marlabs
driving digital agility

# De onde virá o ORÇAMENTO PARA IA

Como financiar a próxima onda de IA com a capacidade que sua empresa já paga

NOVEMBRO DE 2026

WWW.MARLABS.COM

# Índice de conteúdo


Uma pergunta que ninguém quer responder com “mais verba” 3

Um raio-x das oportunidades 5

**PARTE 1**

De onde virá o dinheiro 7

A tempestade perfeita 8

**PARTE 2**

Onde a capacidade está presa 11

A plataforma de dados e o custo que não aparece na fatura 12

Engenharia que entrega mais com o mesmo time 16

Processos, operação e legado 22

**PARTE 3**

O que impede a capacidade de virar orçamento 25

Quando o código acelera, o gargalo muda de lugar 26

Do ganho ao orçamento 29

Roteiro para levar ao seu time 32

De onde virá o seu orçamento? 33

Sobre a Marlabs 33




# Uma pergunta que ninguém quer responder com “mais verba”

**André Ditomaso**
*Country Manager Marlabs Brasil*

Na noite de 11 de novembro, reunimos em São Paulo 38 executivos de tecnologia, dados e engenharia em torno de uma pergunta simples de fazer e difícil de responder. Se a sua empresa precisasse financiar uma agenda agressiva de IA amanhã, de onde viria o dinheiro?

Eu esperava ouvir pedidos de orçamento novo. Ouvi outra coisa. A maior parte da sala apontou para dentro de casa, para a eficiência que ainda não foi capturada na plataforma de dados, no ciclo de desenvolvimento e nos processos. Quem decide sobre tecnologia já entendeu que a próxima onda de IA não vai ser paga com uma linha a mais no orçamento. Vai ser paga com o que hoje se perde no caminho.

Saí da noite com três lições.

A primeira é que o dinheiro existe. Ele está preso em cópias de dados que ninguém precisa, em ferramentas que fazem a mesma coisa, em horas de engenharia gastas em retrabalho e em tarefas que alguém repete todos os dias à mão. Soltar essa capacidade é trabalho de gestão antes de ser trabalho de tecnologia.

A segunda é que saber onde está o desperdício não basta. Quase ninguém na sala disse ignorar onde estavam as

oportunidades. A dificuldade apareceu na hora de provar o valor. Um ganho que não é medido antes e depois vira opinião, e opinião não sustenta orçamento.

A terceira é que, quando a engenharia acelera, o gargalo muda de lugar. Vimos projetos em que agentes de IA reduziram o esforço de desenvolvimento a menos da metade. Nesses projetos, o que passou a definir o ritmo foi a clareza do requisito, a qualidade do contexto e a disciplina das decisões humanas.

Este material leva essas lições adiante. Ele não resume o encontro. Usa o que discutimos naquela noite, os resultados da pesquisa que fizemos com a sala e a experiência da Marlabs em dados, engenharia e IA para responder à pergunta com método. Espero que ele ajude você a encontrar, na sua empresa, o orçamento que já está lá.

**O EVENTO**

## Funding AI at Scale

São Paulo, 11 de novembro de 2026

Um encontro fechado para líderes de tecnologia, dados e engenharia, realizado pela Marlabs em parceria com a Databricks. A noite combinou duas apresentações, um caso executivo e uma mesa redonda com os 38 convidados.

**A pergunta da noite**

Se sua empresa precisasse financiar uma agenda agressiva de IA amanhã, de onde viria o dinheiro?

**Quem conduziu a conversa**

André Ditomaso e Elekmilton Mangino, da Marlabs. Letícia Santos, da Databricks. Rafael Diniz, Product & Innovation da InformaMarket.

**As duas frentes em debate**

O custo invisível da plataforma de dados, e quanto se libera ao simplificar essa base. O uso de IA dentro do próprio ciclo de desenvolvimento, para entregar mais sem ampliar equipes.







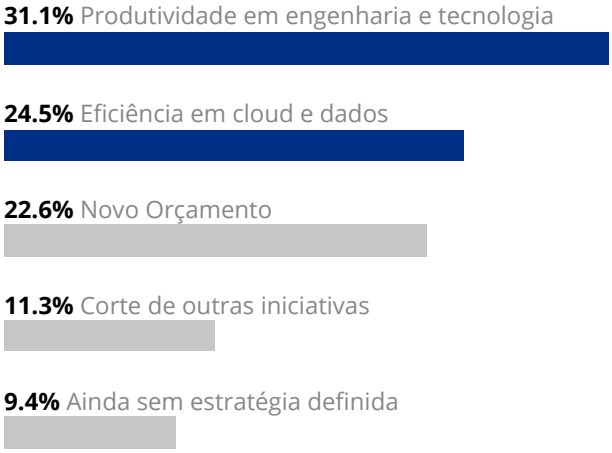
# Um raio-x das oportunidades

Antes das apresentações, pedimos aos convidados que respondessem, de forma anônima, a três perguntas. As respostas não têm pretensão estatística. São um retrato honesto de como líderes de tecnologia de médias e grandes empresas enxergam hoje o financiamento da IA, e esse retrato organiza o restante deste material.

Como ler os números. As perguntas aceitavam mais de uma resposta. Os percentuais indicam a parcela de cada opção no total de menções da pergunta, e não a parcela de executivos.

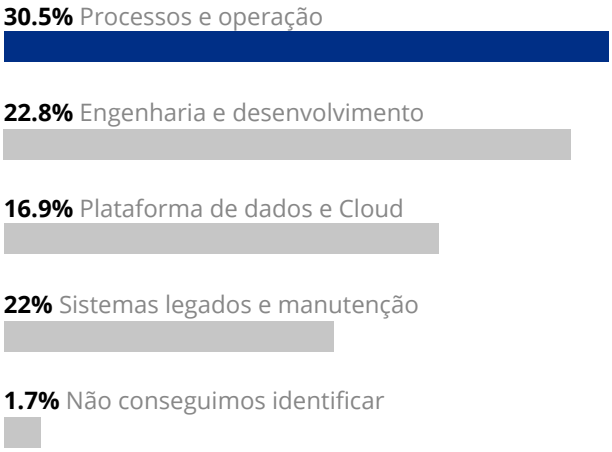
Três leituras se destacam.

### 1. De onde viria o orçamento para ampliar IA?




**A sala já sabe de onde o dinheiro deveria vir.** Na primeira pergunta, produtividade em engenharia e eficiência em cloud e dados somaram 57% das menções. Orçamento novo ficou com 23%, e cortar outras iniciativas, com 11%. Os líderes presentes não esperam que a IA seja paga com dinheiro novo. Esperam pagá-la com eficiência.

