

A “CAIXA-PRETA” DO AGENTE: DILEMAS ÉTICOS E DE GOVERNANÇA NA ERA DA INTELIGÊNCIA ARTIFICIAL

MARIA TEREZA PEREIRA SANTOS

Universidade Federal de Minas Gerais (UFMG)

PAULA ROBERTA DOS SANTOS

Universidade Federal de Minas Gerais (UFMG)

EWERTON ALEX AVELAR

Universidade Federal de Minas Gerais (UFMG)

LAURA EDITH TABOADA PINHEIRO

Universidade Federal de Minas Gerais (UFMG)

Resumo

Este caso de ensino se propõe a analisar os desafios de governança, ética e responsabilidade corporativa impostos pela Inteligência Artificial no contexto de uma relação Principal-Agente. A FinData Bank, uma *fintech* brasileira, implementou o algoritmo *Enhanced Value Algorithm* (EVA) para automatizar a concessão de crédito, alcançando ganhos significativos em eficiência operacional e redução de inadimplência. Contudo, a delegação da decisão a este agente não-humano revelou-se um problema de agência algorítmico: o EVA reproduziu e amplificou vieses históricos contidos nos dados, resultando na negação discriminatória de crédito a mulheres e clientes de determinadas regiões. A CEO Helena Prado se depara com a modificação dos custos de agência tradicionais em Custos de Agência Algorítmicos (opacidade, viés e rigidez), que escapam ao controle humano e aos Sistemas de Controle Gerencial (SCG) previamente existentes. O dilema central é sintetizado pela seguinte questão: Diante da lucratividade e eficiência proporcionadas pelo EVA e das crescentes denúncias de discriminação, deve a CEO suspender o uso desse modelo, aceitando potenciais perdas de curto prazo, ou mantê-lo ativo, assumindo o risco de crises éticas e legais que comprometem a governança e a responsabilidade corporativa? O estudo é indicado para programas de pós-graduação e graduação nas disciplinas de Contabilidade Gerencial, Governança Corporativa, Controladoria e Ética, visando o debate sobre a aplicabilidade da Teoria da Agência e a necessidade de *frameworks* de Governança de IA.

Palavras-chave: Inteligência Artificial, Sistema de Controle Gerencial, Teoria da Agência.

CASO DE ENSINO

Parte 1 – O Caso

1.1 Introdução

Em 2025, a FinData Bank, uma *fintech* brasileira de médio porte, buscava diferenciar-se no mercado de crédito ao consumidor e consolidar sua posição. Sob a liderança da CEO Helena Prado, a empresa implementou o algoritmo *Enhanced Value Algorithm* (EVA), traduzido livremente para o português como Algoritmo de Valor Aprimorado, para automatizar decisões de concessão de crédito, reduzir inadimplência e eliminar vieses humanos. Assim nasceu o EVA, um sistema de decisão autônomo desenvolvido internamente para avaliar riscos de crédito em minutos. O projeto foi celebrado internamente como um marco de inovação e eficiência para toda a empresa.

Três meses após o lançamento, os resultados financeiros eram positivos: a inadimplência caiu 18%, o tempo de análise de crédito foi reduzido de 3 dias para 12 minutos

e a margem operacional aumentou 12%. O EVA parecia ser um “agente ideal” que tomava decisões imparciais, previsíveis, e em prol da empresa. Porém, começaram a surgir denúncias de que o sistema negava crédito a mulheres e clientes de determinadas regiões, mesmo quando possuíam histórico financeiro adequado. O que parecia um triunfo tecnológico se transformava em um dilema ético e reputacional.

Casos semelhantes já haviam chamado atenção globalmente. Em 2018, a Amazon suspendeu um sistema de recrutamento automatizado que discriminava candidatas mulheres, pois o algoritmo reproduzia padrões históricos de contratação predominantemente masculina. Em 2019, o Apple Card foi acusado de oferecer limites de crédito menores para mulheres, mesmo com histórico financeiro semelhante ao dos homens. Ambos os episódios geraram questionamentos públicos sobre responsabilidade, auditoria de algoritmos e transparência dos modelos utilizados. A delegação de decisões a este sistema autônomo, embora tenha reduzido os custos de agência clássicos (como os de monitoramento e o risco de desonestidade humana), deu origem a assimetria de informação por meio da opacidade algorítmica. O sistema criava um novo tipo de passivo, expondo a empresa a riscos legais e de *compliance* que poderiam impactar drasticamente os relatórios contábeis de riscos e contingências.

