# AI Operations

Um framework prático para transformar capacidade de IA em valor de negócio

Roberto Shimizu

Abril de 2026

## Resumo Executivo

A maior parte dos programas corporativos de IA não falha porque o modelo é fraco. Falha porque o sistema de trabalho ao redor é fraco. Os fluxos não são redesenhados, os baselines não são estabelecidos, a responsabilidade continua fragmentada e a governança permanece no nível da política, em vez de ser incorporada ao comportamento em produção. O resultado é um padrão conhecido: demos fortes, pilotos ativos e resultados operacionais decepcionantes.

Este texto defende que as organizações precisam de uma disciplina de gestão específica para transformar capacidade de IA em valor de negócio. Eu chamo essa disciplina de *AI Operations*. Ela combina disciplina de medição, redesenho de workflow, estrutura semântica, desenho de adoção e governança em um único modelo operacional. O objetivo não é ter mais atividade com IA. É obter desempenho confiável em produção.

## 1 Introdução

A IA corporativa tem um problema curioso: capacidade já não é escassa, mas valor de negócio confiável continua sendo. Hoje, a maior parte das grandes organizações consegue acessar bons modelos fundacionais, comprar copilotos prontos e rodar pilotos com relativa facilidade. O que continua difícil é transformar esse acesso em ganhos operacionais repetíveis. A pesquisa global da McKinsey de 2025 mostrou que o uso de IA já é disseminado, enquanto impacto relevante de última linha em escala empresarial ainda permanece concentrado em uma minoria de empresas. A BCG descreveu o mesmo padrão em termos mais operacionais: um pequeno grupo gera valor material, um grupo maior está escalando com retorno parcial, e a maioria ainda vê pouco impacto mensurável apesar do investimento ativo.[5, 1]

Esse gap costuma ser descrito como problema de adoção, problema de governança ou problema de "última milha". As três descrições ajudam, mas nenhuma é suficiente. O que eu continuo vendo na prática é mais simples e mais teimoso: as organizações estão tentando encaixar IA em sistemas operacionais que nunca foram redesenhados para absorvê-la.

No IntelliDoctor, uma implementação hospitalar de IA na qual ajudei a construir, o sinal técnico não foi a parte mais difícil de demonstrar. Em um workflow específico de documentação clínica, a qualidade da conduta gerada por IA foi avaliada em 8,0/10 contra uma linha de base de 3,5/10 para

médicos, e os médicos aceitaram recomendações em 78,5% dos casos. Mesmo assim, a pergunta mais difícil começou depois que o modelo funcionou bem. Qual workflow deveria mudar primeiro? Quem era dono das exceções? Qual métrica realmente importava? Como a liderança deveria julgar valor com confiança, e não com entusiasmo? O problema operacional ficou visível justamente onde a IA já funcionava.

Essa é a tese central deste texto: a maior parte das iniciativas de IA falha operacionalmente antes de falhar tecnicamente. E, por "operacionalmente", eu não quero dizer apenas uptime. Quero dizer o sistema de trabalho ao redor: baseline, desenho do workflow, ownership, consistência semântica, comportamento de adoção e governança em produção. Um modelo pode ser acurado e, ainda assim, destruir valor se adicionar retrabalho, custo de supervisão, custo de escalonamento ou ambiguidade organizacional.

Este texto propõe *AI Operations* como a disciplina de gestão que falta entre ambição com IA e performance de negócio. O termo é útil porque os rótulos vizinhos não cobrem totalmente o trabalho. Linguagem estratégica é abstrata demais. MLOps e LLMOps são tecnicamente estreitos demais. Governança costuma ser estática demais. Gestão da mudança, sozinha, diz pouco sobre instrumentação, semântica ou controle em produção.

O argumento avança em quatro movimentos. Primeiro, explica por que programas de IA normalmente travam no sistema operacional ao redor do modelo, e não no modelo em si. Segundo, define AI Operations como uma disciplina prática de gestão. Terceiro, argumenta que a camada semântica não é arquitetura opcional, mas infraestrutura de reúso, controle e aprendizado institucional. Por fim, traduz o framework em um conjunto curto de testes executivos para decidir se uma iniciativa de IA é real ou apenas teatro.

