CORTEX AI: Agente de Inteligência Artificial para acelerar e apoiar decisões executivas

Kailany Bughi da Silva[1]

Yago Raphael de Melo Mouro[2]

RESUMO

Este artigo apresenta o CORTEX AI, um agente de inteligência artificial corporativo, que encurta o ciclo entre pergunta, evidência e ação. O objetivo é acelerar a tomada de decisão, reunindo em um único fluxo conversacional, a recuperação de conhecimento em documentos, consulta de dados estruturados e geração de materiais executivos prontos para uso, além do disparo de ações operacionais que conectam informação e execução. A proposta se justifica pela fragmentação das fontes informacionais nas organizações, pela dependência recorrente de equipes de dados para responder questões repetitivas e pelo retrabalho na produção de relatórios e comunicações, fatores que retardam decisões, ampliam custos e comprometem a padronização. A metodologia adotada baseou-se em arquitetura modular com serviço de aplicação web e canais de interação corporativa, conectores para armazenamento de objetos em nuvem e bancos de dados, tradução de linguagem natural para consultas estruturadas, busca semântica com indicação de trechos relevantes e composição automática de visualizações e documentos com formatação executiva. O software desenvolvido responde sobre documentos corporativos com citação do trecho de origem, executa consultas explicadas em bases relacionais, gera gráficos como imagens, produz relatórios e apresentações, envia mensagens por e-mail com anexos, cria convites de reunião com pauta e anexos e abre tarefas de acompanhamento a partir das respostas, preservando o contexto da interação. Conclui-se que o CORTEX AI reduz o esforço operacional, aumenta a velocidade e a consistência das entregas informacionais, transformando conversas em artefatos e ações verificáveis com impacto direto e mensurável na gestão e na comunicação executiva.

Palavras-chave: Agente de Inteligência Artificial. Análise de dados. Automação. Relatórios Executivos. Tomada de decisão.

1. Introdução

1.1 Contexto

O cotidiano das organizações é marcado por um imposto da informação, que é o tempo gasto procurando, reunindo e confirmando dados antes de decidir. De acordo com a McKinsey & Company (2023), profissionais do conhecimento despendem parcela relevante da jornada apenas nessa procura, o que dilui foco, alonga ciclos de aprovação e encarece entregas que deveriam ser rotineiras. Na visão da Forrester (2024), o quadro se agrava quando a qualidade dos dados é irregular: retrabalho se multiplicam, a confiança nas fontes oscila e perdas financeiras tornam-se recorrentes. Esses sinais não são apenas estatísticas; traduzem uma fricção estrutural entre a velocidade exigida pela gestão e a dificuldade prática de localizar, validar e padronizar evidências em tempo útil. Em ambientes onde decisões dependem de documentos dispersos, planilhas desconectadas e interlocutores distintos, cada pergunta vira um pequeno projeto e cada resposta, um risco de desalinhamento.

À medida que a adoção de IA cresce no ambiente de trabalho, persiste um hiato entre uso individual e impacto organizacional: segundo o artigo *Work Trend Index* 2024 da Microsoft,

muitas empresas já experimentam a tecnologia, mas carecem de orquestração — processos, padrões e fluxos que transformem conversas em artefatos e ações com rastreabilidade. Paralelamente, estimativas consolidadas pela Gartner (2024) evidenciam o custo da informação pouco confiável, reforçando a necessidade de respostas ancoradas em fontes claras e de materiais executivos padronizados.

É nesse contexto que se posiciona o CORTEX AI: um agente corporativo que opera como camada de decisão, unificando, em um único fluxo conversacional, a recuperação de conhecimento em documentos com citação do trecho de origem, a consulta em linguagem natural a dados estruturados com explicação do que foi consultado e a produção imediata de gráficos e relatórios executivos. A mesma interação também dispara ações operacionais, como o envio de e-mails, criação de convites de reunião e abertura de tarefas, reduzindo *handoffs* e mantendo o contexto entre canais (web, Microsoft Teams e WhatsApp).

