UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA REVISÃO DE BANCO DE PREÇOS DE SANEAMENTO

Nelson Silva Junior [1]
Engenheiro Mecânico pela UNISANTA. MBA em Gestão Estratégica e Econômica de Projetos pela FGV. Mestre em Gestão de Negócios pela FIA. Especialista em eng. no Controle da Poluição pela UNISANTA, em Saúde Pública pela USP, em Gerenciamento de Projetos e em Ferramentas de Gestão Planejamento e Sustentabilidade pela JICA, em Gestão Eficiente em Saneamento pelo MPo e Mcid e em Novas Estratégias para o Controle de Perdas pela ABES. Gerente do Departamento de Engenharia de Custos da SABESP.
Mauricio Vizeu de Castro [2]
Engenheiro eletricista formado pela FEI (Faculdade de Engenharia Industrial), com MBA em Gestão Econômica e Estratégica de Negócios pela FGV (Fundação Getúlio Vargas) e certificado pelo PMI como Project Management Professional (PMP®). Cursando MBA em Artificial Intelligence for Business na Faculdade Impacta. Sócio-diretor da Vizca Engenharia e Consultoria Ltda, onde atua desde 2001.
Edinei do Carmo Lima [3]
Engenheiro Eletricista pela UNICID e em Análise de Sistemas pela FMU, com MBA em Gestão Empresarial pela FGV. Certificado como Project Management Professional (PMP®), Prince 2 Practitioner in Project Management, Scrum Foundation Professional Certificate (SFPC) e Green Belt (Six Sigma). Cursando MBA em Artificial Intelligence for Business na Faculdade Impacta e MBA em Lean Seis Sigma e Excelência a Operacional na UMJ. Sócio consultor da Vizca Engenharia e Consultoria Ltda, onde atua desde 2012.
Flávia Sartorato Pedrotti[4]
Arquiteta e Urbanista pela Universidade Estadual Paulista - UNESP, com mestrado profissional em Gestão e Regulação de Recursos Hídricos – ProfÁgua pela mesma instituição, com foco em Governança do Setor de Saneamento Básico – abastecimento de água e esgotamento sanitário. Especialista em Gestão Estratégica e Econômica de Projetos pela FGV (Fundação Getúlio Vargas) e certificada como Project Management Professional (PMP®). Sócia consultora da Vizca Engenharia e Consultoria Ltda, onde atua desde 2013.
Kauê Guedes Duarte[5]
Engenheiro Eletricista pela Universidade Estadual Paulista – UNESP, MBA em Saneamento Ambiental pela FESPSP - Fundação Escola de Sociologia e Política de São Paulo, MBA em Gestão de Contratos e Licitações pela PUC Minas, com Especialização sobre controle de perdas no abastecimento de água pela Japan International Cooperation Agency – JICA. Coordenador de Projetos Estratégicos do Departamento de Engenharia de Custos da SABESP.

Endereço[1]**:** Avenida Wallace Simonsen, 840 - Nova Petrópolis – São Bernardo do Campo – SP – Brasil – CEP: 09771-210 - Tel: (11) 95059-2366 – e-mail: nsiljunior@sabesp.com.br

RESUMO

Este artigo descreve o desenvolvimento e a aplicação de um sistema baseado em Inteligência Artificial Generativa para acelerar a revisão do banco de preços de obras e serviços de engenharia da Sabesp. Com alta de investimentos e a velocidade necessária para o cumprimento dos compromissos assumidos pela companhia perante a universalização do saneamento, a necessidade de um banco de preços atualizado e condizente com a realidade do mercado, das obras e das novas tecnologias ganhou relevância e urgência. O extenso banco de preços da empresa, que conta com mais de 3.500 itens de serviço e mais de 5 mil insumos e é específico para saneamento, precisava de atualizações e a revisão manual de cada item se mostrou inviável, e o processo de trabalho precisou ser repensado. A solução foi estruturada em cinco ondas evolutivas, entre agosto de 2025 e maio de 2026, combinando Large Language Models (LLMs), técnicas de Geração Aumentada por Recuperação (RAG) e uma plataforma integrada de gestão denominada sistema GPT-Vizca. Os resultados demonstram que a IA atua como copiloto técnico, redirecionando o trabalho dos analistas da busca e compilação de informações para o julgamento crítico especializado, com ganhos expressivos em velocidade, padronização e qualidade das entregas.