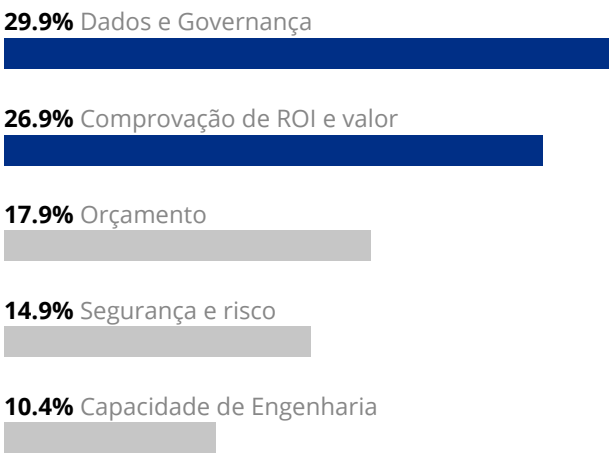
### 2. Onde está a maior oportunidade de liberar capacidade?




**Saber onde está a oportunidade é mais fácil do que provar o valor.** Apenas 1,7% das menções na segunda pergunta disseram não conseguir identificar onde liberar capacidade. Na terceira, comprovação de ROI e valor para o ne-

gócio apareceu como o segundo maior desafio para escalar IA, com 26,9%. As oportunidades estão visíveis. Transformá-las em números que convencem é o trabalho difícil.

### 3. Qual o principal desafio para escalar IA?




**O freio principal não é o dinheiro.** Dados e governança (29,9%) e prova de valor (26,9%) pesaram mais do que orçamento (17,9%) entre os desafios. Capacidade de engenharia foi o menos citado (10,4%), embora a engenharia apareça no topo das fontes de financiamento. O que trava a escala é a base de dados e a capacidade de demonstrar retorno.

Os capítulos seguintes partem dessas três perguntas. A Parte 1 trata de onde o dinheiro pode vir. A Parte 2 percorre as frentes onde a capacidade está presa. A Parte 3 enfrenta o que impede que essa capacidade vire orçamento.

marlabs
driving digital agility

PARTE 1

# De onde virá o dinheiro

Na pesquisa da noite, 57% das menções apontaram eficiência em engenharia ou em cloud e dados como a origem mais provável do orçamento para IA. Orçamento novo ficou com 23%.

CAPÍTULO 1

# A tempestade perfeita




Durante alguns anos, IA ocupou um lugar confortável no orçamento das empresas. Era inovação. Tinha verba de laboratório, prazo de experimento e uma expectativa de retorno que ninguém cobrava com muito rigor. As provas de conceito vieram, os copilots chegaram às mesas dos desenvolvedores, os primeiros casos de uso entraram em produção.

Esse período acabou. IA deixou de ser um projeto de inovação e passou a disputar orçamento com todo o resto da tecnologia. A cobrança agora é por escala, eficiência e impacto que se possa medir. E ela chega num momento em que a conta ficou mais cara dos dois lados.

## A conta subiu por volume e por preço

De um lado, o volume. O Gartner projeta que os gastos mundiais com IA cheguem a US$ 2,67 trilhões em 2026, alta de 49,5% sobre o ano anterior. As empresas processam mais dados, rodam mais modelos e colocam mais pessoas para usar IA todos os dias.

Do outro lado, o preço unitário. Segundo o BCG, o custo médio dos principais modelos de linguagem americanos

passou de menos de US$ 5 para US$ 20 por milhão de tokens entre o início de 2025 e 2026. Na mesma pesquisa, a participação da IA no orçamento de TI mais do que triplicou, de 5% para 17%.

O resultado é uma tempestade perfeita. As empresas precisam rodar volumes muito maiores em sistemas que cobram mais por unidade. E fazem isso com um orçamento de tecnologia que, em sua maior parte, já está comprometido com manter funcionando o que existe.

## Três saídas, e só uma escala

Diante disso, há três maneiras de financiar a próxima onda de IA. A Letícia Santos, da Databricks, colocou as três diante da sala logo no início da noite.

A primeira é conseguir orçamento novo. É possível, mas raro e lento. Depende do ciclo de planejamento, da disputa com outras prioridades do negócio e de um caso de investimento que ainda precisa ser provado.

A segunda é cortar outras iniciativas. Funciona no papel. Na prática, gera atrito com as áreas afetadas e costuma travar a operação justamente onde ela já entregava valor.




*Letícia Santos, da Databricks*

A terceira é liberar o que já é pago. Encontrar a capacidade que a empresa compra todos os meses e não aproveita, na plataforma de dados, no ciclo de desenvolvimento e nos processos, e redirecioná-la para IA. É a única das três que pode se repetir a cada ano, porque cada ganho de eficiência financia o próximo.



*Rafael Diniz, da InformaMarkets*

A sala chegou à mesma conclusão antes de qualquer apresentação. Na pesquisa, as duas fontes de eficiência somaram mais que o dobro das menções a orçamento novo. Os líderes presentes não precisavam ser convencidos de que a terceira saída é a melhor. A dúvida era outra. Como executá-la sem que o ganho se perca no meio do caminho?

## Medir, liberar e realocar

Este material responde a essa dúvida com uma tese. O dinheiro para IA já está dentro da empresa, preso na plataforma de dados, no esforço de engenharia e nos processos. Ele só vira orçamento quando passa por três movimentos deliberados.

**Medir** é saber onde a capacidade está presa e quanto ela custa hoje, antes de mudar qualquer coisa.

**Liberar** é soltá-la sem travar a operação, com mudanças na base de dados, no ciclo de desenvolvimento e nos processos.

**Realocar** é garantir que o ganho volte para IA e não desapareça no orçamento do ano seguinte.

Cada movimento tem método. Para medir e liberar capacidade na plataforma de dados, a Marlabs usa o **MAP** (Modern Analytics Platform), seu framework de estratégia de dados, e o **COMPASS**, o assessment que mostra onde cada empresa está nesse mapa. Para levar IA ao ciclo de desenvolvimento e governá-la, usa os ABCs da AgilityAI, que organizam a transformação em três frentes, Align, Build e Control. Os próximos capítulos mostram esses métodos em ação, com os casos apresentados na noite e outros da experiência da Marlabs.

marlabs
driving digital agility

PARTE 2

# Onde a capacidade está presa

Na pesquisa, processos e operação lideraram as menções sobre onde liberar capacidade para IA (30,5%), seguidos de engenharia (28,8%), sistemas legados (22%) e plataforma de dados e cloud (17%).

CAPÍTULO 2

# A plataforma de dados e o custo que não aparece na fatura

A pesquisa da noite trouxe um contraste curioso. A plataforma de dados foi a fonte de capacidade menos lembrada, com 17% das menções. Ao mesmo tempo, dados e governança foram o desafio mais citado para escalar IA, com quase 30%. A mesma área aparece como o maior obstáculo e como a menor oportunidade.

Há uma explicação simples para isso. O custo da plataforma de dados é difícil de enxergar porque ele não está num lugar só. A fatura de cloud e as licenças aparecem no orçamento. O resto se espalha entre sistemas, equipes e fornecedores, e ninguém soma.

## Onde o custo se esconde

Na maioria das empresas, a plataforma de dados cresceu por camadas. Cada área comprou a ferramenta de que precisava, cada projeto criou seu pipeline, cada nova demanda ganhou uma cópia dos dados. O resultado aparece em alguns padrões conhecidos.

