

# QUAIS OS LIMITES ÉTICOS DA INTELIGÊNCIA ARTIFICIAL E COMO EVITAR VIESES DISCRIMINATÓRIOS?



ENTREVISTA COM SANDRA ÁVILA,  
PROFESSORA E PESQUISADORA DA UNICAMP A  
KRISHMA CARREIRA DA FSB

**Krishma Carreira:** Como você avalia o estágio do ecossistema de Ciência, Tecnologia e Inovação brasileiro? Ele está preparado para esse desafio enorme que temos em termos de pesquisa e capacitação, pensando em relação à inteligência artificial e às aplicações de *machine learning*, por exemplo?

**Sandra Ávila:** É uma visão bastante ampla. Vou responder até onde eu consigo enxergar e também com o que eu estou lidando. Eu acho que, talvez, seja mais interessante começar pela ciência, que é realmente o mundo no que estou bastante imersa. Em relação à ciência, na verdade, a perspectiva é um pouco negativa, pois o investimento está diminuindo drasticamente. Estamos todos, principalmente de dentro das universidades, sofrendo com isso. E a inovação sai bastante de lá. Queremos sempre ultrapassar a fronteira do conhecimento e nunca estamos satisfeitos apenas em resolver um problema. A gente quer resolvê-lo da melhor forma!

Um lado positivo é que, do ponto de vista de inovação das empresas, todo mundo já percebeu que está faltando muita gente para isso. Então, as próprias empresas estão começando a investir nas universidades para formar pessoas. A gente não tem nem programadoras e programadores, muito menos pessoas com um conhecimento mais especializado. É um cenário extremamente difícil e está sendo cada vez mais sucateado.

Mas eu fico muito impressionada com o que as pessoas fazem com tão pouco recurso. A criatividade (pensando nas soluções sem o recurso) é realmente muito impressionante! Você fica se perguntando: e se realmente tivesse investimento, imagina o que a gente faria?



**KC: Fazendo um gancho com o que você falou sobre a questão da redução de investimentos, pensando na Estratégia Brasileira de Inteligência Artificial (EBIA), você acha que ela avançou mesmo ou é ainda meio “fake”? Tem muito pano para manga?**

**SA:** Vou aproveitar sua fala, tem muito pano para manga... Tem muito pano para manga em muitos lugares. Tem alguns países em que as discussões já estão mais avançadas, isso não é uma coisa que está todo mundo sabendo como deve fazer. É muito recente mesmo! As aplicações começaram, de fato, como algo bastante promissor em 2012, 2014, 2015. Então, tem realmente muito pouco tempo. Essa discussão ainda vai durar bastante tempo, envolvendo até o que a gente deveria regular sobre inteligência artificial. Ainda estamos engatinhando.

**KC: É um roteiro de boas intenções, só falta saber como fazer?**


**SA:** Todo mundo sabe que é necessário. E, assim, se você não estiver fazendo isso, vai ficando cada vez mais para trás. Virou realmente decisório.

**KC: Só para pegar um ponto que você já citou quando falou sobre programadoras e programadores. Como atrair mais programadoras? Isso também é uma questão essencial. O que você acha que dá para fazer nesse sentido?**

**SA:** Existem várias iniciativas das próprias mulheres programadoras mostrando que computação é legal. Se você saiu do colégio e vai entrar para uma universidade, ainda é uma pessoa muito nova para tomar essa decisão. E, na própria universidade, quando a mulher começa a ter contato com a programação é que ela descobre. Então, o que tem acontecido muito, inclusive, é mudança de carreira. As pessoas estão percebendo que programação é importante ou pelo menos que deveriam ter esse conhecimento, que trata de um raciocínio lógico de como se deve resolver um problema. E aí as pessoas acabam mudando.

Eu acho que as iniciativas que mostram o que você está fazendo acaba atraindo a atenção por si só. Quando você vai num lugar e vê um tanto de homem, você vai dizer assim: “nossa, eu, Sandra, não gostaria de trabalhar naquele lugar”. Então, na hora que eu vejo um lugar diverso, eu digo: “poxa, aqui deve ser legal de trabalhar”. Você fica com vontade de fazer aquilo. A gente tem (e eu falo “a gente” porque é um grupo) várias iniciativas nesse sentido, começando inclusive com as meninas pequenininhas. Tem um evento chamado Meninas SuperCientistas para meninas do sexto ao nono ano. Estamos sempre trazendo mulheres palestrantes, falando de suas respectivas pesquisas e trabalhos. Não tem nada dizendo assim: “olha, isso é para mulher”. É mais no sentido de “deixa eu mostrar o que eu faço”. Só isso! Eu só estou mostrando o que eu faço. E na hora em que eu mostro o que eu faço, a pessoa olha e diz assim: “eu também posso fazer!” Então, esse mostrar faz parte, mas não só para as mulheres, mas para os homens também.