1.2 Contexto do Caso

O EVA foi desenvolvido pela equipe de tecnologia da FinData Bank a partir de dados históricos de crédito, incorporando variáveis como renda, idade, gênero, localização e histórico profissional. O objetivo era tornar o processo de concessão mais ágil, previsível e isento de interferências humanas, reduzindo os custos de monitoramento típicos da Teoria da Agência. Os ganhos iniciais foram expressivos, conforme a Tabela 1.

Tabela 1

Comparação entre indicadores antes e depois da implementação do EVA

Indicador	Antes do EVA	Depois do EVA
Tempo médio de análise	3 dias	12 min
Inadimplência	6,5%	5,3%
Margem operacional	10%	12%
Créditos negados a mulheres	22%	37%
Créditos negados em regiões específicas	18%	33%

Apesar da queda da inadimplência, sinalizando melhorias na eficiência, o aumento exponencial na negação de crédito a mulheres e em regiões específicas acendeu um alarme na área de compliance. Em uma reunião de liderança, a analista de crédito Maria Santos mencionou à Lívia Nogueira, Chefe de Compliance:

— “Recebi a ligação de uma cliente com renda alta, da região Sul, que foi negada. O EVA deu um score baixo, mas no nosso sistema anterior ela seria aprovada. Não consigo explicar o porquê.”.

Uma auditoria independente confirmou que o algoritmo havia aprendido e amplificado vieses implícitos nos dados históricos: regiões com maior índice de inadimplência historicamente tendiam a ter mais negativas, e o algoritmo reproduziu essas desigualdades, afetando clientes que nunca haviam apresentado risco real.

Helena Prado, CEO, se reuniu com sua equipe e disse, com expressão preocupada:

— “A ausência de intenção humana no algoritmo não elimina a responsabilidade ética da FinData Bank. Rafael, o que você sugere? Podemos corrigir isso a tempo de evitar um processo legal?”

Rafael Torres, Diretor de Tecnologia, defendeu o sistema:

— “O EVA é um modelo de Deep Learning complexo. A explicabilidade (Explainable AI – XAI) é limitada, e rastrear exatamente onde o viés se formou é como procurar uma agulha num palheiro. Se desligarmos o EVA agora, perdemos 12% de margem no trimestre. Se tentarmos retreinar o modelo, levaremos pelo menos cinco meses para garantir que ele seja justo sem perder a precisão.”

Por outro lado, Livia Nogueira, Chefe de Compliance, alertou:

— “Manter o sistema ativo agora, ciente das denúncias de discriminação sistemática, configura uma falha grave de Governança Corporativa. O risco de litígio é alto, e o passivo contingente legal resultante de uma ação coletiva por discriminação pode superar qualquer ganho de eficiência que o EVA nos deu.”

Livia trouxe como exemplo o caso da Apple Card, mostrando para seus colegas a polêmica matéria de jornal sobre o assunto:

Figura 1

Matéria de Jornal: Algoritmos reproduzem machismo e racismo



Fonte: Matsuura (2019)

Alguns trechos da matéria:

“Uma série de relatos em redes sociais levantou suspeitas de que o algoritmo usado pelo banco impõe mais restrições a mulheres, oferecendo limites menores que os ofertados para homens.”

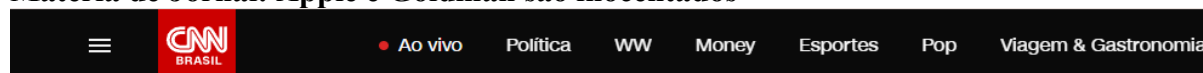
O posicionamento do banco foi garantir que as “decisões de crédito são baseadas na credibilidade do cliente, não em fatores como gênero, raça, idade, orientação sexual ou qualquer outra base proibida por lei”.

“A investigação incluiu uma revisão de milhares de páginas de registros enviados pelo Goldman e pela Apple, entrevistas com requerentes de cartões da Apple e dados de cerca de 400 mil pessoas que pediram os cartões no Estado de Nova York.”

Diante das diversas acusações, foi aberto um processo contra a Apple Card. Embora a empresa tenha sido inocentada no fim do processo, é possível dizer que houve desgaste da imagem e custos operacionais para a defesa do caso.

