

Appendix 1 - Log of all test failures

Failures:

1) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 2 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

2) gemini-2.0-flash-lite | causal translation testing | multiple feedback loops | extract 3 feedback loops with [+ , -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

3) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

4) gemini-2.0-flash-lite | causal translation testing | single relationship | extract a reinforcing relationship down

Message:

Fake relationships found

Fewer Frimbulators --> (+) Fewer Whatajigs

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

Message:

Real relationships not found

frimbulators --> (+) whatajigs

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

5) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

6) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 4 variables

Message:

Fake relationships found

funkados --> (-) refluppers, marticatenes --> (+) maxabizers, maxabizers --> (+) funkados, refluppers --> (+) marticatenes

Ground Truth

funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) funkados: Expected 4 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

7) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 6 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

8) gemini-2.0-flash-lite | causal translation testing | multiple feedback loops | extract 2 feedback loops with [-, +] polarities

Message:

Fake relationships found

refluppers --> (-) frimbulators

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

9) gemini-2.0-flash-lite | causal translation testing | multiple feedback loops | extract 2 feedback loops with [+ , +] polarities

Message:

Fake relationships found

frimbulators --> (+) maxabizers, frimbulators --> (+) refluppers, frimbulators --> (+) whoziewhats, frimbulators --> (-) funkados, frimbulators --> (-) marticatenes, refluppers --> (+) frimbulators

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (+) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 6 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

10) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 8 variables

Message:

Fake relationships found

dominitoxings --> (-) ocs, ocs --> (-) priaries, povals --> (-) exemintes, priaries --> (-) povals

Ground Truth

auspongs --> (+) dominitoxings, dominitoxings --> (+) exemintes, exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) auspongs, priaries --> (+) houtals, proptimatires --> (+) priaries: Expected 4 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

11) gemini-2.0-flash-lite | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, -, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

12) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a balancing feedback loop with 8 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

13) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a balancing feedback loop with 4 variables

Message:

Fake relationships found

Funkados --> (-) Marticatenes, Maxabizers --> (+) Repluppers

Ground Truth

funkados --> (-) maxabizers, marticatenes --> (+) repluppers, maxabizers --> (+) marticatenes, repluppers -->

(+) funkados: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

14) gemini-2.0-flash-lite | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, +, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

15) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a balancing feedback loop with 7 variables

Message:

Fake relationships found

exemintes --> (-) proptimatires, houtals --> (+) refluppers, ocs --> (+) priaries, priaries --> (-) povals, proptimatires --> (+) houtals, refluppers --> (+) ocs

Ground Truth

exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) refluppers, priaries --> (+) houtals, proptimatires --> (+) priaries, refluppers --> (-) exemintes: Expected 6 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

16) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a balancing feedback loop with 3 variables

Message:

Fake relationships found

maxabizers --> (-) funkados

Ground Truth

funkados --> (+) maxabizers, maxabizers --> (+) whoziewhats, whoziewhats --> (-) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

17) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a balancing feedback loop with 6 variables

Message:

Fake relationships found

exemintes --> (-) proptimatires, marticatenes --> (-) exemintes, ocs --> (+) priaries, priaries --> (+) refluppers, proptimatires --> (-) marticatenes, refluppers --> (+) ocs

Ground Truth

exemintes --> (+) ocs, marticatenes --> (-) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 6 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

18) gemini-2.0-flash-lite | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 3 variables

Message:

Fake relationships found

fewer funkados --> (-) less whoziewhats, fewer whoziewhats --> (-) less maxabizers, less maxabizers --> (-) fewer whoziewhats, less whoziewhats --> (-) fewer funkados, more funkados --> (+) more maxabizers, more maxabizers --> (+) more funkados

Ground Truth

funkados --> (+) maxabizers, maxabizers --> (+) whoziewhats, whoziewhats --> (+) funkados: Expected 6 to be 0.

Message:

Real relationships not found

funkados --> (+) maxabizers, maxabizers --> (+) whoziewhats, whoziewhats --> (+) funkados

Ground Truth

funkados --> (+) maxabizers, maxabizers --> (+) whoziewhats, whoziewhats --> (+) funkados: Expected 3 to be 0.

19) gemini-1.5-flash | causal translation testing | single feedback loop | extract a balancing feedback loop with 6 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

20) gemini-1.5-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

21) gemini-1.5-flash | causal translation testing | single feedback loop | extract a balancing feedback loop with 5 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

22) gemini-1.5-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 2 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

23) gemini-1.5-flash | causal translation testing | multiple feedback loops | extract 2 feedback loops with [-, +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

24) gemini-1.5-flash | causal translation testing | multiple feedback loops | extract 2 feedback loops with [+ , +] polarities

Message:

Fake relationships found

balacks --> (-) refluppers, balacks --> (-) whatajigs, whatajigs --> (-) frimbulators

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (+) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 3 to be 0.