O mesmo dado existe em várias cópias, cada uma armazenada, processada e paga separadamente. Parte dos pipelines só move dados de um lugar para outro, sem acrescentar nada. Ferramentas diferentes fazem a mesma coisa para áreas diferentes. As políticas de acesso e de qualidade são refeitas em cada camada, com o risco de não baterem entre si. Capacidade de processamento fica contratada para picos que acontecem poucas horas por dia. E, mais recentemente, modelos e tokens passaram a ser consumidos sem dono, sem limite e sem medida de valor.

Nada disso aparece numa linha única do orçamento. Tudo isso consome dinheiro e pessoas que poderiam estar trabalhando em IA.

O custo de pessoas costuma ser o



mais subestimado. Em uma empresa norte-americana de veículos conectados atendida pela Marlabs, entre 20% e 25% dos recursos da área de dados iam para validar dados, tempo que deixava de ir para iniciativas de crescimento.

## Por que isso trava a IA

O **MAP**, framework de estratégia de dados da Marlabs, parte de uma constatação. A falta de prontidão dos dados é o maior obstáculo para adotar IA. Um modelo alimentado por dados fragmentados, duplicados ou inconsistentes cobra duas vezes. Gasta mais processamento para chegar a uma resposta e entrega uma resposta em que ninguém confia.

Por isso, a plataforma de dados é uma fonte de dinheiro e também a condição para que o dinheiro investido em IA dê retorno. Simplificar a base libera recursos e, ao mesmo tempo, prepara os dados para os casos de uso que vão consumir esses recursos.

## Medir antes de mexer

O primeiro passo é saber onde a empresa está. Parece óbvio, mas é o passo que mais se pula. Muitas iniciativas de modernização começam pela escolha da ferramenta e só depois descobrem que o problema estava na governança, na qualidade ou no jeito como as áreas usam os dados.

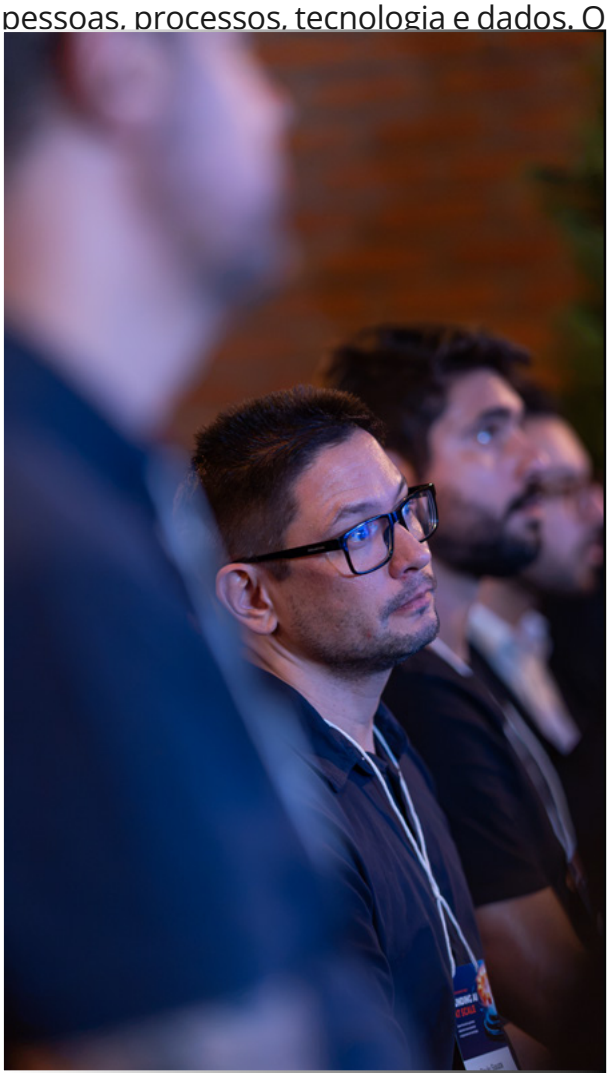
É para isso que existe o **COMPASS** (Comprehensive MAP Assessment). Ele funciona como a bússola do **MAP**. O assessment ouve os stakeholders, analisa os processos envolvidos, faz o inventário completo dos dados e da stack tecnológica e entrega um mapa customizado, que mostra onde a empresa está e quais passos a levam ao destino. Junto com o mapa vêm um plano com objetivos de negócio priorizados, um roadmap de implementação e recomendações.

O **COMPASS** não é só diagnóstico. Durante o assessment, a Marlabs corrige

parte dos problemas que encontra ou orienta o time do cliente sobre como corrigi-los. Conforme o porte e a infraestrutura da empresa, o trabalho leva de seis semanas a três meses.

## Liberar com método

Com a posição clara, o **MAP** organiza a rota. O framework divide a jornada em três seções, Strategy, Execution e Enablement (estratégia, execução e capacitação), sempre olhando para quatro pilares, pessoas, processos, tecnologia e dados. O



destino é o que o **MAP** chama de decision intelligence, decisões melhores sustentadas por dados confiáveis.

Três perguntas da estratégia funcionam também como alavancas de custo.

**Trust.** Os dados refletem a realidade a ponto de sustentar uma decisão? Governança, qualidade e uma fonte única para entidades centrais, como cliente e produto, eliminam boa parte das cópias e do retrabalho de validação.

**Supply.** As pessoas certas acessam os dados certos na hora certa? Arquitetura e integração bem resolvidas consolidam a base, reduzem pipelines redundantes e permitem desligar ferramentas sobrepostas.

**Analyze.** Quem decide consegue chegar ao dado sem passar por uma fila de analistas? Quando a análise chega a quem precisa, o tempo dos especialistas vai para o que realmente exige especialistas.

A empresa de veículos conectados mostra o efeito combinado dessas frentes. O trabalho começou com sessões de descoberta e entrevistas com quem produz e quem consome dados, e terminou num roadmap de lacunas em pessoas, processos, tecnologia e dados. Cinco soluções

on-premise foram unificadas numa fonte única na nuvem. A empresa economizou cerca de 36% dos custos operacionais ao desligar licenças redundantes, e o ciclo de entrega de casos de uso de negócio caiu de três semanas para uma.

Em escala maior, uma companhia industrial global usou o **MAP** para organizar a gestão de dados de uma divisão que operava 11 empresas e mais de 140 sistemas de origem. O trabalho começou por uma fase de assessment e planejamento e seguiu para modernização, governança, visualização e capacitação das equipes.

## Token é a nova fatura de cloud

Há uma linha de custo que ainda não existia quando a maioria das plataformas foi desenhada. O consumo de modelos de IA cresce mais rápido do que o preço por token cai, e quase sempre sem dono. O padrão mais comum é mandar toda requisição para o modelo mais caro, seja ela simples ou complexa.

O **PromptRouter®**, acelerador proprietário da Marlabs, ataca exatamente esse padrão. Ele avalia a complexidade de cada requisição em tempo real e a envia para o modelo mais econômico capaz de responder bem, reservando os modelos mais avançados para o que realmente precisa deles.