PALAVRAS-CHAVE: Inteligência Artificial, Banco de Preços, Saneamento

INTRODUÇÃO

O saneamento básico no Brasil enfrenta o desafio histórico de universalizar os serviços de abastecimento de água e coleta e tratamento de esgoto — missão que demanda investimentos vultosos em infraestrutura e, consequentemente, uma gestão de custos cada vez mais rigorosa e transparente. A elaboração de orçamentos confiáveis e realistas é condição essencial tanto para a viabilização dos investimentos quanto para a contratação bem fundamentada de empreiteiras, para a correta regulação de ativos e regulação tarifária do setor, contribuindo para a boa execução das obras, para a prevenção de pleitos contratuais e para a eficiência na aplicação de recursos.

O Novo Marco Legal do Saneamento Básico - Lei Federal nº 14.026/2020 (BRASIL, 2020) estabeleceu a universalização dos serviços de abastecimento de água e esgotamento sanitário como meta nacional até 2033, condicionando a manutenção das concessões ao cumprimento de metas progressivas de atendimento. No caso da Sabesp, o contrato de concessão firmado com a URAE-1 — Sudeste, se destina a assegurar a universalização dos serviços até 31 de dezembro de 2029 (SÃO PAULO, 2024), antecipando em quatro anos o prazo nacional, e impõe à Sabesp um volume de investimentos sem precedentes em sua história.

Os números em relação ao investimento necessário para a universalização são vultuosos. O Instituto Trata Brasil (2025) apontou recentemente que o montante ainda necessário para se universalizar os serviços de saneamento é da ordem de R$ 454,1 bilhões. Esse contexto impõe às empresas do setor uma gestão de custos crescentemente rigorosa, transparente e ágil — sob pena de comprometer o cumprimento das metas de universalização estabelecidas pelo marco regulatório do saneamento básico.

A Sabesp — Companhia de Saneamento Básico do Estado de São Paulo, maior empresa de água e esgoto do Brasil, atendendo 28,4 milhões de pessoas em 375 municípios do estado (SABESP, 2025) — atravessou, em julho de 2024, seu processo de desestatização, com a e a transferência de seu controle acionário para a iniciativa privada, na maior privatização do ano no país, movimentando R$ 14,7 bilhões. O governo paulista, que detinha 50,3% das ações, passou a uma participação minoritária de cerca de 18,3%, e o Grupo Equatorial ingressou como investidor de referência (Agência Brasil, 2024). Essa mudança de controle inaugurou um novo ciclo estratégico para a companhia.

Os efeitos no ritmo de investimentos foram imediatos e expressivos: em 2025, a companhia realizou aportes recordes de R$ 15,2 bilhões — mais que o dobro do que foi aplicados no ano anterior —, levando água potável a 1,8 milhão de novos paulistas (SABESP, 2026). Para 2026, a projeção é de um novo recorde de aproximadamente R$ 20 bilhões, e no Investor Day 2026 a administração elevou a projeção de Capex para o período 2026-2030 de R$ 64 bilhões para R$ 84 bilhões (INFOMONEY, 2026). Esse ritmo de investimentos sem precedentes na história da companhia exige um aparato de gestão de custos que responda com igual velocidade e precisão.

A engenharia de custos no setor de saneamento apresenta particularidades que a distinguem significativamente da construção civil convencional. As obras de saneamento envolvem desde intervenções lineares extensas — redes de distribuição de água e coletores de esgoto — até estações de tratamento de complexidade elevada, passando por reservatórios, elevatórias, adutoras, ligações prediais e obras de reabilitação. Cada tipologia apresenta metodologias executivas específicas, fortemente influenciadas por variáveis como a presença e o volume de tráfego sobre a via, o tipo e a resistência do solo (argiloso, arenoso, rochoso), a profundidade e a altura do lençol freático, entre outros. Todos esses fatores influenciam diretamente os custos unitários e precisam ser refletidos nas composições de preços para que os orçamentos tenham aderência à realidade executiva. Destaca-se que outros bancos de preços de referência convencionais não costumam contemplar a especificidade técnica do saneamento. Serviços como assentamento de adutoras e coletores de grande diâmetro ou não integram essas bases ou figuram de forma insuficiente para fundamentar contratações com o nível de detalhamento exigido pelo setor.

