



Figure 1: Average number of calls to each agent among solved problems. Op-tiMUS only requires one call per agent on the simple problems of NL4OPT. On the more complex datasets, it relies more heavily on the programmer to fix errors identified by the evaluator, but rarely improves by fixing formulation errors.