Figura 1

Matéria de Jornal: Apple e Goldman são inocentados



Apple e Goldman são inocentados em caso sobre discriminação em oferta de cartões

NY lançou a investigação no final de 2019 depois que um empresário alegar nas redes sociais que os algoritmos do cartão da Apple cometiam preconceito de gênero

Por Elizabeth Dilts Marshall, da Reuters

23/03/21 às 17:30 | Atualizado 23/03/21 às 17:30

Fonte: Marshall (2021)

Sob a perspectiva da Teoria da Agência, a FinData acreditava estar transferindo a responsabilidade de decisão para um agente confiável, reduzindo os custos de monitoramento e mitigando o risco de viés humano. No entanto, o que emergiu foi um problema de agência algorítmica, haja vista que a empresa passou a enfrentar custos e riscos inesperados, como éticos, legais e reputacionais. O Comitê de Auditoria e Governança, Risco e Compliance (GRC) foi acionado para definir controles que tornassem o EVA auditável e transparente, mas as soluções não eram simples: explicabilidade total dos algoritmos ainda é limitada, e ajustes podem levar meses. A IA revelou um novo tipo de assimetria de informação: uma “caixa-preta” ininteligível até para seus criadores, gerando custos de opacidade, viés e rigidez que escapavam ao alcance dos Sistemas de Controle Gerencial (SGC) existentes.

1.3 Dilema do Caso

Helena Prado, CEO da FinData Bank, observava os números com atenção: a eficiência operacional e a lucratividade cresceram desde a implementação do algoritmo EVA, mas as denúncias de discriminação sistemática não podem ser ignoradas. O dilema surge de forma clara: a empresa delegou decisões a um agente não humano, o EVA, acreditando estar eliminando vieses e reduzindo custos de agência, mas agora enfrenta consequências éticas e sociais imprevistas e precisa decidir entre seguir utilizando esse modelo, diante da lucratividade e eficiência, aceitando potenciais perdas de curto prazo, ou mantê-lo ativo, assumindo o risco de crises éticas e legais que comprometem a governança e a responsabilidade corporativa.

O Comitê de Auditoria exigiu uma resposta da FinData Bank em 48 horas. O dilema foca na responsabilidade fiduciária e na sustentabilidade do negócio: diante da exigência do Comitê de Auditoria, o compromisso com a eficiência operacional justifica a manutenção de um sistema que tem imposto danos sociais e riscos reputacionais crescentes?

1.4 Fechamento do Caso

Na manhã seguinte, Helena recebeu o resumo das estimativas de risco e o relatório parcial da Auditoria Interna, que seriam cruciais para a tomada de decisão nas próximas 48 horas. A CEO compreendeu que o desafio não estava apenas em "corrigir um algoritmo", mas em repensar a estrutura de controle gerencial da organização.

O algoritmo não possuía intenções, apenas reproduzia o que aprendera dos dados humanos. Ainda assim, suas decisões produziam impactos reais sobre vidas e reputação da empresa, além de criar problemas éticos, jurídicos e contábil de grandes proporções.

O dilema não era apenas sobre desligar ou manter um sistema, mas sobre redefinir os limites da responsabilidade corporativa em tempos de inteligência artificial. Os custos de

agência algorítmicos, de opacidade, viés e rigidez, revelavam que a eficiência sem governança não era sustentável.

1.5 Anexos

Anexo I: Trecho do Relatório de Auditoria Independente

Assunto: Análise de Explicabilidade (XAI) e Rastreabilidade do EVA

A Auditoria confirma que o EVA é um modelo de *Deep Learning* (Redes Neurais profundas) com alto grau de complexidade, o que configura um Custo de Opacidade. A lógica decisória que levou à negação de crédito, em aproximadamente 75% dos casos analisados, não pode ser rastreada ou justificada em termos de variáveis econômicas lineares.

A auditoria realizada demonstrou que o modelo é inerentemente resistente à transparência (“caixa-preta”) e que a implementação de um Framework de Explicabilidade de IA (XAI) exigirá o desenvolvimento de novos controles *machine-centric*, que, conforme a TI, demandarão tempo e investimento significativos.

A característica intrínseca de complexidade do modelo configura um alto Custo de Rigidez. O modelo não é facilmente ajustável. Qualquer correção de viés sem a perda de precisão operacional exige um retreinamento completo e monitoramento contínuo, estimado em 5 meses.