A relevância dessas variáveis reflete-se diretamente na estruturação dos bancos de preços, instrumentos indispensáveis para balizar contratações públicas e privadas, contabilizar ativos de infraestrutura, elaborar estudos de viabilidade e subsidiar a regulação tarifária do setor. Um banco de preços bem estruturado e permanentemente atualizado é, portanto, um ativo estratégico para qualquer empresa de saneamento. Contudo, a vasta extensão e a heterogeneidade dessas bases de dados tornam os processos tradicionais de atualização

extremamente morosos e onerosos, comprometendo a agilidade exigida por um ambiente de negócios em aceleração.

Em meados de 2025, foi iniciado o processo de revisão do banco de preços da companhia, contemplando o estudo, a reestruturação e a atualização dos bancos de preços de obras e serviços de engenharia, estudos e projetos, serviços operacionais e insumos. O escopo incluiu pesquisas de mercado para definição de preços de referência, verificações de produtividade por meio de apropriações de campo e levantamentos bibliográficos, revisão de categorias profissionais e elaboração de Especificações Técnicas (ET), Composições de Preços Unitários (CPU) e Regulamentos de Preços e Critérios de Medição (RPCM).

O banco de preços da Sabesp é composto por quatro grupos: Obras e Serviços de Engenharia, com aproximadamente 2.300 serviços; Estudos, Projetos e Serviços de Apoio, com cerca de 460 serviços; Serviços Operacionais, com aproximadamente 1.600 serviços; e Insumos, com mais de 5.300 itens — totalizando quase 10 mil referências de preços cujos valores precisam ser permanentemente atualizados para refletir as condições reais de mercado. Nesse contexto, a engenharia de custos consolida-se como uma área estratégica, e inovar em sua metodologia de revisão tornou-se uma necessidade operacional. A busca por metodologias confiáveis, mas ao mesmo tempo velozes, de revisão do banco representa uma oportunidade de garantir maior precisão orçamentária, completude de informações e agilidade de resposta frente às demandas do mercado e às exigências regulatórias.

Um banco com milhares de serviços que precisa espelhar fielmente a prática executiva para gerar orçamentos confiáveis, aliado a um ritmo de investimentos sem precedentes —, a revisão manual de todo esse universo tornou-se praticamente inviável. A velocidade necessária para revisar os preços, deixá-los aderentes às normas técnicas vigentes e às condições reais de campo, exige a capacidade de cruzar grandes volumes de dados de fontes distintas, extrair informações relevantes de documentos técnicos não estruturados e gerar subsídios organizados que orientem a tomada de decisão dos analistas com agilidade e segurança.

Nesse cenário desafiador, que a Inteligência Artificial Generativa, por meio de *Large Language Models* (LLMs) e técnicas como a Geração Aumentada por Recuperação (RAG — *Retrieval-Augmented Generation*), surgem como uma solução inovadora na Sabesp. Os LLMs combinam velocidade de processamento com interpretação avançada da linguagem natural técnica, permitindo extrair, comparar e estruturar informações de múltiplas fontes de maneira sistemática e replicável.
A técnica vai além, ela ancora as respostas do modelo exclusivamente em bases documentais predefinidas e controladas. Isso não apenas mitiga o risco de alucinações, mas também garante a rastreabilidade das informações geradas.

Dessa forma, essa arquitetura não substitui – nem poderia substituir – o julgamento crítico do especialista. Em vez disso, ela atua como um instrumento de potencialização da produtividade e da qualidade das entregas ao automatizar etapas cognitivas complexas, liberando os analistas para focar em decisões estratégicas.