1.2 Objetivos

1.2.1 Objetivo Geral

Desenvolver um agente de inteligência artificial corporativo que unifique, em um único fluxo conversacional, a recuperação de conhecimento em documentos, a consulta a dados estruturados e a geração de artefatos e ações, reduzindo o tempo entre pergunta, evidência e decisão executiva.

1.2.2 Objetivos específicos

Implementar perguntas e respostas em documentos armazenados na nuvem.

Integrar perguntas em linguagem natural sobre bancos de dados, convertendo-as em consultas de leitura e apresentando uma explicação clara do que foi consultado.

Disponibilizar um documento executivo que, a partir de um prompt, monta uma apresentação com narrativa, gráficos e referências e poder salvar em PDF ou PPTX.

Enviar e-mails com anexos e a criação de convites de reunião online com pauta e materiais anexados.

Habilitar *follow-ups* automáticos, criando tarefas com prazo e anexos em ferramenta de gestão de trabalho.

Reduzir em, no mínimo, 30% o tempo para produzir materiais executivos e responder perguntas recorrentes.

1.3 Justificativa

Como citado anteriormente, muitas organizações enfrentam fontes informacionais dispersas, sobrecarga de comunicação e retrabalho na produção de relatórios e apresentações, condições que retardam decisões e aumentam custos. A centralização de perguntas e ações em um agente de IA corporativo elimina buscas manuais, padroniza conteúdos executivos e viabiliza respostas auditáveis com citações e explicações de consultas, além de transformar insights em ações imediatas (e-mail, agenda e tarefas). O CORTEX AI endereça esses pontos ao integrar documentos, dados e execução em um fluxo único, elevando agilidade, consistência e clareza na gestão.

2. Referencial teórico

2.1 Sobrecarga informacional e pressão por decisões mais rápidas

Organizações operam sob fluxo contínuo de mensagens, reuniões e documentos distribuídos, gerando custo cognitivo para procurar, validar e consolidar informação. De acordo com o *Work Trend Index* 2024 da Microsoft, esse acúmulo como trabalho que desvia a atenção da análise para o simples ato de localizar evidências, atrasando decisões e ampliando retrabalho.

A diretriz é tornar o conhecimento encontrável no momento da decisão, com mínimo atrito. Em vez de navegar por pastas, e-mails e chats, a interface conversacional orquestra a descoberta e já apresenta evidências no formato utilizável. Respostas acompanhadas do trecho de origem e indicação clara da fonte aceleram a validação e reduzem disputas sobre a “versão correta”. Assim, um agente que centraliza perguntas e devolve evidências verificáveis atua como antídoto à sobrecarga informacional, reduz custos de coordenação e aumenta a velocidade de resposta.

2.2 Resposta à ação em atividades intensivas em linguagem

Grande parte do trabalho do conhecimento é linguagem: sintetizar documentos, redigir relatórios, preparar apresentações e alinhar comunicações. De acordo com o artigo *The economic potential of generative AI* da McKinsey & Company (2023), a IA generativa pode automatizar parcela expressiva desse esforço, sobretudo quando pergunta, evidência e produção do material ocorrem no mesmo fluxo. O ganho vai além de escrever mais rápido: trata-se de encadear pergunta, evidência e material executivo, sem trocar de ferramenta nem perder contexto.

O passo seguinte é colar execução na saída: da mesma conversa, enviar o relatório por e-mail, criar um convite com pauta e anexos ou abrir uma tarefa com responsáveis e prazos. Tornar explícito o que foi consultado em bases estruturadas e citar o trecho em documentos eleva a confiança e facilita aprovação executiva. Ao integrar descoberta, produção e orquestração em um único fluxo conversacional, o agente converte ganhos individuais em resultados organizacionais e acelera decisões em escala.