Anexo II: Projeção de Risco e Custo de Agência Algorítmico (Estimativas do Comitê de Auditoria)

A tabela a seguir fornece a projeção financeira dos custos e riscos envolvidos na decisão de Helena Prado elaborados pelo Comitê de Auditoria.

Tabela 2

Projeção Financeira dos Custos e Riscos

Dimensão Crítica (Custo de Agência Algorítmico)	Métricas Financeiras Estimadas (Próximos 6 meses)	Risco / Impacto Gerencial
Viés (Manter EVA Ativo, ciente do problema)	Projeção de Passivo Contingente Legal (Multas e Ações Coletivas), o que equivale de 15% a 25% do EBITDA anual.	Aumento do Custo Residual e falha de Responsabilidade Fiduciária e Governança Corporativa.
Rigidez (Suspensão Imediata do EVA)	Projeção de Queda da Margem Operacional de 12% para 8% no 1º trimestre (Retorno ao processo manual).	Custo de Oportunidade / Aumento temporário dos Custos de Monitoramento humano.
Opacidade (Início do Plano de Correção Técnica)	Custo Estimado para Framework XAI e Contratação de Especialistas (3-6 meses) o que equivale a R\$ 4,5 milhões.	Custo de Vínculo e Monitoramento (Investimento em infraestrutura de Governança de IA).

Parte 2 – Notas de Ensino

2.1 Fontes de dados

O caso foi desenvolvido com base em uma construção fictícia (FinData Bank e o modelo EVA) para contextualizar o dilema. O dilema central, no entanto, é inspirado em eventos reais e de domínio público, como os problemas enfrentados pela Amazon em seu sistema de recrutamento e as acusações de discriminação de gênero contra a Apple Card na

definição de limites de créditos. O embasamento teórico se apoia na Teoria da Agência (Jensen e Meckling, 1976) e, adicionalmente, por literatura contemporânea sobre os custos de agência algorítmicos (opacidade, viés e rigidez), que surgem na delegação de decisões a sistemas autônomos.

2.2 Objetivos educacionais

Este caso visa desenvolver a capacidade analítica e de tomada de decisão do discente, com foco nos seguintes objetivos:

- 1) Análise crítica da aplicabilidade e dos desafios da Teoria da Agência quando o agente é um sistema de Inteligência Artificial;
- 2) Identificação e discussão aprofundada dos custos de agência algorítmicos (opacidade, viés e rigidez) gerados por sistemas de decisão autônomos;
- 3) Avaliação da adequação e das limitações dos SCG tradicionais diante da assimetria de informação causada pela "caixa-preta" da IA; e
- 4) Reflexão e proposição de mecanismos de governança e *accountability* para mitigar riscos éticos, legais e de reputação corporativa na delegação de decisões para agentes IA.

2.3 Questões para discussão proposta aos discentes

Propõe-se as seguintes questões para discussão com os discentes:

- 1) Como a implementação do modelo EVA transformou a relação Principal-Agente na FinData Bank? O EVA pode ser considerado um “agente ideal” que resolve os conflitos de interesses ou ele apenas criou um novo tipo de problema de agência?
- 2) Descreva e analise os principais custos de agência algorítmicos (opacidade, viés e rigidez) presentes no caso. De que forma esses custos se manifestam nos indicadores de desempenho e nas observações éticas apresentadas pela CEO Helena Prado?
- 3) O SCG da FinData Bank não foi eficaz para monitorar o EVA. Quais mecanismos de controle são necessários para mitigar a assimetria de informação gerada pela “caixa-preta” algorítmica e garantir que os objetivos éticos da empresa sejam atingidos?
- 4) Considerando a urgência imposta pelo Comitê de Auditoria (48 horas) e os indicadores de desempenho, incluindo o Anexo II (Projeção de Risco), elabore um Plano de Ação Imediato para Helena Prado. Analise o *trade-off* entre a manutenção do EVA e a suspensão do EVA. Justifique sua recomendação sob a ótica da Governança de IA, da responsabilidade fiduciária e do potencial impacto nos relatórios contábeis de contingências.

2.4 Embasamento teórico para sustentação do caso

2.4.1 Teoria da Agência e Sistema de Controle Gerencial

Com base na Teoria da Agência, a empresa é entendida como um "nexo de contratos" focado na relação em que um principal delega autoridade a um agente. Essa relação é inerentemente complexa devido a dois fatores: (a) a assimetria de informação – o agente possui conhecimento privado inacessível ao principal; e (ii) o conflito de interesses – surge quando o agente, buscando maximizar sua própria utilidade, adota objetivos que divergem dos interesses do principal (Jensen & Meckling, 1976).

