(a) Type-wise accuracy (baseline vs v3/v4)

- Baseline
- v3
- v4

Algebraic n=575
Numerical n=144
Conceptual n=50

Accuracy (%)

(b) Net gain / loss by type

Algebraic +0.87 / -0.17
Numerical +0.69 / +3.47
Conceptual +0.00 / -14.00

Δ Accuracy (pp)

(c) v4 gating mechanism

MCQ question
+ options

Type classifier
(Conceptual / Numerical / Algebraic)

Selective gate
(if eligible → SymPy; else → Zero-shot)

Answer extraction
& correctness
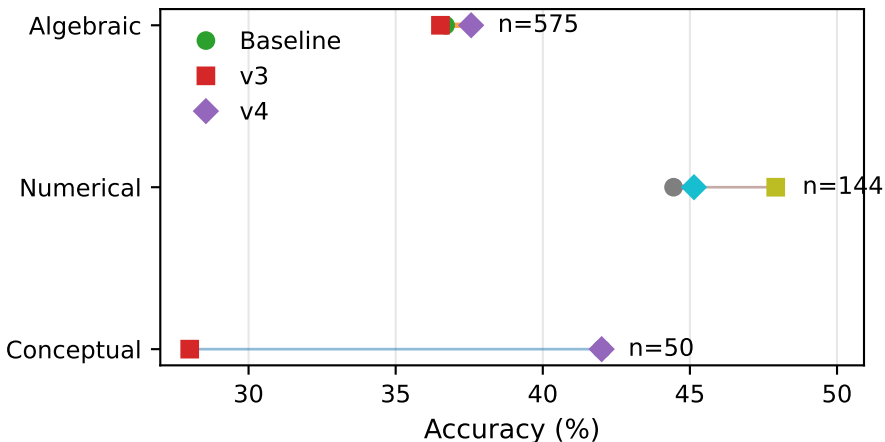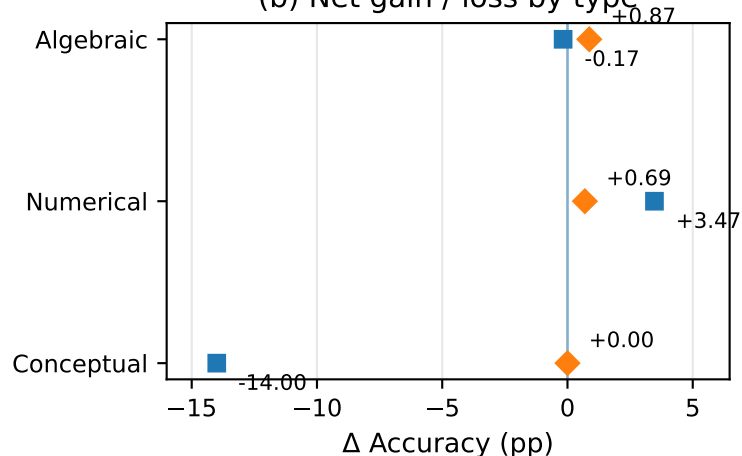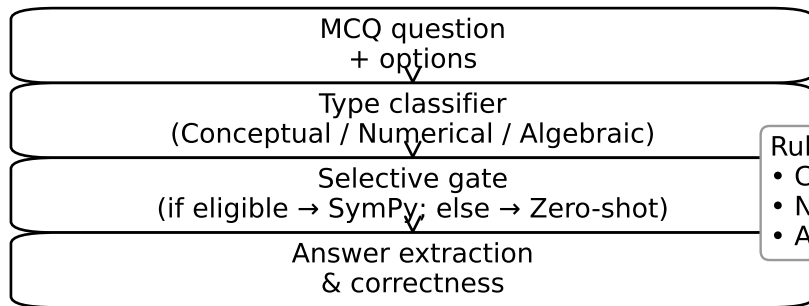
Rules (v4):
- Conceptual → Zero-shot
- Numerical → SymPy
- Algebraic → SymPy in {Math, QC, QChem, QM}