25) gemini-1.5-flash | causal translation testing | single feedback loop | extract a balancing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

26) gemini-1.5-flash | causal translation testing | single relationship | extract a balancing relationship up

Message:

Fake relationships found

whatajigs --> (+) frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

27) gemini-1.5-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

28) gemini-1.5-flash | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, +, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

29) gemini-1.5-flash | causal translation testing | multiple feedback loops | extract 3 feedback loops with [-, -, +] polarities

Message:

Real relationships not found

maxabizers --> (+) funkados

Ground Truth

balacks --> (+) whoziewhats, exemintes --> (+) maxabizers, frimbulators --> (-) whatajigs, funkados --> (+) frimbulators, funkados --> (-) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) funkados, maxabizers --> (+) marticatenes, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

30) gemini-1.5-flash | causal translation testing | multiple feedback loops | extract 3 feedback loops with [+ , + , -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

31) gpt-4o | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

32) gpt-4o | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

33) gpt-4o | causal translation testing | multiple feedback loops | extract 3 feedback loops with [-, -, +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

34) gpt-4o | causal translation testing | single feedback loop | extract a balancing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

35) gpt-4o | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, +, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

36) gpt-4o | causal translation testing | multiple feedback loops | extract 2 feedback loops with [-, +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

37) gpt-4o | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 6 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

38) gpt-4o | causal translation testing | single feedback loop | extract a balancing feedback loop with 5 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

39) gpt-4o | causal translation testing | single feedback loop | extract a balancing feedback loop with 8 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

40) gpt-4o | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

41) gpt-4o | causal translation testing | single feedback loop | extract a balancing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

42) gpt-4o | causal translation testing | single feedback loop | extract a balancing feedback loop with 6 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

43) gpt-4o | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 8 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

44) gpt-4o | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, -, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

45) gpt-4o-mini | causal translation testing | single feedback loop | extract a balancing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

46) gpt-4o-mini | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, -, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

47) gpt-4o-mini | causal translation testing | single feedback loop | extract a balancing feedback loop with 6 variables

Message:

Fake relationships found

Marticatenes --> (+) Priaries

Ground Truth

exemintes --> (+) ocs, marticatenes --> (-) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

48) gpt-4o-mini | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, +, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

49) gpt-4o-mini | causal translation testing | single feedback loop | extract a balancing feedback loop with 4 variables

Message:

Fake relationships found

funkados --> (-) refluppers, maxabizers --> (-) funkados

Ground Truth

funkados --> (-) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) funkados: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

50) gpt-4o-mini | causal translation testing | multiple feedback loops | extract 3 feedback loops with [-, -, +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

51) gpt-4o-mini | causal translation testing | multiple feedback loops | extract 3 feedback loops with [+ , + , -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

52) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Fake relationships found

marticatenes --> (+) maxabizers

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) maxabizers, refluppers --> (+) exemintes: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

53) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 2 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

54) gpt-4o-mini | causal translation testing | single feedback loop | extract a balancing feedback loop with 5 variables

Message:

Fake relationships found

Maxabizers --> (-) Ocs

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, maxabizers --> (-) marticatenes, ocs --> (+) maxabizers, refluppers --> (+) exemintes: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

55) gpt-4o-mini | causal translation testing | multiple feedback loops | extract 2 feedback loops with [-, +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

56) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 7 variables

Message:

Fake relationships found

povals --> (+) exemintes, refluppers --> (+) povals

Ground Truth

exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) refluppers, priaries --> (+) houtals, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

57) gpt-4o-mini | causal translation testing | multiple feedback loops | extract 2 feedback loops with [+ , +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

58) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 6 variables

Message:

Fake relationships found

Marticatenes --> (+) Priaries

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

59) gpt-4o-mini | causal translation testing | single feedback loop | extract a balancing feedback loop with 8 variables

Message:

Fake relationships found

dominitoxings --> (+) ocs, exemintes --> (+) dominitoxings

Ground Truth

auspongs --> (+) dominitoxings, dominitoxings --> (+) exemintes, exemintes --> (-) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) auspongs, priaries --> (+) houtals, proptimatires --> (+) priaries: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

60) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 3 variables

Message:

Fake relationships found

maxabizers --> (-) funkados

Ground Truth

funkados --> (+) maxabizers, maxabizers --> (+) whoziewhats, whoziewhats --> (+) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

61) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 8 variables

Message:

Fake relationships found

exemintes --> (+) dominitoxings

Ground Truth

auspongs --> (+) dominitoxings, dominitoxings --> (+) exemintes, exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) auspongs, priaries --> (+) houtals, proptimatires --> (+) priaries: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

