

Upraszczanie rozwoju generatywnej sztucznej inteligencji

Stacje robocze Dell Precision AI
i NVIDIA AI Workbench

Rozpocznij



Spis treści

Wyzwania związane z rozwojem generatywnej sztucznej inteligencji	3
Rozwiązanie: NVIDIA AI Workbench na stacjach roboczych Dell Precision	4-7
Przykład wdrożenia: Invoke i Invoke Enterprise	8
NVIDIA AI Workbench i NVIDIA AI Enterprise	9
Dalsze działania i zasoby	10

Wyzwania związane z rozwojem generatywnej sztucznej inteligencji

Codzienne wyzwania związane z rozwojem generatywnej sztucznej inteligencji (konfigurowanie procesorów graficznych i środowiska Jupyter Notebook, przenoszenie obciążeń, zarządzanie przepływami pracy itp.) mogą mieć negatywny wpływ na postępy realizowanych projektów. Czas i wiedza techniczna niezbędne do obsługi tej technologii mogą zniechęcać danetyków, inżynierów i deweloperów do zajmowania się tą technologią. Na szczęście dzięki odpowiednim partnerstwom i narzędziom technicznym można te wyzwania pokonać i przyspieszyć rozwój sztucznej inteligencji.

Typowe przeszkody dla generatywnej sztucznej inteligencji



Złożona konfiguracja
procesorów graficznych



Ograniczona przenośność
obciążeń roboczych



Czasochłonne zarządzanie
przepływami pracy

Rozwiązanie: NVIDIA AI Workbench na przystosowanych do obsługi sztucznej inteligencji stacjach roboczych Dell Precision

Dzięki wspomnianemu rozwiązaniu zaawansowani i początkujący deweloperzy mogą szybko i łatwo tworzyć, testować i dostosowywać projekty generatywnej sztucznej inteligencji na lokalnych stacjach roboczych, a następnie skalować je w górę lub w dół w odniesieniu do dowolnej platformy. Przystosowane do obsługi sztucznej inteligencji stacje robocze Dell Precision zapewniają przy tym wydajność niezbędną do przetwarzania dużych zbiorów danych i obsługi złożonych sieci neuronowych.

NVIDIA AI Workbench

- ✓ Usprawniona konfiguracja
Szybsze kodowanie i dostosowanie modeli dzięki zautomatyzowanej konfiguracji procesorów graficznych
- ✓ Realizacja projektów z dowolnego miejsca
Tworzenie obciążeń lokalnie i łatwe ich przenoszenie na najlepszą platformę pod względem kosztów, szybkości i skalowalności
- ✓ Przyspieszenie przepływów pracy sztucznej inteligencji
Zwiększenie produktywności dzięki automatyzacji zarządzania kontenerami i wersjami

Stacje robocze Dell Precision

- ✓ Rozwój generatywnej sztucznej inteligencji lokalnie
Dostosowywanie i rozwijanie modeli generatywnej sztucznej inteligencji na lokalnych stacjach roboczych z akceleracją GPU
- ✓ Skalowalność
Skalowanie od jednego do czterech procesorów graficznych NVIDIA RTX™ Ada Generation
- ✓ Przystosowanie do obsługi sztucznej inteligencji
Rozwiązanie opracowane i skonfigurowane specjalnie pod kątem obsługi obciążeń związanych ze sztuczną inteligencją

Usprawnienie konfiguracji sprzętu

Ilekość deweloperzy muszą zajmować się zawiłą konfiguracją komputera z procesorami graficznymi na potrzeby głębokiego uczenia i generatywnej sztucznej inteligencji, cierpi na tym produktywność. NVIDIA AI Workbench eliminuje typowe problemy związane z konfiguracją, na przykład niedopasowanie systemu operacyjnego i sterowników oraz niezgodność sterowników CUDA i oprogramowania wewnętrznego. Dzięki zautomatyzowanej instalacji i konfiguracji oprogramowania oraz sterowników deweloperzy mogą natychmiast szybko rozpocząć kodowanie. NVIDIA AI Workbench pomaga również zapewnić, że kontenery są skonfigurowane pod kątem optymalnej wydajności procesorów graficznych w zależności od obciążenia.



Większa produktywność dzięki zautomatyzowanej konfiguracji



Eliminacja typowych problemów z konfiguracją sprzętu



Zapewnienie poprawnej konfiguracji kontenerów

Realizacja projektów z dowolnego miejsca

NVIDIA AI Workbench oferuje deweloperom możliwość łatwego tworzenia, dostosowywania i współpracy przy jednoczesnej bezproblemowej przenośności obciążeń roboczych. Projekty NVIDIA AI Workbench można bezpiecznie przenieść do najlepszego pod względem kosztów, szybkości i skalowalności systemu obsługującego procesory graficzne. Wszystko to można robić szybko i bez większych umiejętności technicznych. Wszelkie zmiany i wersje środowiska, danych, modeli i lokalizacji mogą być już z łatwością obsługiwane, podczas gdy NVIDIA AI Workbench bezpiecznie zarządza uwierzytelnianiem i kluczami tajnymi potrzebnymi użytkownikom do uzyskiwania dostępu do różnych zasobów.



