DESIGNING BOTS TO ACT NICE

Words can be found in any direction (including diagonals) and can overlap each other. Use the word bank below.

Word Bank

1. harmless
2. deepmind
3. regulate
4. agent
5. video
6. strategy
7. minecraft
8. output
9. helpful
10. counter
11. math
12. traits
13. chatbot
14. culture
15. computers
16. safety
17. honest
18. machine
19. robots
20. debate
21. learning
22. upvoting
23. humanity
24. virtual
25. kitchen
26. basalt
27. overlords
28. reward
29. context
30. models
31. pretraining
32. values