RIFF:

You give the AI a problem that is meant to have two answers and it tries to find a third.

A person could give the AI information via text or maybe image.

Things like the car problem where you have to hit either a group of people or a single person.

It will first be like: hey are the people moving? I am going to project where they are going to be and see if there is any way to work with that information to come up with an answer that will result in no casualties.

I think the best base starting place will just be: try to come up with outside the box answers or think outside the obvious.

So maybe the AI will come up with the third possible answer: hitting EVERYONE instead of having to choose.

Or maybe the AI can be given a problem, and come up with a way to prevent the problem from happening?