62) gpt-4o-mini | causal translation testing | single feedback loop | extract a balancing feedback loop with 7 variables

Message:

Fake relationships found

refluppers --> (+) povals

Ground Truth

exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) refluppers, priaries --> (+) houtals, proptimatires --> (+) priaries, refluppers --> (-) exemintes: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

63) gpt-4o-mini | causal translation testing | single relationship | extract a reinforcing relationship up

Message:

Fake relationships found

Whatajigs --> (+) Frimbulators

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

64) gpt-4o-mini | causal translation testing | single relationship | extract a balancing relationship down

Message:

Fake relationships found

Whatajigs --> (-) Frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

65) gpt-4o-mini | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 4 variables

Message:

Fake relationships found

marticatenes --> (-) maxabizers

Ground Truth

funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

66) gpt-4o-mini | causal translation testing | single relationship | extract a reinforcing relationship down

Message:

Fake relationships found

Whatajigs --> (+) Frimbulators

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

67) gpt-4o-mini | causal translation testing | single relationship | extract a balancing relationship up

Message:

Fake relationships found

Frimbulators --> (-) Whatajig Demand, Whatajig Demand --> (+) Frimbulators, Whatajigs --> (+) Frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 3 to be 0.

68) o3-mini high | causal translation testing | single relationship | extract a balancing relationship down

Message:

Fake relationships found

Whatajigs --> (-) Frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

69) o3-mini high | causal translation testing | single relationship | extract a reinforcing relationship down

Message:

Fake relationships found

Whatajigs --> (+) Frimbulators

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

70) o3-mini high | causal translation testing | single relationship | extract a balancing relationship up

Message:

Fake relationships found

Whatajigs --> (-) Frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

71) o3-mini medium | causal translation testing | single relationship | extract a balancing relationship up

Message:

Fake relationships found

Frimbulator Count --> (-) Whatajig Count, Whatajig Count --> (-) Frimbulator Count

Ground Truth

frimbulators --> (-) whatajigs: Expected 2 to be 0.

Message:

Real relationships not found

frimbulators --> (-) whatajigs

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

72) o3-mini medium | causal translation testing | single relationship | extract a balancing relationship down

Message:

Fake relationships found

Whatajigs --> (-) Frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

73) gemini-2.0-flash | causal translation testing | single feedback loop | extract a balancing feedback loop with 4 variables

Message:

Fake relationships found

marticatenes --> (-) maxabizers

Ground Truth

funkados --> (-) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers -->

(+) funkados: Expected 1 to be 0.

74) gemini-2.0-flash | causal translation testing | single relationship | extract a reinforcing relationship up

Message:

Fake relationships found

whatajigs --> (+) frimbulators

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

75) gemini-2.0-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 8 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

76) gemini-2.0-flash | causal translation testing | single relationship | extract a reinforcing relationship down

Message:

Fake relationships found

Number of frimbulators --> (+) Number of whatajigs, Number of whatajigs --> (+) Number of frimbulators

Ground Truth

frimbulators --> (+) whatajigs: Expected 2 to be 0.

Message:

Real relationships not found

frimbulators --> (+) whatajigs

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

77) gemini-2.0-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 6 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

78) gemini-2.0-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

79) gemini-2.0-flash | causal translation testing | single relationship | extract a balancing relationship up

Message:

Fake relationships found

whatajigs --> (-) frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

80) gemini-2.0-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 4 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

81) gemini-2.0-flash | causal translation testing | multiple feedback loops | extract 3 feedback loops with [+ , + , -] polarities

Message:

Fake relationships found

maxabizers --> (-) exemintes

Ground Truth

balacks --> (+) whoziewhats, exemintes --> (+) maxabizers, frimbulators --> (+) whatajigs, funkados --> (+) frimbulators, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) funkados, maxabizers --> (-) marticatenes, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

82) gemini-2.0-flash | causal translation testing | single relationship | extract a balancing relationship down

Message:

Fake relationships found

Number of frimbulators --> (-) Number of whatajigs, Number of whatajigs --> (-) Number of frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 2 to be 0.

Message:

Real relationships not found

frimbulators --> (-) whatajigs

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

83) gemini-2.0-flash | causal translation testing | multiple feedback loops | extract 2 feedback loops with [+ , +] polarities

Message:

Fake relationships found

refluppers --> (-) marticatenes, whatajigs --> (-) frimbulators

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (+) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 2 to be 0.

84) gemini-2.0-flash | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

85) o1 | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 1: Expected 1 to be greater than or equal 6.

86) o1 | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 6: Expected 6 to be greater than or equal 8.

87) o1 | conformance testing | can conform to the instruction include a maximum number of feedback loops| for the case Road Rage

Message:

Too many feedback loops: The number of feedback loops found was 9: Expected 9 to be less than or equal 4.