## 2 Por que a IA falha operacionalmente antes de falhar tecnicamente

A maior parte dos programas de IA não quebra no ponto da inferência. Quebra no ponto da operação. Um demo convincente gera crença. Um piloto prova utilidade local. Depois, a escala expõe as condições reais do trabalho empresarial: handoffs ambíguos, papéis legados, restrições de compliance, sistemas fragmentados, caminhos de exceção escondidos e gestores que nunca foram forçados a definir como seria, de fato, o sucesso. A *Harvard Business Review* descreveu isso recentemente como o problema da "última milha" da transformação com IA e capturou bem o padrão: muitas organizações estão ficando ricas em pilotos e pobres em transformação.[3]

A pergunta decisiva, portanto, não é se o modelo performa isoladamente. É se o workflow performa sob condições reais de operação. Esse é um padrão mais duro. Ele pergunta se o sistema melhora throughput, qualidade, compliance, economia ou julgamento quando entra na bagunça do trabalho real. Muitos programas de IA nunca atingem esse padrão porque confundem sucesso de protótipo com desenho operacional.

### A armadilha do piloto

Pilotos são úteis. Eles reduzem incerteza, revelam edge cases e constroem confiança interna. Mas um piloto não é um modelo operacional. Em geral, ele roda com usuários escolhidos a dedo, atenção

concentrada e heroísmo implícito do time mais próximo do sistema. Essas condições não sobrevivem à escala.

É por isso que modelos melhores, sozinhos, raramente resolvem o problema de implantação. Um modelo mais forte pode melhorar a qualidade das respostas na margem, mas não decide quem age sobre a resposta, quando a resposta é confiável, como as exceções são tratadas ou qual processo downstream absorve a mudança. A pesquisa da McKinsey de 2025 mostrou que as organizações com maior impacto de IA eram muito mais propensas a redesenhar workflows em vez de simplesmente inserir ferramentas em rotinas já existentes.[5] Isso não é nota de rodapé técnica. É a diferença entre uma feature e uma capacidade operacional.

## É no desalinhamento do workflow que a economia vira

Um assistente tecnicamente impressionante ainda pode ser economicamente inútil. Se ele adiciona mais uma etapa de revisão, mais uma incerteza ou mais um ponto em que um humano precisa traduzir entre sistemas, o workflow pode desacelerar mesmo quando o modelo parece rápido. Essa é a aritmética oculta da IA corporativa: velocidade de geração é óbvia; custo de supervisão é mais fácil de ignorar.

Na prática, a IA muitas vezes desloca trabalho, em vez de simplesmente eliminá-lo. Ela move esforço para verificação, escalonamento e julgamento. Isso pode ser benéfico quando o workflow é redesenhado em torno desses novos pontos de controle. Torna-se destrutivo quando o processo antigo permanece intacto e a IA é apenas colocada por cima. O volume de saída sobe. O trabalho oculto de revisão sobe junto. A confiança cai.

Em saúde, isso fica visível rapidamente porque o desalinhamento do workflow não permanece teórico por muito tempo. Uma recomendação preliminar só é útil se o caminho clínico ao redor souber quando adotá-la, quando questioná-la e quem é responsável quando o caso sai do caminho esperado. No IntelliDoctor, essa distinção apareceu repetidamente. O problema raramente era se o sistema conseguia gerar uma sugestão clinicamente relevante. O problema era se a sugestão aterrissava dentro de um workflow desenhado para absorvê-la.

## Capacidade não é a mesma coisa que absorção

Uma das distinções mais úteis para times executivos é a diferença entre *capacidade* e *absorção*. Capacidade é o que o modelo consegue gerar. Absorção é o que a organização consegue receber, rotear, verificar e reutilizar de forma confiável.

