

Data Generation

A base instruction

Let's think step by step.

Generate

More instructions

Solve the problem step by step and provide the final answer.

Please solve the problem in a structured manner, outlining each step clearly.

⋮

Please solve the problem in detail, explaining each step thoroughly.

a

Extract
(Ignorability,
Consistency)

Proxy features

T_1

Directness

T_2

Clarity

T_3

Conciseness

⋮

T_n

Specification
of Detail Level

n

Generate
(Positivity)

Counterfactual instructions

Solve the problem all at once and provide the final answer directly.

Can you please walk me through the solution to the problem in detail?

⋮

Provide an overview of the problem without going into detail.

$n(a+1)$

Labeled instructions

Label

Instruction	T_1	T_2	...	T_n
Solve the problem step ...	10	9	...	8
⋮	⋮	⋮	...	⋮
Please solve the problem ...	4	3	...	2

$n(a+1)$
 $+(a+1)$

Cartesian product
with questions

High-quality and low-quantity sample data

Instruction	T_1	T_2	...	T_n	Question (O)	Y
Solve the problem step ...	10	9	...	8	Earl will have \$90...	1
⋮	⋮	⋮	...	⋮	⋮	⋮
Please solve the problem ...	4	3	...	2	The proportion of ...	0

$b(n+1)(a+1)$

Aggregation/generalization

Causal Effect Estimation

T-Learner as the demonstration

LLMs



Code
Interpreter



Generated
data



Enhanced Instructions Generation

T_1

T_2

⋮

T_5

⋮

T_n

$ATE_1(-)$

$ATE_2(+)$

⋮

$ATE_n(+)$

Y

Generate

Instructions with optimized $ATE_{overall}$

Ensure clarity and maintain a positive, engaging tone. Proceed methodically, articulating each thought step by step with clear, precise language.

Class

T_1

Directness

T_2

Clarity

⋮

T_n

Specification
of Detail Level

Generate

OR Module

Enhanced instruction

Please provide a clear, step-by-step solution to the following problem.....or concluding remarks