OBJETIVOS

O objetivo principal foi acelerar o processo de revisão do banco de preços de saneamento da Sabesp, com embasamento e segurança da informação. Para isso, foram admitidos os seguintes objetivos específicos:

- Identificação de atividades da equipe técnica que poderiam ser aceleradas com a utilização de inteligência artificial;
- Desenvolvimento de testes para garantia da qualidade e funcionamento das ferramentas;
- Desenvolvimento de sistema para aplicação de inteligência artificial em larga escala.

METODOLOGIA UTILIZADA

Do ponto de vista de seus objetivos, esta pesquisa classifica-se predominantemente como exploratória-descritiva: exploratória por investigar uma aplicação sem precedentes relevantes conhecidos no domínio do saneamento brasileiro, em que a viabilidade e os condicionantes do uso de IA Generativa precisavam ser empiricamente estabelecidos; e descritiva à medida que o sistema se consolidou e seus resultados passaram a ser sistematizados e analisados (GIL, 2008).

Quanto aos procedimentos técnicos, trata-se de uma pesquisa-ação, uma vez que os pesquisadores estiveram diretamente implicados na implementação e no aprimoramento contínuo da solução ao longo de todo o processo (THIOLLENT, 2025). A abordagem é qualitativa, privilegiando a análise aprofundada dos fenômenos observados — como a qualidade das composições geradas, a aderência às normas técnicas e a acurácia das comparações entre bases (MARCONI; LAKATOS, 2021).

O desenvolvimento da pesquisa foi estruturado em ondas evolutivas, sucessivas e interdependentes, entre agosto de 2025 e maio de 2026. Essa estratégia metodológica, inspirada no conceito de maturidade tecnológica progressiva, permitiu testar, validar e refinar incrementalmente cada etapa antes de avançar para a seguinte, reduzindo riscos técnicos e assegurando o aprendizado contínuo. Cada onda cumpriu um papel específico na construção do conhecimento:

- **1ª onda —** Prova de Conceito: validação da hipótese central em escala reduzida;
- **2ª onda —** Estruturação em Larga Escala: desenvolvimento de agentes, sistematização de prompts e processamento do primeiro lote;
- **3ª onda —** Uniformização e Validação Comparativa: padronização de bases, testes cruzados e refinamento dos instrumentos;
- **4ª onda —** Consolidação da Infraestrutura: desenvolvimento do sistema GPT-Vizca, banco de dados e geração automatizada de documentos;
- **5ª onda —** Aprimoramento e Expansão de Escopo: refinamento contínuo, tratamento de casos especiais (MND – Métodos não destrutivos) e desenvolvimento do chatbot com RAG.

O processo de calibração da IA baseou-se em uma estratégia de aprendizado por referência: itens já revisados manualmente pela equipe, com resultado conhecido e validado, foram utilizados como padrão-ouro para o ajuste dos *prompts*. Esse método ancorou o refinamento em evidências concretas — o que o especialista efetivamente produz e considera correto — em vez de critérios abstratos ou de difícil operacionalização. A partir dessa referência, os *prompts* foram ajustados iterativamente até que o output do sistema se aproximasse do resultado esperado em termos de cobertura de fontes, nível de detalhe, estruturação das informações e adequação ao formato padrão de entrega.

Foram adotados controles de segurança da informação desde a concepção do sistema, incluindo a definição de fontes autorizadas para a arquitetura RAG, controles de acesso baseados em perfis de usuário, armazenamento em servidor próprio e registro de todas as operações realizadas (*logs*). A confidencialidade dos dados estratégicos da organização foi tratada como premissa inegociável ao longo de todo o processo, influenciando diretamente as escolhas de arquitetura tecnológica.

A elaboração deste artigo contou com o uso de ferramentas de Inteligência Artificial Generativa AdaptaOne26® e Claude.ai® (ic) para auxílio na organização de ideias e revisão textual. O material gerado foi analisado e revisado pelos autores, que se responsabilizam pela redação final do trabalho. O uso da ferramenta não substituiu a atividade intelectual, a análise científica, a formulação de conclusões nem a autoria do presente estudo.