A maioria das empresas superestima a primeira e subinveste na segunda. Conhecimento operacional crítico continua vivendo na cabeça das pessoas: qual especialista costuma receber certo encaminhamento, qual diagnóstico muda a exposição de faturamento, qual protocolo importa para segurança da alta, qual exceção exige escalonamento e qual saída é segura o suficiente para seguir sem revisão. Quando esse conhecimento permanece tácito, a IA só consegue operar como um assistente fino nas bordas. Ela não participa de forma confiável do sistema real de decisão.

É por isso que tantas implantações parecem promissoras no laboratório e frágeis na produção. O modelo é capaz, mas a organização ainda não tornou sua própria lógica de julgamento legível. As

pessoas então interpretam as falhas resultantes como imaturidade do modelo, quando o problema mais profundo é opacidade institucional.

### Vácuos de medição criam teatro de IA

A ausência de medição é uma das maiores razões pelas quais programas fracos de IA sobrevivem tempo demais e programas promissores têm dificuldade para conquistar patrocínio. Se ninguém estabelece a linha de base do workflow atual, quase qualquer piloto pode alegar progresso. Se ninguém define a métrica econômica, ninguém consegue dizer se a melhora na qualidade de saída realmente importou. Se ninguém separa uso experimental de mudança comportamental duradoura, a adoção vira exagero.

É aí que começa o teatro de IA. As organizações reportam número de pilotos, número de usuários, nomes de modelos, contratos com vendors e entusiasmo anedótico. O que elas não conseguem reportar com clareza é se o workflow melhorou e quanto melhorou.

A experiência com o IntelliDoctor tornou isso concreto para mim. Os ganhos técnicos eram visíveis. O que a gestão ainda precisava era uma linguagem operacional compartilhada para falar de valor. O comportamento médico mudou de forma duradoura? A qualidade da documentação melhorou de um jeito que reduzisse perda downstream? A captura de encaminhamentos melhorou? A visibilidade sobre compliance de protocolo aumentou? Quais dessas mudanças eram repetíveis, e em que condições? Essas são perguntas mais difíceis do que "o modelo funciona?", mas são as únicas que importam quando dinheiro real, risco e accountability entram em cena.

### Adoção e ownership determinam se o valor se acumula

Sistemas de IA não têm sucesso porque os usuários ficaram impressionados. Têm sucesso porque as pessoas sabem quando confiar, quando verificar e quando escalar. Um usuário a quem se diz apenas que o sistema é "inteligente" não recebeu orientação operacional. Recebeu ambiguidade.

Ownership importa tanto quanto. Em muitas organizações, ninguém é dono do comportamento ponta a ponta do workflow habilitado por IA. Tecnologia é dona da ferramenta. Uma área de negócio é dona do processo nominal. Risco é dono da política. Ninguém é dono da performance em produção. Quando os resultados decepcionam, o modelo leva a culpa porque é o artefato mais visível. A falha real era accountability fragmentado.

Um pequeno exemplo clínico ilustra isso. Em um caso relacionado a cefaleia, o protocolo relevante estava disponível e o sistema trouxe a orientação operacional correta, mas a recomendação não foi seguida e o paciente retornou em 24 horas. A lição não foi que o modelo precisava de mais benchmark. A lição foi que adoção, controle de workflow e disciplina de escalonamento estavam incompletos.

É por isso que a maior parte das falhas sérias de IA é operacional antes de ser técnica. Se desenho de workflow, medição, captura de conhecimento tácito, regras de adoção e ownership permanecem sem gestão, modelos melhores basicamente ampliam complexidade sem gestão. Esse é o problema que AI Operations se propõe a resolver.

# 3 AI Operations como disciplina de gestão

Se a IA vai importar operacionalmente, ela precisa de uma disciplina operacional, e não apenas de uma estratégia de modelos. AI Operations é essa disciplina: a prática de desenhar, medir, governar e melhorar continuamente trabalho habilitado por IA para que capacidade técnica se transforme em desempenho confiável de negócio.

O termo merece existir porque os rótulos vizinhos deixam lacunas importantes. MLOps e LLMOps importam, mas não cobrem por inteiro redesenho de workflow, comportamento de adoção, consistência semântica, instrumentação de negócio ou controle executivo. "Transformação" é abstrato demais. "Automação" é estreito demais. "Governança" costuma ser estática demais e distante da produção. AI Operations nomeia a camada em que a intenção executiva ou vira realidade operacional ou se dissolve em clutter de pilotos.

