

ZERVIDJANNA

Skrócona dokumentacja projektowa

Robot kroczący inspirowany naturą, z lokalną warstwą AI

Status: prototyp głowy, lokalna rozmowa STT/LLM/TTS oraz stanowisko testowe napędu capstan



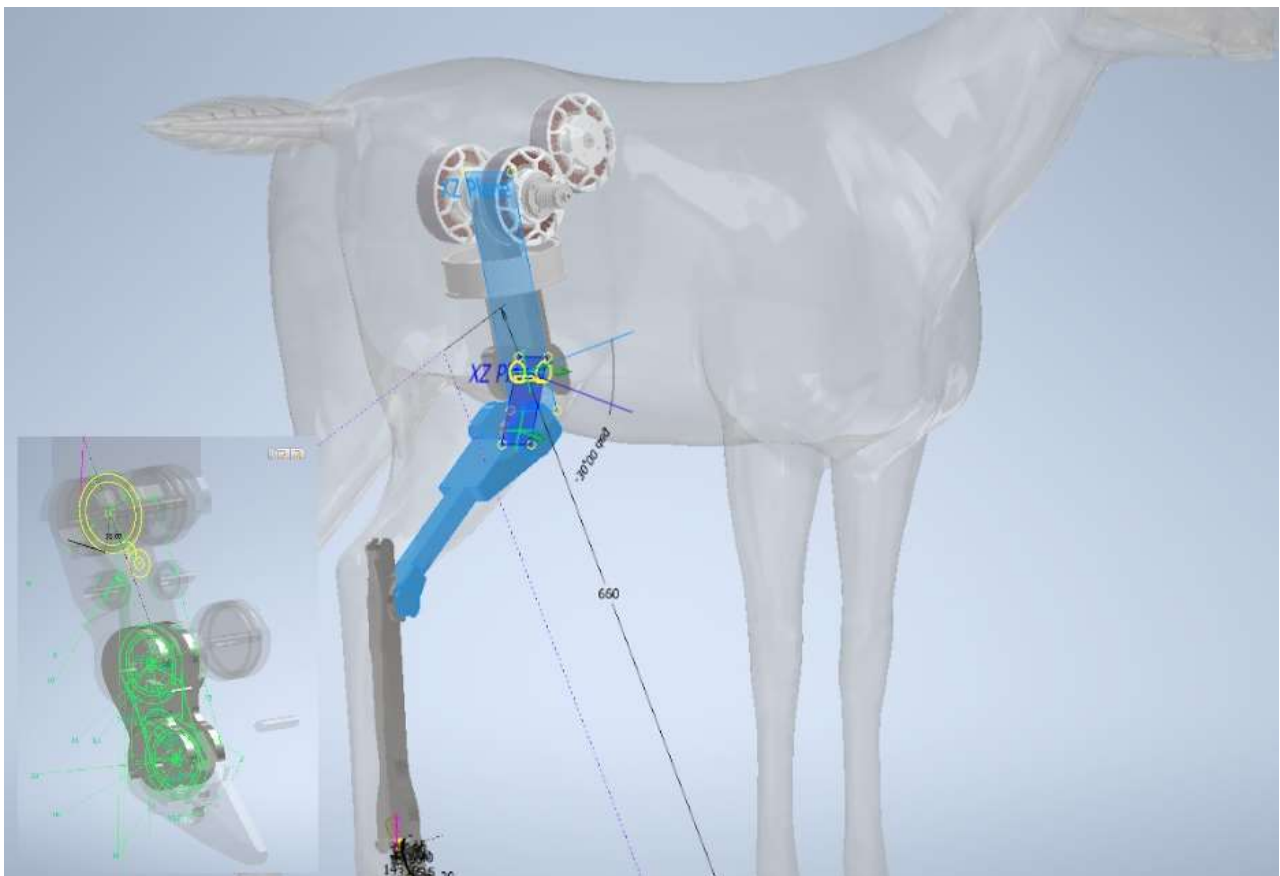
1. Charakterystyka projektu

Zervidjanna to autorski projekt robota krocącego inspirowanego budową i sposobem poruszania się zwierzęcia kopytnego. Docelowo konstrukcja ma łączyć biomechaniczny układ kończyn, lokalną sztuczną inteligencję oraz naturalną komunikację głosową. Aktualnie projekt rozwijany jest etapowo: najpierw powstają niezależne stanowiska i moduły testowe, które pozwalają sprawdzić mechanikę, elektronikę oraz oprogramowanie przed integracją w pełnym robocie.

