

dr inż. Michał Czarkowski
Dyrektor ds. Rozwoju Oprogramowania
AMD

„Aspekty lokalnych wdrożeń chmur AI dla ekosystemów z pełną kontrolą danych”

Prezentacja przedstawia koncepcję budowy lokalnych ekosystemów chmurowych AI, które łączą pełną kontrolę nad danymi z możliwością skalowania infrastruktury i świadomym zarządzaniem kosztami. Punktem wyjścia jest dobór miejsca uruchomienia każdego obciążenia na podstawie oczekiwanego rezultatu, profilu zadania, wrażliwości danych, opóźnień, dostępności oraz ekonomiki. Zamiast stosować jeden kosztowny poziom infrastruktury dla wszystkich zastosowań, rekomendowany jest model warstwowy: od zasobów lokalnych, przez współdzieloną infrastrukturę zespołową, po rozwiązania klasy enterprise i zasoby chmurowe. Zadanie powinno trafiać do najniższego poziomu, który spełnia wymagania jakościowe, wydajnościowe i regulacyjne, z możliwością eskalacji w przypadku potrzeb pojemnościowych lub SLA.

Materiał podkreśla, że koszty AI zależą nie tylko od ceny modelu, lecz także od wolumenu tokenów wyjściowych, wykorzystania zasobów, opóźnień i ponownego użycia kontekstu. Kluczowe znaczenie ma spójny model zarządzania obejmujący polityki, lokalizację danych, ochronę informacji, tożsamość, audyt i odpowiedzialność operacyjną. Docelowy ekosystem umożliwia bezpieczne rozpoczęcie pracy lokalnie, współdzielenie mocy w zespołach i strategiczne skalowanie w centrum danych.