3 Trabalhos correlatos

3.1 Microsoft Copilot para Microsoft 365

O Microsoft Copilot está integrado aos aplicativos do Microsoft 365 (Word, Excel, PowerPoint, Outlook e Teams), auxiliando na criação de documentos, síntese de e-mails e resumos de reuniões dentro do ecossistema M365, com base no Microsoft Graph. Seu foco é produtividade em conteúdo e comunicação já armazenados no ambiente (por exemplo, arquivos no OneDrive/SharePoint e threads no Outlook/Teams), com recursos como geração e organização de materiais e “Copilot Notebooks”. É poderoso para conhecimento e colaboração dentro do M365, porém não é orientado, por padrão, a consultas de linguagem natural para SQL em bancos corporativos arbitrários fora desse ecossistema.

3.2 Gemini para Google Workspace

O Gemini oferece recursos nativos no Gmail, Docs, Sheets, Slides e Meet, incluindo redação e revisão de textos, transcrição e sumarização de reuniões, e geração de slides e gráficos a partir de prompts e conteúdo do Drive. No Sheets, o Gemini ajuda a encontrar respostas nos dados e a construir visualizações rapidamente. O escopo é consolidar criação e síntese dentro do Google Workspace, e a interação com bases relacionais corporativas ocorre indiretamente (por exemplo, via dados trazidos para Sheets), não como um mecanismo padrão de linguagem natural para SQL sobre bancos externos.

3.3 Amazon Q Business

O Amazon Q Business é um assistente corporativo voltado a dados e documentos empresariais, com conectores prontos (incluindo Amazon S3) e plugins para integrar aplicativos (como Salesforce) e executar ações diretamente a partir do chat. Pode operar apenas com dados da empresa e se integra a Slack/Teams e ao QuickSight para perguntas em linguagem natural sobre painéis e geração de insights, aproximando o uso de IA ao contexto operacional do cliente.

3.4 Síntese Comparativa

Embora Microsoft Copilot, Gemini for Workspace e Amazon Q Business acelerem a produção de conteúdo e a descoberta de informação, esses recursos tendem a entregar seu maior valor dentro de seus próprios ecossistemas. O CORTEX AI, por outro lado, foi concebido para operar entre ambientes e encurtar o caminho entre pergunta, evidência e ação em um único fluxo conversacional. Ao responder sobre documentos corporativos com citação do trecho de origem, consultar dados estruturados em linguagem natural com explicação do que foi executado e, na mesma interação, produzir gráficos como imagem e relatórios executivos em PDF e Power Point, o agente elimina etapas manuais, reduz retrabalho e padroniza a comunicação usada na tomada de decisão.

Além de integrar fontes e formatos, o CORTEX AI transforma insight em execução ao enviar e-mails com anexos diretamente do chat, criar convites de reunião no Google Meet com pauta e materiais, disparar follow-ups como tarefas no Azure Devops e manter o contexto contínuo entre Web, Microsoft Teams e WhatsApp. Esse encadeamento — da pergunta à evidência citada, da evidência ao material pronto e do material às ações — encurta ciclos decisórios, melhora a clareza das entregas e amplia a capacidade de resposta de áreas executivas. O resultado prático é uma experiência unificada que combina descoberta, análise e orquestração em um único agente, diferenciando o CORTEX AI em cenários reais onde empresas utilizam simultaneamente soluções Microsoft, Google e AWS e precisam de velocidade com consistência para apresentar, alinhar e agir.

4. PESQUISA e discussões

4.1 Resultados Obtidos

Evidências recentes indicam adoção ampla de IA no trabalho e apontam gargalos claros de produtividade que o CORTEX AI busca endereçar. O Work Trend Index da Microsoft indica que 75% dos profissionais do conhecimento já utilizam IA, embora muitas organizações ainda careçam de um plano para converter uso individual em impacto mensurável. Em paralelo, a McKinsey registra que trabalhadores do conhecimento gastam cerca de um quinto do tempo buscando e reunindo informações e projeta ganhos relevantes com IA generativa. No eixo da confiabilidade informacional, estudos reportam perdas superiores a US$ 5 milhões por ano por empresa, com 7% acima de US$ 25 milhões (Forrester, 2024), além de custo médio anual estimado em US$ 12,9 milhões (Gartner, 2024).

