PERGUNTAS E RESPOSTAS
Volume 3 — Fevereiro 2026

AI FOR AMERICANS FIRST

Compreendendo o Ratio CACI de 7–12:1

Por que a Europa não pode "simplesmente pagar mais"
e por que usar a nuvem americana não é a solução

Fabrice Pizzi
Université Paris Sorbonne — Inteligência Econômica & Cibersegurança

Sobre este documento

Este terceiro volume responde à **pergunta mais frequente** sobre o estudo: "Por que a Europa tem 7 a 12 vezes menos compute efetivo que os EUA, sendo que poderia simplesmente investir mais — ou usar a nuvem americana?" Este documento decompõe a resposta em **10 perguntas progressivas**, do mais intuitivo ao mais estratégico. Pode ser lido independentemente dos Volumes 1 e 2.

Repositório: https://mo0ogly.github.io/America-First-IA/

PARTE A — Compreendendo o ratio: por que 7–12:1 e não 1,5:1?

Q1. O ratio CACI EUA/UE é de 7–12:1. Mas os EUA são mais ricos que a Europa. Isso não é normal? "Você paga, você recebe"? *[Público geral / Decisor]*

Esta é a objeção mais natural — e a mais disseminada. Mas ela se baseia em uma confusão entre compute bruto e o CACI.

O CACI é normalizado pelo PIB. Essa é toda a sutileza do índice. A fórmula é: CACI(r) = [F(r) × E(r)$^{-1}$] / [PIB(r) × L(r)]. O ratio de 7–12:1 não significa "os EUA têm 7–12× mais compute em termos absolutos". Esse é o ratio bruto, que é de ~15:1 (Epoch AI). O CACI diz algo muito mais grave:

> **"Com tamanho econômico comparável e capital humano comparável, os atores americanos dispõem de 7 a 12 vezes mais compute efetivo que os europeus."**

Vamos fazer a conta simples. O PIB dos EUA é de ~US$ 28 trilhões, o da UE de ~US$ 18 trilhões. O PIB americano é ~1,5× maior. Se fosse simplesmente uma questão de investimento proporcional ao PIB ("você paga, você recebe"), o ratio de compute deveria ser de 1,5:1.

Mas ele é de 7–12:1. A pergunta é: de onde vêm os 5–10× restantes? Essa é a contribuição central do estudo: esses 5–10× vêm de fatores estruturais que o investimento sozinho não consegue compensar.

Q2. Certo. Então de onde vêm esses 5–10× adicionais? *[Público geral / Decisor]*

Três fatores estruturais explicam a diferença entre o ratio "esperado" (1,5:1) e o "observado" (7–12:1). Nenhum dos três se resolve "gastando mais".

Fator 1: custo da energia (×2–3 contra a UE)

Um FLOP de computação de IA na Europa custa 2 a 3 vezes mais que nos Estados Unidos. A eletricidade industrial custa US$ 110–145/MWh na Europa contra US$ 50–65/MWh nos EUA. Concretamente: mesmo que a Europa investisse a mesma quantia, obteria 2 a 3 vezes menos compute útil. Um data center de 500 MW na Europa custa ~60% mais em eletricidade que um idêntico no Texas ou na Virgínia.

Isso não é uma escolha de investimento — é uma restrição estrutural ligada ao mix energético, à tributação de carbono e à regulação dos mercados de eletricidade na Europa. É exatamente por isso que a recomendação nº 1 são as Special Compute Zones com energia a US$ 50–60/MWh via PPAs nucleares.

Fator 2: controles de exportação + Seção 232 (prioridade de entrega)

Mesmo com orçamento, você não pode comprar o que não é entregue. A Nvidia produz um número finito de GPUs H100/H200/B200 por trimestre. As isenções domésticas americanas significam que os hyperscalers dos EUA (Microsoft, Google, Amazon, Meta,

xAI) são atendidos primeiro. Em período de escassez estrutural — o que é o caso desde 2023 — os pedidos europeus ficam para depois, com atrasos de 3 a 6 meses.

Além disso, a Seção 232 (janeiro 2026) impõe tarifas de 25% sobre semicondutores de IA avançados importados. **Mesmo que as empresas europeias comprando para uso doméstico não paguem diretamente essa tarifa, o efeito de anúncio cria incerteza regulatória**: a Proclamação prevê explicitamente uma ampliação possível até julho de 2026. Qual investidor vai comprometer €2 bilhões em um data center europeu se as GPUs podem custar 25% a mais em 6 meses?

Fator 3: efeito de gravidade do capital (fuga de investimentos)

Este é o fator mais pernicioso. Os investidores racionais — incluindo europeus e aliados — investem nos EUA em vez da UE, porque o retorno sobre o investimento é melhor: energia mais barata, compute mais denso, mercado maior, ecossistema de talentos mais concentrado.

O Japão investiu US$ 550 bilhões em solo americano — dinheiro que poderia ter construído compute japonês (ou europeu por meio de parcerias). Os Emirados convergem para os EUA. Até as empresas europeias (**72–80% dos workloads de IA da UE em infraestrutura dos EUA**) votam com os pés.