Numa consultoria especializada em Workday, o roteamento reduziu os custos com APIs de modelos de linguagem em mais de 50%. Numa empresa global de testes, inspeção e certificação, o projeto começou por um baseline de custo da carga de extração de dados. Com o roteamento, o uso de tokens caiu 30%, com a qualidade do resultado preservada.

## O que medir

Antes de qualquer mudança na plataforma, vale ter cinco números em mãos. O custo por workload e por consulta. Quantas cópias existem do mesmo dado. Quantas licenças e contratos podem ser desligados. Quanto custa, por mês, cada caso de uso de IA em produção. E quanto tempo leva para uma pergunta de negócio virar resposta.

CAPÍTULO 3

# Engenharia que entrega mais com o mesmo time




Elekmilton Mangino abriu sua apresentação com uma provocação. Todo desenvolvedor já tem um copilot. A capacidade de entrega da sua equipe dobrou?

A pergunta incomoda porque a adoção já aconteceu. Segundo o relatório de benchmarks de 2026 da LinearB, que analisou mais de 8,1 milhões de pull requests, 88,3% das organizações usam IA para programar todos os dias ou toda semana. O mesmo relatório mostra que 44,7% delas não medem formalmente o impacto dessas ferramentas. A maioria das empresas sente que a equipe ficou mais produtiva. Poucas sabem dizer quanto.

A sala do encontro também apostou na engenharia. Produtividade em engenharia e tecnologia foi a fonte de orçamento mais citada na pesquisa, com 32% das menções, e capacidade de engenharia foi o desafio menos citado, com 10,4%. Os líderes presentes enxergam a engenharia como lugar de ganho. A questão é como transformar esse ganho em capacidade real de entrega.

## Do copilot ao ciclo inteiro

Um copilot ajuda o desenvolvedor a escrever código mais rápido. Mas escrever código é só uma parte do trabalho. Antes dele há requisitos, decisões de arquitetura e planejamento. Depois dele há revisão, testes, integração e correções. Se só a escrita acelera, o ganho se dilui no resto do ciclo.

A Marlabs trabalha com IA ao longo de todo o ciclo de desenvolvimento, e avalia qualquer modelo de trabalho por cinco elementos.

1. **Funcionalidades.** Requisitos claros viram entregas previsíveis.
2. **Arquitetura e contexto.** Agentes rendem mais quando entendem o sistema em que trabalham.
3. **Qualidade.** Testes, processos e políticas ficam dentro do fluxo de entrega, e não depois dele.
4. **Governança e segurança.** Regras de uso, dados protegidos e rastreabilidade de cada etapa.
5. **Custo.** Gestão e previsibilidade do que cada entrega consome em horas e em tokens.

No encontro, Elekmilton apresentou dois modelos que a Marlabs opera hoje com squads no Brasil. Eles partem dos mesmos cinco elementos e chegam a desenhos diferentes.



MODELO A

# Agentes no ambiente do desenvolvedor




O primeiro modelo roda dentro do ambiente de cada desenvolvedor, em uma empresa norte-americana de software para a área da saúde. Cada tipo de trabalho tem um comando, e cada comando aciona apenas os agentes daquele trabalho. Há comandos para construir, corrigir, melhorar, revisar e acompanhar. Um comando de construção de história, por exemplo, chama de sete a nove agentes, conforme o que a história pede. Ao todo, são 35 skills e 13 agentes, integrados ao Jira, ao Confluence, ao Figma e ao Playwright.

Toda história passa por três quality gates antes de virar pull request. O primeiro valida o requisito, a partir do card do Jira, e o desenvolvedor decide se ele está claro o bastante para seguir. O segundo produz um plano de arquitetura, que o desenvolvedor aprova. Só então os agentes de banco, backend e frontend trabalham em paralelo, com os testes escritos antes do código e uma revisão de segurança no caminho. O terceiro gate verifica a história critério por critério, com testes de ponta a ponta no navegador. Se uma etapa falha, o trabalho volta ao agente responsável, no máximo três vezes. O pull request sai com o tempo economizado registrado.

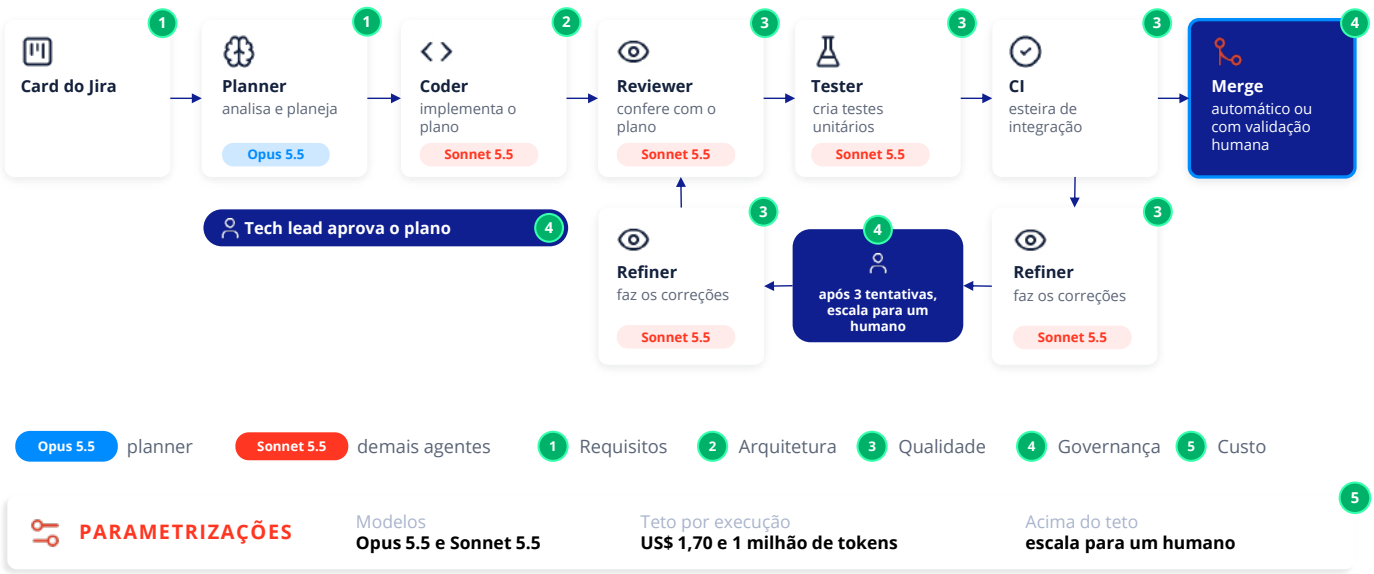
Os números mostram o tamanho

do efeito. Em 359 tickets estimados em 13.679 horas sem IA, a equipe economizou 9.575 horas, uma redução de 70% no esforço. O projeto principal chegou a 90% de conclusão em 13 meses, quando a estimativa sem IA passava de quatro anos. Um ticket de criptografia de dados sensíveis estimado em 64 horas foi entregue em 5. Uma restrição de exclusão de registros estimada em 40 horas levou 2.