Para confrontar essas tendências com a nossa realidade, realizamos uma pesquisa piloto do tipo levantamento via Google Forms, por amostragem de conveniência entre profissionais de educação e tecnologia. A coleta ocorreu entre 10 e 15 de setembro de 2025 e totalizou 40 respondentes. O questionário reuniu itens fechados (1–5 e múltipla escolha) e uma questão aberta, cobrindo: esforço de busca/consolidação, canais de uso, utilidade de um agente que una documentos, dados e ações, importância de citações/explicações e frequência de produção de materiais executivos. Os principais achados reproduzem o padrão observado:

70% concordam que gastam tempo excessivo buscando e consolidando informações; 60% relatam dificuldade em unir dados e documentos de múltiplas fontes; 80% usam Microsoft Teams diariamente e 60% utilizam WhatsApp no trabalho; 90% consideram útil um agente que reúna consulta a documentos e dados e também execute ações (e-mail, convites, tarefas); 85% atribuem alta importância à citação do trecho de origem e 80% à explicação do que foi consultado; 62% produzem materiais ao menos quinzenalmente e 75% projetam redução de tempo igual ou superior a 30% nas entregas.

4.2 Discussão dos Resultados

. Os resultados do levantamento reforçam três diretrizes do projeto. EM primeiro lugar, a sobrecarga informacional é recorrente e legitima a necessidade de encurtar o caminho entre localizar evidências e decidir. Depois, rastreabilidade é condição de confiança: quando a resposta traz o trecho citado e explicita o que foi consultado, reduzem-se retrabalhos e disputas de narrativa, e ganham-se agilidade e padronização. Terceiro, operar nos canais de maior uso (Teams e WhatsApp) favorece adoção e continuidade do contexto entre ambientes. Observa-se ainda coerência entre frequência de produção e expectativa de ganho: quem entrega com maior cadência concentra projeções de economia acima de 30%, o que orienta a priorização do MVP para casos repetitivos e de alto volume (relatórios com gráficos padronizados, pareceres executivos e comunicações com anexos e agenda). Esses sinais validam as escolhas do CORTEX AI — respostas com citação, linguagem humana para SQL com explicação, geração imediata de gráficos/relatórios e acoplamento de ações — e sustentam a meta de reduzir o tempo médio por entregável, mantendo clareza e consistência nas entregas. Limitações do estudo incluem amostra por conveniência e autorrelato; ainda assim, os indícios são consistentes com as referências e apoiam o foco do projeto em padronização, rastreabilidade e orquestração no mesmo fluxo.

5. desenvolvimento do projeto

Para o desenvolvimento do CORTEX AI, adotou-se uma abordagem ágil com iterações curtas e incrementais, priorizando entregas utilizáveis ao final de cada ciclo. A lista de trabalho foi organizada em “fatias verticais” que percorrem toda a jornada de uso, da pergunta à resposta citada, da evidência ao material pronto e do material às ações, permitindo validar cedo o encadeamento completo. Nas fases iniciais, utilizou-se túnel seguro apenas para homologações rápidas; na implantação-alvo, a API em Flask é empacotada e executada no AWS Lambda, exposta pelo Amazon API Gateway, o que simplifica a operação e garante elasticidade sob demanda. A pesquisa mencionada na Seção 4 (questionário aplicado no Google Forms, com 40 participantes) foi utilizada como insumo de desenho: os resultados orientaram a priorização de casos de uso com maior cadência e impacto esperado (redução igual ou superior a 30% no tempo por entregável), além de critérios de rastreabilidade (citações e explicações) e de adoção em canais já empregados pelas equipes. A cada incremento, testes internos e o retorno do questionário guiaram ajustes de usabilidade (clareza das respostas, padronização dos artefatos, fluidez das ações) e decisões técnicas (consultas somente leitura em dados estruturados, limites e tempos de execução).