DESENVOLVIMENTO E RESULTADOS OBTIDOS

Para compreender o valor agregado pela IA neste contexto, é necessário primeiro compreender o processo manual anteriormente utilizado, e verificar em quais atividades a IA seria benéfica.

O Processo Manual de Revisão

De forma didática e simplificada, pode-se afirmar que a revisão de um serviço do banco de preços inicia-se por uma etapa preparatória de levantamento de referências, que envolve a compilação de dados e informações técnicas dispersos em múltiplas fontes. Essa base apoia a elaboração da memória de cálculo e das premissas adotadas. A etapa subsequente consiste na estruturação da CPU (Composição de Preços Unitária) e da elaboração das Especificações Técnicas (ET) e da Regulamentação de Preços e Critérios de Medição (RPCM).

Quando há dados de apropriação em campo, existe uma etapa a mais, de equalização de informações que antecede a estruturação da CPU. A seguir, estão listadas as atividades da etapa preparatória (1 a 4) e da revisão dos documentos (5 e 6).

1. Verificação da existência do serviço em bases de referência externas (bancos estaduais e municipais, entre outros);
2. Comparação com a composição vigente da Sabesp — insumos, coeficientes de produtividade, unidade de medida e outros parâmetros;
3. Identificação e leitura das normas técnicas nacionais, internacionais e dos procedimentos internos da Sabesp (NTS) pertinentes ao serviço, avaliando quais exigências se refletem no procedimento executivo e na composição de preços;
4. Realizar pesquisa de mercado e entender a dinâmica do serviço – pontos críticos, variáveis, metodologia executiva;
5. Estruturação da CPU (Composição de Preços Unitária), com memória de cálculo justificada – teórica e cálculos envolvidos;
6. Redação das Especificações Técnicas (ET) e dos Regulamentos de Preços e Critérios de Medição (RPCM) no formato padrão da Sabesp.

Cada uma dessas atividades apresentava desafios específicos no processo manual. A busca em bases externas era dificultada pela ausência de nomenclatura padronizada entre fontes: um mesmo serviço pode ser denominado de formas distintas em diferentes bases, exigindo do analista um trabalho de interpretação técnica para identificar equivalências que nem sempre são evidentes. A pesquisa normativa demandava acesso a múltiplos portais e sistemas, com risco de omissão de documentos relevantes. A consolidação das informações era realizada em formatos heterogêneos, variáveis de analista para analista, o que dificultava a revisão por pares e o controle de qualidade. E a redação das ETs e RPCMs exigia habilidades de comunicação técnica escrita que nem sempre estavam disponíveis em igual medida em todos os membros da equipe.

A tabela a seguir apresenta a análise das atividades frente à utilização da IA:

Tabela 1 – Atividades e potencial de uso de IA

Atividade	*Potencial de uso de IA*
1: Consulta a outras bases	Sim, para posterior análise humana
2: Comparação com serviço Banco Sabesp	Sim, para posterior análise humana
3: Identificação de normas pertinentes	Sim, para posterior análise humana
4: Dinâmica do Serviço	Parcial – possíveis insights, com diretrizes para apropriações em campo e posterior análise humana
5: Estruturação da CPU	Parcial – possível proposta inicial, mediante justificativa, para análise humana
6: Redação ET e RPCM	Parcial – versão inicial

Fonte: Autores.

A etapa preparatória resultava num referencial teórico produzido a custo elevado de tempo e atenção especializada — tempo que, idealmente, deveria ser investido na etapa seguinte: a análise técnica propriamente dita, com o julgamento crítico sobre a pertinência dos insumos, a correção dos coeficientes e a adequação das especificações às condições reais de execução.

A partir da identificação deste conjunto de atividades, iniciou-se de forma exploratória o primeiro ciclo para investigar a possibilidade de uso de inteligência artificial para aceleração. Em seguida, foram realizadas ondas evolutivas como ciclos sucessivos de teste, validação e refinamento — cada onda com seu papel específico no processo de construção do conhecimento: prova de conceito (1ª), estruturação do método (2ª), validação comparativa (3ª), consolidação dos dados (4ª) e aprimoramento com expansão (5ª).

