



AI PART III

GLOBAL PERSPECTIVES AND INSIGHTS

Framework de Auditoria de Inteligência Artificial do The IIA

Aplicações Práticas, Parte B

Edição Especial



The Institute of
Internal Auditors

Global



Conselho Consultivo

Nur Hayati Baharuddin, CIA, CCSA,
CFSA, CGAP, CRMA –
Membro do *IIA–Malásia*

Lesedi Lesetedi, CIA, QIAL –
IIA Federação Africana

Hans Nieuwlands, CIA, CCSA, CGAP –
IIA–Países Baixos

Karem Obeid, CIA, CCSA, CRMA –
Membro do *IIA–Emirados Árabes
Unidos*

Carolyn Saint, CIA, CRMA, CPA –
IIA–América do Norte

Ana Cristina Zambrano Preciado, CIA,
CCSA, CRMA – *IIA–Colômbia*

Edições Anteriores

Para acessar as edições anteriores do
Global Perspectives and Insights,
acesse www.theiia.org/gpi.

Feedback dos Leitores

Envie perguntas ou comentários para
globalperspectives@theiia.org.

Índice

Framework de Auditoria de IA do The IIA	1
O Fator Humano.....	2
Ética	
Parcialidade do Algoritmo	
<i>Meaning Making</i>	
A Caixa Preta	3
Governança	4
Mensurando o Desempenho	4
Reflexões de Encerramento	5

Introdução

Imagine. Um robô – ou, ao menos, seu cérebro - perante um comitê de *ética* ou investigação. Poderia acontecer. De acordo com o [The Guardian](#), acadêmicos argumentam que, conforme os robôs começam a entrar em espaços públicos e a trabalhar com humanos, a necessidade de medidas de segurança se torna mais urgente. Cientistas estão tentando defender que robôs sejam equipados com uma "caixa preta ética", para manter registro de suas decisões e permitir a eles que expliquem - sim, expliquem - suas ações quando houver *acidentes*.

Professores da Universidade de Oxford argumentam que firmas de robótica devem seguir os exemplos e regulamentos definidos pela indústria da aviação, que incluem a adoção de caixas pretas e gravadores de voz no *cockpit* para investigar acidentes aéreos, garantindo que lições essenciais de segurança sejam aprendidas após tais eventos trágicos. Se instalada em um robô, a caixa preta ética gravaria as decisões do robô, sua base decisória, seus movimentos e as informações de sensores, como câmeras, microfones e telêmetros.

Essas ações surgem por ocasião de incidentes recentes, como quando "Steve", um robô de patrulha K5 [Knightscope](#), caiu da escada e [dentro de uma fonte](#) enquanto estava a serviço, ou quando outro robô K5 [envolveu-se em uma briga em um estacionamento](#) com um homem de 41 anos enquanto patrulhava as ruas de Mountain View, na Califórnia, ou quando um robô foi acusado de atropelar uma criança de 16 meses em um shopping em Stanford. Todos os ferimentos dos humanos envolvidos foram mínimos.

Há centenas de relatórios sobre a ética da inteligência artificial; no entanto, a maioria é *superficial* e cheia de clichês sobre colocar as pessoas em primeiro lugar, escreve Scott Rosenberg, editor do *Backchannel*. Rosenberg cita [relatórios](#) publicados recentemente pelo AI Now Institute, da New York University, sobre uma indústria tecnológica que tenta reformular a sociedade a partir da IA, sem qualquer garantia de resultados confiáveis ou justos. Um [relatório](#) conclui que "os esforços para submeter a IA a normas éticas, até hoje, foram fracassados e os novos frameworks éticos para IA precisam ir além da responsabilidade individual, para responsabilizar os interesses poderosos da indústria, do governo e das forças armadas, que desenvolvem e empregam a IA". Na opinião do autor, os sistemas de IA estão sendo apresentados em áreas diversas e vulneráveis, como policiamento, educação, saúde e outros ambientes, em que a "falha de um algoritmo pode arruinar uma vida".