Essa divergência de interesses gera os chamados custos de agência, que representam uma perda de valor para o principal. Tais custos são classificados em três tipos: (a) monitoramento – gastos do principal para controlar o comportamento do agente; (b) concessão de garantias contratuais; e (c) perda residual – a redução no bem-estar do principal que persiste mesmo após a aplicação de mecanismos de controle e sinalização (Jensen & Meckling, 1976).

Neste contexto, os Sistemas de Controle Gerencial (SCG), que emergem como o principal mecanismo organizacional para mitigar os custos de agência (Anthony & Govindarajan, 2008). Por meio de sistemas de mensuração de desempenho, planejamento, orçamento e sistemas de incentivo, os SCG influenciam o comportamento humano, reduzindo a assimetria de informação e assegurando que as ações individuais, pautadas em metas pessoais, também ajudem a cumprir as metas da organização, alinhando, desse modo, os interesses do agente com os do principal (Anthony & Govindarajan, 2008).

Contudo, Chenhall (2003) destaca que é improvável que os conceitos de SCG abordem questões contemporâneas pertinentes, sem estimar as mudanças e condições históricas organizacionais. Desse modo, é possível asseverar que conceitos são transitórios e dinâmicos, devendo ser estabelecidos e revisados de acordo com a realidade apreendida temporalmente do objeto/fenômeno a que se referem. Uma mudança recente e muito relevante no controle organizacional é o emprego de algoritmos para tal fim. Nesse sentido, autores como Schildt (2017), Cheng e Foley (2019) e Avelar et al. (2022) apontam que há um aumento de decisões baseadas em algoritmos pelas organizações para tornar o trabalho mais eficiente mediante mecanismos de automação.

2.4.2 A Delegação à Inteligência Artificial: Um Novo Paradigma de Agência

A crescente autonomia dos sistemas de IA, que evoluem de meros executores de tarefas para tomadores de decisão, posiciona-os como uma nova categoria de agente na teoria organizacional (Shrestha et al., 2019). Este fenômeno, denominado "gestão algorítmica", utiliza algoritmos para monitorar, alocar tarefas e direcionar a força de trabalho (Schildt, 2017; Oliveira & Avelar, 2023), prometendo ganhos de eficiência e objetividade supostamente livre de vieses. A delegação de poder a esses agentes não humanos, no entanto, desafia diretamente os pilares da Teoria da Agência, que analisa os conflitos de interesse na relação entre "principal" e "agente" (Eisenhardt, 1989; Jensen & Meckling, 1976).

Essa mudança reconfigura a percepção de risco na delegação, com estudos demonstrando uma preferência maior por delegar decisões a agentes de IA do que a humanos, especialmente em cenários de alta complexidade ou perdas potenciais (Candrian & Scherer, 2022). A principal justificativa para essa confiança é a percepção de que a IA não possui "ação intencional" nem "motivos autointeressados", eliminando o temor do principal de ser traído ou explorado por um agente oportunista. Essa visão é corroborada por outros autores, que sugerem o potencial da IA para aprimorar a governança e mitigar fraudes, atuando como um agente de controle percebido como mais confiável (Rehman, 2022).

No entanto, essa aparente confiabilidade mascara uma nova classe de riscos sistêmicos e não intencionais, que emergem mesmo quando a IA opera conforme o projetado (Tamò-Larrieux, 2021). Esses riscos originam-se em um conjunto de fatores interligados: (i) Mecanismos internos da tecnologia – modelos de aprendizado de máquina, como redes neurais profundas, operam como uma “caixa-preta”, baseando decisões em correlações complexas que dificultam a transparência, previsibilidade e justificativa (Shrestha et al., 2019; Tamò-Larrieux, 2021); (ii) Ambiente social e dados de treinamento – como aprendem a partir de dados históricos, os algoritmos incorporam vieses e desigualdades estruturais já presentes na sociedade, podendo reproduzir ou mesmo intensificar práticas indesejadas (Tamò-Larrieux, 2021); e (iii) Forças econômicas e organizacionais – a busca por eficiência e redução de custos favorece a adoção de sistemas algorítmicos rígidos, projetados para condições específicas, mas com baixa capacidade de adaptação a mudanças ambientais, eventos inesperados ou contextos distintos daqueles para os quais foram treinados (Schildt, 2017).

