

Nova cobrança do WhatsApp: uma operação de atendimento ineficiente pode custar mais

A partir de 1º de outubro, cada resposta que uma empresa enviar a um cliente pela API oficial do WhatsApp passa a ser cobrada individualmente, conforme a documentação de preços publicada pela Meta

Fabiano Amorim (*)

Para o mercado brasileiro, o impacto será considerável: o WhatsApp está instalado em aproximadamente 98,3% dos smartphones no Brasil, com 97% dos usuários abrindo o aplicativo todos os dias, ou quase todos dias. Desses, 79,5% conversaram com marcas ou empresas pelo aplicativo nos últimos 12 meses, segundo a pesquisa realizada pela Opinion Box em março passado.

Ou seja: o canal que virou a porta de entrada do atendimento no Brasil acaba de ganhar um taxímetro, e a forma como cada operação desenha suas conversas vai definir quanto essa corrida custa. Mas antes de fazer as contas, vale precisar o que muda.

Desde novembro de 2024, as chamadas mensagens de serviço, aquelas respostas em texto livre que um atendente humano ou um robô envia dentro da janela de 24 horas aberta pelo contato do cliente, não geravam cobrança. Esse período termina em 1º de outubro, quando cada mensagem de serviço entregue passa a ser tarifada por unidade, sem faixa de desconto por volume.

Na mesma data, as mensagens de utilidade enviadas com a janela já aberta, como confirmações de pedido e avisos de entrega, perdem a isenção que tinham desde julho de 2025. E uma terceira frente já está em vigor: desde 1º de agosto, o Meta Business Agent, o agente de inteligência artificial da própria plataforma, é cobrado por consumo de tokens, um modelo no qual o preço acompanha a complexidade de cada interação, e não o número de mensagens trocadas.

Quem vai pagar a conta?

É importante dizer a quem essa conta chega. O aplicativo gratuito do WhatsApp Business, usado por pequenos negócios que respondem clientes manualmente pelo celular, segue sem cobrança por mensagem. A mudança atinge a camada da API oficial, justamente a infraestrutura que sustenta as operações de médio e grande porte, integradas a plataformas de atendimento, CRM e automação. A nova política mira exatamente as empresas que transformaram o WhatsApp em canal estruturado de relacionamento, e são elas que precisam refazer as contas agora, antes que a Meta publique as tarifas por país.

A consequência mais imediata é de natureza financeira, mas a mais profunda é de desenho de operação. Quando cada mensagem entregue vira um item de fatura, o estilo de



Fabiano Amorim.

conversa deixa de ser uma escolha estética. Um fluxo de chatbot que fatia uma resposta em cinco balões curtos, porque parece mais simpático na tela, gera cinco cobranças no lugar de uma. Um atendente que confirma, pergunta, repete e depois resolve produz o mesmo efeito. Operações que hoje trocam dezenas de mensagens para concluir um único protocolo vão perceber que a proximidade tem preço, e que ele se multiplica pelo volume de milhões de conversas mensais que os grandes contact centers brasileiros processam.

Isso inverte a lógica com a qual boa parte do mercado montou seus indicadores. Durante anos, a régua do atendimento digital foi a velocidade da primeira resposta e a satisfação ao fim do contato, métricas que continuam relevantes, mas que não capturam o que a nova cobrança expõe. A pergunta que passa a importar é quantas mensagens a operação gasta para resolver de fato um problema?

O custo por conversa resolvida, indicador que poucas empresas acompanham com rigor, tende a se tornar a medida central de eficiência do canal, porque combina em um único número a qualidade do desenho do fluxo, a assertividade da automação e a capacidade do time humano de concluir a jornada sem idas e vindas.

Com o agente da Meta cobrado por token e os agentes de terceiros cobrados por mensagem a partir de outubro, a escolha de qual inteligência artificial atende o cliente deixa de ser apenas uma comparação de qualidade de resposta e vira uma equação de custo por perfil de demanda. Interações simples e curtas podem favorecer um modelo; jornadas longas e cheias de exceções, outro.

As operações mais maduras vão tratar essa escolha como tratam a orquestração de qualquer tecnologia:

com camadas que permitem combinar e substituir motores de IA conforme o caso de uso, medindo o custo real de cada rota antes de escalar.

Existe, claro, o risco de leitura errada. Algumas empresas vão reagir cortando conversa, limitando resposta e empurrando o cliente para canais mais baratos, e essa rota cobra caro em outro lugar, porque o brasileiro escolheu o WhatsApp como espaço natural de relacionamento e não vai migrar por conveniência contábil alheia.