Ao longo do processo foram identificadas também as seguintes atividades que demandavam tempo excessivo da equipe:

- Controle de serviços em produção, por qual analista;

- Apoio às equipes de apropriadores em campo, em especial para os pontos de atenção a serem checados em obra.

Inteligência Artificial Generativa e Large Language Models

Os *Large Language Models* (LLMs) são modelos de linguagem de grande escala treinados em vastos corpos textuais por meio de técnicas de aprendizado profundo, fundamentalmente baseada na arquitetura *Transformer*, introduzida por Vaswani et al. (2017). Essa arquitetura, baseada em mecanismos de atenção (*attention*), permitiu avanços significativos na capacidade dos modelos de capturar relações semânticas complexas em textos extensos. A partir desse fundamento, trabalhos como o GPT-3 (Brown et al., 2020) demonstraram que modelos suficientemente grandes e bem treinados são capazes de realizar tarefas de linguagem complexas a partir de poucos exemplos (*few-shot learning*) ou mesmo sem exemplos explícitos (*zero-shot learning*).

Para o contexto deste trabalho, a capacidade mais relevante dos LLMs é a interpretação e geração de linguagem natural técnica: o modelo é capaz de compreender nomenclaturas e conceitos específicos do domínio de saneamento, identificar equivalências semânticas entre serviços com denominações distintas em bases diferentes, estruturar informações em formatos padronizados e redigir textos técnicos coerentes com base em documentação de referência. Essas capacidades, que normalmente demandariam horas de trabalho especializado, podem ser realizadas em minutos quando adequadamente orientadas por *prompts* bem calibrados.

Retrieval-Augmented Generation (RAG)

A arqutietura RAG, sistematizada por Lewis et al. (2020), combina dois componentes: um módulo de recuperação de informações (*retriever*), que seleciona documentos relevantes de uma base de conhecimento controlada, e um modelo generativo (*generator*), que produz a resposta com base nos documentos recuperados. Essa arquitetura oferece vantagens significativas sobre o uso isolado de LLMs em contextos corporativos: as respostas são fundamentadas exclusivamente nas fontes autorizadas fornecidas, o que reduz o risco de alucinações e garante rastreabilidade; a base de conhecimento pode ser atualizada independentemente do modelo, sem necessidade de retreinamento; e o acesso às informações pode ser controlado por políticas de segurança corporativas.

No contexto deste trabalho, o RAG foi adotado tanto para o sistema de revisão de preços quanto para o chatbot destinado aos apropriadores em campo. As bases reúnem bancos de referência externos, o Banco de Preços Sabesp, normas, procedimentos (empresariais, operacionais e comerciais) e catálogos de fornecedores. Em ambos os casos, a restrição do conhecimento do sistema a fontes predefinidas foi simultaneamente uma decisão técnica (qualidade das respostas) e uma decisão de governança (segurança da informação).

Primeira Onda: Prova de Conceito com LLMs

A primeira onda constituiu uma prova de conceito (*proof of concept*), realizada a partir de um experimento exploratório com LLMs para extração e análise comparativa de dados de fontes não estruturadas – bancos de preços externos e o banco de preços vigente da Sabesp. O procedimento consistiu em submeter composições de custos ao modelo, instruído por *prompts* específicos, e avaliar sua capacidade de identificar discrepâncias em insumos, coeficientes de produtividade e especificações técnicas.

Esta onda ocorreu entre agosto e setembro de 2025. Nesta onda foram observadas as primeiras ocorrências de alucinação — respostas geradas pelo modelo sem correspondência factual nas fontes consultadas, um comportamento bem documentado na literatura sobre LLMs. Esse fenômeno demandou um trabalho de calibração técnica: foi necessário estabelecer critérios objetivos para definir o que constituía um serviço similar para fins de comparação de composições. Sem esses critérios explicitados nos *prompts*, o modelo tendia a apresentar serviços superficialmente relacionados por denominação, mas tecnicamente incomparáveis em termos de metodologia executiva e estrutura de custos. O refinamento dos *prompts* com base nesses