**KC:** Com a pandemia, houve um processo de aceleração do uso de algumas ferramentas. Os *chatbots* e outras interfaces com os clientes são alguns exemplos mais visíveis. Você acha que as pessoas já estão confiando mais nessas ferramentas? Acredita que elas venceram alguns preconceitos e medos em relação ao uso de robôs e tecnologias mais disruptivas?

**SA:** Se você está lidando com *chatbot*, tem muitas pessoas, inclusive, que se sentem mais à vontade de fazer determinadas perguntas. Mas há coisas diferentes do que eu posso tirar da sua pergunta.

Sobre as pessoas confiarem em usar, eu acho que sim. Menos quando querem cancelar algum serviço. Aí você não quer falar com um *chatbot*, você quer falar com uma pessoa.

No Brasil, sei que tem, por exemplo, máquinas de sorvete em *shoppings*. Você interage com o robô e ele dá tchauzinho para você e as pessoas acham: “oh, que lindo! Que fofo!” As pessoas gostam. Agora, isso é uma tarefa muito objetiva, que eu tenho uma pergunta e uma resposta objetiva para quem está em frente a uma máquina que faz sorvete, por exemplo.

***Agora, quando você está lidando com coisas que têm uma subjetividade ou então que estão fazendo um diagnóstico, você não espera que a máquina faça aquilo sozinha, você está esperando que seja a máquina, que é uma inteligência aumentada, como a gente chama, e não inteligência artificial. Inteligência aumentada é quando você está juntando o ser humano com a máquina e aquilo está dando uma resposta melhor para uma determinada tarefa.***

***Também tem a confiança até dos próprios algoritmos em relação à resposta. E as técnicas, hoje, não são totalmente interpretáveis. Na hora que você não sabe como aquela resposta aconteceu, trazendo para um outro ponto de transparência, perde a confiança. Então, tem um caminho longo pela frente. E as pessoas estão muito mais preocupadas com isso hoje. Como que eu deixo tudo mais transparente, confiável e auditável?***

As pessoas estão muito predispostas realmente a utilizarem essas tecnologias. Eu trabalho bastante na área da saúde e eu nunca me deparei com uma situação de chegar para conversar com alguém e a pessoa falar bem assim: “ah, mas eu acho que a gente não deveria aplicar esse negócio porque não vai dar certo”. Ao contrário, a recepção é sempre muito boa. As pessoas já perceberam de fato que aquilo é uma necessidade.

Vamos falar agora sobre como a gente deveria enxergar essa questão do que tem sido proposto em relação à IA.

***A IA é hoje uma tecnologia de propósito geral e vai se tornar cada vez mais presente. E uma tecnologia de propósito geral, fazendo um paralelo aqui, é como a gente tem hoje a eletricidade e o computador, que estão presentes em todos os lugares. Não perguntamos porque isso ocorre. Na verdade, é o contrário. Quando você chega num lugar e não tem energia, as pessoas ficam surpresas. A IA vai estar assim também, permeada em todos os lugares. E a gente vai ficar se perguntando: “não está aplicando IA aqui? Deveria!” Ela não vai resolver todos os problemas do mundo, esse não é o propósito, mas com certeza ela vai estar bem mais presente e a gente vai começar a achar estranho quando não estiver.***

**KC:** Mas você não acha que aí entra um lado que é mais complexo e traz outros problemas do ponto de vista social? Porque, com a eletricidade, você percebe a presença dela... Você não raciocina no seu dia a dia que você já tem, mas você a vê. Agora, quando você tem uma IA, você muitas vezes não faz ideia de que foi uma inteligência artificial por trás. De novo, a gente volta para a questão da transparência. Como é possível fazer para que as pessoas entendam? Faz parte de uma literacia, de uma educação, de um treinamento digital, mas como você consegue de fato explicar, traduzir uma coisa para que as pessoas a entendam, estejam mais abertas e saibam usá-la melhor?

**SA:** Uma boa parte realmente vem da educação, para que as pessoas entendam que aquilo vai estar presente em todos os lugares e elas têm que se questionar. Tem uma parte também que é responsabilidade...