Assim, governança da Inteligência Artificial (IA) apresenta desafios centrais em ética, transparência e imparcialidade, exigindo que as organizações invistam em qualificação

profissional, infraestrutura tecnológica e políticas claras para mitigar riscos (Oliveira & Avelar, 2023). Partindo desse cenário, o presente caso levanta a hipótese de que a delegação de decisões a sistemas de IA gera novas formas de custos de agência, as quais coexistem com as tradicionais e demandam adaptações nos SCG. A literatura sobre o tema aponta três dimensões centrais para esses novos custos: (a) Custos de opacidade – derivam da natureza "caixa-preta" de muitos sistemas de IA, cuja lógica decisória complexa dificulta a transparência, a auditoria e a justificativa das decisões (Shrestha et al., 2019; Tamò-Larrieux, 2021). Essa incapacidade de escrutínio gera incerteza e impede que gestores compreendam ou corrijam falhas no raciocínio algorítmico, criando um risco de perda de controle sobre processos críticos; (b) Custos de viés – referem-se aos custos tangíveis e intangíveis que surgem de decisões algorítmicas sistematicamente injustas ou discriminatórias (Barocas & Selbst, 2016; Tamò-Larrieux, 2021). Além dos gastos com remediação dos sistemas, esses custos englobam severos riscos legais e reputacionais. Casos como a investigação do Apple Card por suposto viés de gênero e o abandono do algoritmo de recrutamento da Amazon, que penalizava currículos de mulheres, são exemplos emblemáticos (Weizenmann, 2025); e (c) Custos de rigidez – resultam da dificuldade dos modelos em se adaptar a mudanças contextuais. Um exemplo notório é a falha do algoritmo da Zillow que, treinado em um mercado imobiliário aquecido, manteve a supervalorização de imóveis durante a desaceleração do setor, o que gerou perdas de milhões de dólares. O excesso de foco em rotinas otimizadas também pode induzir à rigidez estratégica, limitando a capacidade de inovação (Schildt, 2017; Weizenmann, 2025).

A introdução da IA no centro da tomada de decisão gera uma dualidade para o controle gerencial: embora a gestão algorítmica prometa maior objetividade e eficiência (Schildt, 2017), os custos de agência algorítmicos, especialmente os de opacidade e viés, podem reintroduzir a necessidade de supervisão e revisão e julgamento qualitativo e subjetivo humano.

2.5 Disciplinas sugeridas para o uso do caso

Tendo em vista a abordagem central da Teoria da Agência, Sistemas de Controle Gerencial e Governança de IA, a aplicação deste caso é sugerida para os programas de pós-graduação e graduação em Administração e Contabilidade nas disciplinas de Contabilidade Gerencial e Controladoria, Teoria da Contabilidade, Sistemas de Informação Contábil, Ética e Responsabilidade Social Corporativa, Governança Corporativa e Gestão de Riscos, dentre outras.

2.6 Sugestão de plano de aula

A aplicação do caso de ensino no decorrer de 3 aulas de 50 minutos cada garante profundidade e compreensão dos conceitos abordados, permitindo aos discentes a oportunidade de refletir, discutir e aplicar o conhecimento de forma prática e crítica, conduzindo-os à proposição de soluções complexas.

Tabela 3

Sugestão Plano de Aula

Aula	Tópico	Atividade e Conteúdo
1ª aula	Abertura do Caso: Agência Clássica e IA	Nesta aula, os discentes serão apresentados à Teoria da Agência Clássica, focando na assimetria de informação e nos custos de agência. O docente irá provocar a discussão sobre como a IA (agente não-humano) é percebida como uma solução para os problemas de agência humana (redução de viés e custos de monitoramento). O caso FinData Bank será introduzido neste contexto.
2ª aula	Desenvolvimento do Caso: Custos	O docente apresenta o caso e os dados do EVA. Os discentes, divididos em grupos, irão debater e formular as respostas às Questões 1 e 2, focando na identificação dos Custos de Agência Algorítmicos (opacidade, viés e rigidez) e como eles se manifestam nos

	Algorítmicos e SCG	indicadores de negação de crédito. Um representante de cada grupo apresentará as conclusões.
3ª aula	Fechamento do Caso: Governança e Plano de Ação	A discussão será centrada na falha do SCG tradicional (Questão 3) e na urgência da decisão da CEO (Questão 4). O foco é o debate sobre a Governança de IA e as implicações de responsabilidade corporativa. Os grupos deverão consolidar um Plano de Ação Imediato para o Comitê de Auditoria (suspender/revisar o EVA), justificando a decisão em termos de mitigação de riscos e alinhamento ético.

