— Bakalárska práca · úvodná prezentácia

TalkToWalk

Vizuálna navigácia humanoidného robota podľa textových inštrukcií

Autor
Dávid Moniš

Školiteľ
RNDr. Andrej Lúčny, PhD.

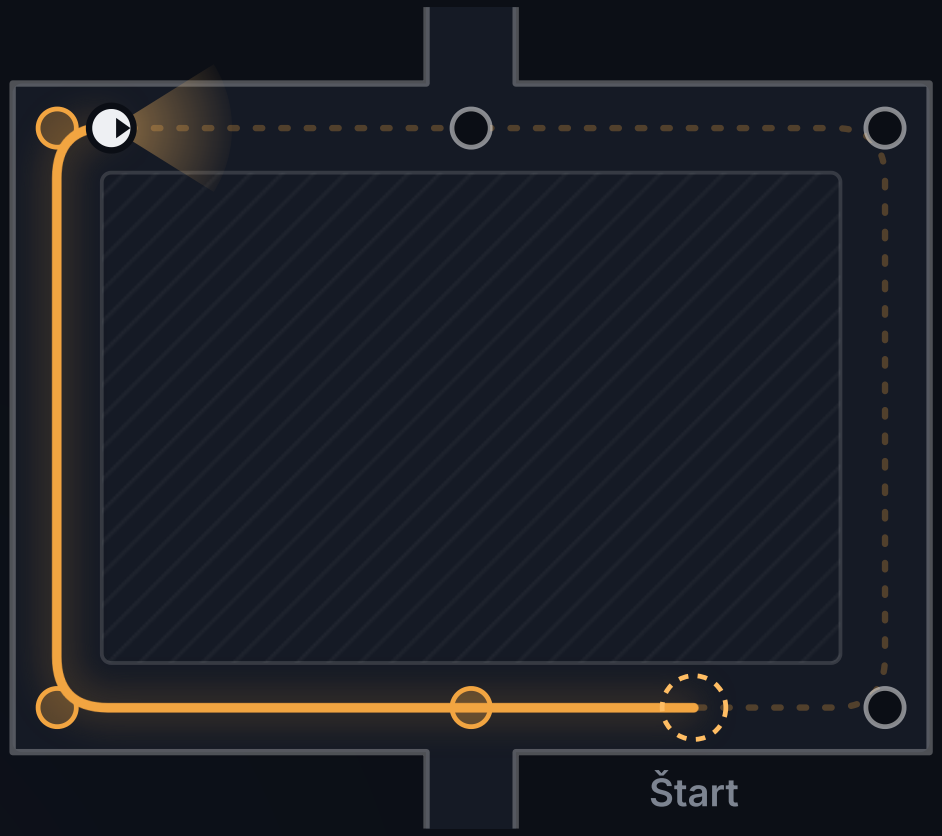
Pracovisko
FMFI UK, Katedra aplikovanej informatiky

Rok
2026/2027



— 01 · Úloha

Povedať robotovi cestu tak ako človeku



Ilustrácia cieľovej úlohy, nie záznam skutočného pokusu.

Inštrukcia pre robota

1 Choď rovno.

2 Na ďalšej križovatke doprava,

3 potom znova doprava,

4 rovno,

5 doprava,

6 doprava

7 a zastav tam, kde si začal.