5.1 Arquitetura da solução

A arquitetura foi concebida em três camadas. Na borda, os canais de interação oferecem a interface conversacional, o histórico de mensagens e o acesso aos artefatos gerados. Todos os canais contam com integração de áudio: o usuário pode enviar mensagens de voz para transcrição automática e receber respostas em áudio sintetizado, além de anexar arquivos

de áudio para extração de conteúdo relevante. No centro, o orquestrador desenvolvido em Flask roda no AWS Lambda e recebe as mensagens pelo API Gateway, identifica a intenção e aciona módulos especializados: perguntas sobre documentos corporativos no AWS S3 retornam respostas com citação do trecho de origem; perguntas sobre dados estruturados são traduzidas para consultas somente leitura, executadas sobre bases relacionais e acompanhadas de explicação do que foi consultado; solicitações de visualização geram gráficos em PNG; e a composição de relatórios executivos organiza texto, tabelas e visualizações em PDF (e, quando necessário, exporta também PPTX para apresentação). Ainda no orquestrador, os módulos de e-mail e agenda integram-se ao Gmail e ao Google Calendar para envio de mensagens com anexos e criação de eventos com link do Google Meet, enquanto o módulo de follow-ups registra tarefas no Azure DevOps. Na base, os serviços de dados incluem PostgreSQL para persistir conversas e metadados e o S3 para armazenar artefatos (PNGs, PDFs, PPTXs), mantendo a separação entre conteúdo conversacional e arquivos gerados e permitindo evolução modular dos conectores sem impacto nos canais.

5.2 Desenvolvimento e integração

O desenvolvimento percorreu a cadeia completa de valor. Para documentos, implementou-se ingestão e extração de texto, indexação e busca, retornando respostas sempre ancoradas em trechos citados e indicando arquivos relevantes. Para dados estruturados, a interação em linguagem natural é mapeada ao esquema disponível, a consulta é gerada e executada com limites e tempo de execução controlados, e a resposta combina tabela-resumo com uma explicação clara do que foi feito. A geração de gráficos foi padronizada para aceitar parâmetros de negócio (período, recorte, segmento) e produzir imagens consistentes com o visual executivo.

A produção de relatórios executivos organiza capa, seções, tabelas e gráficos, incluindo referências de fonte no rodapé para reforçar rastreabilidade, além da opção de exportar a estrutura para apresentação quando aplicável. Na camada de ações, o envio de e-mails com anexos ocorre diretamente a partir do contexto da conversa; a criação de convites gera eventos no Google Calendar com pauta e materiais; e os follow-ups abrem tarefas com prazos, responsáveis e anexos no Azure DevOps, devolvendo os links para acompanhamento.

Em todos os canais (web, Teams e WhatsApp) o contexto de conversa, inclusive áudio transcrito, é preservado, de modo que uma análise iniciada em um ambiente possa ser concluída em outro sem perda de continuidade. Essa integração contínua, combinada às evidências da pesquisa, garantiu que o CORTEX AI evoluísse como um fluxo único.

6. Considerações finais

6.1 Contribuições e Impacto do CORTEX AI

O CORTEX AI endereça um problema recorrente nas organizações: a distância entre pergunta, evidência e ação. Ao unificar, em um mesmo fluxo conversacional, a recuperação de conhecimento em documentos, a interação em linguagem natural com dados estruturados e a produção de materiais executivos, o agente padroniza a comunicação e acelera decisões. A validação com fontes externas e o questionário aplicado (Google Forms, 40 participantes) reforçam esse direcionamento: a maioria relatou sobrecarga na busca e consolidação de informações, valorizou respostas com citação e explicação do que foi consultado e projetou redução de tempo igual ou superior a 30% nas entregas.

A implantação em AWS Lambda com Flask e a atuação em Web/React, Microsoft Teams e WhatsApp, com integração de áudio (transcrição e síntese), reforçam viabilidade técnica e aderência aos hábitos de trabalho. O projeto contribui com um modelo operacional conversacional que conecta descoberta, análise e execução: responde sobre documentos do S3 com citação do trecho de origem; traduz perguntas em consultas somente leitura a bases relacionais com explicação do que foi consultado; gera gráficos em PNG e relatórios executivos em PDF; e executa ações de e-mail, convites no Google Meet e follow-ups em tarefas.