A forma mais clara de descrever a disciplina é por três movimentos: o que ela mantém, o que descarta e o que adiciona.

### O que AI Operations mantém

Os melhores instintos das disciplinas operacionais anteriores continuam válidos.

- Definir o resultado de negócio antes de celebrar atividade local.
- Estabelecer baseline antes de reivindicar melhoria.
- Procurar o gargalo no sistema, e não apenas o defeito no artefato.
- Tratar aprendizado em produção como ativo operacional.
- Exigir controle depois do deploy, e não apenas esforço antes do deploy.

Nada disso é nostalgia. É apenas verdade operacional que continua valendo.

### O que AI Operations descarta

O que não deve ser carregado adiante importa tanto quanto. Eu não estou defendendo uma volta de hierarquias de belts, teatro de certificação ou tollgates burocráticos disfarçados de rigor. Muito do antigo mundo de processos degenerou em scores abstratos de maturidade, documentação ritualizada e jogos de métrica desconectados da criação de valor. Esses hábitos criaram anticorpos merecidos.

Programas de IA já estão vulneráveis ao mesmo padrão de falha. É fácil produzir policy binders, templates de avaliação e comitês de governança que parecem sérios enquanto permanecem desconectados do comportamento do workflow. AI Operations rejeita esse teatro. O trabalho da disciplina é tornar implantações de IA operacionalmente legíveis e controláveis.

### O que a era da IA adiciona

A IA muda a forma do controle operacional, então a disciplina precisa adicionar capacidades que sistemas operacionais anteriores não gerenciavam explicitamente.

- **Desenho de tarefa e prompt como desenho de workflow.** Em muitos workflows, o prompt não é apresentação. É parte da instrução de trabalho.
- **Fronteiras de revisão.** A pergunta central não é se humanos continuam envolvidos, mas onde seu julgamento agrega valor e onde sua revisão vira desperdício.
- **Qualidade de contexto.** A performance da IA é muitas vezes limitada menos pela força bruta do modelo do que por receber ou não o contexto certo, no momento certo.
- **Roteamento de exceções.** Nenhum workflow sério é feito só do caso médio. Edge cases precisam de caminhos explícitos, donos e gatilhos de revisão.
- **Estrutura semântica.** Significado compartilhado não é adorno. É a camada que torna o comportamento da IA reutilizável, auditável e melhorável entre workflows.
- **Telemetria operacional.** Acurácia em isolamento é insuficiente. Times precisam de visibilidade sobre overrides, escalonamentos, latência, retrabalho downstream, exposição de compliance e qualidade da adoção.

O objetivo não é autonomia máxima. É desempenho confiável sob restrição.

### Os cinco requisitos de AI Operations

Para fins executivos, AI Operations pode ser reduzido a cinco requisitos. Se um deles faltar, a iniciativa ainda pode produzir demos, mas dificilmente produzirá valor durável.

1. **Um workflow definido.** A iniciativa precisa atacar um workflow específico, e não uma capacidade genérica.
2. **Um baseline e uma métrica de valor.** Alguém precisa saber como a performance está hoje e o que deve melhorar.
3. **Um desenho de controle.** Ownership humano, fronteiras de revisão e caminhos de exceção precisam estar explícitos.
4. **Uma estrutura semântica compartilhada.** O sistema precisa saber o que seus conceitos-chave significam entre contextos e sistemas, e em quais fontes esses significados se apoiam.
5. **Um loop de aprendizado em produção.** A organização precisa conseguir observar comportamento, aprender com drift e exceções, e melhorar o workflow em produção.

Esse é o modelo executivo compacto. Benchmark de modelo pode dizer algo sobre capacidade. Política de governança pode expressar princípios. Nenhum dos dois substitui definição de workflow, desenho de controle, estrutura semântica ou instrumentação dentro do trabalho.