***Cabe às pessoas que estão desenvolvendo ou aplicando uma solução deixá-la pelo menos de uma forma um pouco mais clara, porque você não está esperando que todo mundo tenha conhecimento sobre isso.***

***Tem uma outra coisa também em relação à questão da IA, que é em relação aos dados. As pessoas não sabem da importância dos dados, sobre seus respectivos dados, sobre como aquilo está sendo usado e como aquilo pode estar está sendo usado contra ela. As pessoas precisam realmente de uma educação, precisam saber que aquele dado está sendo usado para influenciar diversas tomadas de decisões na vida dela.***

Eu não tenho resposta, realmente, pois é muito difícil dar essa resposta sobre o que a gente deveria fazer, porque há várias coisas que podem ser feitas, mas ainda são coisas muito pequenas.

**KC: O programador e a programadora também têm vieses que são incorporados na hora de criarem modelos. Como é possível minimizar isso na hora de programarem?**

***SA: Primeiro, a gente tem que perguntar: qual é o problema que a gente está resolvendo e qual é o público-alvo? Geralmente, o que tem acontecido é que as pessoas estão propondo modelos e estão dizendo que aqueles modelos podem ser aplicados para qualquer coisa. Eu acho que os nossos modelos deveriam sair com uma bula de remédio: isso aqui é indicado para isso, isso é contraindicado para aquilo, não usa esse porque isso vai ter problemas ....***

E isso começa com os dados. A questão é que esses modelos hoje precisam de muitos dados que, muitas vezes, você não sabe que estão sendo usados ali. Mas há algumas coisas simples que podem ser feitas e as pessoas não estão fazendo. Um exemplo: gerei meu modelo, agora vou aplicar testes aleatórios para avaliar se meu modelo está realmente funcionando, pensando em casos mais diversos possíveis, tendo uma equipe diversa que consegue fazer essas perguntas. Na hora que a gente faz uma prova dos 9, em relação ao nosso modelo, você vê: "opa, isso não está funcionando".

***Vamos puxar o gancho do reconhecimento facial, que é muito problemático. Quantos modelos a gente tem de reconhecimento facial que não funcionam para pessoas de pele negra? Se você disponibiliza, tem que verificar se o negócio está funcionando. Não? Faça o seu trabalho! Vá lá e questione. E aí esse questionamento geralmente vem quando você tem uma equipe diversa, para quem é possível perguntar se acha que o que foi proposto é adequado e se funcionaria para todo mundo.***

Vou pegar um exemplo bem bobo: filtros no Instagram que você muda a cor do cabelo. Vi uma pessoa que tinha cabelo liso e aquilo funcionou maravilhosamente bem para ela. Fui testar e ficou um caos. Não funcionava por causa do cabelo cacheado. Estou pegando aqui um exemplo realmente bobo.

Mas vamos dar outros exemplos, como a questão do câncer de pele. As imagens que a gente tem são essencialmente de pessoas brancas. Aquilo não vai funcionar como a gente espera para população brasileira. E eu não posso entregar esse modelo e achar que vai funcionar com o mesmo resultado de outros países. Não vai funcionar de jeito nenhum, porque o padrão é diferente do tipo de lesão. Vai dar uma resposta possivelmente errada. Mas acho que se questionar e ter equipe diversa já são grandes passos.

**KC: O que você pode falar mais relação à questão do racismo algorítmico?**

***SA: O que eu acho que está acontecendo é que a gente está automatizando o racismo e, muitas vezes, quem está desenvolvendo isso são pessoas que não estão preocupadas (mesmo que de forma não intencional) em se questionarem sobre isso.***

Então, novamente, precisamos de outras pessoas que estão questionando os modelos e vendo coisas que não deveriam ser aplicadas. Reconhecimento facial não deveria estar sendo usado! Sistemas de reconhecimento facial foram banidos e julgados no mundo. A gente não deveria estar fazendo reconhecimento facial, porque isso, na verdade, só arrebenta para o lado de quem já está sendo prejudicado e está só piorando uma situação que já existe.



**KC: Pegando do ponto de vista de segurança, se você tem um sistema de reconhecimento facial que não reconhece direito a pele negra, mas no fundo há mais pessoas sendo presas em função disso, o que leva a esse resultado?**

**SA:** Exatamente, ele está sendo vitimado duas vezes, pelo menos. Primeiro porque a pessoa só está sendo presa por causa da pele, isso já é uma agressividade enorme. E, segundo, porque aquela pessoa nem deveria estar sendo presa. Foi um erro realmente em relação ao sistema, de quem está usando aquilo, implantando aquilo [reconhecimento facial] que não deveria estar sendo implantado. Por isso que hoje há uma discussão muito grande em relação ao banimento do reconhecimento facial e não ser aplicado para tomar decisões para nada.