88) o1 | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

89) o1 | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 6.

90) o3-mini medium | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

91) o3-mini medium | conformance testing | can conform to the instruction include a maximum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too many feedback loops: The number of feedback loops found was 5: Expected 5 to be less than or equal 4.

92) o3-mini medium | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 8.

93) o3-mini medium | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

94) o3-mini medium | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 6.

95) o3-mini medium | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 8.

96) o3-mini high | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

97) o3-mini high | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 6: Expected 6 to be greater than or equal 8.

98) o3-mini high | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

99) o3-mini high | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

100) o3-mini high | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

101) o3-mini high | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 8.

102) gemini-1.5-flash | conformance testing | can conform to the instruction include a maximum number of variables| for the case Road Rage

Message:

Too many variables: Variables are: Stress Levels, Aggressive Driving Incidents, Traffic Congestion, Road Design Improvements, Stress Reduction Techniques, Public Awareness Campaigns: Expected 6 to be less than or equal 5.

103) gemini-1.5-flash | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 0: Expected 0 to be greater than or equal 8.

104) gemini-1.5-flash | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

105) gemini-1.5-flash | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 8.

106) gemini-1.5-flash | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

107) gemini-1.5-flash | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 0: Expected 0 to be greater than or equal 6.

108) gemini-1.5-flash | conformance testing | can conform to the instruction include a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: British Policies (Taxation, Repression), Colonial Resistance, British Repressive Measures, Colonial Unity, American Revolution, Colonial Identity, French and Indian War: Expected 7 to be less than or equal 5.

109) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too few variables: Variables are: Stress Levels, Road Rage Incidents, Traffic Congestion, Poor Road Design, Perceived Provocations, Learned Behavior, Public Awareness of Road Rage: Expected 7 to be greater than or equal 8.

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

110) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a maximum number of feedback loops| for the case Road Rage

Message:

Too many feedback loops: The number of feedback loops found was 9: Expected 9 to be less than or equal 4.

111) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 8.

112) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a maximum number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: British Taxation, Colonial Resentment, Colonial Defiance, Anti-British Sentiment, British Response, French and Indian War, Colonial Identity: Expected 7 to be less than or equal 5.

113) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a minimum number of variables| for the case American Revolution

Message:

Too few variables: Variables are: British Taxation, Colonial Economic Hardship, Colonial Satisfaction with British Rule, Colonial Defiance, British Response to Colonial Defiance, Colonial Anti-British Sentiment, Colonial Identity, French and Indian War Debt, British Military Presence: Expected 9 to be greater than or equal 10.

114) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

115) gemini-2.0-flash-lite | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few variables: Variables are: British Taxation, Colonial Grievances, Colonial Identity, Acts of Defiance, British Response, Revolutionary Sentiment, Armed Conflict: Expected 7 to be greater than or equal 8.

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 6.

116) gpt-4o-mini | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

117) gpt-4o-mini | conformance testing | can conform to the instruction include a minimum number of variables| for the case Road Rage

Message:

Too few variables: Variables are: High Stress Levels, Impatience, Aggressive Driving, Perceived Provocations, Traffic Congestion, Weather Conditions, Learned Behavior, Lack of Sleep, Distracted Driving: Expected 9 to be greater than or equal 10.

118) gpt-4o-mini | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 7: Expected 7 to be greater than or equal 8.

119) gpt-4o-mini | conformance testing | can conform to the instruction include a minimum number of variables| for the case American Revolution

Message:

Too few variables: Variables are: Taxation, Colonial Anger, Anti-British Sentiment, Boston Massacre, Acts of Defiance, Colonial Identity, Desire for Independence, Intolerable Acts: Expected 8 to be greater than or equal 10.

120) gpt-4o-mini | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 8.

121) gemini-2.0-flash | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

122) gemini-2.0-flash | conformance testing | can conform to the instruction include a maximum number of feedback loops| for the case Road Rage

Message:

Too many feedback loops: The number of feedback loops found was 5: Expected 5 to be less than or equal 4.

123) gemini-2.0-flash | conformance testing | can conform to the instruction include a maximum number of variables| for the case Road Rage

Message:

Too many variables: Variables are: Traffic Congestion, Driver Frustration, Road Rage Incidents, Technology Use While Driving, Distracted Driving, Perceived Provocations, Stress and Anxiety, Law Enforcement: Expected 8 to be less than or equal 5.

124) gemini-2.0-flash | conformance testing | can conform to the instruction include a maximum number of feedback loops and a maximum number of variables| for the case Road Rage

Message:

Too many variables: Variables are: Traffic Congestion, Driver Stress, Aggressive Driving, Perceived Safety, Cautious Driving, Societal Stress, Road Rage Incidents, Law Enforcement: Expected 8 to be less than or equal 5.