Esse encadeamento reduz retrabalho, dá rastreabilidade às evidências, melhora a clareza dos materiais e sustenta a meta do MVP de reduzir, de forma mensurável, o tempo para produzir entregáveis e responder perguntas recorrentes. A neutralidade de ecossistema e o suporte a áudio ampliam a adoção em rotinas reais, inclusive móveis.

6.2 Desafios e Melhorias Futuras

O impacto do CORTEX AI depende da qualidade e atualidade dos dados e da curadoria de documentos. Há desafios na evolução do mapeamento de esquemas para processamento de linguagem natural (mudanças de estrutura e semântica), na ampliação de modelos de relatório sem perder padronização e na orquestração de áudio com baixa latência em diferentes canais. Também são necessários mecanismos sistemáticos de avaliação contínua (precisão percebida, utilidade, tempo economizado) e práticas de adoção organizacional (onboarding, boas práticas de pergunta, catálogo de exemplos). No plano técnico, há espaço para otimização de custo/latência no Lambda, cache de resultados e melhoria de robustez a consultas ambíguas. Os achados do questionário servem como referência para priorizar melhorias em casos de uso de maior recorrência e para ajustar métricas de sucesso alinhadas à redução de tempo e à clareza das evidências.

6.3 Recomendações para Trabalhos Futuros

Recomenda-se conduzir pilotos controlados por área, com comparação entre linha de base e uso do agente, medindo tempo por solicitação, taxa de reutilização de consultas e satisfação do usuário. É pertinente expandir conectores (por exemplo, SharePoint e Confluence), incorporar briefings executivos agendados, explorar simulações “e se…?” para cenários de negócio e evoluir o catálogo de KPIs e modelos de relatório.

Em pesquisa aplicada, sugerem-se estudos sobre explicabilidade em processamento de linguagem natural, avaliação de RAG com conjuntos internos e interação por voz em contextos ruidosos, além de estratégias de governança de conteúdo que mantenham consistência ao longo do tempo.

Considerando o questionário, recomenda-se também manter a medição contínua de ganhos de tempo e de qualidade percebida, ajustando os modelos e prompts a partir dos casos de maior impacto. Esses caminhos tendem a consolidar o CORTEX AI como um agente corporativo de alto impacto, transformando conversas em decisões apoiadas por evidências e em ações executáveis.

7. Referências bibliográficas

ACTIAN. The Costly Consequences of Poor Data Quality. Disponível em: https://www.actian.com/blog/data-management/the-costly-consequences-of-poor-data-quality. Acesso em: 18 set. 2025.

CORTEX AI. Projeto em andamento, mas o Frontend em React.Js já está disponível para pré-visualização do resultado. Disponível em: https://cortex-ai-agent.vercel.app. Acesso em: 30 set. 2025.

FORRESTER. Millions Lost in 2023 Due to Poor Data Quality: Potential for Billions to Be Lost with AI Without Intervention. Disponível em: https://www.forrester.com/report/millions-lost-in-2023-due-to-poor-data-quality-potential-for-billions-to-be-lost-with-ai-without-intervention/RES181258. Acesso em: 19 set. 2025.

GARTNER. Data Quality: Best Practices for Accurate Insights. Disponível em: https://www.gartner.com/en/data-analytics/topics/data-quality. Acesso em: 20 set. 2025.

IBM RESEARCH. What is retrieval-augmented generation (RAG)? Disponível em: https://research.ibm.com/blog/retrieval-augmented-generation-RAG. Acesso em: 22 set. 2025.

MCKINSEY & COMPANY. The economic potential of generative AI: the next productivity frontier. Disponível em: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier. Acesso em: 20 set. 2025.

MICROSOFT. Work Trend Index 2024 — AI at Work Is Here. Now Comes the Hard Part. Disponível em: https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part. Acesso em: 22 set. 2025.