alguns pares/moedas. Onde antes discriminamos essas perspectivas como Dólar, Euro, Ouro, e Real, hoje resolvemos pontuar os pares que olhamos diretamente de forma explícita: "USD/DXY" (ie. a DXY é uma cesta de moedas globais, na qual balizamos as nossas perspectivas para o Dólar), "USD/EUR", "BRL/USD", "Ouro/USD".



TAG | 20

INVESTIMENTOS ANOS

Este material não deve ser considerado como material de venda ou divulgação, e pode ser usado para simular resultados futuros com base em informações passadas, sem qualquer garantia de que os resultados simulados serão obtidos ao longo do tempo.