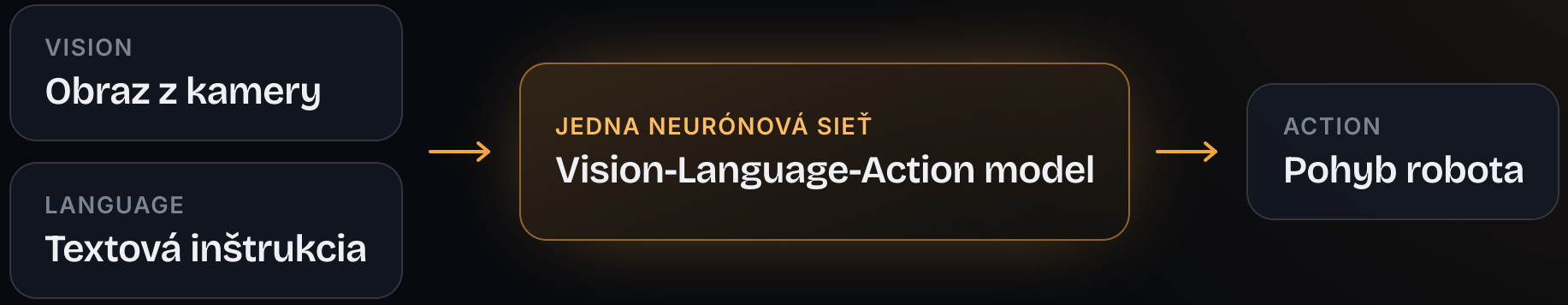
Vstup: len kamera a text

Skutočná budova

Bez kolízií

— 02 · Problematika

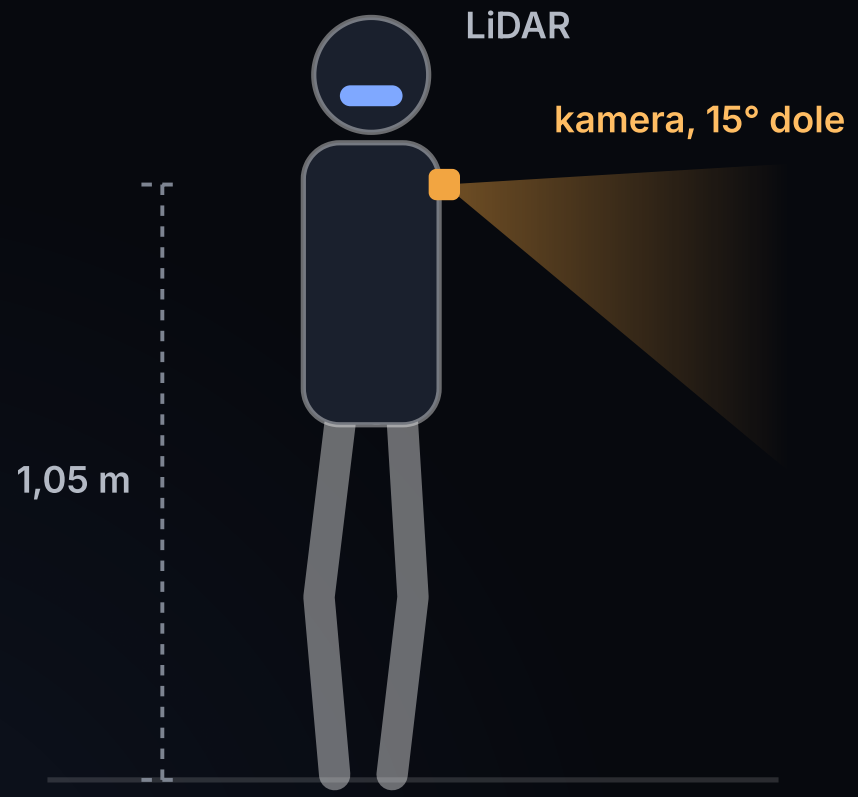
VLA model: vidí, číta a koná



- **GR00T N1 od NVIDIE** je otvorený VLA model pre humanoidy. Pomalá časť (VLM) pochopí scénu a úlohu, rýchla časť (diffusion transformer) generuje pohyby. [1]
- **Háčik:** takéto modely sa učili hlavne manipuláciu rukami. Ja potrebujem chôdzu po budove podľa textu.
- **Dá sa to:** ViNT ukázal, že predtrénovaný model sa dá lacno doladiť na nové prostredie aj na slovné pokyny. FocusNav ukázal nekolíznu navigáciu priamo na G1. [3] [2]

— 03 · Robot

Unitree G1: model rozhoduje, kontrolér kráča



29	1,32 m	RGB	LiDAR
kĺbov	výška	kamera na hrudi	a Jetson

```
move.py → Move(vx, vy, vyaw)
```

`vx`	dopredu a dozadu, najviac 0,4 m/s
`vy`	do strán, najviac 0,3 m/s
`vyaw`	otáčanie, najviac 0,6 rad/s

Rovnováhu a kroky rieši kontrolér od Unitree. Môj model hovorí len **kam a ako rýchlo**. Skript `move.py` je na robote overený.