### Um método operacional prático

A disciplina fica mais útil quando traduzida em um ritmo operacional:

1. escolher um workflow em que dor e valor já são visíveis,

2. estabelecer a linha de base da performance atual, da adoção e dos modos de falha,
3. localizar o gargalo, os pontos de decisão e o conhecimento tácito de que o workflow depende,
4. definir a estrutura semântica: conceitos-chave, relações, regras de proveniência e fronteiras de exceção,
5. redesenhar o workflow em torno dos pontos fortes da IA, do julgamento humano e de controles explícitos,
6. implantar, monitorar e melhorar em produção.

Esse ciclo é simples de propósito. As organizações querem a aceleração da IA sem o incômodo do redesenho. Querem decisões melhores sem tomar decisões sobre ownership, medição, contexto e tratamento de exceções. AI Operations insiste que essas escolhas sejam feitas explicitamente.

O método também deve produzir artefatos, não apenas reuniões de alinhamento: um mapa de workflow, um scorecard de baseline, uma canvas semântica ou ontológica, regras de exceção, uma visão de adoção e uma cadência de revisão de valor e drift. Com o tempo, esses artefatos podem ser apoiados por ferramentas reutilizáveis em vez de serem recriados a cada iniciativa.

É aí que a camada semântica deixa de ser apenas um componente técnico. Ela faz parte do trabalho de *Definir*, porque explicita o domínio operacional, e do trabalho de *Controlar*, porque oferece uma superfície estável para regras de revisão, proveniência e memória institucional. Quando essa camada existe, novos casos de uso de IA passam a se parecer menos com projetos inéditos e mais com novas configurações construídas sobre significado compartilhado.

### Por que isso é um sistema de gestão

Um sistema de gestão define o que importa, como a performance é vista, quem é dono do resultado e como o aprendizado retorna à operação. É exatamente isso que falta à maior parte dos programas de IA. A pesquisa de Thomas Davenport e Laks Srinivasan, em 2026, é útil aqui porque mostra que empresas com maior disciplina em avaliação, agregação e reporte de valor são mais propensas a relatar retornos relevantes com IA. Eles também encontraram melhores resultados quando a função financeira ajuda a possuir ou certificar valor.[2] A cadeia causal pode ser debatida, mas a lição gerencial é clara: realização de valor melhora quando alguém tem autoridade e rigor para medi-lo.

Isso também esclarece o que AI Operations não é. Não é benchmark de modelo, automação genérica nem teatro de transformação. E não é puramente MLOps, porque um modelo perfeitamente gerido ainda pode estar dentro de um sistema operacional mal gerido.

AI Operations é a camada de gestão que falta entre possibilidade técnica e valor operacional.

## 4 Significado compartilhado como infraestrutura operacional

A IA corporativa se torna frágil no momento em que cada time inventa seu próprio significado em isolamento. Um prompt chama um caso de urgente. Outro chama de complexo. Um dashboard agrupa encaminhamentos de um jeito, um gestor operacional agrupa de outro, e o modelo recupera

contexto a partir de documentos que nunca foram escritos para sustentar interpretação consistente. O resultado é conhecido: lógica duplicada, integrações frágeis, saídas inconsistentes e discussões intermináveis sobre por que o sistema produziu determinada resposta.

É por isso que a camada semântica importa. Em termos práticos, ela é a estrutura compartilhada de significado de negócio que fica por baixo do trabalho habilitado por IA: definições comuns, relações, fronteiras de contexto, proveniência e memória reutilizável. Isso pode soar arquitetural, mas as consequências são gerenciais. Siddhant Pardeshi descreve problema semelhante em sistemas de agentes de software: memória textual plana não escala, enquanto uma estrutura em grafo funciona mais como um GPS para navegar a complexidade. Peter McCrory aponta o lado organizacional do mesmo problema: capacidade não vira adoção se o conhecimento operacional relevante permanece tácito e indisponível no ponto de uso.[6, 4]

A camada semântica, portanto, não é apenas um glossário para IA. Na linguagem de AI Operations, ela faz quatro trabalhos ao mesmo tempo: ajuda a *definir* o domínio operacional, ajuda a *contextualizar* o trabalho, ajuda a *controlar* o comportamento da IA e ajuda a organização a *lembrar*, transformando lições de produção em memória operacional reutilizável.

