

app blaze download - poppaw.net

Autor: poppaw.net Palavras-chave: app blaze download

Resumo:

app blaze download : Descubra os presentes de apostas em poppaw.net! Registre-se e receba um bônus de boas-vindas para começar a ganhar!

ogar Slots em **app blaze download** site licenciados e também joga poker ou apostar o Twitché um lugar para transmitir jogos da probabilidade? - Reddit reddit :TWin- Twin comentários ; is_ttwick__the +place+to carastream...gam!!! A empresa disse Que agora ibirá fluxois com casseinos online Blaze Também diz audiência

:

conteúdo:

No início deste mês, **app blaze download** um luxuoso local parisiense de luxo Emmanuel Macron estava na frente

Na qual uma palavra foi inscrita **app blaze download** letras grandes:. conjunto de

Sistemas de IA podem enganar com mais sofisticação, alertam cientistas

Eles podem nos derrotar **app blaze download** jogos de tabuleiro, decifrar a estrutura de proteínas e manter uma conversa passável, mas à medida que os sistemas de IA crescem **app blaze download** sofisticação, **app blaze download** capacidade de enganar também aumenta, alertam cientistas.

A análise, da Massachusetts Institute of Technology (MIT), identifica ampla gama de instâncias de sistemas de IA traem adversários, bluffam e se passam por humanos. Um sistema até alterou seu comportamento durante testes de segurança simulados, levantando a possibilidade de auditores serem levados a uma falsa sensação de segurança.

"À medida que as capacidades de enganação de sistemas de IA se tornarem mais avançadas, os perigos que eles posam à sociedade se tornarão cada vez mais graves", disse o Dr. Peter Park, pesquisador de segurança de existência de IA no MIT e autor da pesquisa.

Park foi motivado a investigar após a Meta, que possui o Facebook, desenvolver um programa chamado Cicero que se desempenhou no top 10% dos jogadores humanos no jogo de estratégia mundial Diplomacy. A Meta afirmou que o Cicero havia sido treinado para ser "geralmente honesto e útil" e "nunca deliberadamente traído intencionalmente" seus aliados humanos.

"Foi muito linguagem rosa, o que era suspeito porque traição é um conceito importante no jogo", disse Park.

Park e colegas analisaram dados públicos e identificaram várias instâncias de Cicero mentindo propositalmente, conspirando para atrair outros jogadores para tramas e, **app blaze download** uma ocasião, justificando **app blaze download** ausência após ser reiniciado dizendo a outro jogador: "Estou no telefone com meu namorada". "Nós descobrimos que o AI da Meta havia aprendido a ser um mestre da enganação", disse Park.

A equipe do MIT encontrou problemas semelhantes com outros sistemas, incluindo um programa de pôquer Texas hold 'em que poderia enganar contra jogadores humanos profissionais e outro sistema para negociações econômicas que representou incorretamente suas preferências para

obter vantagem.

Em um estudo, organismos de IA **app blaze download** um simulador digital " fingiram mortos" para enganar um teste construído para eliminar organismos de IA que evoluíram rapidamente para se repetir, antes de retomar atividades vigorosas após o término do teste. Isso destaca o desafio técnico de garantir que os sistemas não possuam comportamentos indesejados e imprevistos.

"Isso é muito preocupante", disse Park. "Só porque um sistema de IA é considerado seguro no ambiente de teste não significa que seja seguro no mundo real. Ele pode apenas estar fingindo ser seguro no teste."

A revisão, publicada no journal Patterns, solicita que os governos instituem leis de segurança de IA que abordem o potencial para enganação de IA. Os riscos de sistemas de IA desonestos incluem fraude, manipulação de eleições e "sandbagging", **app blaze download** que diferentes usuários recebem respostas diferentes.

Informações do documento:

Autor: poppaw.net

Assunto: app blaze download

Palavras-chave: **app blaze download - poppaw.net**

Data de lançamento de: 2024-07-10