— 04 · Návrh riešenia

Od obrazu k chôdzi



Rovnaké rozhranie v simulácii aj na skutočnom robote.

Model beží na notebooku s RTX 5090, robot je klient.

— 05 · Voľba modelu

Dve možnosti, medzi ktorými sa rozhodujem

A — GR00T N1.7

hotový VLA model od NVIDIE

- \+ Presne podľa zadania, existujú verzie pre G1.
- \+ Akčná časť je už natrénovaná na robotoch.
- – Na 24 GB sa dá doladiť len malá časť modelu.

B — Qwen3.5 2B + vlastná hlava

VLM, ku ktorému pridám malú výstupnú hlavu

- \+ Hlava vracia priamo **(vx, vy, vyaw)**, teda vstup pre move.py.
- \+ Tréning cez LoRA sa na 24 GB zmestí s rezervou.
- – Akčnú časť musím natrénovať od nuly.

	A · GR00T N1.7	B · Qwen3.5 2B
Parametre	3,14 mld.	2,21 mld.
Jeden krok	158 ms (6 Hz)	50 ms (20 Hz)
Pamäť pri behu	6,9 GB	4,2 GB
Doladenie	22,6 GB, len projektor	9,7 GB, LoRA

Namerané na notebooku s RTX 5090 (24 GB)

GR00T N1.7 má v sebe tiež VLM z rodiny Qwen. Začínam možnosťou B, lebo je jednoduchšia. GR00T ostáva v hre.

— 06 · Trasa a križovatky

Trasa je zoznam rozhodnutí na križovatkách

Čo je križovatka

Miesto, kde sa robot musí naozaj rozhodnúť:

1. dá sa pokračovať aspoň dvoma smermi, alebo
2. nedá sa ísť rovno.



Bočná chodba



Kríženie



Roh

Ako robot vie, kde v trase je

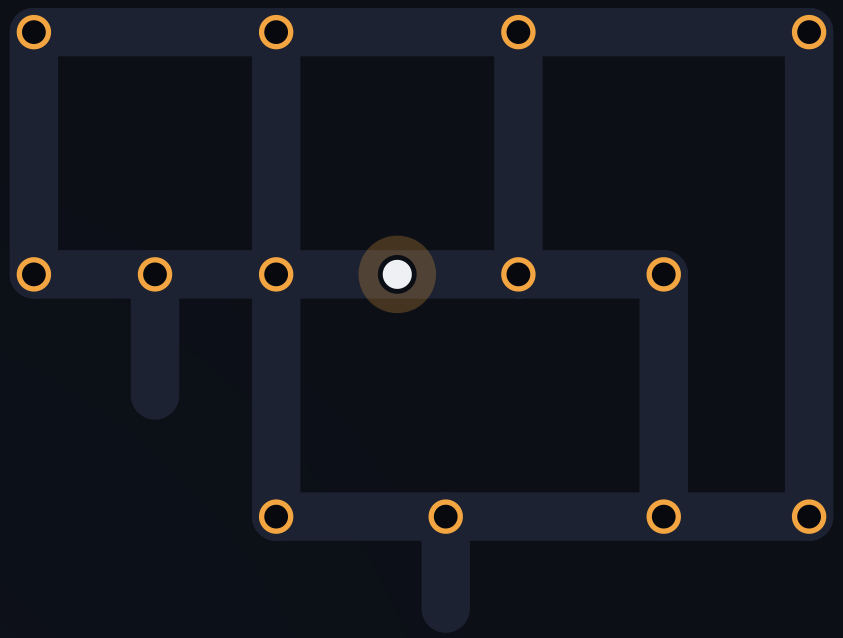
1. Choď rovno.
2. **Na ďalšej križovatke doprava.**
3. Potom znova doprava.
4. Zastav.