É por isso também que craft de prompt, sozinho, não escala. Sem significado compartilhado, cada novo caso de uso vira um projeto customizado de integração mais uma nova rodada de improvisação de prompts. A lógica é duplicada, as exceções passam a ser tratadas de forma diferente entre times e a auditabilidade enfraquece porque o sistema não tem uma representação estável daquilo sobre o qual está raciocinando.

Com uma camada semântica, a economia da implantação de IA muda. Novos workflows podem reutilizar definições compartilhadas em vez de inventá-las. A recuperação de contexto melhora porque ele pode ser montado em torno de relações significativas. A governança ganha uma superfície de controle porque os mesmos conceitos podem ser monitorados e restringidos em diferentes casos de uso. Melhorar fica mais barato porque um esclarecimento feito uma vez pode se propagar por muitos caminhos operacionais. Em outras palavras: sem camada semântica, cada caso de uso de IA parece um projeto novo. Com ela, mais deles passam a se parecer com novas configurações sobre significado compartilhado.

Uma camada semântica prática não começa com metafísica corporativa. Começa com perguntas operacionais: quais entidades importam neste workflow? De quais decisões elas participam? Quais relações mudam o desfecho? Quais sistemas contêm a evidência? Quais termos precisam significar a mesma coisa entre times? Quais exceções exigem roteamento especial? Essas perguntas produzem artefatos práticos: mapa de conceitos, modelo de relações, regras de proveniência, fronteiras de recuperação e lógica de exceção. É por isso que a camada semântica pertence à metodologia, e não a um apêndice de arquitetura.

O caso do IntelliDoctor é o exemplo mais claro que conheço. O problema central não era falta de dado clínico. Era fragmentação. O significado clínico e operacional estava espalhado por cinco sistemas desconectados e múltiplas rotinas humanas. Uma resposta convencional seria esperar por um grande programa centralizado de dados antes de tentar inteligência cross-system. Seguimos outro caminho: construir uma ontologia operacional capaz de conectar diagnósticos, ordens, protocolos, encaminhamentos, especialidades, implicações de faturamento e contexto do paciente de forma

suficientemente firme para que workflows conseguissem raciocinar através deles.

Esse trabalho semântico destravou capacidades que, de outro modo, ficariam como features isoladas. Ele permitiu uma narrativa clínica mais unificada entre registros. Sustentou scoring de encaminhamento ao conectar diagnóstico, aderência de especialidade, lógica de capacidade e relevância econômica. Tornou possível enxergar compliance de protocolo porque contexto de alta, regras de protocolo e eventos de retorno podiam ser interpretados em conjunto, e não como fragmentos separados de dado. Isso não eliminou a necessidade de engenharia. Mudou a economia de reúso e raciocínio cross-system.

Essa distinção importa. A camada semântica costuma ser mal interpretada de duas formas opostas. Um grupo a trata como exercício acadêmico de ontologia com pouca consequência operacional. Outro a trata como infraestrutura mágica que elimina a necessidade de trabalho de integração. As duas visões estão erradas. Uma camada semântica não elimina problemas de qualidade de dados, desenho de workflow ou esforço de engenharia. Ela concentra a lógica de interpretação em uma camada reutilizável, para que cada nova implantação de IA não precise reconstruir do zero o mesmo significado de negócio.

Para executivos, as consequências práticas são diretas:

- **Controle:** significado compartilhado facilita definir o que o sistema pode inferir, recomendar ou acionar;
- **Reúso:** uma vez que os conceitos de negócio ficam explícitos, workflows podem construir uns sobre os outros em vez de permanecerem ferramentas isoladas;
- **Velocidade:** novos casos de uso podem ser montados mais rápido porque a organização reutiliza significado, em vez de renegociá-lo toda vez;
- **Aprendizado institucional:** lições de produção podem ser codificadas uma vez e reutilizadas muitas vezes.