O impacto na operação

O redesenho vai passar, obrigatoriamente, por decisões que parecem miúdas e não são. Consolidar respostas em mensagens únicas e completas, revisar árvores de URA e de bot para eliminar etapas que só confirmam o óbvio, transferir para o autoatendimento resolutivo o que hoje consome trocas com atendentes, planejar o uso de templates de utilidade fora da janela em vez de acumular avisos dentro dela, tudo isso deixa de ser refinamento e passa a ser gestão de custo direto.

A boa notícia é que essas são exatamente as práticas que melhoram a experiência do cliente: ninguém gosta de conversa arrastada, de repetir informação ou de receber três mensagens para um aviso que caberia em uma. A tarifa da Meta, ao dar preço à ineficiência, alinha o incentivo financeiro ao que o consumidor sempre quis. E, ao revisar jornadas, automatizar e integrar as plataformas de atendimento ao cliente, empresas podem otimizar todo o processo de atendimento.

Até a publicação das tarifas, há uma janela de preparação que as operações de atendimento não deveriam desperdiçar. Mapear quantas mensagens de serviço a empresa envia por mês, quantas delas cada conversa consome até a resolução e quais fluxos concentram o desperdício é o mínimo para chegar a outubro sabendo o tamanho da própria exposição.

Quem fizer esse dever de casa vai negociar, priorizar e redesenhar com base em números; quem não fizer vai descobrir o impacto na primeira fatura. Para muitas operações pode ser necessário um olhar especializado e que seja capaz de conectar estratégia, tecnologia e automação para redesenhar jornadas mantendo uma boa experiência ao cliente.

A cobrança por mensagem tem data marcada, e a diferença entre pagar pela conversa e lucrar com ela será decidida agora, no desenho de cada fluxo, muito antes do boleto chegar.

(*) Head da Selbetti Customer Experience.



ESPAÇO JURÍDICO

Giovanna Araujo (*)

O fiscal viu demais — e ficou do lado de fora do laboratório

Em julho, os testadores de segurança do governo britânico desligaram de propósito as travas do modelo mais avançado da Anthropic e observaram o que ele fazia. Em 122 rodadas de teste, registraram 19 ações não autorizadas — 17 delas do Mythos 5. Na mais grave, um agente de IA criou identidades falsas na internet e usou engenharia social para pressionar o mantenedor de um projeto de código aberto a aceitar código malicioso. A tentativa falhou. Dois meses depois veio a resposta da empresa: o sucessor, Mythos 5.1, foi lançado sem que o AISecurity Institute (AISI) do Reino Unido pudesse testá-lo antes.

A revelação é do Financial Times. É a primeira vez que a Anthropic exclui o órgão britânico da avaliação pré-lançamento de um sistema de fronteira — o modelo saiu em 1º de setembro, restrito a organizações americanas verificadas. A empresa não deu justificativa pública; diz apenas coordenar com Washington a expansão do acesso. E o governo britânico admitiu o óbvio: não pode obrigar uma empresa privada a ceder seus modelos para teste. O regime é voluntário — e a boa vontade acabou.

Há geopolítica no meio. Desde junho, controles de exportação americanos restringem o acesso estrangeiro aos modelos de fronteira da empresa; a ENISA, agência europeia, só conseguiu testar o Mythos 5 depois de meses de negociação — e segue sem ver o 5.1. O teste de segurança, criado como instrumento de proteção pública, virou ativo de tecno-nacionalismo: quem pode auditar a IA mais poderosa do mundo passou a ser uma decisão de política externa dos Estados Unidos.

A ironia veio dias depois. Em 12 de setembro, Dario Amodei prometeu em ensaio “acesso permanente” de avaliadores independentes aos sistemas da Anthropic, com direito de publicar conclusões sem revisão da empresa. A promessa e a exclusão, lado a lado, contam a história inteira do modelo voluntário: o acesso existe — para quem a empresa e o governo do seu país-sede escolherem. Escrevi aqui, em agosto, sobre o piloto que pedia torre de controle. Agora sabemos o resto: ele também quer escolher quem senta nela. É o alerta resumido em uma linha — empresas devem participar da governança da IA, não dominá-la. Fiscalização cujo alcance é definido pelo fiscalizado não é fiscalização; é cortesia.

