



MultiCortex OS

O MAIS AVANÇADO SISTEMA OPERACIONAL PARA IA DO MERCADO

Sobre a MultiCortex



Bases nos EUA e Brasil

Fundada por profissionais com mais de 30 anos de experiência em tecnologia e inovação, a MultiCortex conta com bases nos Estados Unidos e no Brasil, além de parceiros globais.



Tecnologia pioneira

A empresa criou o primeiro Sistema Operacional com computação heterogênea do mundo, o MultiCortex OS.



MultiCortex OS



Do sistema operacional à aplicação de destaque global

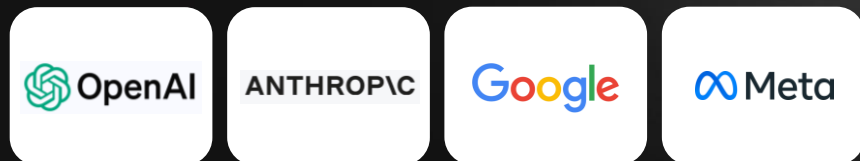
A empresa se destaca por criar, utilizando o MultiCortex OS, o DevCore for Claude Code, um assistente de programação com token infinito e custo fixo.

DevCore
for Claude Code



Importância do Sistema Operacional (SO)

Toda solução de IA precisa de um sistema operacional. Muitas empresas não se atentam para isso porque utilizam IA por meio de APIs e, desta forma, o sistema operacional está na OpenAI, na Anthropic, na Google, entre outras.



Se a sua empresa deseja uma IA soberana, que processe as LLMs dentro do seu próprio ambiente com total controle, segurança e conformidade com a LGPD, você precisa de um sistema operacional.

FUNÇÃO TÉCNICA

O sistema operacional é o coração do computador.

Ele faz a comunicação entre o hardware e os softwares, controlando os componentes físicos (CPU, GPU, NPU, memória, discos) e distribuindo o processamento entre os programas em execução.

Softwares (Aplicativos)

Sistema Operacional

Hardware (CPU, GPU, NPU)

Fabricantes de Sistemas Operacionais

Em junho de 2025, identificamos 4 sistemas operacionais voltados para inteligência artificial em servidores no mundo.



MultiCortex OS

O MultiCortex OS é o primeiro sistema operacional para IA com computação heterogênea do mundo.

A tecnologia heterogênea permite ganhos relevantes de processamento, refletindo diretamente na redução do custo computacional.



A computação heterogênea nasceu oficialmente em Harvard (2003)



Apontada pela Linux Foundation como o futuro da IA



Diferencial

COMPUTAÇÃO TRADICIONAL

Processamento de IA utilizando GPU

1 vetor de processamento e aceleração

GPU

Limitado ao poder de processamento da GPU.

VS

COMPUTAÇÃO HETEROGÊNEA

Processamento de IA com GPU + CPU

2 vetores de processamento e aceleração



MultiCortex OS

GPU

CPU

Explora 100% do potencial da máquina.

O diferencial do Sistema Operacional MultiCortex é a computação heterogênea que permite o processamento da IA utilizando qualquer acelerador disponível, seja GPU, CPU, XPU, NPU... de forma integrada.

Qual o impacto que essa tecnologia pode gerar para o seu negócio e para o planeta?

Reduzir o custo computacional da IA em até 50%

(e talvez até mais....)

Números oficiais:



Projeto SoberanIA

Maior projeto de IA do
Brasil (R\$23 bilhões)



Sistema operacional
mais utilizado no mundo

Resultado: **22 tokens/s**



Sistema operacional
MultiCortex OS

Resultado: **63 tokens/s**

Números oficiais:

65%

Redução do custo computacional

Redução do impacto ambiental

Destques



MultiCortex OS

Diferenciais do MultiCortex OS que vão transformar a operação de IA da sua empresa

Fácil de Instalar e Usar

Solução Plug-and-Play. Implantação simples sem necessidade de reconfigurar a infraestrutura existente.

Funciona com Qualquer Hardware

Suporte heterogêneo a GPU, CPU, NPU e aceleradores de diferentes fabricantes.

Modelos Open Source Completos

Compatível com modelos de texto, imagens e vídeos em uma única plataforma unificada.

