MSc Project Plan

Title: *Cooperative Learning Between a Large LLM and a Small VLA*
Subtitle: *Evaluating How Linguistic Diversity Affects SmolVLA Manipulation Performance*
Student: Daniel Bryars
Date: December 2025

---

1. Project Aim

This project evaluates how **LLM-generated linguistic diversity**—including paraphrase quantity, linguistic style, persona, and reasoning-chain augmentation—affects the performance and generalisation of a **small Vision-Language-Action model (SmolVLA)** on manipulation tasks.

A large LLM (QWEN/GPT) is used to **rewrite, expand, paraphrase, stylise, and add reasoning chains** to instructions paired with demonstration episodes. The resulting datasets are used to train **multiple SmolVLA variants**, whose behaviour is compared in simulation and on a real SO100 robot.

This is a **controlled study** of how text influences robot policy learning. No prior work has directly measured this.

---

2. Novelty

This project provides the **first systematic ablation study** investigating:

- how **paraphrase quantity** affects VLA performance,
- how **linguistic style** (concise, verbose, symbolic, hierarchical, reasoning) influences results,
- whether **persona-based rewriting** changes learned behaviour, and
- whether **reasoning-chain augmentation** improves grasping or planning.

The study is:

- **controlled** (only the text changes; demos & visuals stay fixed),
- **reproducible**,
- **scalable** (LLM-generated language),

- **simulation-first** (allowing statistically meaningful evaluation),
- and grounded with a **real-robot validation** using the SO100.

---

3. Research Questions

1. How does **paraphrase quantity** (1, 5, 20, 100 per demonstration) affect SmolVLA's generalisation?
2. Do different **linguistic styles** (concise, verbose, symbolic, hierarchical, reasoning-based) change policy performance?
3. Does **persona-based rewriting** (terse, verbose, safety-focused, technical, optimistic) bias the learned behaviour?
4. Does adding **reasoning chains** (step-by-step or causal reasoning) improve manipulation success?
5. Do these effects seen in simulation transfer to the **real SO100 robot**?

---

4. System Architecture

pgsql
Copy code

```
┌──────────────────────────────┐
│ Real SO100 Teleoperation │
│ • Human-controlled demos │
│ • RGB camera data / actions│
└──────────────┬───────────────┘
│
▼
┌────────────────────────────┐
│ MuJoCo LeRobot-Gym │
│ (github.com/DanielBryars/ │
│ lerobot-gym) │
│ • SO100 simulation env │
│ • Teleop + auto-reset │
│ • Calibration debugging │
└──────────────┬──────────────┘
│
▼
┌──────────────────────────────────────┐
```

```
│ Base Dataset: │
│ • Images (real + sim) │
│ • Actions (joint deltas / EE pose) │
│ • Single task description per demo │

│
▼

│ LLM Linguistic Augmentation │
│ • Paraphrase counts: 1,5,20,100 │
│ • Styles: concise, verbose, │
│ symbolic, hierarchical, │
│ reasoning-based │
│ • Persona-based rewriting │
│ • Reasoning-chain augmentation │

│

▼ ▼ ▼

│ Model Variant │ │ Model Variant │ │ Model Variant │
│ (1 paraphrase) │ │ (20 hierarchical) │ │ (100 mixed styles) │

│ │ │
▼ ▼ ▼

│ Train SmolVLA │ │ Train SmolVLA │ │ Train SmolVLA │
│ (LeRobot) │ │ │ │ │

│ │ │
▼ ▼ ▼

│ Evaluation Pipeline (Simulation + Real) │
│ • Simulation: 200 randomised trials per model │
│ • Real-world SO100: external validity tests │
```

│ • Metrics: success rate, robustness, error types │
└───
──────┘

markdown
Copy code

Simulation provides the **main** experimental dataset.
The real SO100 robot provides a **small but meaningful external validity check**.

---

5. Experimental Design

5.1 Paraphrase Count Conditions

For each demonstration:

- **1** (baseline)
- **5**
- **20**
- **100**

5.2 Linguistic Style Conditions

- **Concise**
- **Verbose/narrative**
- **Symbolic** (e.g., "MOVE A → B; VERIFY STATE = OK")
- **Hierarchical / Stepwise**
- **Reasoning-Based** ("I should pick up the block because…")

5.3 Persona-Based Conditions

LLM rewrites instructions using personas such as:

- terse engineer
- overly verbose assistant
- safety-focused
- optimistic motivator
- highly procedural technician

5.4 Reasoning-Chain Augmentation

Chain-of-thought–like expansions:

- step-by-step plans
- causal reasoning ("to avoid knocking the other block…")
- decomposed subgoals

---

6. Experiment Matrix

A simplified version (full matrix adjustable depending on time):

Count	Style	Persona	Reasoning	Model ID
1	baseline	none	no	M1
5	concise	none	no	M2
20	hierarchical	technician	no	M3
20	reasoning-based	none	yes	M4
100	mixed styles	mixed	mixed	M5

These models differ **only in language**; visuals/actions remain identical.

---

7. Evaluation

Simulation (Primary)

- MuJoCo SO100 gym
- 200 trials per model
- Randomised object placement
- Metrics:
    - success rate
    - robustness to unseen phrasing
    - wrong-object / wrong-goal errors
    - trajectory stability

Real SO100 Robot (External Validity)

- Test 1–2 of the strongest/weakest conditions
- 20–30 trials

- Confirm whether trends from simulation hold in real hardware

---

8. Task List

- Finalise calibration in MuJoCo gym
- Collect clean baseline demonstrations
- Generate paraphrases across counts, styles, personas, and reasoning modes
- Assemble all augmented datasets
- Train each SmolVLA variant
- Evaluate all models in simulation
- Conduct real-world validation
- Analyse performance differences and error patterns
- Write the dissertation (methods, results, discussion, limitations, future work)

---

9. Summary

This MSc project delivers a **rigorous, controlled, and novel** investigation into how language affects robot manipulation policies.

I will:

1. Collect demonstration data
2. Generate broad linguistic variation with a large LLM
3. Train multiple SmolVLA models
4. Evaluate them across simulation and real-world tasks
5. Identify which linguistic properties most improve performance

The study advances understanding of **how LLM-generated language shapes the behaviour of small VLA models**, providing actionable insights for future robot learning systems.

---

10. Stretch Goals

RGB-D Sensing

- Add a RealSense-style RGB-D camera
- Compare RGB-only vs RGB-D training

Reinforcement Learning Integration

- Use PPO/SAC in MuJoCo to fine-tune the policy
- Explore IL+RL hybrid training
- Learn recovery behaviours

Additional Manipulation Tasks

- Occlusion ("digging") tasks
- Multi-object tasks
- Shape/colour variation

Multi-Camera Fusion

- Wrist + overhead + side view experiments

Representation Experiments

- Joint deltas vs end-effector pose (FK/IK)
- IK remapping at inference

Generalisation & Adversarial Tests

- Distractors, lighting changes, clutter
- Zero-shot language stress-tests

---

11. Repository References

- **SmolVLA / LeRobot:** https://github.com/huggingface/lerobot
- **SO100 MuJoCo Gym:** https://github.com/DanielBryars/lerobot-gym
- **QWEN Models:** https://huggingface.co/Qwen