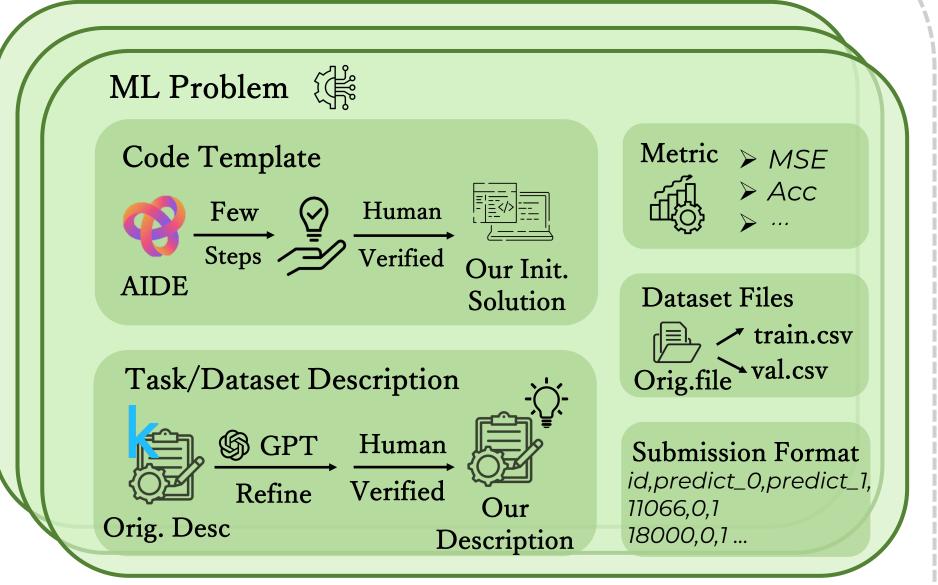
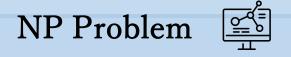
OPT-BENCH





Validation Script



Human

Manual Generate



Refine



validation.py

Description



Generated ₩ By GPT and Human Revised

Dataset File



Human

Assist



gen question script



Sub. Format

The answer should be an ordered sequence of city IDs ... Example Input: ... Example Output: ...

OPT-Agent





NP Additional Files 🗐 question.json

Draft



By N Steps



Solution



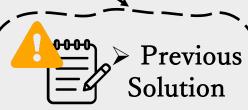
Full Solution Solution Method

**** ML: ```python import.. NP: Method: first we... Answer: [1, 2, 4...



Current Metric

ML: MSE: 1.23 ... NP: color_num: 20 path length: 50 ...



- Fail to run properly
- Cannot extract answer

Error Analysis by LLM/Rule

Rule: path is not cycle: start and end not ... LLM: This error may ...

Think & Optimization





ML: MSE: 1.05; Acc: 78% ...

NP: color_num: 14; path length: 60