Tem uma outra questão que a gente está entendendo sobre o viés. Aí vem muito a parte computacional. Estamos tentando entender melhor as técnicas que aplicamos.

***Há diversos tipos de vieses e isso é importante comentar. Tem um viés que é histórico. Se a gente voltar para o reconhecimento facial, que é bastante problemático, o viés histórico é o que a gente tem hoje na população carcerária do Brasil, que tem mais pessoas de pele negra do que de branca. Então, se você treinar o algoritmo com um conjunto de dados desse, você vai dizer que, provavelmente, ele indicaria uma pessoa de pele negra se eu tivesse que identificar alguém para prender ou não. Se a gente traz exatamente esse dado, o algoritmo aprendeu errado, porque a gente ensinou errado. É necessário balancear de alguma forma para esse aprendizado não acontecer desta forma. Tem outros vieses, como o que acontece quando extrai correlações espúrias do dado***

Isso também aconteceu muito com a Covid agora. Há alguns artefatos na imagem que estavam sendo usados, em algumas técnicas, para dar certas respostas para pessoas que fizeram raio-X. Mas a técnica do modelo não estava prestando atenção na imagem do pulmão, e sim nos artefatos que estavam na imagem. Há vários problemas de vieses, em vários lugares, em relação a isso.

***Há outros tipos de vieses ainda, como o da própria técnica e o de interpretação. É preciso ter cuidado na própria interpretação dos modelos. E, novamente, estamos estudando todas essas coisas, mas enquanto nós estivermos estudando, não podemos sair cometendo erros. Isso é importante. Se tem alguma coisa que não está funcionando, não dá para deixar o negócio operando, enquanto ele vai melhorando. Não! Tira esse negócio! Essa é a história do reconhecimento facial. Tem que parar com isso, porque não está funcionando. Tem que ser estudado antes de sair aplicando nas pessoas. Estuda, mas não no modelo implantado.***


**KC: Estamos entrando na questão da ética nas tecnologias.**

***SA: O que eu tenho falado recentemente é que a gente precisa falar mais de ética e menos de técnica.***

E se nós falarmos um pouco mais sobre ética, começamos a refletir e colocá-la no processo. Eu vou pegar um exemplo de um trabalho que eu fiz com uns alunos de uma disciplina e que acabei discutindo uma questão de ética em relação ao processo e pedindo tarefas muito simples. Um grupo deveria definir o que é ética. Com o outro grupo, eu queria que eles pegassem ética em inteligência artificial e procurassem uma reportagem sobre como essa questão de ética (ou falta de ética na verdade) estava acontecendo no processo de IA. Foi realmente um trabalho muito simples. Mas eles foram procurar e se depararam com vários problemas, começaram a se questionar sobre os problemas que estavam resolvendo, sobre os dados que estavam usando e o que as pessoas estavam fazendo. Com isso, o projeto final, que tinha que começar de um jeito, terminou de outro, por causa desse questionamento.

Então, uma coisa que falta (e falta mesmo!) é a gente começar a falar, pois a gente não está discutindo sobre a ética.





**Muitas coisas nesses modelos estão sendo gerados e somos todos responsáveis por ele. E a gente não está discutindo como aquele modelo está sendo gerado de fato, quais são os dados que estão sendo usados.**

Hoje, existem as grandes conferências (falando do ponto de vista científico) que no processo de submissão dos artigos já estão fazendo questionamento em relação à ética do *paper* que você está propondo. Qual o problema que você está resolvendo? Quais são os seus dados? Como estão os seus dados? Onde ele foi testado? Essa discussão já está vindo para conferências, conferências grandes na parte de *machine learning*, que tem um público razoável.

