

D



Notícias do Mundo



Divulgação

Castelfranchi afirma que máquinas já substituem humanos na tomada de decisão

ENTREVISTA

Máquinas ainda não compreendem intenção humana: algo a ser conquistado a longo prazo

O motivo que fez com que Cristiano Castelfranchi, um dos maiores especialistas europeus na área de inteligência artificial (IA), abandonasse seu trabalho no Conselho Nacional de Pesquisa italiano (CNR, na sigla em italiano) para passar alguns dias no Brasil é nobre: o nascimento de seu neto. Em meio às tarefas de avô, Castelfranchi falou sobre seu trabalho com agentes artificiais, pequenas máquinas independentes com capacidades limitadas e específicas, mas que resolvem problemas complexos graças à coordenação, comu-

nicação e distribuição de informação entre elas. Segundo ele, os agentes são frutos de uma nova direção dada aos estudos com IA nos últimos 30 anos, que busca modelar uma inteligência social em entidades artificiais. Mas tal formalização se depara com obstáculos muito mais complexos do que a capacidade de cálculo matemático ou a habilidade de jogar xadrez que as máquinas há tempos já possuem. Entretanto, uma vez vencidos esses obstáculos, teremos em nosso meio máquinas conscientes de si, de seus ob-

jetivos e dos objetivos de terceiros com quem interagem. Teremos até mesmo, máquinas capazes de experimentar sentimentos e emoções.

O que são agentes artificiais?

A idéia central dos agentes é que uma inteligência centralizada não é suficiente. Os primeiros 30 anos da inteligência artificial foram dedicados a construir um único cérebro racional, uma inteligência com todas as representações, decisões, planejamentos, todo o conhecimento. Mas isso não funciona. Surgiu então uma nova onda que diz que inteligência é algo cooperativo, distributivo. O que realmente importa para a resolução de problemas é uma certa quantidade de inteligências distribuídas na qual todas as informações, banco de dados locais e experiências sejam compartilhadas entre essas inteligências. Aí sim se resolve o problema. Dessa maneira, os agentes são pequenas entidades artificiais onde cada uma sabe fazer algo específico e simples. Através da cooperação entre os agentes é que emerge o comportamento inteligente deles.

Que questões precisam ser resolvidas para que tenhamos máquinas com inteligência social?

Para existir cooperação, em qualquer sociedade, é necessário que existam normas e que estas sejam respeitadas pelos sujeitos. Mas, onde estão as



normas na nossa mente? Como elas consolidam nosso comportamento ou nos obrigam a fazer algo? As entidades artificiais podem respeitar normas? Este é um trabalho difícil. Outra questão trata de como é possível modelar emoções, como inveja, simpatia e culpa em criaturas artificiais. Um último aspecto importante a ser trabalhado é muito conhecido pela psicologia e pela filosofia e se chama teoria da mente. Para uma boa cooperação social é necessário que um agente saiba o objetivo, a intenção do outro e não apenas faça o que lhe é pedido. Sem o conhecimento prévio de uma teoria da mente alheia a interação homem-máquina, ou até mesmo homem-homem se torna impossível.

Artefatos artificialmente inteligentes atuais possuem uma boa noção de teoria da mente?

Uma cooperação eficiente não é aquela em que um indivíduo faz exatamente o que o outro lhe pede. É preciso que ele entenda primordialmente o objetivo do outro, pois de outro modo a cooperação se torna estúpida. Por exemplo. Um homem pergunta ao outro. “Você tem fósforo?” e o outro lhe entrega um isqueiro. A cooperação humana funciona 90% desse jeito. A pessoa entendeu a intenção de acender um cigarro. Acontece que as máquinas atuais não conseguem compreender o objetivo do outro. O ideal mesmo seria que ela se

antecipasse e tomasse espontaneamente a iniciativa na cooperação com o ser humano. Como objetivo último disso tudo teríamos uma máquina possuidora de sentimentos, mas isso ainda é algo a longo prazo. A curto prazo o que se pode esperar das máquinas é que compreendam melhor a intenção humana.

As máquinas serão algum dia conscientes?

Criar uma máquina com consciência é uma afirmação mal posta. Pois consciência é um termo muito amplo e envolve uma série de processos fenomenológicos complexos e independentes. Eu, pessoalmente, gosto bastante da noção de consciência como estar ciente de si, o *self*. Não falo necessariamente de aspectos fenomenológicos ou de experiências subjetivas, mas de meta-representações, o modo como eu represento minha mente para mim mesmo. O grande desafio não é a criação de um artefato com inteligência explícita, simbólica, mas o fato das máquinas serem capazes de realizar cálculos complexos, mas não conseguirem encontrar e abrir uma porta para sair de uma sala.

Existe a possibilidade de termos guerras contra máquinas racionalmente mais sofisticadas que os humanos?

Sou um pouco cético quanto à idéia da possibilidade real da evolução

biológica das máquinas, da capacidade delas virem a se reproduzir, etc. Por outro lado, acredito que o perigo real e mais próximo seja a profecia *Unabomber*, que dizia que, no surgimento das máquinas artificialmente inteligentes, estas seriam totalmente dependentes das nossas decisões, por levarem a resultados melhores que os obtidos por decisões humanas. Eventualmente, chegaríamos a um momento que, de tantas decisões delegadas às máquinas, mesmo que tenhamos o poder de desligá-las, acabaríamos por levar nossa espécie ao suicídio. Este cenário previsto por *Unabomber* já começa a se tornar realidade. Hoje temos muitos sistemas especialistas que tomam decisões nos setores financeiro, militar e até mesmo médico. Testes mostram que sistemas de diagnóstico médico artificiais são capazes de tomar decisões mais acertadas que as de um cirurgião humano. Hoje, ainda tenho a liberdade de perguntar algo a esse sistema, ele me dá a resposta e eu decido. Mas suponhamos que daqui a dez anos esse sistema esteja aperfeiçoado, com um desempenho excelente. O que acontecerá legalmente a um médico que, confiando na sua intuição profissional, desrespeitar a sugestão da máquina e o seu paciente vier a falecer?

Luiz Paulo Juttel