Łatwe przenoszenie
obciążeń



Bezpieczne przenoszenie kluczy
tajnych i danych uwierzytelniających



Łatwe zarządzanie
zmianami i wersjami

Przyspieszenie przepływów pracy sztucznej inteligencji



Automatyzacja instalacji projektów i zarządzania nimi



Automatyczne śledzenie zależności projektów



Uproszczenie zarządzania projektami dzięki intuicyjnym interfejsom użytkownika i wiersza polecenia

Dzięki NVIDIA AI Workbench początkujący i zaawansowani deweloperzy mogą szybko identyfikować i instalować elementy projektów związanych ze sztuczną inteligencją i głębokim uczeniem oraz zarządzać nimi. Możliwość automatycznej instalacji różnych zasobów projektowych zwiększa produktywność. Zależności projektowe dla danego modelu i procesora graficznego można łatwo śledzić, co pozwala deweloperom skupiać się na realizacji zadań. NVIDIA AI Workbench upraszcza zarządzanie przepływami pracy dzięki intuicyjnemu interfejsowi użytkownika i zaawansowanemu interfejsowi wiersza poleceń.

Zobacz przykładowe przepływy pracy na platformie GitHub:

➞ [Projekt Hybrid RAG](#)

➞ [Projekt NIM Anywhere](#)

➞ [Projekt Stable Diffusion](#)

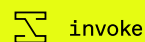
➞ [Projekt Fine-Tuning](#)

Przykład wdrożenia: Invoke Enterprise

Na bazie rozwiązania NVIDIA AI Workbench firma Invoke opracowała rozwiązanie pozwalające przedsiębiorstwom na bezpieczne i oparte na współpracy trenowanie modeli sztucznej inteligencji do przetwarzania obrazów. Ponadto wdrożyła rozwiązania Invoke Enterprise i NVIDIA AI Workbench lokalnie, umożliwiając wykonywanie obliczeń związanych ze sztuczną inteligencją na [mobilnych stacjach roboczych Dell Precision 7780](#) wyposażonych w procesory graficzne NVIDIA RTX™ 5000 Ada Generation.

„Wdrażanie generatywnych modeli sztucznej inteligencji na stacjach roboczych Dell Precision przystosowanych do obsługi sztucznej inteligencji, w szczególności Precision 7780, wyznaczyło nowy standard wydajności. Starszy sprzęt może mieć trudności z obsługą zapotrzebowania na zasoby obliczeniowe sztucznej inteligencji, ale funkcja Invoke na stacjach 7780 działa płynnie i wydajnie oraz obsługuje wszystkie nasze najbardziej zaawansowane przepływy pracy związane ze sztuczną inteligencją”.

- Kent Keirse, założyciel, prezes i dyrektor generalny Invoke



[Przeczytaj analizę przypadku](#)





NVIDIA AI Workbench i NVIDIA AI Enterprise



Dzięki połączeniu rozwiązań NVIDIA AI Workbench z NVIDIA AI Enterprise można uzyskać dostęp do kompleksowej, natywnej dla chmury platformy oprogramowania, która przyspiesza działanie związanych z danymi kanałów przepływu danych oraz usprawnia opracowywanie i wdrażanie produkcyjnych aplikacji sztucznej inteligencji. NVIDIA AI Workbench to rozwiązanie bezpłatne, które obejmuje wsparcie dla posiadaczy licencji NVIDIA AI Enterprise, w tym funkcje bezpieczeństwa, funkcje NIM (NVIDIA Infrastructure Manager) oraz odpowiednie wersje produkcyjne.



Dalsze działania i zasoby

Rozpocznij

- ➔ Pobierz rozwiązanie AI Workbench bezpłatnie
- ➔ Znajdź konfiguracje stacji roboczych przystosowanych do obsługi sztucznej inteligencji
- ➔ Skontaktuj się z nami, aby kupić rozwiązanie NVIDIA AI Enterprise za pośrednictwem Dell
- ➔ Wypróbuj bezpłatnie rozwiązanie NVIDIA AI Enterprise przez 90 dni

Więcej informacji

- ➔ NVIDIA Tools for Developers z przystosowanymi do obsługi sztucznej inteligencji stacjami roboczymi Precision (artykuł na blogu)
- ➔ NVIDIA AI Workbench wspomaga rozwój aplikacji (artykuł na blogu)
- ➔ Zoptymalizuj wydajność modelu sztucznej inteligencji i zachowaj prywatność danych dzięki hybrydowej funkcji RAG (artykuł na blogu)
- ➔ Bezproblemowe tworzenie i wdrażanie skalowalnych modeli generatywnej sztucznej inteligencji z technologią NVIDIA AI Workbench (artykuł na blogu)