**Precisamos discutir e trazer isso para os cursos que são técnicos, como a matemática, estatística, computação, engenharias. Mas a gente não está discutindo a ética, a gente está falando muito das técnicas! Eu tenho certeza de que, em outros cursos, essa preocupação não é algo que está começando, é uma discussão permanente, mas, para muitas áreas, não é. Isso dá um problema... Estamos automatizando o problema e a gente tem que solucioná-lo.**

**KC: Como garantir transparência de fato e uma espécie de prestação de contas algorítmica? A gente sabe que os sistemas são tão complexos, com vários algoritmos e pessoas envolvidas, como dá para garantir, então, transparência no processo?**

**SA:** É uma pergunta bem difícil. Essa discussão está acontecendo agora em vários setores, setores de governo, nas próprias empresas. Você tem que garantir essa transparência e informar como é aquele seu dado. Eu estou coletando alguma coisa? Onde aquele dado vai ser usado? Para que problema aquilo vai ser resolvido? Não é no sentido de impedir que aquele problema seja solucionado, mas se eu estou disposta a fornecer meu dado, eu quero saber onde aquele dado vai ser utilizado. Começando por aí, realmente existem várias perguntas que a gente poderia fazer.

***Inclusive a questão ética e a transparência estão muito juntas para você poder auditar uma resposta. Eu entrei com um dado, saiu uma resposta e preciso entender aquele processo. Como funciona isso? As melhores técnicas não estão prontas para isso. Elas não fazem isso. Tem muita coisa que precisa ser feita.***

**KC: Algumas grandes empresas de tecnologia têm técnicas para detectar se tem viés no algoritmo. Esse tipo de técnica funciona?**

**SA:** Funciona, mas só que sempre para um escopo. Tem um escopo de dado, um escopo de programa, tem sempre restrições. Não é uma técnica que funciona para tudo. E não funciona mesmo. A gente tem vários problemas que estão sendo resolvidos. Muitas vezes, a saída vai ser um número, que pode ser um número variando no espaço real (não é um número, 1, 2, 3, mas pode ser um 1.1, 1.2, 2.75) e isso já muda todo o processo, porque a gente está resolvendo problemas diferentes. Pode ser uma imagem, pode ser um vídeo. Então, como a gente dá essa resposta? É um conjunto imenso de técnicas e aquilo não funciona realmente para todas as coisas.

Mas eu vou dizer que tem muita gente preocupada, que sabe que, se isso não acontecer, algumas aplicações não poderão ser executadas, porque elas vão passar por uma regulação e não serão aprovadas. Então, a gente vai ter que resolver aquele problema. E é melhor que a gente comece a resolver, já desenvolvendo as técnicas, do que a gente ser cobrado depois. Porque seremos cobrados depois, pois somos todos responsáveis por isso e a gente precisa realmente estudar esse processo.

**KC: Para finalizar, em relação ao aprendizado de máquina, o que está por vir no curto e no médio prazo, em termos de aplicação?**

**SA:** Tem uma coisa que acho que vai ter bastante mudança (e já tem acontecido até), que é uma medicina mais precisa, voltada para cada pessoa. Hoje, já tem coisas desse tipo (um remédio pode funcionar para uma pessoa e de outro jeito para outra). Hoje, a gente usa o que dá para fazer no momento, mas poderia talvez ter resultados bem mais interessantes. Outras aplicações já existentes são relacionadas ao processo de produção de alimentos. Produzir alimento com menos desperdício e garantindo uma produção maior, por exemplo. São tarefas que, sendo automatizadas (e elas são bastante objetivas), têm um impacto muito forte para a população mundial.



## Sandra Ávila

Sandra Ávila é professora no Instituto de Computação (IC), da Universidade Estadual de Campinas (UNICAMP), desde Fevereiro de 2017. É doutora em Ciência da Computação, com duplo diploma, pela Universidade Federal de Minas Gerais (UFMG) e pela Sorbonne Université (UPMC Sorbonne, Paris 6, França), em 2013; mestre em Ciência da Computação pela UFMG, em 2008; e bacharel em Ciência da Computação pela Universidade Federal de Sergipe (UFS), em 2006. Foi pesquisadora científica em Pós-Doutorado na UNICAMP (2013 a 2016), na Faculdade de Engenharia Elétrica e Computação (FEEC), em convênio com a SAMSUNG e pelo PNPd/CAPES. Em 2018, 2019 e 2020, foi agraciada com o prêmio Google Latin America Research Awards. Suas pesquisas estão voltadas para Inteligência Artificial, mais especificamente Aprendizado de Máquina e Visão Computacional, com ênfase na Saúde, Análise de conteúdo sensível e Agricultura de precisão. Em 2020, Sandra Ávila foi selecionada para representar o Brasil no BRICS Young Scientists Forum, na área de Inteligência Artificial. Sua paixão é incentivar meninas e mulheres a apostarem em carreiras nas áreas de STEM (Ciências, Tecnologia, Engenharias e Matemática).