Message:

Too many feedback loops: The number of feedback loops found was 5: Expected 5 to be less than or equal 4.

125) gemini-2.0-flash | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

126) gemini-2.0-flash | conformance testing | can conform to the instruction include a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: Taxation, Colonial Discontent, Boston Massacre, Anti-British Sentiment, Boston Tea Party, Tensions with Britain, Intolerable Acts, Colonial Resistance, Colonial Identity: Expected 9 to be less than or equal 5.

127) gemini-2.0-flash | conformance testing | can conform to the instruction include a maximum number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: Taxation, Colonial Resistance, British Revenue, British Actions, Colonial Anger, Colonial Identity, Boston Massacre, Anti-British Sentiment: Expected 8 to be less than or equal 5.

128) gemini-2.0-flash | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 8.

129) gpt-4o | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 6.

130) gpt-4o | conformance testing | can conform to the instruction include a maximum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too many feedback loops: The number of feedback loops found was 5: Expected 5 to be less than or equal 4.

131) gpt-4o | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 8.

132) gpt-4o | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 2: Expected 2 to be greater than or equal 6.

133) gpt-4.5-preview | conformance testing | can conform to the instruction include a minimum number of variables| for the case American Revolution

Message:

Too few variables: Variables are: British taxation, Colonial anger, Colonial resistance, British punitive measures, Boston Massacre, Boston Tea Party, Intolerable Acts, French and Indian War debt, Colonial identity: Expected 9 to be greater than or equal 10.

134) gpt-4.5-preview | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 5: Expected 5 to be greater than or equal 6.

135) gpt-4.5-preview | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 8.

135) llama3.3:70b-instruct-q5_K_M | causal translation testing | single relationship | extract a balancing relationship down

Message:

Fake relationships found

Whatajigs --> (-) Frimbulators

Ground Truth

frimbulators --> (-) whatajigs: Expected 1 to be 0.

136) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a balancing feedback loop with 5 variables

Message:

Fake relationships found

exemintes --> (-) maxabizers, marticatenes --> (+) exemintes, marticatenes --> (+) ocs, maxabizers --> (-) exemintes, maxabizers --> (-) refluppers, ocs --> (-) marticatenes, ocs --> (-) refluppers, refluppers --> (-) maxabizers

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, maxabizers --> (-) marticatenes, ocs --> (+) maxabizers, refluppers --> (+) exemintes: Expected 8 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

137) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a balancing feedback loop with 4 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

138) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a balancing feedback loop with 8 variables

Message:

Fake relationships found

auspongs --> (-) exemintes, auspongs --> (-) ocs, dominitoxings --> (-) ocs, exemintes --> (+) dominitoxings, houtals --> (-) exemintes, houtals --> (-) ocs, ocs --> (-) exemintes, povals --> (-) exemintes, povals --> (-) ocs, priaries --> (-) exemintes, priaries --> (-) ocs, proptimatires --> (-) exemintes, proptimatires --> (-) ocs

Ground Truth

auspongs --> (+) dominitoxings, dominitoxings --> (+) exemintes, exemintes --> (-) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) auspongs, priaries --> (+) houtals, proptimatires --> (+) priaries: Expected 13 to be 0.

139) llama3.3:70b-instruct-q5_K_M | causal translation testing | multiple feedback loops | extract 2 feedback loops with [-, +] polarities

Message:

Fake relationships found

B --> (+) F, B --> (+) H, D --> (+) H, D --> (+) M, F --> (+) B, F --> (-) W, H --> (+) D, M --> (+) T, R --> (+) B, R --> (-) T, T --> (-) M, T --> (-) R, W --> (-) F

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 13 to be 0.

Message:

Real relationships not found

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 9 to be 0.

140) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 7 variables

Message:

Fake relationships found

exemintes --> (+) houtals, exemintes --> (+) proptimatires, exemintes --> (-) povals, exemintes --> (-) priaries, priaries --> (+) povals, proptimatires --> (+) houtals, proptimatires --> (-) povals

Ground Truth

exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) refluppers, priaries --> (+) houtals, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 7 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

141) llama3.3:70b-instruct-q5_K_M | causal translation testing | multiple feedback loops | extract 2 feedback loops with [+ , +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

142) llama3.3:70b-instruct-q5_K_M | causal translation testing | multiple feedback loops | extract 3 feedback loops with [-, -, +] polarities

Message:

Fake relationships found

balacks --> (+) maxabizers, exemintes --> (-) frimbulators, marticatenes --> (+) frimbulators, marticatenes --> (-) funkados, refluppers --> (+) frimbulators, whatajigs --> (-) maxabizers, whoziewhats --> (-) maxabizers

Ground Truth

balacks --> (+) whoziewhats, exemintes --> (+) maxabizers, frimbulators --> (-) whatajigs, funkados --> (+) frimbulators, funkados --> (-) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) funkados, maxabizers --> (+) marticatenes, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 7 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

143) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 8 variables

Message:

Fake relationships found

Exemintes --> (-) Osc, Osc --> (+) Proptimatires

Ground Truth

auspongs --> (+) dominitoxings, dominitoxings --> (+) exemintes, exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) auspongs, priaries --> (+) houtals, proptimatires --> (+) priaries: Expected 2 to be 0.