Główna idea rozwoju: najpierw potwierdzić działanie pojedynczych mechanizmów i algorytmów, a dopiero później łączyć je w większą, bardziej złożoną konstrukcję. Takie podejście ogranicza ryzyko projektowe i ułatwia wykrywanie błędów na wczesnym etapie.

Aktualny zakres prac

Obszar	Aktualny stan / rola w projekcie
Warstwa AI	Lokalne przetwarzanie rozmowy: VAD, STT, lokalny LLM i TTS. Warstwa jest rozszerzana o obraz z kamery, detekcję twarzy oraz kontekst osób przekazywany do LLM.
Głowa robota	Prototyp wykonany ze sklejki do testów mechanizmu zamykania szczęki. Obecny etap służy ocenie geometrii, napędu serwem i praktycznej pracy zuchwy.
Napęd kończyn	Stanowisko testowe capstan z silnikiem BLDC 9235-100 KV, linką Dyneema DM20, sterownikiem ODrive 3.6 i enkoderem AMT 102-V do pomiaru pozycji przyszłego stawu nogi.
Kierunek dalszy	Integracja mechaniki, sterowania, lokalnej AI i percepcji wizyjnej w spójny system robota krocącego, zdolnego do interakcji z otoczeniem.



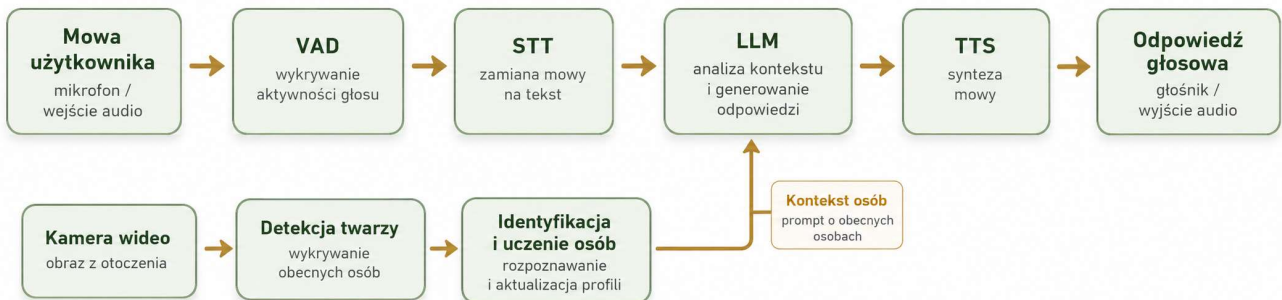
2. Warstwa AI i komunikacja głosowa

W projekcie działa lokalny tor rozmowy rozszerzony o kontekst wizyjny. Użytkownik mówi do robota, VAD wykrywa aktywność głosu, STT zamienia mowę na tekst, a lokalny LLM analizuje wypowiedź i generuje odpowiedź. Równolegle kamera dostarcza obraz, z którego system wykrywa twarze i rozpoznaje obecne osoby; ta informacja trafia do LLM jako krótki kontekst rozmowy.

Najważniejsza cecha obecnego rozwiązania: przetwarzanie odbywa się lokalnie, bez konieczności opierania działania na chmurze. Daje to większą kontrolę nad systemem i pozwala łączyć rozmowę z informacją o tym, kto znajduje się przed robotem.