Na saúde, ambiguidade é cara porque aparece como handoffs ruins, baixa visibilidade de compliance, perda de receita e risco para o paciente. Mas a mesma lógica vale muito além do hospital. Qualquer empresa que tente implantar IA através de múltiplas funções enfrentará o mesmo problema: escala exige significado compartilhado. Se AI Operations é a disciplina de gestão, a camada semântica é um de seus mecanismos operacionais centrais.

## 5 Governança, controle e prontidão executiva

Governança que vive apenas em documentos de política vai falhar no primeiro contato com a produção. A pergunta executiva não é se a organização possui uma política de IA. É se ela consegue controlar a IA em operação: onde ela é usada, o que ela pode influenciar, como o valor é medido e quem intervém quando o comportamento deriva.

Isso parece básico, mas é justamente onde muitas equipes de liderança são mais fracas. Elas conseguem aprovar pilotos, negociar licenças e pedir princípios de IA responsável. Muitas vezes não conseguem distinguir promessa técnica de credibilidade operacional. Davenport e Srinivasan

mostram isso indiretamente em sua pesquisa de 2026: o diferencial não é simplesmente quais ferramentas de IA a empresa implanta, mas quão deliberadamente ela mede, gerencia, agrega e reporta o valor que essas ferramentas criam.[2]

Um bom padrão de governança, portanto, não é elegância de política, mas qualidade das respostas operacionais. Líderes seniores deveriam ser capazes de fazer um conjunto curto de perguntas e receber respostas claras e testáveis.

1. **Que workflow está sendo mudado?**
2. **Que métrica deve se mover, e qual é a linha de base?**
3. **Quais decisões continuam sendo humanas?**
4. **Quais exceções disparam revisão ou escalonamento?**
5. **Quais controles semânticos e de dados limitam o sistema?**
6. **Quem é dono da performance ponta a ponta em produção?**

Essas perguntas são deliberadamente pouco glamourosas. Aí está sua força. Conselhos não precisam de mais entusiasmo sobre modelos. Precisam de filtros melhores para separar teatro de IA de substância operacional.

### Teatro de IA versus credibilidade operacional

Teatro de IA celebra demos, relações com vendors, quantidade de pilotos e nomes de modelos. Ele reporta atividade, não controle. Trata adoção como logins, não como mudança comportamental. Fala com confiança sobre capacidade enquanto permanece vago sobre a economia do workflow.

Credibilidade operacional parece outra coisa.

- Ela mostra melhoria medida em um workflow definido.
- Ela nomeia a linha de base e os modos de falha.
- Ela torna ownership visível.
- Ela explica a fronteira de revisão humana.
- Ela consegue rastrear como dados e regras semânticas moldam o comportamento do sistema.
- Ela tem um ritmo de revisão para drift, exceções e mudança de política.

Essa é a linguagem de que times de liderança precisam porque a IA agora atravessa silos demais para ser governada função por função. A pesquisa da McKinsey de 2025 mostrou que organizações com maior impacto eram mais propensas tanto a redesenhar workflows quanto a colocar líderes seniores em papéis substantivos de governança.[5] Implantação séria de IA expõe falhas de handoff que nenhum time, isoladamente, possui por completo. Sem patrocínio executivo, a organização compra o software, mas evita o redesenho.

### O controle precisa ser desenhado dentro do workflow

Controle real não começa com um memo de política. Começa quando o próprio workflow contém checkpoints mensuráveis, regras de escalonamento e intervenções com dono claro. Uma recomendação de modelo que ninguém é obrigado a revisar não está governada. Uma etapa de assinatura humana que existe apenas para absorver ansiedade também não está governada. Ambas são formas de ambiguidade.

O princípio de desenho mais forte é colocar revisão onde ela muda o resultado. Decisões de baixo risco e alto volume podem justificar mais automação e amostragem. Casos de alto risco ou baixa frequência podem exigir ownership humano explícito. Muitos workflows vão precisar de controle em camadas, em que a IA cuida do caminho padrão e as pessoas concentram energia em exceções, disputas e casos limítrofes. A governança se torna real quando essas escolhas ficam visíveis dentro do trabalho.