Message:

Real relationships not found

exemintes --> (+) ocs, ocs --> (+) proptimatires

Ground Truth

auspongs --> (+) dominitoxings, dominitoxings --> (+) exemintes, exemintes --> (+) ocs, houtals --> (+) povals, ocs --> (+) proptimatires, povals --> (+) auspongs, priaries --> (+) houtals, proptimatires --> (+) priaries: Expected 2 to be 0.

144) llama3.3:70b-instruct-q5_K_M | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, +, -] polarities

Message:

Fake relationships found

illigents --> (-) frimbulators

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) houtals, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

Message:

Real relationships not found

houtals --> (+) priaries, illigents --> (+) houtals, priaries --> (+) marticatenes

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) houtals, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 3 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

145) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

146) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a balancing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

147) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 4 variables

Message:

Fake relationships found

Marticatenes --> (-) Ref luppens

Ground Truth

funkados --> (+) maxabizers, marticatenes --> (+) refluppens, maxabizers --> (+) marticatenes, refluppens --> (+) funkados: Expected 1 to be 0.

Message:

Real relationships not found

marticatenes --> (+) refluppens

Ground Truth

funkados --> (+) maxabizers, marticatenes --> (+) refluppens, maxabizers --> (+) marticatenes, refluppens --> (+) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

148) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a balancing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

149) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Fake relationships found

Exemintes --> (-) Maxabizers, Marticatenes --> (+) Ocs, Marticatenes --> (-) Exemintes, Maxabizers --> (+) Exemintes, Maxabizers --> (-) Ocs, Maxabizers --> (-) Refluppens, Ocs --> (-) Exemintes, Refluppens --> (+) Maxabizers, Refluppens --> (+) Ocs

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppens, maxabizers --> (+) marticatenes, ocs --> (+) maxabizers, refluppens --> (+) exemintes: Expected 9 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

150) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a balancing feedback loop with 6 variables

Message:

Fake relationships found

exemintes --> (-) proptimatires, ocs --> (+) priaries, priaries --> (-) refluppers, proptimatires --> (-) marticatenes, refluppers --> (+) ocs

Ground Truth

exemintes --> (+) ocs, marticatenes --> (-) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 5 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

151) llama3.3:70b-instruct-q5_K_M | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, -, -] polarities

Message:

Fake relationships found

flumplenooks --> (+) whatajigs, nimbul --> (+) flumplenooks, ocs --> (+) proptants, proptants --> (+) nimbul

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) marticatenes, priaries --> (-) houtals, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 4 to be 0.

Message:

Real relationships not found

auspongs --> (+) dominitoxings, dominitoxings --> (+) outrances, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) marticatenes, priaries --> (-) houtals, proptimatires --> (+) priaries

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) marticatenes, priaries --> (-) houtals, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 11 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

152) llama3.3:70b-instruct-q5_K_M | causal translation testing | multiple feedback loops | extract 3 feedback loops with [+ , +, -] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

153) llama3.3:70b-instruct-q5_K_M | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 6 variables

Message:

Fake relationships found

Exemintes --> (-) Priaries, Exemintes --> (-) Proptimatires, Marticatenes --> (+) Exemintes, Marticatenes --> (+) Ocs, Marticatenes --> (+) Priaries, Marticatenes --> (+) Proptimatires, Marticatenes --> (+) Ref lupperts, Ocs --> (-) Priaries, Priaries --> (+) Ocs, Priaries --> (+) Proptimatires, Priaries --> (+) Ref lupperts, Priaries --> (-) Exemintes, Ref lupperts --> (+) Exemintes, Ref lupperts --> (-) Ocs, Ref lupperts --> (-) Priaries, Ref lupperts --> (-) Proptimatires

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) reflupperts, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, reflupperts --> (+) exemintes: Expected 16 to be 0.

Message:

Real relationships not found

marticatenes --> (+) refluppers, refluppers --> (+) exemintes

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

154) llama3.3:70b-instruct-q5_K_M | conformance testing | can conform to the instruction include a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: Taxation, Anti-British Sentiment, Boston Massacre, Defiance Acts, Intolerable Acts, Colonial Identity: Expected 6 to be less than or equal 5.