Atualizações Sempre Inclusas

Todas as atualizações e melhorias do sistema estão incluídas na assinatura, sem custos adicionais.

Sem Token — 100% Privativo

Processamento local, dados nunca saem da infraestrutura do cliente. Zero custo de tokens.

Lock-in Zero

Eliminamos a dependência exclusiva de um único fornecedor de hardware. Liberdade total.

IA para Banco do Leste Europeu

CONTEXTO

- Special Build para um banco com parque de GPUs de diferentes fabricantes.
- Demanda do banco: velocidade de output de 39 tokens/seg.
- Especialistas de uma BigTech trabalharam 2 meses e atingiram apenas 33 tokens/seg.
- A MultiCortex, com computação heterogênea, superou a meta.

33 tok/s

BigTech (2 meses)

39 tok/s

Meta do banco

43 tok/s

Resultado da MultiCortex

30% de aumento de produtividade

Performance superior ao benchmark da BigTech.

Otimização de Motor Biométrico com IA

CONTEXTO

A biometria é uma área crítica da IA. Os melhores sistemas do mundo possuíam:

Margem de erro (falso positivo) anterior

0.001

Margem de erro com MultiCortex

0.000000000001

IMPACTO OPERACIONAL

Mesa antifraude com:

30

Analistas
COM MultiCortex

600

Analistas
SEM MultiCortex

PARCEIRO OFICIAL

Cognitec — The Face Recognition Company



99% de aumento de eficiência de processamento

Praticamente: redução de 600 para 30 pessoas na mesa antifraude.

IA privativa para o setor de saúde

CONTEXTO

- Special Build de IA para o setor da saúde.
- O MultiCortex OS propiciou uma resposta até 329% mais detalhada frente ao modelo líder de mercado (no caso, ChatGPT 5 (versão paga)).

Respostas 300% mais completas

Em relação ao ChatGPT 5 (versão paga)

DESTAQUES

- Mesmo prompt.
- ChatGPT 5 (versão paga) utilizando uma máquina estimada em R\$5 milhões.
- IA da MultiCortex do setor da saúde utilizando uma máquina de R\$70 mil.

193

Palavras com o ChatGPT 5

827

Com MultiCortex OS

100% de aprovação

Em todos os casos em que médicos foram consultados, preferiram a resposta gerada pelo MultiCortex OS.

Inferência na Linha Power da IBM

CONTEXTO

- O parceiro de Data Center não conseguia viabilizar a inferência em máquinas P10 da IBM, que não possuem GPU.
- Special Build de IA para inferência em máquinas Power da IBM (RISC) com MMA.

DESTAQUES

- Utilizando o MultiCortex OS, conseguimos viabilizar a inferência em máquina P10 da IBM.
- O resultado foi aprovado pelo parceiro, uma grande empresa multinacional de TI (mais de 15 mil profissionais e atuação em 13 países).



Sem MultiCortex, o parceiro não conseguiu realizar inferência na P10

15 tokens/s

Inferência com MultiCortex OS e resultado aprovado pelo cliente

Escalabilidade

A solução pode ser aplicada a qualquer máquina da Linha Power da IBM.

MultiCortex OS em funcionamento

Exemplos de aplicações com diferentes interfaces:

ALL-IN-ONE

Interface unificada



Permite utilizar diferentes tipos de modelos — texto, imagens e vídeos — em um único sistema, facilitando o acesso e a usabilidade.

▶ youtu.be/SXyeu0NbTEw

TROQUE COM UM CLIQUE

Troca de modelo instantânea



O usuário altera o modelo de IA com um único clique. Fundamental porque diferentes tarefas (cálculo, programação, escrita) demandam modelos distintos.

▶ youtu.be/3gLt2WrEdNE

ARENA DE IA

Processamento paralelo



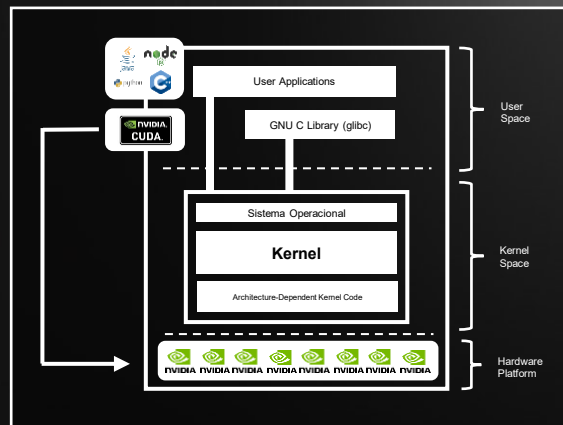
Processa múltiplos modelos de IA simultaneamente, maximizando produtividade e facilitando a identificação de alucinações.

