DIRTY DATA EXPERT

Structure Before Intelligence™

POR QUE A IA FALHA ANTES DA IMPLANTAÇÃO

Executive Brief

Gary Segal
www.dirtydataexpert.com
gary@dirtydataexpert.com

A maioria das pessoas acha que a IA falha porque os modelos ainda não são bons o suficiente. Eu acho que ela falha antes mesmo de um único modelo ser construído — porque ninguém interrogou os dados. E a IA amplifica dados sujos em vez de corrigi-los.

Considere o agente descontrolado. Chatbots descrevem seus dados; agentes agem sobre eles. Uma unidade de medida errada e um agente de IA pede doze vezes mais estoque do que deveria, paga a fatura e cancela suas linhas de Natal — sem que nenhuma pessoa aprove nada disso. Sem dados limpos, o agente alucina, e alucinações estrangulam o fluxo de caixa. Este é um exemplo muito real, não hipotético. (Veja o briefing "When AI Agents Go Rogue".)

Construir uma plataforma de IA sobre dados sujos é construir sobre areia: não sustenta nenhuma estrutura que você coloque em cima. Como escrevi em *Dirty Data*: "Technology does not cleanse strategy. It amplifies whatever foundations already exist." A Gartner prevê que, até 2026, as organizações abandonarão 60% dos projetos de IA que não forem sustentados por dados prontos para IA.

Como são os dados sujos? Nomes de clientes duplicados e com erros de grafia. Endereços de clientes errados — ou inexistentes. O mesmo SKU atribuído a fornecedores diferentes.

O que causa isso? A busca pela velocidade. Todo sistema é otimizado para registrar a venda o mais rápido possível — e dados que não passam por validação de inconsistências no ponto de captura permanecem sujos para sempre. (*Dirty Data*, Capítulo 7: "What Caused the Problem".)

Antes de implantar, gere dois relatórios. O primeiro — o relatório mais importante desta intervenção — é uma avaliação do quão sujos seus dados realmente estão. O segundo é uma análise de outliers. Outliers são preços fora da curva de distribuição normal que distorcem valores e montantes acumulados. (*Dirty Data*, Capítulo 11: "Outliers: The Data That Refuses to Fit".)

O que quebra em seguida, em ordem de dano:

1. Agentes de IA saem do controle.. Sem dados limpos, o agente alucina — e alucinações atingem o fluxo de caixa diretamente.

2. A segmentação desmorona.. Se você não sabe em que negócio está, não consegue segmentar seus dados. Os relatórios ficam impossíveis de ler ou interpretar.

3. A previsão perde o sentido.. Exercícios de previsão construídos sobre dados corrompidos.

4. Premissas não são testadas.. Premissas precisam ser testadas para serem confiáveis — caso contrário, toda interpretação futura da informação é suspeita.

5. A governança humana falha.. Vendi o TM1, um produto OLAP, para preparar pacotes gerenciais — e foi só para isso que ele foi usado. Nunca para obter insight real sobre as contas de resultado. Construí uma análise de vendas de imóveis segmentada por casa, imobiliária e amenities. Mostrou quais tipos de casas estavam vendendo, quem eram os corretores mais ativos e o que os novos empreendimentos estavam fazendo com água, esgoto, tráfego e eletricidade à medida que a população da região crescia. Todo esse insight, deixado na mesa — porque as pessoas não tinham aptidão para fazer perguntas à ferramenta. Sem o pessoal certo, num ambiente de IA elas vão falhar.

Eis a parte desconfortável. Você pode não saber que seus dados estão sujos — e está cansado de gastar dinheiro com TI que entrega retornos questionáveis. Há cinquenta anos, os CEOs não

assumem responsabilidade pelos projetos de TI. Com a IA viralizando, é hora de acordar.

Os vencedores não serão os que têm mais dados. Serão os que têm os dados mais confiáveis.