Tor komunikacji głosowej i kontekst wizyjny AI

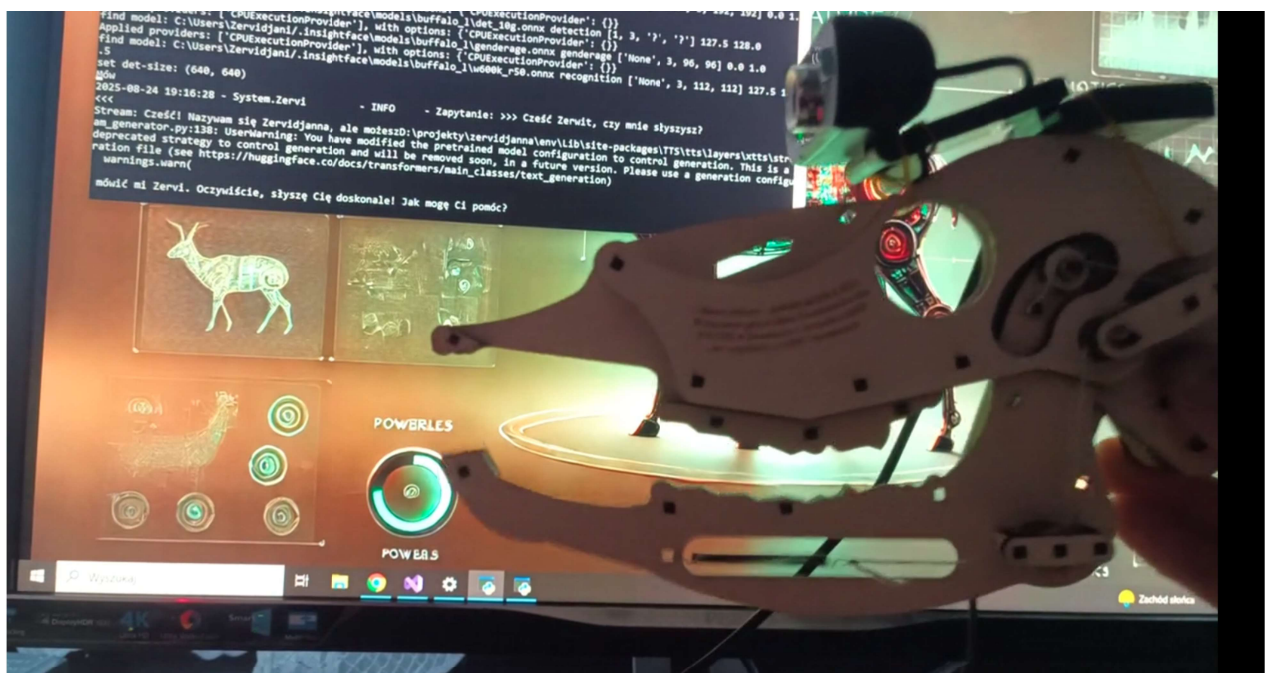
VAD, rozpoznawanie mowy, analiza LLM, synteza odpowiedzi oraz dane o osobach z kamery.



Rys. 1. Uproszczony przepływ informacji w module rozmowy i kontekstu wizyjnego.

Elementy funkcjonalne

- **VAD/STT** – wykrywanie aktywności głosu oraz zamiana mowy na tekst wejściowy dla systemu.
- **LLM lokalny** – analiza wypowiedzi, kontekstu osób oraz generowanie odpowiedzi i przyszłe planowanie reakcji robota.
- **Wizja i uczenie osób** – detekcja twarzy, zapis profili jako embeddingów oraz porównywanie ich odległością euklidesową w przestrzeni latentnej.
- **TTS i integracja z robotem** – zamiana odpowiedzi na mowę oraz przyszłe akcje: ruch głowy, szczęki, kończyn lub zmiana trybu pracy.



3. Prototyp głowy i mechanizm szczęki

Głowa jest pierwszym fizycznym modułem interakcji robota. Wykonana ze sklejki konstrukcja służy do sprawdzenia mechanizmu zamykania szczęki, rozmieszczenia napędu oraz ograniczeń przestrzennych. Na tym etapie nie jest to finalna obudowa robota, lecz stanowisko mechaniczne pozwalające szybko testować geometrię, przeniesienie ruchu i zachowanie serwomechanizmu.