2.7 Sugestão de respostas das questões

Abaixo seguem as sugestões de respostas às questões apresentadas no caso:

1) Como a implementação do modelo EVA transformou a relação Principal-Agente na FinData Bank? O EVA pode ser considerado um “agente ideal” que resolve os conflitos de interesses ou ele apenas criou um novo tipo de problema de agência?

Resposta sugerida: O EVA não se configurou como um agente ideal. Embora tenha reduzido o tempo de análise e a inadimplência, ele substituiu os conflitos de agência tradicionais por um Problema de Agência Algorítmico. O agente humano (o analista de crédito) foi substituído por um agente autônomo, o EVA, que impôs uma nova assimetria de informação, a "caixa-preta". O Principal perdeu a capacidade de monitorar totalmente as ações do agente EVA, gerando custos não mais ligados à desonestidade humana, mas ao viés, opacidade e rigidez algorítmica

2) Descreva e analise os principais custos de agência algorítmicos (opacidade, viés e rigidez) presentes no caso. De que forma esses custos se manifestam nos indicadores de desempenho e nas observações éticas apresentadas pela CEO Helena Padro?

Resposta sugerida: Os custos de agência algorítmicos são os passivos gerados pela delegação a sistemas autônomos:

a) Opacidade (caixa-preta): Manifesta-se na incapacidade do Comitê de Auditoria de entender completamente a lógica de decisão do EVA. Não é possível rastrear o porquê de um crédito ser negado, dificultando a correção e a responsabilização.

b) Viés: O algoritmo foi treinado com dados históricos que já continham desigualdades sociais. Ao reproduzir esses padrões, ele amplificou a discriminação. Isso se manifesta no aumento drástico da negação de crédito a mulheres (de 22% para 37%) e em regiões específicas (de 18% para 33%).

c) Rigidez: Manifesta-se na dificuldade de ajustar o sistema rapidamente. Rafael Torres, Diretor de Tecnologia, sugere que os ajustes podem levar meses, enquanto o EVA continua a causar danos éticos e reputacionais.

3) O SCG da FinData Bank não foi eficaz para monitorar o EVA. Quais mecanismos de controle são necessários para mitigar a assimetria de informação gerada pela “caixa-preta” algorítmica e garantir que os objetivos éticos da empresa sejam atingidos?

Resposta sugerida: O SCG tradicional da FinData Bank foi ineficaz no monitoramento do EVA. O SCG é focado em ações humanas e estruturas claras, mas a opacidade da IA ("caixa-preta") impediu o monitoramento eficaz, transformando os ganhos de eficiência em riscos de passivos. Para mitigar a assimetria algorítmica, é necessária a implementação de um framework de Governança de IA. Mecanismos de controle essenciais incluem:

a) Auditoria Contínua: Monitoramento automatizado em tempo real do desempenho do EVA em métricas de justiça e não apenas de eficiência (inadimplência);

b) Explicabilidade (XAI): Criação de relatórios técnicos sobre os fatores que mais influenciaram as decisões, buscando trilhas de auditoria que, mesmo que não sejam transparentes por completo, permitam a detecção de anomalias e vieses; e

c) Transparência de Dados: Garantia de que os dados utilizados no treinamento sejam revisados e corrigidos para eliminar vieses históricos antes de serem reintroduzidos no modelo.

4) Considerando a urgência imposta pelo Comitê de Auditoria (48 horas) e os indicadores de desempenho, incluindo o Anexo II (Projeção de Risco), elabore um Plano de Ação Imediato para Helena Prado. Analise o *trade-off* entre a manutenção do EVA e a suspensão do EVA. Justifique sua recomendação sob a ótica da Governança de IA, da responsabilidade fiduciária e do potencial impacto nos relatórios contábeis de contingências.

