

RESUMO SIMPLES - CSAP - CIÊNCIAS SOCIAIS APLICADAS

NVIDIA NA INFRAESTRUTURA DA IA (2019–2024)

José Victor Rodrigues Catalano (falecomjosevictor@gmail.com)

A difusão do ChatGPT, em 2022, desencadeou um novo ciclo de expansão da inteligência artificial (IA) generativa e evidenciou os requisitos materiais e organizacionais necessários para sua operação em escala — hardware especializado, plataformas de software, cadeias de suprimento e capacidade de data center. Este artigo examina a transformação estratégica da NVIDIA, no período 2019–2024, de fabricante originalmente orientada ao mercado de jogos em um ator central da infraestrutura global de IA. Argumenta-se que essa inflexão decorre menos de ganhos incrementais de desempenho e mais de uma arquitetura competitiva construída pela combinação entre

inovação tecnológica, governança de plataforma e expansão ao longo da cadeia de

valor. Metodologicamente, o estudo mobiliza análise setorial e estratégica e se concentra em três vetores: (i) a diferenciação de GPUs para aprendizado de máquina;

(ii) a consolidação do CUDA como plataforma proprietária e padrão de fato para desenvolvimento em IA; e (iii) uma estratégia dirigida de fusões e aquisições ao longo

de 2019–2024. Sustenta-se que a articulação desses elementos produz um lock-in

tecnológico robusto: ao reduzir custos de adoção e elevar custos de migração, a empresa

reforça externalidades de rede, amplia barreiras à entrada e consolida sua posição por

meio da integração entre hardware, bibliotecas, ferramentas e soluções de data center

em um ecossistema dominante. Os achados indicam que a dominância da NVIDIA

deriva menos de uma vantagem isolada de produto e mais da capacidade de orquestrar

um ecossistema, definir interfaces críticas e capturar rendas associadas ao controle

infraestrutural; em 2024, a empresa concentrava 92% do mercado de aceleradores de IA

e obtinha 88% de suas receitas na divisão de data center. O artigo também identifica

tensões que condicionam a sustentabilidade dessa posição, incluindo vulnerabilidades

em cadeias de suprimento, intensificação de restrições tecnológicas no eixo Estados

Unidos–China, sinais de arrefecimento relativo da intensidade de pesquisa e desenvolvimento, e pressões competitivas associadas a alternativas open source, além

do dilema inovativo de trajetórias de crescimento ancoradas em aquisições. Conclui-se

que a posição da NVIDIA é melhor explicada como resultado de uma arquitetura competitiva baseada em plataforma e integração infraestrutural, e não como simples

liderança incremental de desempenho.

Palavras-chave: nvidia; inteligência artificial; data center; fusões e aquisições; infraestrutura digital.