O detalhe mais importante está no que não mudou. A squad não foi reduzida e o escopo não foi cortado. O cliente decidiu usar o ganho para entregar antes.

MODELO B

## Uma plataforma de agentes orquestrada




O segundo modelo é uma plataforma que conduz o trabalho por fases independentes, em uma empresa global de eventos e feiras B2B. Cada fase tem entrada, etapas e saída próprias. A ideação transforma transcrições de reuniões e documentos em requisito de negócio. O planejamento transforma o requisito em épico e histórias no Jira, com documento de produto, design, arquitetura e plano de entrega. O desenvolvimento leva o card do Jira até o merge. A validação verifica a aplicação integrada em testes de ponta a ponta, usabilidade, segurança, acessibili-

dade e performance. A fase de desenvolvimento já está em uso, e as demais estão em teste ou em construção.

Na fase de desenvolvimento, seis agentes levam o card até o merge, e uma pessoa decide em dois pontos. Um agente planejador analisa o card e propõe um plano, que o tech lead aprova antes de qualquer linha de código. Em seguida, agentes de codificação, revisão e testes implementam, conferem o resultado contra o plano e criam os testes unitários. Um agente de refinamento corrige o que falhar. Depois de três tentativas sem sucesso, o trabalho escala para um humano. O merge pode ser automático ou exigir validação humana.

A plataforma tem regras de custo explícitas. Cada execução tem teto de US$ 1,70 e de 1 milhão de tokens. Acima disso, a decisão passa para uma pessoa. Os modelos mais avançados ficam reservados ao planejamento, e os demais agentes usam modelos mais leves.

Em 29 histórias, os resultados foram os que estão apresentados na tabela abaixo.

Cada R$ 1 estimado custou R$ 0,49. A IA respondeu por R$ 808 do total, cerca de R$ 28 por história, e metade desse valor foi consumida pelo agente planejador. O dado é revelador. O trabalho mais caro da plataforma é justamente o de pensar antes de construir.

## O método por trás dos dois modelos

Os dois modelos são diferentes na forma e iguais na lógica. Essa lógica corresponde à forma como a AgilityAI, suíte de transformação em IA da Marlabs, organiza o trabalho em seus ABCs.

O **Build** é a frente de construção. A AgilityAI a define como um ciclo disciplinado de engenharia de IA, apoiado em

| Indicador | Estimativa sem IA | Realizado | Diferença |
|---|---|---|---|
| **Horas de desenvolvimento** | 190H | 89,2H | 2.1X mais rápido |
| **Dias úteis** | ~24 | 14 | 1.7X mais rápido |
| **Custo total, desenvolvedor e IA** | R$32,3 mil | R$16 mil | 2X mais barato |

aceleradores, para encurtar o caminho do conceito à execução. Nos dois modelos, a disciplina aparece no desenho do fluxo. Cada tipo de trabalho tem seu caminho, cada agente tem seu papel e cada etapa tem critério de saída.

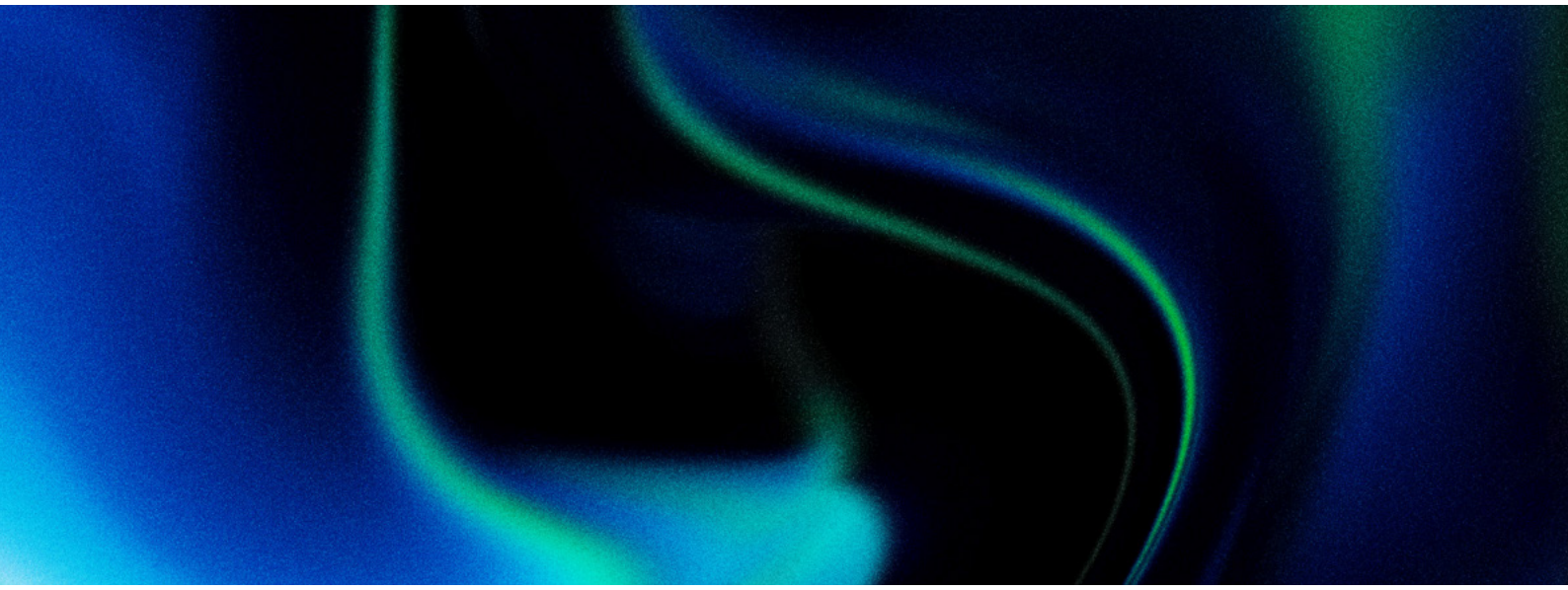
O **Control** é a frente de governança, e nos ABCs ela não espera o fim do projeto. Está presente desde o primeiro comando. Nos dois modelos, ela aparece nos quality gates, na aprovação humana do plano e do merge, no limite de três tentativas antes de escalar para uma pessoa, no teto de custo por execução e no registro do tempo economizado em cada entrega. É o Control que permite acelerar sem perder qualidade, e é ele que produz os números que provam o ganho.

## O que medir

Em engenharia, quatro números mostram se a IA está gerando capacidade ou apenas sensação de produtividade. O esforço real comparado ao estimado. O custo por história, somando horas e tokens. O volume de retrabalho e de correções. E a cobertura de testes de cada entrega.

CAPÍTULO 4

# Processos, operação e legado



A maior oportunidade apontada pela sala não estava em nenhuma das duas frentes discutidas na noite. Processos e operação receberam 30,5% das menções como o lugar onde mais se pode liberar capacidade para IA. Sistemas legados e manutenção vieram logo atrás, com 22%. Somadas, as duas frentes passam da metade das respostas.