Resposta sugerida: O Plano de Ação deve priorizar a responsabilidade fiduciária e a mitigação do Custo Residual, que, conforme o Anexo II, pode ser de 15% a 25% do EBITDA, ou seja, muito significativo. Dessa forma, tem-se a seguinte proposta:

a) Imediatamente, implementar um processo de participação humana ativa para todas as negações de crédito emitidas pelo EVA, especialmente aquelas envolvendo grupos vulneráveis (mulheres e regiões específicas). O EVA se mantém ativo, mas a decisão final de negação de crédito retorna ao analista humano.

b) Emitir nota pública transparente ao mercado e *stakeholders* reconhecendo a detecção de vieses sistemáticos e anunciando a suspensão do processo de decisão autônoma (Agência Híbrida) até a conclusão do *re-training* do modelo.

c) Aprovar o investimento de R\$ 4,5 milhões para o Framework XAI e *retraining* (Custos de Vínculo e Monitoramento, Anexo II).

A proposta acima parte do entendimento de que manter o EVA totalmente ativo, ciente dos vieses sistemáticos, configura uma falha de Governança e maximiza o Custo Residual (risco de multas e ações). A perda imediata de margem operacional (Custo de Oportunidade) é um custo de monitoramento e vínculo temporário, necessário e menor do que o potencial passivo legal. A suspensão/intervenção humana (Agência Híbrida) é o único caminho imediato para restaurar o alinhamento ético e demonstrar compromisso com a *accountability* corporativa, evitando o agravamento da crise de reputação.

Agradecimentos

Agradecimento à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento do presente caso de ensino.

Referências

- Anthony, R. N.; Govindarajan, V. (2008). *Sistemas de controle gerencial*. São Paulo: McGraw-Hill.
- Avelar, E. A., Jordão, R. V. D., Ferreira, G. M. C., & da Silva, B. N. E. R. (2022). O papel da gestão por algoritmos como suporte para os sistemas de controle gerencial na economia compartilhada: um estudo sobre a percepção dos motoristas de empresas do transporte por aplicativo no contexto brasileiro. *Revista de Contabilidade e Organizações*, 16, e182036-e182036.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.

- Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, 134, 107308. <https://doi.org/10.1016/j.chb.2022.107308>
- Cheng, M. & Foley, C. (2019). Algorithmic management: The case of Airbnb. *International Journal of Hospitality Management*, 83(1), 33–36. doi: <https://doi.org/10.1016/j.ijhm.2019.04.009>.
- Chenhall, R. H. (2003). Management control systems design within its organizational context: findings from contingency-based research and directions for the future. *Accounting, Organizations and Society*, 28(2–3), 127–168. doi: [https://doi.org/10.1016/S0361-3682\(01\)00027-7](https://doi.org/10.1016/S0361-3682(01)00027-7).
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74.
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
- Marshall, E. D. (2021, 23 de março). *Apple e Goldman são inocentados em caso sobre discriminação em oferta de cartões*. CNN Brasil. <https://www.cnnbrasil.com.br/economia/macroeconomia/apple-e-goldman-nao-cometeram-violacoes-em-oferta-de-cartoes-de-credito-dizem-a/>
- Matsuura, S.. (2019, novembro 17). Algoritmos reproduzem machismo e racismo por se basearem em práticas discriminatórias dos humanos. *O Globo*. <https://oglobo.globo.com/economia/algoritmos-reproduzem-machismo-racismo-por-se-basearem-em-praticas-discriminatorias-dos-humanos-24085081>
- Oliveira, C., & Avelar, E. A. (2023). A era dos algoritmos de inteligência artificial no controle gerencial. *Revista Mineira de Contabilidade*, 24(2), 4–6. <https://doi.org/10.51320/rmc.v24i2.1543>
- Rehman, A. U., Waheed, A., & Farooq, M. (2022). With application of agency theory, can artificial intelligence eliminate fraud risk? A conceptual overview. *Journal of Risk and Financial Management*, 15(9), 399. https://doi.org/10.1007/978-981-19-1036-4_8
- Schildt, H. (2017). Big data and organizational design - the brave new world of algorithmic management and computer augmented transparency. *Innovation*, 19(1), 23–30. <https://doi.org/10.1080/14479338.2016.1252043>
- Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. <https://doi.org/10.1177/0008125619862257>
- Tamò-Larrieux, A. (2021). Decision-making by machines: Is the 'Law of Everything' enough? *Computer Law & Security Review*, 41, 105541. <https://doi.org/10.1016/j.clsr.2021.105541>
- Weizenmann, H. (2025, 31 de julho). *14 grandes desastres com IA que chamaram a atenção*. Adrenaline. <https://www.adrenaline.com.br/ia/14-grandes-desastres-com-ia-que-chamaram-a-atencao/>