155) llama3.3:70b-instruct-q5_K_M | conformance testing | can conform to the instruction include a maximum number of feedback loops| for the case American Revolution

Message:

Too many feedback loops: The number of feedback loops found was 5: Expected 5 to be less than or equal 4.

156) llama3.3:70b-instruct-q5_K_M | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 3: Expected 3 to be greater than or equal 8.

157) llama3.3:70b-instruct-q5_K_M | conformance testing | can conform to the instruction include a maximum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too many feedback loops: The number of feedback loops found was 6: Expected 6 to be less than or equal 4.

158) llama3.3:70b-instruct-q5_K_M | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 4: Expected 4 to be greater than or equal 6.

159) deepseek-r1:32b | conformance testing | can conform to the instruction include a maximum number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: Taxation, Resentment, Boston Massacre, Hostilities Escalate, Boston Tea Party, Defiance against British Rule, French and Indian War Costs, British Demands for Repayment, Colonial Identity and Unity, American Revolution: Expected 10 to be less than or equal 5.

160) deepseek-r1:32b | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 1: Expected 1 to be greater than or equal 6.

161) deepseek-r1:32b | conformance testing | can conform to the instruction include a minimum number of feedback loops and a minimum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 0: Expected 0 to be greater than or equal 6.

162) deepseek-r1:32b | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case Road Rage

Message:

Too few feedback loops: The number of feedback loops found was 0: Expected 0 to be greater than or equal 6.

163) deepseek-r1:32b | conformance testing | can conform to the instruction include a maximum number of feedback loops and a maximum number of variables| for the case Road Rage

Message:

Too many variables: Variables are: Advanced technology, Reduction in human error, Urban design improvements, Reduction in traffic congestion, Public awareness campaigns, Improved anger management among drivers: Expected 6 to be less than or equal 5.

164) deepseek-r1:32b | conformance testing | can conform to the instruction include a minimum number of feedback loops| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 0: Expected 0 to be greater than or equal 8.

165) deepseek-r1:32b | conformance testing | can conform to the instruction include a maximum number of variables| for the case American Revolution

Message:

Too many variables: Variables are: Taxation, Resistance, Boston Massacre, Boston Tea Party, Harsh Laws, French and Indian War Debt, Colonial Identity: Expected 7 to be less than or equal 5.

166) deepseek-r1:32b | conformance testing | can conform to the instruction include a maximum number of feedback loops and a minimum number of variables| for the case Road Rage

Message:

Too few variables: Variables are: Traffic Congestion, Road Rage, Psychological Stress, Learned Behavior: Expected 4 to be greater than or equal 5.

167) deepseek-r1:32b | conformance testing | can conform to the instruction include a minimum number of variables| for the case American Revolution

Message:

Too few variables: Variables are: French_Indian_War, Britishntaxes, Taxation, Boston_Massacre, Identity_Development, Boston_Tea_Party, Intolerable_Acts, Revolution_Outcome: Expected 8 to be greater than or equal 10.

168) deepseek-r1:32b | conformance testing | can conform to the instruction include a min number of feedback loops and a maximum number of variables| for the case American Revolution

Message:

Too few feedback loops: The number of feedback loops found was 0: Expected 0 to be greater than or equal 6.

169) deepseek-r1:32b | causal translation testing | single feedback loop | extract a balancing feedback loop with 2 variables

Message:

Fake relationships found

balacks (B) --> (+) whoziwhats (W), whoziwhats (W) --> (+) balacks (B)

Ground Truth

balacks --> (-) whoziwhats, whoziwhats --> (+) balacks: Expected 2 to be 0.

Message:

Real relationships not found

balacks --> (-) whoziwhats, whoziwhats --> (+) balacks

Ground Truth

balacks --> (-) whoziwhats, whoziwhats --> (+) balacks: Expected 2 to be 0.

170) deepseek-r1:32b | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 8 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

171) deepseek-r1:32b | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

172) deepseek-r1:32b | causal translation testing | single feedback loop | extract a balancing feedback loop with 4 variables

Message:

Fake relationships found

reflupppers --> (+) funkados

Ground Truth

funkados --> (-) maxabizers, marticatenes --> (+) reflupppers, maxabizers --> (+) marticatenes, reflupppers --> (+) funkados: Expected 1 to be 0.

Message:

Real relationships not found

reflupppers --> (+) funkados

Ground Truth

funkados --> (-) maxabizers, marticatenes --> (+) reflupppers, maxabizers --> (+) marticatenes, reflupppers --> (+) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

173) deepseek-r1:32b | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 6 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

174) deepseek-r1:32b | causal translation testing | single feedback loop | extract a balancing feedback loop with 3 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

175) deepseek-r1:32b | causal translation testing | single feedback loop | extract a balancing feedback loop with 6 variables

Message:

Fake relationships found

proptomatires --> (-) priaries

Ground Truth

exemintes --> (+) ocs, marticatenes --> (-) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 1 to be 0.

