

Przemysław SUJECKI*

* Wojskowa Akademia Techniczna,
DKWOC
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa
e-mail: przemo.sujecki@gmail.com

MODELE WIZUALNO-JĘZYKOWE W WIELOMODELOWYM SILNIKU PERCEPCJI AI

Współczesne systemy percepcji sztucznej inteligencji muszą interpretować nie tylko pojedyncze obrazy, lecz także dokumenty, strumienie wideo oraz polecenia językowe. Celem pracy jest przedstawienie koncepcji wykorzystania modeli wizualno-językowych jako warstwy łączącej analizę obrazu i języka w systemach wspierających rozumienie sytuacji w czasie rzeczywistym.

W prezentowanym ujęciu pokazano przejście od klasycznych rozwiązań multimodalnych, w których detektor wizualny, reprezentacja cech i model językowy działały jako oddzielne komponenty, do nowoczesnych modeli vision-language. Obraz, dokument lub klatki wideo są przetwarzane przez encoder wizualny, a następnie reprezentowane jako tokeny wizualne analizowane wspólnie z tekstem przez model językowy.

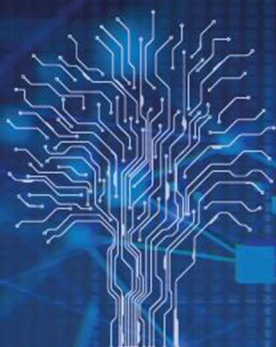
Drugim elementem pracy jest koncepcja wielomodelowego silnika percepcji AI. Strumień z kamery lub drona może być przetwarzany równolegle przez moduły detekcji i śledzenia obiektów, oceny sytuacji, krótkoterminowej predykcji oraz automatycznego raportowania. Pozwala to łączyć szybką analizę niskiego poziomu z interpretacją kontekstu i generowaniem komunikatów dla operatora.

Wnioskiem z pracy jest wskazanie, że modele wizualno-językowe mogą pełnić funkcję komponentu integrującego percepcję, interpretację i komunikację człowiek-system. Takie podejście wspiera przejście od samego wykrywania obiektów do tworzenia opisów, alertów i raportów użytecznych w systemach wymagających świadomości sytuacyjnej.



Wojskowa
Akademia
Techniczna

Wydział
Elektroniki



SECON 2026

TWOJA PIERWSZA KONFERENCJA
TWÓJ POCZĄTEK W ŚWIECIE NAUKI