Framework de Auditoria de IA do The IIA

Conforme explicado em [Inteligência Artificial – Considerações para a Profissão de Auditoria Interna](#), o papel da auditoria interna na IA é "ajudar a organização a avaliar, entender e comunicar até que nível a inteligência artificial terá efeito (negativo ou positivo) sobre a habilidade da organização de agregar valor a curto, médio e longo prazo".

Para ajudar a auditoria interna a cumprir com esse papel, os auditores internos podem alavancar o Framework de Auditoria de IA do The IIA para prestar serviços de consultoria, avaliação ou combinações desses serviços em relação à IA, conforme apropriado para a organização. O Framework é constituído de

Observação: Este é o terceiro relatório de uma série de três partes. Para mais informações, consulte [Inteligência Artificial – Considerações para a Profissão de Auditoria Interna e Framework de Auditoria de Inteligência Artificial do The IIA: Aplicações Práticas Parte A](#).



três componentes gerais – Estratégia de IA, Governança e o Fator Humano – e sete elementos: Ciber-resiliência; Competências de IA; Qualidade dos Dados; Arquitetura e Infraestrutura dos Dados; Mensuração do Desempenho; Ética; e a Caixa Preta.

A auditoria interna deve considerar inúmeros objetivos de trabalho ou controle, e atividades ou procedimentos, na implementação do Framework e na prestação de serviços de consultoria, avaliação ou combinações desses serviços em relação às atividades de IA da organização. Os objetivos e atividades ou procedimentos relevantes que abordam a Estratégia (elementos Ciber-resiliência e Competências de IA) e Governança (elementos Arquitetura e Infraestrutura de Dados e Qualidade dos Dados) do Framework foram apresentados em Framework de Auditoria de Inteligência Artificial do The IIA: Aplicações Práticas Parte A. Este documento oferece objetivos e atividades ou procedimentos relevantes que abordam o Fator Humano (elementos Ética e Caixa Preta) e Governança (elemento Mensuração do Desempenho).

O Fator Humano

O componente Fator Humano, que inclui os elementos Ética e Caixa Preta, aborda o risco de erro humano que comprometa a habilidade da IA de entregar os resultados esperados.

Ética

Algoritmos desenvolvidos por humanos e que incluam erros e parcialidades humanas (intencionais ou não) impactarão o desempenho do algoritmo. O componente do Fator Humano considera:

- Se o risco de parcialidades humanas não intencionais estiverem incluídas no desenvolvimento da IA é identificado e gerenciado.
- Se a IA foi testada devidamente, para garantir que os resultados reflitam o objetivo original.
- Se as tecnologias de IA podem ser transparentes, considerando a complexidade envolvida.
- Se os produtos da IA estão sendo usados legal, ética e responsabilmente.

Parcialidade do Algoritmo

De acordo com um relatório recente da McKinsey & Company, as empresas estão aplicando rapidamente o *machine learning* na tomada de decisões comerciais. Os programas colocam algoritmos complexos para trabalhar com conjuntos de dados enormes e frequentemente atualizados. No entanto, a parcialidade do algoritmo é um negócio arriscado, porque pode comprometer o próprio propósito do *machine learning* se deixado de lado e não verificado (leia [Controlling machine-learning algorithms and their biases](#)).





Por exemplo, na pontuação de crédito, o cliente com longo histórico de empréstimos sem atrasos ou inadimplência é geralmente determinado como "baixo risco". No entanto, o que pode não ser visto é que as hipotecas desse cliente foram mantidas e apoiadas por benefícios tributários que estão para expirar. Um pico na inadimplência pode estar no horizonte, não previsto no modelo estatístico da instituição de crédito. Com acesso aos dados certos e a orientações de especialistas, modelos preditivos de *machine learning* poderiam encontrar os padrões escondidos nos dados e considerar tais picos. Ainda mais, fora das decisões de *negócios*, a parcialidade do algoritmo pode causar erros que poderiam dar início a problemas reais e a certa agitação nos cidadãos. Por exemplo, o serviço Google Photos e outros serviços como tal são usados para identificar pessoas, objetos e cenas, mas podem dar terrivelmente errado, como quando uma câmera não detectou sensibilidade racial ou quando um software usado para avaliações de risco para [prever futuros criminosos](#) demonstrou parcialidade.