Faz sentido. Processos manuais e sistemas antigos consomem gente e dinheiro todos os meses, e raramente aparecem como custo de tecnologia. O tempo gasto em conferências, lançamentos e reconciliações fica no orçamento das áreas de negócio. O custo de manter um sistema que ninguém documentou se dilui em horas de sustentação. Nos dois casos, a capacidade está presa num lugar que o orçamento de TI não enxerga direito.

## Processos. Mapear antes de automatizar

A Cury, construtora brasileira com mais de seis décadas de história, enfrentava um problema comum. As áreas administrativa e financeira dependiam de tarefas manuais e repetitivas, com alto índice de erros, custo opera-

cional elevado e atrasos que se espalhavam para outros departamentos. A Marlabs implantou automação robótica de processos nas tarefas que mais consumiam tempo. O retorno sobre o investimento veio em poucos meses. A capacidade operacional cresceu 150%, e o tempo gasto em tarefas manuais na área financeira caiu 84%. As pessoas que antes conferiam e lançavam dados passaram a trabalhar em análise e em iniciativas de crescimento.

O primeiro passo, em processos, é o mesmo da plataforma de dados. Medir antes de mexer. Uma fabricante global com mais de 5 mil funcionários em 20 localidades pediu à Marlabs que avaliasse suas áreas de finanças, contabilidade e relatórios. A equipe analisou 39 processos de ponta a ponta e encontrou mais de US$ 40 milhões em custos que a automação poderia endereçar. Ao aprofundar os seis processos mais críticos, chegou a uma economia anual estimada de US$ 5,6 milhões. Esse levantamento mudou a conversa. Em vez de automatizar o que parecia mais fácil, a empresa passou a priorizar o que valia mais.

Com IA, a automação alcança processos que antes exigiam julgamento humano em cada caso. Uma agência norte-americana de faturamento de serviços médicos de emergência processa cerca de 85 mil pagamentos por mês, cada um com regras específicas da operadora. A Marlabs construiu um fluxo com IA que recebe, verifica e lança os pagamentos, mostra cada alteração feita e sinaliza as exceções para análise mais cuidadosa. A maior parte dos lançamentos passou a ser automática, com uma pessoa revisando cada caso e muito menos erros.

O retorno pode vir rápido. Numa empresa global de serviços corporativos, a automação do processamento e do lançamento de faturas gerou ROI de 60% já no primeiro ano.

## Legado. Entender antes de reescrever

Sistemas legados prendem capacidade de duas formas. Consomem horas de manutenção e tornam cada melhoria mais lenta, porque toda mudança começa por alguém tentando entender o que o código faz. Quando a tecnologia

envelhece, o problema piora, porque fica mais difícil encontrar quem a conheça.

A IA muda a conta da modernização justamente nesse ponto. Uma provedora norte-americana de ensino superior online dependia de uma aplicação em ColdFusion sem nenhuma documentação funcional ou técnica. A decisão foi migrar para Next.js sem interromper os usuários. Dois engenheiros e dois especialistas em qualidade usaram uma ferramenta de codificação com agentes de IA para decifrar o código, planejar a migração de cada funcionalidade e criar mais de mil testes automatizados. Uma pessoa revisou e aprovou o trabalho da IA em cada etapa. Em dois meses e meio, 7 das 17 funcionalidades chegaram ao teste de aceitação, sem nenhuma alteração no banco de dados, e o esforço por funcionalidade caiu entre 55% e 60%.

## O padrão comum

Os casos são de setores e tecnologias diferentes, mas seguem a mesma sequência. Primeiro, entender e medir o processo ou o sistema como ele é. Depois, automatizar ou modernizar com uma pessoa revisando o que a máquina faz. Por fim, registrar o ganho de forma que ele possa ser mostrado.

Esse último passo é o que separa um projeto bem-sucedido de um orçamento liberado. A Parte 3 trata dele.

## O que medir

Em processos e legado, três números sustentam a conversa sobre orçamento. As horas gastas em tarefas manuais e repetitivas antes e depois. O custo anual de manter cada sistema legado, incluindo pessoas. E o tempo que uma melhoria leva para sair do pedido e chegar à produção.

marlabs
driving digital agility

PARTE 3

# O que impede a capacidade de virar orçamento

Na pesquisa, dados e governança (29,9%) e comprovação de ROI (26,9%) foram os desafios mais citados para escalar IA. Orçamento veio em terceiro (17,9%), e capacidade de engenharia, por último (10,4%).

CAPÍTULO 5

# Quando o código acelera, o gargalo muda de lugar

Os dois modelos de engenharia apresentados no encontro reduziram o esforço de desenvolvimento a menos da metade. Esse ganho traz uma consequência que nem sempre entra no plano. Quando o desenvolvimento encolhe, o tempo e o risco do projeto se deslocam para as etapas que vêm antes e depois dele.

Antes do código, o novo gargalo é de contexto e decisão. Depois do código, é de risco e produção. Uma equipe que constrói duas vezes mais rápido também pode construir a coisa errada duas vezes mais rápido.

## O requisito vira o insumo crítico

Não é por acaso que, nos dois modelos da Marlabs, a primeira etapa confere o requisito antes de qualquer agente escrever código. No Modelo A, um quality gate verifica se os critérios de aceite são claros e testáveis. No Modelo B, o plano só segue depois da aprovação do tech lead. E foi o agente planejador que consumiu metade do custo de IA da plataforma.

Quando o código deixa de ser o gargalo, a qualidade do que entra no ciclo passa a definir a qualidade do que sai. Um requisito vago custa pouco quando o desenvolvimento leva semanas, porque há tempo de corrigir o rumo no caminho. Quando o desenvolvimento leva dias, o mesmo requisito vago vira retrabalho em escala.

## Agentes precisam de contexto

Agentes de IA trabalham bem quando sabem em que sistema estão, quais regras precisam seguir e o que já foi decidido antes. Por isso os modelos da Marlabs mantêm uma base de conhecimento do projeto, atualizada a cada entrega, e planos de arquitetura aprovados antes do código. Processos e políticas bem definidos deixam de ser burocracia e passam a ser as instruções que guiam a execução dos agentes.




## Os papéis da equipe mudam

Nesse arranjo, o trabalho das pessoas muda de natureza. Elas planejam, aprovam e decidem nos pontos de controle. O desenvolvedor analisa requisitos, toma decisões de arquitetura, orquestra os agentes e avalia o que eles produzem. Para isso, precisa de mais fundamento técnico, e não de menos. Quem não entende arquitetura não consegue avaliar uma proposta de arquitetura, venha ela de uma pessoa ou de um agente.

Há também um equilíbrio a ser calibrado. Revisão humana demais anula a velocidade dos agentes. Revisão de menos compromete a qualidade. E manter a evolução da IA em sintonia com a evolução da engenharia de software é, por si só, um trabalho contínuo.

## O que a mesa redonda acrescentou

Na mesa redonda que fechou a noite, os convidados levaram a discussão além do ciclo de desenvolvimento. Cinco ideias apareceram com frequência.