### Prontidão executiva é uma decisão de alocação de capital

Muitos executivos dizem querer valor com IA enquanto financiam apenas a camada mais visível da stack. Eles aprovam licenças, experimentos e budgets de inovação. Subfinanciam instrumentação, ativos semânticos, redesenho de workflow, treinamento de usuários e revisão operacional. Depois se surpreendem quando o programa continua preso entre entusiasmo e prova.

Prontidão executiva significa estar disposto a financiar as partes menos glamourosas da implantação. Significa aceitar que realização de valor pode depender menos de mais um upgrade de modelo e mais de telemetria melhor, ownership mais claro e mudança de processo mais disciplinada. Significa pedir evidência de antes e depois, e não confiança narrativa.

A experiência com o IntelliDoctor também foi instrutiva aqui. A barreira não era simplesmente se a IA conseguia produzir saídas úteis. Era se a liderança tinha uma linguagem robusta para avaliar um sistema já funcional em termos de adoção, impacto financeiro, controle de protocolo e mudança de workflow. É por isso que AI Operations importa no nível executivo. Ele dá aos líderes algo melhor do que hype e algo mais prático do que frameworks abstratos de risco.

Daí decorre um padrão simples. Se um conselho ou time executivo não consegue responder onde a IA está atuando, que resultado ela deveria mudar, como o sistema é limitado e quem é accountable pelo comportamento em produção, então ele ainda não está governando IA. Ainda está financiando experimentação e esperando que o valor se acumule.

## 6 Conclusão

As organizações que vão vencer com IA não serão as que têm mais demos. Serão as que tiverem melhor disciplina operacional.

Na maior parte das empresas, o valor da IA é limitado menos por fraqueza do modelo do que por fraqueza operacional: linhas de base pouco claras, workflows não redesenhados, ausência de estrutura semântica, desenho fraco de adoção e governança desligada da produção. AI Operations é a camada de gestão que reúne esses elementos.

O próximo passo não precisa ser grandioso. Precisa ser sério. Escolha um workflow de alta fricção. Estabeleça sua linha de base. Defina os termos-chave e os pontos de controle. Construa a camada semântica apenas até o ponto em que o workflow se torne legível. Redesenhe o workflow em torno de IA combinada com julgamento humano. Meça o que mudou. Depois melhore em produção.

Com o tempo, esse trabalho deve virar método prático, e não intervenção isolada: um playbook, uma canvas semântica, um scorecard de baseline, um desenho de controle e ferramentas de apoio que façam a próxima implantação ser mais fácil do que a anterior.

Esse é o teste executivo real. Não se a organização está "fazendo IA", mas se está construindo a disciplina necessária para transformar capacidade de IA em valor operacional durável.

## References


[1] Boston Consulting Group. The widening AI value gap. https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap, 2025. BCG study on enterprise AI value realization.

[2] Thomas H. Davenport and Laks Srinivasan. 7 factors that drive returns on AI investments, according to a new survey. *Harvard Business Review*, March 2026. Based on a survey of senior executives and enterprise leader interviews.

[3] Karim R. Lakhani, Jared Spataro, and Jen Stave. The "last mile" problem slowing AI transformation. *Harvard Business Review*, March 2026. Seven-friction framework for AI scaling.

[4] Peter McCrory and Azeem Azhar. Anthropic's head of economics on AI adoption data, claude code, the burden of knowledge, and the next generation of experts. https://www.exponentialview.co/p/anthropics-head-of-economics-on-ai, 2026. Exponential View interview.

[5] McKinsey & Company. The state of AI: Global survey 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai, 2025. McKinsey Global Survey on AI.

[6] Siddhant Pardeshi and Sam Charrington. Agent swarms and knowledge graphs for autonomous software development with siddhant pardeshi. https://twimlai.com/podcast/twimlai/agent-swarms-knowledge-graphs-autonomous-software-development, March 2026. The TWIML AI Podcast, episode 763.