Meaning Making

O que as máquinas podem fazer tem limites e as *peessoas* devem ser capazes de entender os produtos da IA. Conforme descrito em um relatório da McKinsey & Company, "o *meaning-making* [atribuição de sentido] na era da IA começa com a apreciação do que as máquinas podem e não podem fazer. Pode ser possível, por exemplo, que uma máquina faça certos tipos de diagnósticos mais precisamente do que uma pessoa conseguiria. Mas depende dos enfermeiros, médicos e terapeutas ajudar os pacientes a entender e gerenciar as consequências. É a diferença entre conhecimento e sentido". De acordo com o relatório, habilidades formais como programação, análise e ciência de dados são fundamentais para a IA, mas habilidades informais, como colaboração, empatia e atribuição de sentido também são (veja [McKinsey Quarterly pp. 56-61](#)).

Objetivos e Atividades ou Procedimentos Relevantes à Ética

Objetivo(s) do Trabalho ou de Controle	Atividades ou Procedimentos
Prestar avaliação de que os resultados das atividades de IA da organização estão livres de parcialidades não intencionais.	Revisar os resultados desejados das atividades de IA (objetivos estratégicos) e comparar com resultados reais. Se for detectada uma variação, determinar se a causa foi a parcialidade.
A organização consegue "atribuir sentido" aos produtos da IA.	Revisar os produtos da IA e o sentido que foi atribuído a partir deles.

A Caixa Preta

De acordo com o dicionário online *Merriam-Webster*, uma caixa preta é "um dispositivo eletrônico normalmente complicado, cujo mecanismo interno é normalmente escondido ou misterioso para o usuário; *amplamente*: qualquer coisa que tenha funções ou mecanismos internos misteriosos ou desconhecidos." Conforme as organizações progredem na implementação de tecnologias de IA Tipo III e IV – usando máquinas ou plataformas que possam aprender por conta própria ou se comunicar entre si –, a forma como os algoritmos estão operando se torna menos transparente ou compreensível. O

Os objetivos relevantes e atividades ou procedimentos identificados pelo The IIA não incluem um plano de auditoria previsto, mas são exemplos que devem ser úteis na identificação dos objetivos do trabalho ou de controle e na condução dos trabalhos de auditoria de IA.

Trabalhos de auditoria de IA devem estar em conformidade com a Norma 2200 do IIA: Planejamento do Trabalho de Auditoria. Os planos de auditoria de IA e objetivos e procedimentos do trabalho de IA devem sempre ser personalizados para atender às necessidades da organização.

fator caixa preta torna-se cada vez mais um desafio, conforme as atividades de IA da organização se tornam mais sofisticadas.

Objetivos e Atividades ou Procedimentos Relevantes à Caixa Preta

Objetivo(s) do Trabalho ou de Controle	Atividades ou Procedimentos
Avaliar o entendimento da organização quanto aos dados da "caixa preta" (isto é, os algoritmos inerentes, funções internas ou mecanismos que capacitam a IA).	<p>Revisar o desenvolvimento da IA e as políticas, processos e procedimentos de implementação, e verificar se os dados da caixa preta foram identificados.</p> <p>Entrevistar os responsáveis pelos resultados da IA e verificar se entendem e poderiam explicar os dados da caixa preta.</p>



Governança

A governança de IA refere-se às estruturas, processos e procedimentos implementados para dirigir, gerir e monitorar as atividades de IA da organização. O nível de formalidade e estrutura da governança variará com base nas características específicas da organização. A governança de IA:

- Estabelece a prestação de contas, responsabilidade e supervisão.
- Ajuda a garantir que as pessoas com responsabilidades de IA tenham as habilidades e expertise necessárias.
- Ajuda a garantir que as atividades de IA e decisões relativas à IA sejam consistentes com os valores e responsabilidades éticas, sociais e legais da organização.