**Escolher a tecnologia pelo problema**. Nem todo caso de uso precisa de IA generativa. Modelos clássicos de aprendizado de máquina, ou mesmo soluções mais simples, podem resolver o problema com menos custo e menos risco. Conhecer as alternativas é parte de gastar bem.

**Saber quantos agentes existem.** Com a IA se espalhando pelas áreas, muitas empresas já não sabem quantos agentes estão rodando, quem responde por eles e que resultado entregam. Um catálogo central, com capacidade, resul-

tado e responsável de cada agente, ajuda a avaliar, controlar e reaproveitar.

**Errar rápido, com critério.** Ciclos curtos de protótipo e validação evitam investimentos grandes em funcionalidades que não atendem ao negócio. Mas só funcionam quando objetivo, métrica e critério de sucesso estão definidos antes do primeiro teste.

**Aproximar tecnologia e negócio.** Workshops e maratonas com áreas como RH e Finanças tornam os problemas concretos e ajudam a justificar investimentos. Quem vive o problema costuma ser quem melhor descreve o que precisa ser resolvido.

**Governança no ritmo do setor.** Empresas com menos restrições avançam mais rápido. Setores regulados, como saúde e finanças, precisam de ambientes controlados e salvaguardas mais rígidas. Em todos os casos, segurança e conformidade precisam entrar no planejamento, e não na véspera da produção.

## Align. Decidir antes de construir

Todas essas ideias apontam para o mesmo lugar. Antes de construir mais rápido, é preciso decidir melhor o que construir. Nos **ABCs** da **AgilityAI**, essa é a função do Align. Antes de comprometer tempo e recursos, o Align coloca liderança, dados e equipes apontando na mesma direção.

Na prática, isso significa priorizar casos de uso por critério de valor, e não por entusiasmo. Significa avaliar se processos e dados estão prontos para o caso escolhido. E significa tratar desde o início políticas de uso de IA, risco e preparação das pessoas. A AgilityAI considera três condições inegociáveis para a IA dar resultado. Governança de dados madura, estratégia e indicadores bem definidos, e gestão da mudança com letramento em IA.

O Align responde diretamente aos dois desafios que a sala mais citou. Dados e governança lideraram a lista, e é nesse ponto que o **MAP** e o **COMPASS** entram, preparando os dados para IA. A comprovação de ROI veio logo depois, e ela começa aqui. Um caso de uso sem critério de sucesso definido no início dificilmente terá retorno demonstrável no fim. O próximo capítulo trata de como esse retorno se transforma em orçamento.

CAPÍTULO 6

# Do ganho ao orçamento

Uma frase dita no encontro resume o problema que ainda falta resolver. Sem métrica, *saving* é opinião.

Os capítulos anteriores mostraram que há capacidade presa na plataforma de dados, na engenharia, nos processos e nos sistemas legados, e que existem métodos para soltá-la. Mas capacidade liberada não vira orçamento sozinha. Sem medição, o ganho não convence quem aprova investimentos. Sem uma decisão explícita sobre seu destino, ele se dissolve. As horas economizadas viram mais demandas na fila, a economia em cloud some na próxima revisão de orçamento, e a IA continua disputando verba como se nada tivesse mudado.

É por isso que a comprovação de ROI apareceu como o segundo maior desafio na pesquisa da noite, à frente do próprio orçamento. As empresas sabem onde está a oportunidade. A dificuldade está em transformá-la em um número que sustente uma decisão.

## Três regras para o ganho virar orçamento

**Medir antes de mudar.** Todo ganho é relativo a um ponto de partida. Os casos deste material têm isso em comum. O projeto de roteamento de prompts começou por um baseline de custo. A fabricante global mapeou 39 processos antes de automatizar o primeiro. Os modelos de engenharia comparam cada entrega com a estimativa sem IA. Sem a foto do antes, não há como provar o depois.

**Medir por unidade de valor, e não só pela fatura.** A fatura total de cloud ou de tokens pode subir enquanto a empresa fica mais eficiente, simplesmente porque está fazendo mais. O que interessa é o custo de cada consulta, de cada caso de uso, de cada história entregue. No Modelo B, a medida que importava era quanto custou cada real estimado. A resposta foi R$ 0,49.

**Decidir o destino do ganho.** *Saving* só conta quando sai do orçamento ou volta para IA. Se ninguém decide, o ganho se espalha e desaparece.

## Orçamento, prazo ou backlog

Elekmilton fechou sua apresentação com uma pergunta que todo líder de tecnologia vai precisar responder. Quando a engenharia fica mais rápida, o ganho vira orçamento, prazo ou backlog?

Não há resposta certa. No Modelo A, o cliente escolheu prazo. Manteve a equipe e o escopo e entregou em 13 meses o que levaria mais de quatro anos. Outra empresa pode preferir reduzir custo e redirecionar a diferença para novos casos de IA. Uma terceira pode usar a capacidade para atacar um backlog acumulado. O erro está em não escolher, porque a escolha que ninguém faz costuma ser o backlog, e o backlog não aparece como retorno em lugar nenhum.

A segunda pergunta do Elekmilton completa a primeira. Quem, na sua empresa, é o dono do custo por entrega, somando horas e tokens? Hoje, na maioria das organizações, o custo das pessoas está num orçamento e o custo de IA está espalhado por vários outros. Sem um responsável pela soma, ninguém consegue dizer se a IA está barateando a entrega ou apenas mudando o custo de lugar.

## Controle desde o primeiro dia

Nos ABCs da AgilityAI, o Control é a frente que protege o investimento com governança. Ele gera confiança, gerencia riscos e cria um ciclo de valor que se acumula com o tempo. O ponto central é que ele não chega no fim. Está presente desde o início, ao lado do Align e do Build.

Nos modelos de engenharia apresentados na noite, o Control aparece no teto de US$ 1,70 por execução do Modelo B e no registro do tempo economizado em cada pull request do Modelo A. A mesma lógica de governança está nos guardrails aplicados antes de cada requisição no roteamento de prompts e na revisão humana dos pagamentos automatizados. Em todos esses casos, a governança faz duas coisas ao mesmo tempo. Reduz o risco e produz os dados que provam o retorno.