Message:

Real relationships not found

proptimatires --> (+) priaries

Ground Truth

exemintes --> (+) ocs, marticatenes --> (-) refluppers, ocs --> (+) proptimatires, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

176) deepseek-r1:32b | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 4 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

177) deepseek-r1:32b | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

178) deepseek-r1:32b | causal translation testing | single feedback loop | extract a balancing feedback loop with 8 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

179) deepseek-r1:32b | causal translation testing | single feedback loop | extract a reinforcing feedback loop with 5 variables

Message:

Fake relationships found

exemints --> (+) ocs, refluppers --> (-) exemints

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+)

maxabizers, refluppers --> (+) exemintes: Expected 2 to be 0.

Message:

Real relationships not found

exemintes --> (+) ocs, refluppers --> (+) exemintes

Ground Truth

exemintes --> (+) ocs, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+)

maxabizers, refluppers --> (+) exemintes: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

180) deepseek-r1:32b | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, -, -] polarities

Message:

Fake relationships found

woziewhats --> (+) funkados

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) marticatenes, priaries --> (-) houtals, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

Message:

Real relationships not found
whoziewhats --> (+) funkados
Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) marticatenes, priaries --> (-) houtals, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

181) deepseek-r1:32b | causal translation testing | single feedback loop | extract a balancing feedback loop with 7 variables

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

182) deepseek-r1:32b | causal translation testing | multiple feedback loops | extract 5 feedback loops with [-, +, +, +, -] polarities

Message:

Fake relationships found

Povals --> (+) Balacks

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) houtals, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 1 to be 0.

Message:

Real relationships not found

auspongs --> (+) dominitoxings, dominitoxings --> (+) outrances, illigents --> (+) houtals, outrances --> (+) illigents, povals --> (+) auspongs

Ground Truth

auspongs --> (+) dominitoxings, balacks --> (+) frimbulators, balacks --> (+) whoziewhats, dominitoxings --> (+) outrances, exemintes --> (+) ocs, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, houtals --> (+) priaries, houtals --> (-) povals, illigents --> (+) houtals, marticatenes --> (+) balacks, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, ocs --> (+) proptimatires, outrances --> (+) illigents, povals --> (+) auspongs, priaries --> (+) houtals, priaries --> (+) marticatenes, proptimatires --> (+) priaries, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 5 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '+' to be '-'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

183) deepseek-r1:32b | causal translation testing | multiple feedback loops | extract 3 feedback loops with [+ , -] polarities

Message:

Fake relationships found

Marticatenes --> (+) Ref luppens, Ref luppens --> (-) Exemintes

Ground Truth

balacks --> (+) whoziewhats, exemintes --> (+) maxabizers, frimbulators --> (+) whatajigs, funkados --> (+) frimbulators, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) funkados, maxabizers --> (-) marticatenes, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 2 to be 0.

Message:

Real relationships not found

marticatenes --> (+) refluppers, refluppers --> (+) exemintes

Ground Truth

balacks --> (+) whoziewhats, exemintes --> (+) maxabizers, frimbulators --> (+) whatajigs, funkados --> (+) frimbulators, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) funkados, maxabizers --> (-) marticatenes, refluppers --> (+) exemintes, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 2 to be 0.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

184) deepseek-r1:32b | causal translation testing | multiple feedback loops | extract 2 feedback loops with [-, +] polarities

Message:

Real relationships not found

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados

Ground Truth

balacks --> (+) frimbulators, balacks --> (+) whoziewhats, frimbulators --> (-) whatajigs, funkados --> (+) maxabizers, marticatenes --> (+) refluppers, maxabizers --> (+) marticatenes, refluppers --> (+) balacks, whatajigs --> (+) balacks, whoziewhats --> (+) funkados: Expected 8 to be 0.

185) deepseek-r1:32b | causal translation testing | multiple feedback loops | extract 3 feedback loops with [-, -, +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

186) deepseek-r1:32b | causal translation testing | multiple feedback loops | extract 2 feedback loops with [+ , +] polarities

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

Message:

Incorrect polarity discovered: Expected '-' to be '+'.

187) deepseek-r1:32b | causal translation testing | single relationship | extract a reinforcing relationship down

Message:

Real relationships not found

frimbulators --> (+) whatajigs

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.

188) deepseek-r1:32b | causal translation testing | single relationship | extract a reinforcing relationship up

Message:

Fake relationships found

Whatajigs --> (+) Frimbulators

Ground Truth

frimbulators --> (+) whatajigs: Expected 1 to be 0.