■ Micro-LLM Mapping Recipe (Reusable Across Domains)

Goal: Build a deterministic, domain-specific mapper (prompt → class) that plugs into NGF rails (Warp → Detect → Denoise) and achieves high accuracy with controllable abstention.

1. Collect & Label

- **Prompts JSONL** — natural-language prompts, one per line.
- **Labels CSV** — gold labels aligned with prompts.
- Target at least **1k samples / 8+ classes** for balance.

2. Train Initial Mapper

Run:

```
python3 milestones/defi_milestone7.py \
  --mapper_path .artifacts/domain_mapper_embed.joblib \
  --prompts_jsonl data/prompts.jsonl \
  --labels_csv data/labels.csv
```

Produces an embed/joblib mapper model. Small set (≈30) suffices for sanity checks; scale to ≈5k for production.

3. Sweep Thresholds

Run:

```
python3 milestones/defi_milestone8.py \
  --mapper_path .artifacts/domain_mapper_embed.joblib \
  --prompts_jsonl data/prompts.jsonl \
  --labels_csv data/labels.csv \
  --thresholds "0.2,0.25,0.3,0.35,0.4" \
  --per_class_thresholds fixtures/m8_per_class_thresholds.json \
  --rows_csv .artifacts/m8_rows.csv \
  --out_summary .artifacts/m8_sum.json \
  --out_csv .artifacts/m8_metrics.csv
```

Runs evaluation across thresholds and produces metrics, rows, and summary artifacts.

4. Inspect & Choose

Check metrics and chosen threshold:

```
column -s, -t < .artifacts/m8_metrics.csv | head -15
jq '.chosen' .artifacts/m8_sum.json
```

Look for high accuracy with acceptable abstain rate. Adjust per-class thresholds if needed.

5. Verify Class Balance

Compute per-class accuracy:

```
awk -F, 'NR>1{g[$2]++; if($2==$3) c[$2]++} \
END{for(k in g) printf "%-15s %5d acc=%.4f\n", k, g[k], c[k]/g[k]}' \
.artifacts/m8_rows.csv | sort
```

Ensures each class is represented and performance is uniform.

6. Lock in & Reuse

- Pin chosen thresholds in **per_class_thresholds.json**.
- Commit **rows.csv**, **metrics.csv**, and **summary.json** as reproducible artifacts.
- Mapper is now domain-ready and can serve as input to higher-tier rails (Stage-11).

■ This recipe is **domain-agnostic** — only your prompt set and class labels change. The rails, scripts, and workflow stay identical.