É o dilema do prisioneiro: individualmente, cada ator otimiza investindo nos EUA. Coletivamente, o resultado é uma perda de soberania tecnológica para todos os aliados.

Q3. Em resumo, por que “simplesmente investir mais” não funciona? *[Público geral]*

Porque o problema não é de volume de investimento. É um problema estrutural. Aqui está a decomposição:

Componente	Ratio	Causa	Investimento sozinho basta?
PIB relativo	1,5:1	EUA maior que UE	Sim, proporcional
Energia	×2–3	Mix energético, taxa carbono, mercado elétrico	Não — estrutural
Prioridade de entrega	×1,5–2	Controles exportação, isenções EUA	Não — geopolítico
Gravidade do capital	×1,5–2	ROI superior nos EUA	Não — sistêmico
Total acumulado	**7–12:1**	**Multiplicativo**	**NÃO**

Tabela. Decomposição do ratio CACI EUA/UE. Fonte: calibração do autor (2024).

Os fatores são multiplicativos, não aditivos. 1,5 (PIB) × 2–3 (energia) × 1,5–2 (entrega) × 1,5–2 (gravidade) = 7 a 12. Mesmo resolvendo um fator, os outros mantêm uma lacuna considerável.

PARTE B — “Mas podemos usar a nuvem americana, não?”

Q4. A Europa já usa massivamente a nuvem dos EUA. O ratio não é um falso problema? *[Decisor / Industrial]*

Essa é a objeção de 90% dos decisores europeus. E no curto prazo, eles têm razão. Por que investir em compute soberano quando se pode alugar dos americanos? AWS, Azure e GCP estão disponíveis, performantes e imediatamente acessíveis. 72% dos workloads de IA europeus já rodam neles.

Mas isso é exatamente a definição de dependência estratégica. Você aluga a infraestrutura crítica de outra parte, que pode: (1) ler seus dados, (2) cortar seu acesso, (3) cobrar mais caro, e (4) atendê-lo depois dos próprios clientes. Os quatro riscos são documentados e concretos.

Q5. Risco 1: o CLOUD Act. Nossos dados são realmente acessíveis aos americanos? *[Jurista / DPO]*

Sim. O Clarifying Lawful Overseas Use of Data Act (CLOUD Act, 2018) autoriza as autoridades americanas a acessar dados armazenados em servidores de empresas americanas, **mesmo que o servidor esteja fisicamente fora do solo dos EUA.**

Concretamente: um banco francês que treina seu modelo de credit scoring no Azure (servidor na Irlanda) — o DoJ americano pode legalmente exigir acesso. Um modelo de defesa treinado na AWS — a NSA pode acessá-lo. Um algoritmo de trading proprietário no GCP — a SEC pode solicitá-lo.

Isso não é teoria da conspiração: é o direito americano vigente. O CLOUD Act não faz distinção entre dados americanos e europeus — ele se aplica à empresa (AWS, Microsoft, Google), não ao território.

> Implicação: qualquer modelo de IA treinado por uma empresa europeia em infraestrutura de nuvem americana é potencialmente acessível às autoridades americanas. É um risco de soberania, não um risco teórico.

Q6. Risco 2: o acesso à nuvem pode realmente ser cortado ou restringido? *[Geopoliticista / Estrategista]*

Já aconteceu — e o mecanismo está pronto para ser ampliado. É o “chokepoint effect” de Farrell & Newman (2019) aplicado à nuvem. Os EUA já:

• **Cortaram o acesso à nuvem para entidades chinesas** (Entity List, BIS). As empresas chinesas listadas não podem mais acessar AWS, Azure ou GCP. Da noite para o dia.

• **Restringiram o acesso à nuvem por país** (AI Diffusion Rule, cotas Tier 2). Países classificados como Tier 2 têm limites no volume de compute em nuvem que podem consumir.

• **Previram ampliação das restrições** (Seção 232, Proclamação prevê possível ampliação em julho de 2026).

Hoje, a Europa é Tier 1 — acesso livre. Mas o mecanismo jurídico está pronto. No dia em que um conflito comercial estourar (retaliação AI Act, taxa digital 2.0, desacordo OTAN), basta uma ordem executiva para restringir o acesso europeu à nuvem. O Cenário B do estudo ("vassalização digital") modela exatamente esse caso.

> Lembrete histórico: a Europa comprava gás russo porque era mais barato e simples do que construir alternativas. Até fevereiro de 2022. O compute americano é o gás russo da IA — exceto que a dependência é ainda mais profunda, porque não se estoca compute como se estoca gás.

Q7. Risco 3: mesmo sem corte, a nuvem dos EUA não custa mais caro para os europeus? *[Industrial / CFO]*

Sim. Um FLOP custa US$ 0,5/TFlop nos EUA e US$ 1,2–1,8/TFlop na UE (Tabela 9, Capítulo IV). Esse diferencial de 2,4–3,6× vem de três fatores:

Latência de rede: um modelo treinado na Virgínia a partir de Paris sofre atrasos de transferência de dados que reduzem a eficiência do treinamento distribuído. Para aplicações em tempo real (manufatura autônoma, veículos conectados, trading), a latência é proibitiva.