Znaczenie prototypu: szczeka może docelowo pełnić kilka funkcji: chwytak do prostych interakcji z obiektami, element ekspresji robota oraz mechanizm synchronizowany z mową TTS. Dlatego już na etapie prób istotne są szybkość ruchu, powtarzalność, odporność na przeciążenia i możliwość dalszej miniaturyzacji.

Wnioski z dotychczasowego etapu

- konstrukcja ze sklejki pozwala szybko iterować kształt elementów i sprawdzać mechanikę w rzeczywistej skali;
- napęd serwowym umożliwia praktyczne testy otwierania i zamykania szczęki bez budowy finalnej głowy;
- najważniejszym ograniczeniem pozostaje ilość miejsca wewnątrz głowy oraz konieczność ochrony napędu przed przeciążeniem;
- moduł szczęki jest dobrym punktem przejściowym między częścią mechaniczną, oprogramowaniem rozmowy i przyszłą ekspresją robota.



4. Stanowisko testowe napędu capstan

Najnowszym etapem projektu jest stanowisko do testu przeniesienia napędu typu capstan. Układ wykorzystuje silnik BLDC 9235-100 KV, sterownik ODrive 3.6, linkę Dyneema DM20 oraz enkoder pojemnościowy AMT 102-V. Silnik napędza bęben, a linka przenosi siłę na drugi bęben, który docelowo może odpowiadać pracy stawu nogi robota. Enkoder dostarcza informację zwrotną o położeniu, dlatego stanowisko pozwala oceniać dokładność i powtarzalność ruchu.

Dlaczego capstan: napęd linkowy pozwala odsunąć cięższy silnik od samego stawu, uzyskać przełożenie poprzez różnicę średnic bębnow oraz testować rozwiązania bliższe lekkiej konstrukcji kończyn. Połączenie z enkoderem ułatwia pracę w pętli sprzężenia zwrotnego i porównywanie pozycji zadanej z rzeczywistą.



Rys. 2. Uproszczona struktura stanowiska dla testu przyszłego stawu nogi.

Elementy stanowiska

Obszar	Aktualny stan / rola w projekcie
Silnik BLDC	9235-100 KV – napęd główny stanowiska, sterowany przez ODrive.
Sterownik	ODrive 3.6 – platforma sterowania BLDC, użyta do testów pozycji/prędkości/momentu oraz pracy z informacją zwrotną z enkodera.
Przeniesienie napędu	Linka Dyneema DM20 i bębny capstan – sprawdzenie naciągu, poślizgu i możliwości przenoszenia momentu.
Cel pomiarów	Ocena przydatności rozwiązania jako napędu stawu kończyny oraz sprawdzenie dokładności ruchu z enkoderem AMT 102-V.

5. Zakres testów i dalszy rozwój

Dotychczas sprawdzane funkcje

- komunikacja głosowa w lokalnym torze VAD → STT → LLM → TTS oraz kontekst z kamery;
- mechanika prototypowej szczęki: otwieranie, zamykanie, praca z serwym i ocena ograniczeń konstrukcyjnych;
- podstawy napędu capstan: BLDC, bębny, linka Dyneema i pomiar pozycji enkoderem;
- dalsza integracja mechaniki, sterowania, rozmowy lokalnej i rozpoznawania osób.

Najbliższe kroki

- sterowanie ODrive 3.6 z enkoderem AMT 102-V i zapis parametrów testów;
- wpływ naciągu linki, średnic bębnow, poślizgu, elastyczności i odczytu pozycji;
- opracowanie pierwszego modelu stawu nogi połączonego z bębniem wyjściowym stanowiska;
- reakcje TTS z ruchem szczęki i kontekstem osób widzianych przez kamerę;
- przygotowanie dokumentacji fotograficznej i krótkich nagrań do prezentacji projektu.



Podsumowanie

Projekt Zervidjanna osiągnął etap, w którym istnieją już kluczowe fundamenty: lokalna komunikacja AI z kontekstem wizyjnym, prototyp głowy oraz stanowisko capstan z pomiarem pozycji. Moduły te tworzą praktyczną bazę do dalszej integracji pełnego robota krocącego inspirowanego naturą.