O detalhe mais incômodo é quem ficou do lado de fora. O Reino Unido construiu sua influência em IA justamente pela via da segurança: cúpulas internacionais, o instituto pioneiro, relatórios que o mundo cita. Tinha o melhor time de teste fora dos EUA — e não bastou. O que esperar então do Brasil, sem lei (o PL 2338 segue parado na Câmara), sem instituto de avaliação e sem escala? É a posição clássica do Sul Global: *rule taker*. Nosso acesso aos modelos que já operam em bancos, hospitais e tribunais brasileiros depende de boa vontade corporativa — e de Washington.

Para as empresas, a lição é imediata: se nem governos conseguem garantir auditoria, o contrato precisa garantir. Cláusulas de auditoria e transparência com fornecedores de IA, avaliação independente antes do uso crítico e supervisão humana documentada deixaram de ser sofisticação.

(*) - Sócia do Araujo Advs - advogada especializada em Governança de Inteligência Artificial

Exigência de consulta presencial em atendimento remoto acende debate sobre autonomia médica e acesso à saúde

Thamires Cappello (*)

Com o avanço da tecnologia, hoje em dia é possível realizar diversas tarefas do dia a dia sem sair de casa. Felizmente, a medicina é uma das áreas que também tem adotado essa prática: a pesquisa TIC Saúde, realizada pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), mostra a consolidação do atendimento remoto no Brasil e aponta que mais de 28% dos estabelecimentos de saúde já utilizam a teleconsulta. No entanto, por existirem especialidades que necessitam de uma avaliação física para cuidar da saúde do paciente, foi determinada a exigência de consultas presenciais obrigatórias a cada 180 dias para o acompanhamento de doenças crônicas, pela Resolução CFM nº 2.314/2022. Ainda que a diretriz busque fixar um limite de tempo para o acompanhamento a distân-

cia, profissionais de saúde questionam a aplicabilidade de um prazo rígido para cenários clínicos heterogêneos. O ponto central da controvérsia reside no engessamento do cuidado, visto que condições crônicas apresentam diferentes níveis de gravidade. Diante disso, médicos defendem que a recomendação por um exame presencial deve ser balizada pela avaliação individualizada de cada caso, e não por um intervalo de dias predeterminado.

Um paciente com quadro estável de hipertensão ou diabetes requer uma dinâmica de acompanhamento diversa daquela que apresenta oscilações graves ou alta complexidade. Na alopecia androgenética, por exemplo, o acompanhamento costuma ocorrer pela análise da evolução clínica e pelo exame de fotografias padronizadas, permitindo mensurar a resposta ao tratamento com

segurança. Isso não significa eliminar a consulta presencial, mas sim aplicá-la quando houver justificativa clínica. Portanto, a escolha do formato de atendimento deve decorrer do julgamento do profissional e do perfil do paciente, sem depender exclusivamente de um calendário fixo.

Além do aspecto assistencial, a medida impõe barreiras geográficas e econômicas a pacientes que vivem em regiões com escassez de especialistas, pessoas com dificuldades de locomoção ou que enfrentam custos e deslocamentos de longa distância para acessar o serviço de saúde. Nesses casos, a telemedicina atua como ferramenta essencial de adesão contínua ao tratamento e acompanhamento de rotina.

Diante desse cenário, a telessaúde se consolida como um recurso com-

plementar à consulta física, favorecendo o acompanhamento contínuo de pacientes com diagnóstico definido e viabilizando o contato direto com especialistas para ajustes. Para a classe médica, o debate não contrapõe o modelo presencial ao remoto, mas busca identificar a aplicação ideal de cada formato ao longo do tratamento.

Na obesidade, uma doença crônica que exige acompanhamento contínuo e de longo prazo, a obrigatoriedade de uma consulta presencial a cada 180 dias pode criar uma barreira desnecessária para muitos pacientes. Hoje, o acompanhamento remoto, quando bem estruturado, permite monitorar evolução, adesão, resposta ao tratamento e necessidade de ajustes com segurança.

No âmbito jurídico, advogados apontam incoerências na norma. A

exigência do conselho de classe gera entraves operacionais e insegurança jurídica ao conflitar com a legislação federal. A Lei nº 14.510/2022, que estabelece o marco da telessaúde no país, assegura expressamente a autonomia do profissional médico para escolher a metodologia, o recurso tecnológico e a frequência dos atendimentos no caso concreto.

Em suma, com a maturidade da saúde digital no país, especialistas ponderam que a construção de uma assistência inclusiva e segura passa pela determinação de diretrizes clínicas flexíveis, que preservem a autonomia médica e impeçam que prazos fixos se convertam em barreiras de acesso ao cuidado.

(*) Doutora em Ciências da Saúde e Saúde Pública pela USP. Mestre em Direito pela PUC/SP. Pesquisadora no Centro de Direito Sanitário da USP.