Quanto ao componente Governança, a Parte II desta série, Framework de Auditoria de Inteligência Artificial do The IIA: Aplicações Práticas Parte A, abordou a prestação de contas, responsabilidade e supervisão; conformidade regulatória; arquitetura e infraestrutura de dados; e qualidade dos dados. Este documento aborda a mensuração do desempenho.

Mensurando o Desempenho

A auditoria interna está posicionada para se tornar vital para a habilidade da organização de mensurar o desempenho de suas iniciativas de IA. Na fase de planejamento, a auditoria interna pode oferecer conselhos sobre como estabelecer métricas que deem à administração e ao conselho informações suficientes, confiáveis, relevantes e úteis. No entanto, a auditoria interna não pode ser responsável por estabelecer ou ser proprietária de métricas de desempenho de IA. Nas organizações em que a IA foi implementada, a auditoria interna deve prestar avaliação sobre os controles da primeira linha de defesa e sobre a supervisão da segunda linha de defesa quanto à IA. Não há, talvez, melhor forma de demonstrar a competência da auditoria interna na IA do que usando tecnologias de IA, como a automação robótica de processos para auditar a IA.

Objetivos e Atividades ou Procedimentos Relevantes à Mensuração do Desempenho

Objetivo(s) do Trabalho ou de Controle	Atividades ou Procedimentos
Oferecer orientações sobre como estabelecer métricas de IA.	Facilitar sessões de trabalho para os responsáveis por estabelecer as métricas de IA. Transmitir a importância e significado dos termos <i>suficiente, confiável, relevante e útil</i> no que tange às informações.
Fazer testes de estresse com as vulnerabilidades de IA.	Aplicar técnicas de teste de estresse usadas pela indústria bancária para determinar como as atividades de IA se desempenharão em cenários extremos.
Comunicar os resultados dos trabalhos relativos à IA.	Comunicar os resultados de trabalhos relativos à IA em conformidade com: <ul style="list-style-type: none"> ■ Norma 2400 do IIA: Comunicação dos Resultados. ■ Norma 2410 do IIA: Critérios para a Comunicação. ■ Norma 2420 do IIA: Qualidade das Comunicações. ■ Norma 2421 do IIA: Erros e Omissões. ■ Norma 2430 do IIA: Uso de “Conduzido em Conformidade com as <i>Normas Internacionais para a Prática Profissional de Auditoria Interna</i>”. ■ Norma 2431 do IIA: Declaração de Não Conformidade do Trabalho da Auditoria . ■ Norma 2440 do IIA: Divulgação dos Resultados.
Prestar avaliação sobre os controles da primeira linha de defesa e a supervisão da segunda linha de defesa relativos à IA.	Adote a robótica e outras formas de IA para conduzir trabalhos relativos à IA.

Reflexões de Encerramento

A profissão da auditoria interna não pode ser deixada para trás no que pode ser a próxima fronteira digital - a inteligência artificial. Para se preparar, os auditores internos devem entender os conceitos básicos de IA, os papéis que a auditoria interna pode e deve desempenhar e os riscos e oportunidades da IA.

Sejam as tecnologias de IA da organização desenvolvidas internamente, por meio de uma tecnologia facilitadora, ou terceirizadas, a auditoria interna deve estar preparada para aconselhar o conselho e a alta administração, trabalhar em conjunto com a primeira e segunda linhas de defesa e prestar avaliação sobre o gerenciamento de riscos, a governança e os controles de IA. Para vencer esses desafios, os auditores internos devem alavancar o Framework de Auditoria de Inteligência Artificial do The IIA.

Sobre o The IIA

The Institute of Internal Auditors (The IIA) é o mais reconhecido advogado, educador e fornecedor de normas, orientações e certificações da profissão de auditoria interna. Fundado em 1941, o The IIA atende, atualmente, mais de 190.000 membros de mais de 170 países e territórios. A sede global da associação fica em Lake Mary, na Flórida, EUA. Para mais informações, visite www.globaliia.org.

Isenção de Responsabilidade

As opiniões expressas no Global Perspectives and Insights não são necessariamente as dos contribuintes individuais ou dos funcionários dos contribuintes.

Copyright

Copyright © 2018 The Institute of Internal Auditors, Inc. Todos os direitos reservados.