Model dostane vždy len **aktuálny úsek**. Keď hlava ohlási križovatku, robot príkaz vykoná a na vstup ide ďalší úsek.

Toto je môj návrh, nie hotová vec. Rozpoznávanie križovatiek je preto prvá vec, ktorú trénujem.

— 07 · Simulácia a dáta

Najprv simulácia, potom skutočná chodba



Ilustrácia. Generátor pozná polohu každej križovatky.

1 **Mapy v Isaac Sim**

Generátor náhodných poschodí, najväčšie má asi 100 × 100 m a vyše 50 chodieb. Ku každej mape vznikne aj súbor so správnymi odpoveďami.

2 **Dáta „kde som“**

Robot prejde mapu, ukladám snímky z kamery a značky. Model testujem na mape, ktorú nikdy nevidel.

3 **Skutočný robot**

Trasy prejdené pri ručnom ovládaní. Tam, kde sa model mýli, ho opravím a opravy idú späť do tréningu.

— 08 · Spätná väzba

LiDAR a človek ako poistka aj učiteľ



LiDAR

- Prekážka bližšie ako asi 0,5 m v smere pohybu: rýchlosť tým smerom sa vynuluje.
- Keď dáta z LiDARu neprídu, robot zastane.
- V simulácii pokusy s nárazom z dát vyradím.



Človek

- Robota viem kedykoľvek prevziať klávesnicou.
- Moje opravy ukazujú modelu, ako vyzerá správny prechod.
- Príkaz platí len 1 s. Keď model alebo sieť vypadne, robot sám zastane.

— 09 · Stav a plán

Čo je hotové a čo ma čaká

Hotovo

Rešerš a zadanie

Tri kľúčové články, téma schválená v októbri 2026.

Hotovo

Ovládanie robota

move.py, kamera, režimy robota.

Hotovo

Simulácia

G1 s kamerou v Isaac Sim, generátory máp.

Hotovo

Meranie modelov

Pamäť a rýchlosť oboch kandidátov.

Teraz

Prvý tréning

Model sa učí spoznať, kde je: križovatka a typ miesta.

Ďalej

Chôdza v simulácii

Robot ide podľa výstupu modelu, ochrana z LiDARu.

Ďalej

Skutočný robot

Zber dát, doladenie a test na chodbách fakulty.

Jar 2027

Text práce

Vyhodnotenie a odovzdanie.

— 10 · Úskalia

Kde to môže zlyhať

Simulácia nie je realita

Model zo simulácie nemusí fungovať na skutočnej kamere. Pomôcť majú reálne dáta a náhodné textúry a svetlo.

Pamäť grafickej karty

GR00T sa na 24 GB nedá doladiť celý. Preto zvažujem menší model.

Výpočet mimo robota

Jetson v robotovi má starý systém. Model beží na notebooku cez Wi-Fi, čo znamená oneskorenie.

Križovatiek je málo

Robot ide väčšinu času rovno chodbou. Dáta musím vyvážiť.

Stratená trasa

Keď model križovatku prehliadne alebo ju započíta dvakrát, pokazí sa celý zvyšok trasy.

Bezpečnosť

Humanoid môže spadnúť alebo do niekoho naraziť. Preto nízke rýchlosti a brzda z LiDARu.

— Zhrnutie

Ďakujem za pozornosť

Úloha: G1 prejde trasu opísanú textom, len podľa kamery a bez kolízií.

Ako: doladený model dáva (vx, vy, vyaw) a hlási križovatky, chôdzu rieši kontrolér robota.

Stav: robot, simulácia a merania sú pripravené, začínam trénovať.

Literatúra

[1] Björck, J. a kol.: GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. 2025.

[2] Zhang, Y. a kol.: FocusNav: Spatial Selective Attention with Waypoint Guidance for Humanoid Local Navigation. 2026.

[3] Shah, D. a kol.: ViNT: A Foundation Model for Visual Navigation. CoRL 2023.

Kód na GitHube

david-monis.com ↗

talk-to-walk.david-monis.com