O objetivo final é que a IA deixe de ser uma sequência de projetos e passe a funcionar como uma capacidade contínua do negócio. Por isso a AgilityAI trabalha num modelo build-to-run, ajudando o cliente a estabelecer os papéis, as fundações de dados e os comitês de supervisão que mantêm a IA operando, medida e governada depois que o projeto termina. É o

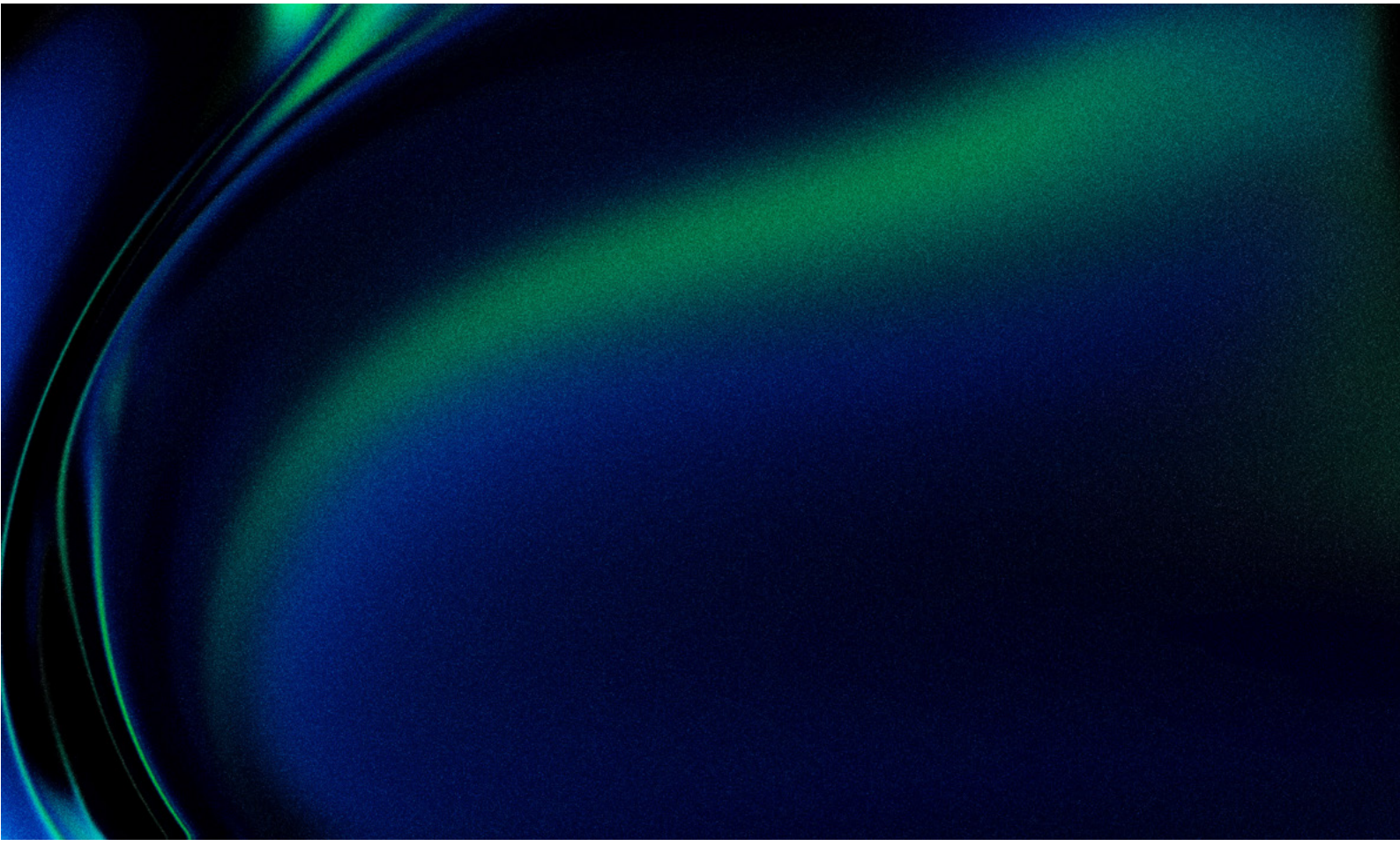
que a própria AgilityAI resume em inglês como Faster to value. Built to scale. Governed from day one.

## A IA que se autofinancia

Com medição e governança, os três movimentos deste material formam um ciclo. Medir o custo que não aparece. Ganhar eficiência na plataforma, na engenharia e nos processos. Capturar o ganho e realocá-lo. Financiar novos casos de IA. E voltar a medir, agora sobre uma base mais simples e com dados mais confiáveis.

Cada frente contribui de um jeito. A plataforma de dados libera dinheiro. A engenharia libera capacidade de entrega. Os processos e a modernização do legado liberam pessoas. Juntas, financiam a próxima onda de IA, e a governança garante que o novo consumo não recrie o desperdício que acabou de ser eliminado.

A primeira onda de eficiência paga a próxima onda de IA. Essa é a resposta mais sólida para a pergunta da noite.



# Roteiro para levar ao seu time

As perguntas abaixo reúnem as provocações da noite e os pontos de medição de cada capítulo. Foram pensadas para uma reunião de uma hora com as lideranças de dados, engenharia e operação. Nenhuma delas tem resposta certa. O valor está em descobrir quais delas a sua empresa ainda não consegue responder com um número.12. Quantos agentes de IA estão rodando hoje, e quem responde por cada um?

## Medir

1. Quanto custa, por mês, cada caso de uso de IA que temos em produção?
2. Quantas vezes o mesmo dado é copiado, governado e pago dentro da empresa?
3. Quanto do tempo da equipe de dados vai para validar e corrigir dados?
4. Sabemos quanto a IA mudou o esforço real da engenharia em relação ao estimado?

## Liberar

5. Que ferramentas, licenças e pipelines fazem a mesma coisa para áreas diferentes?
6. Quais processos ainda dependem de conferência e lançamento manual todos os dias?
7. Quais sistemas legados consomem manutenção sem gerar nenhum valor novo?
8. Nossos requisitos estão claros o bastante para uma engenharia que entrega duas vezes mais rápido?

## Realocar

9. Quando surge um saving, ele volta para IA ou desaparece no orçamento?
10. Quando a engenharia acelerar, o ganho vai virar orçamento, prazo ou backlog, e quem decide?
11. Quem é o dono do custo por entrega, somando horas e tokens?
12. Quantos agentes de IA estão rodando hoje, e quem responde por cada um?

# De onde virá o seu orçamento?

Na noite do encontro, 38 executivos responderam a essa pergunta sobre o próprio mercado. Agora ela vale para a sua empresa.

Propomos uma conversa de uma hora com especialistas da Marlabs em dados e engenharia. Levamos as perguntas deste material para a sua realidade. Onde está a capacidade presa, o que ela custa hoje e o que seria preciso medir para transformá-la em orçamento para IA.

Você sai com uma leitura inicial das suas oportunidades e com clareza sobre o próximo passo, seja um diagnóstico da plataforma de dados com o **COMPASS**, seja uma conversa sobre IA no ciclo de desenvolvimento com a AgilityAI.

**Uma conversa estratégica sobre dados, eficiência e IA**

Leve os desafios da sua operação. Nós ajudamos a identificar os próximos movimentos com maior impacto para o negócio.

**Agende aqui**

Ou fale diretamente com nossa equipe no Whatsapp.

# Sobre a Marlabs

Há mais de 25 anos, a Marlabs atua em Digital Product Engineering, Dados e IA como parceira estratégica de grandes empresas em diversos mercados. Ajudamos nossos clientes a acelerar entregas, modernizar sistemas legados, extrair valor de dados e transformar tecnologia em vantagem competitiva. A Marlabs tem cerca de 3 mil profissionais e opera nos Estados Unidos, Canadá, Alemanha, Bulgária, Suíça, Índia e Brasil.

**www.marlabs.com/pt-br**

marlabs

driving digital agility

www.marlabs.com/pt-br

Marlabs Brasil

@marlabsbrasil