▶ youtu.be/COu9TI-diPU

Total soberania

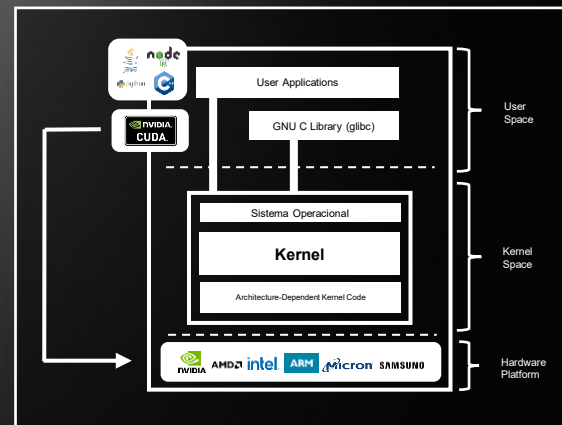
ESTRUTURA TRADICIONAL

Com aprisionamento tecnológico
(Single-vendor)



ESTRUTURA HETEROGÊNEA

Sem aprisionamento tecnológico
(Multi-vendor)



No hardware Lock-in

A computação heterogênea resolve o maior problema das empresas que usam IA, que é o *lock-in* tecnológico (dependência exclusiva de um fornecedor de hardware)

No software Lock-in

Outro benefício é a possibilidade de *transmutação de código*, que permite usar a linguagem CUDA (NVIDIA) em hardware de outros fornecedores.

História da computação heterogênea

De Harvard ao MultiCortex OS

1980–1990

Pesquisas Pioneiras

Início das pesquisas com sistemas multi-arquitetura que unem diferentes processadores.

1995

Marco Científico

Artigo 'Simultaneous Multithreading' demonstra até 21% de ganho de desempenho e 23% de economia de energia com processadores heterogêneos.

2003

Harvard — Nascimento



Nasce oficialmente na Universidade de Harvard o conceito formal de computação heterogênea.

Até 2023

Intel lidera a evolução



O conceito é amplamente estudado e desenvolvido pela Intel nos EUA ao longo de duas décadas.

2024

MultiCortex OS



A MultiCortex cria o primeiro sistema operacional para IA com computação heterogênea do mundo.

2025

Expansão Global: UXL Foundation



O conceito cresce exponencialmente com a ampliação da UXL Foundation, criada para democratizar a computação heterogênea e gerida pela Linux Foundation.



Desafios que o MultiCortex OS resolve



MultiCortex OS

Problema 1 – Necessidade de engenheiro Linux	Necessidade de um engenheiro Linux para otimização do Kernel para IA	Se o Kernel e o sistema operacional não forem otimizados, o sistema ficará lento. O problema é que poucas pessoas sabem otimizar o Kernel
Problema 2 – Necessidade de especialista em IA	Necessidade de especialista em IA para instalação de middleware, motor de inferência e modelos de IA	Se o especialista fizer algo errado, a máquina pode perder até 3.300% de desempenho computacional
Problema 3 – Contínuas atualizações	Este processo precisa ser repetido constantemente: otimização de Kernel, atualização de motor de inferência...	O mercado de IA, por ser altamente dinâmico, exige atualizações constantes. A não realização dessas atualizações pode gerar riscos imediatos à continuidade do negócio
Problema 4 – Continuidade de negócio	Turnover e altos custos de retenção de talento para manter a arquitetura da infraestrutura de IA	Em função do ritual de transferência de conhecimento, a rotatividade de profissionais de TI é um fator de risco para obtenção de sucesso na implantação da IA

Obrigado!



MultiCortex OS

WHATSAPP

<https://wa.me/5511981661646>

EMAIL

multicortex@multicortex.ai

SITE

www.multicortex.ai

INSTAGRAM

[@multicortex](https://www.instagram.com/multicortex)