Taxas de transferência (egress fees): os hyperscalers cobram pela transferência de dados para fora de suas regiões. Repatriar um modelo treinado nos EUA para a UE custa dinheiro. Essas taxas se acumulam.

Margens regionais: AWS, Azure e GCP cobram 15–25% a mais em regiões da UE do que em regiões dos EUA pela mesma GPU-hora. Documentado pela Bruegel.

Resultado: startups francesas pagam 2 a 3× mais pela mesma GPU-hora que suas concorrentes do Vale do Silício. Mesmo usando a nuvem americana, o campo de jogo não é igual.

Q8. Risco 4: qual o impacto concreto no time-to-market das empresas europeias? *[Industrial / Startup]*

McKinsey documenta uma extensão de 25–40% no time-to-market para empresas europeias em comparação com concorrentes americanas. Esse atraso vem de três fatores cumulativos:

1) Acesso ao compute mais lento: filas para GPUs, atrasos de entrega, cotas de capacidade em nuvem nas regiões da UE.

2) Custo mais elevado: startups da UE precisam captar mais recursos para o mesmo volume de compute, o que alonga os ciclos de captação e atrasa projetos.

3) Conformidade com o AI Act: 3–5% de orçamento adicional em conformidade regulatória (Accenture), que se soma ao sobrecuço do compute.

Em IA, 6 meses de atraso = um ciclo de modelo de atraso = morte competitiva para uma startup. Quando o GPT-5 é lançado em março e seu concorrente americano o integra em abril, você não pode esperar até setembro. O mercado já terá sido tomado.

PARTE C — Então, o que fazemos?

Q9. Se “pagar mais” não é suficiente e a nuvem americana é uma armadilha, qual é a solução? *[Decisor / Político]*

O objetivo não é a autarquia tecnológica, mas a capacidade de escolha. Não 0% de nuvem americana, mas 30–40% dos workloads sensíveis em nuvem soberana certificada até 2029. Ter um Plano B para o dia em que o Plano A se fechar.

As cinco alavancas são detalhadas no Capítulo VII:

1) Special Compute Zones com energia a US$ 50–60/MWh via PPAs nucleares. É a alavanca nº 1: sem energia competitiva, nada mais funciona.

2) Integração nuclear-IA: a França tem 63 GW nucleares, a EDF pode dedicar 2 GW imediatamente. O único mix energético da UE que permite uma trajetória credível. 250 MW até o final de 2026, 6 EPR 2 em construção, SMRs até 2030+.

3) Reservas estratégicas de GPU: contratos-quadro da UE com Nvidia, AMD, Intel Foundry por 18–36 meses. Garantir volumes antes que as restrições se ampliem.

4) AI Act como alavanca ofensiva: condicionar o acesso ao mercado da UE à localização do compute. Reconhecimento mútuo com Japão, Coreia, Singapura. CLOUD Act Shield europeu.

5) Modelo Mistral (analogia Airbus): não substituir a OpenAI, mas construir uma alternativa credível. A ASML investiu €1,3 bilhão (11% do capital) — o sinal industrial está lá. A pergunta não é “A Mistral pode vencer a OpenAI?” mas “A Europa pode ficar sem a Mistral?” A resposta é não.

Q10. Em uma frase, por que esse ratio de 7–12:1 é um problema existencial e não apenas um atraso recuperável? *[Público geral]*

Porque o compute produz efeitos de aglomeração autorreforçantes. Quanto mais compute você tem, mais talentos vêm, mais investidores vêm, mais compute aumenta. É um círculo virtuoso para os EUA e um círculo vicioso para a UE. Cada ano de inação amplia o ratio em vez de reduzi-lo.

Por isso a janela 2026–2028 é crítica. Após 2028, as posições estarão consolidadas: as AI Gigafactories dos EUA estarão operacionais, os ecossistemas de nuvem congelados, os talentos instalados. O Cenário D (“Resposta maciça da UE”) ainda é possível hoje. Não será mais em 2030.

O ratio de 7–12:1 não é um atraso temporário: é uma divergência estrutural que, sem intervenção, se torna irreversível. É a diferença entre “estamos atrasados” e “perdemos a capacidade de alcançar.”

> **“A questão não é mais se a recomposição da ordem tecnológica mundial acontecerá — ela está em andamento — mas se seremos seus arquitetos ou seus súditos.”**

— Fabrice Pizzi, AI for Americans First, 2026

— FIM DO VOLUME 3 —

Este documento pode ser lido independentemente do Volume 1 (tese, CACI, metodologia, cenários, limitações) e do Volume 2 (25 perguntas técnicas, geopolíticas e operacionais).

Repositório: https://mo0ogly.github.io/America-First-IA/