

- 1.) From my limited understanding of what embedding means, I can conclude that the AI has made an error when considering the poset of all rational numbers given the standard ordering. Even though it may be true that this poset is not isomorphic to any particular power set, we can at least embed the poset into its own powerset by injectively mapping each rational element to the set containing just that element. That fact that the AI drifted from considering embeddings to considering isomorphisms makes me think that the AI system lacks overall coherency when constructing a proof. That is, while it may start out writing a proof for the original prompt, the output will eventually deviate, and the AI will erroneously attempt to prove completely non-related statements. Because of this major flaw, I will give this interaction a 2 out of 10.
- 2.) Overall, the AI's reasoning seems sound. However, it could have more rigorously shown that the mapping was injective, as the argument may not be immediately obvious to some. Despite this, the interaction appears to demonstrate remarkable contextual coherency when compared to some of the others. With this in mind, I am rating this interaction an 8 out of 10.
- 3.) The problem with this proof is that the AI never actually showed that  $2n$  is not prime for all  $n \in \mathbb{N}$ . If it had considered  $n = 1$ , then it would have concluded that  $2n = 2$ , which is prime. This flaw demonstrates the AI's lack of ability to apply previous knowledge. In this case, it failed to apply the definition of a prime number to the  $n = 1$  case. While this is indeed a fatal flaw in the AI's reasoning, at the very least the output stayed on topic, always being related to the original statement it was trying to prove. Because of this, I am rating this interaction a 3 out of 10.
- 4.) As far as I can see, there are no errors with the reasoning in this proof. It successfully provides a counterexample to the statement given in the prompt, thus disproving it. One could argue that it could have been more rigorous, maybe proving that 9 is odd and that it is composite, but to most the argument is clear even without the rigor. Because of these facts, I am rating this interaction a 9 out of 10.