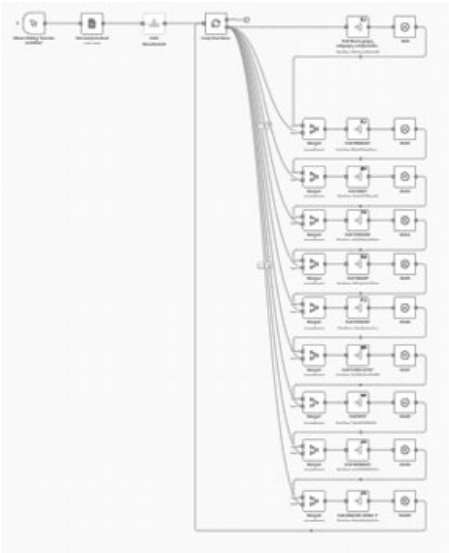
critérios técnicos objetivos foi a resposta metodológica central desta fase, e os resultados positivos obtidos forneceram o suporte empírico necessário para a continuidade do estudo.

Segunda Onda: Estruturação em Larga Escala

Com a hipótese validada em escala reduzida, passou-se à estruturação do sistema de coleta e análise de dados. Foram desenvolvidos agentes de IA — módulos especializados em tarefas específicas do processo de revisão — e um conjunto sistematizado de *prompts* especializados para processamento em lote. Paralelamente, ampliou-se o universo de bancos de preços externos consultados, incorporando novas fontes de evidência e aumentando a cobertura do sistema.

Em outubro de 2025, processou-se o primeiro lote de 50 serviços, demarcando a transição do ambiente controlado do experimento inicial para a operação em condições reais. Esse lote funcionou como uma amostra piloto para avaliar a consistência e a replicabilidade do método. O principal desafio desta onda foi o ajuste do nível de detalhe das informações entregues: os primeiros resultados apresentavam informações ora excessivamente sucintas, ora desnecessariamente prolixas. O objetivo passou a ser garantir que o output fornecesse o embasamento necessário e suficiente para apoiar a análise do analista humano — um equilíbrio que só seria atingido mediante refinamento iterativo, comparando os resultados do sistema com itens já revisados manualmente pela equipe.

Figura 1 – Agentes de IA – Segunda Onda – Teste em escala



Fonte: Autores.

Terceira Onda: Uniformização de Bases

Esta onda enfrentou o desafio da heterogeneidade das fontes: cada base de dados consultada possuía estrutura, nomenclatura e granularidade próprias. Apesar do resultado inicialmente satisfatório da onda anterior, observou-se que existia um potencial de melhoria do resultado com a melhoria da qualidade das fontes, ou seja, bases de mais fácil leitura para a IA. Assim, realizou-se um trabalho sistemático de padronização terminológica para aumentar a comparabilidade dos dados — etapa essencial para a validade interna da pesquisa.

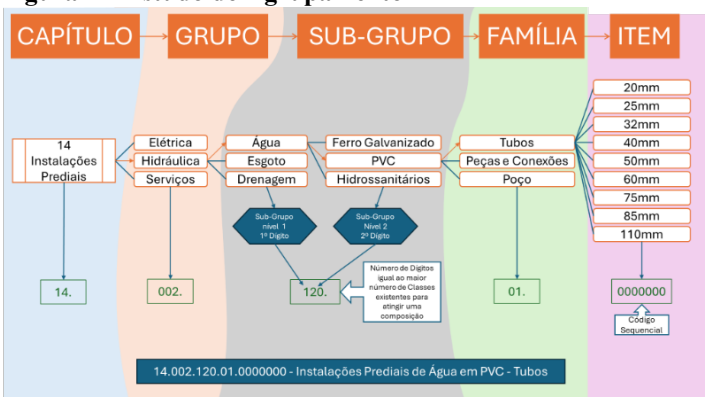
Além disso, foram realizados testes a partir de cada melhoria – tanto de bases como de prompts. Os resultados eram comparados com os resultados pretendidos em serviços realizados manualmente e de pleno conhecimento da equipe. Tal processo permitiu identificar lacunas, vieses e oportunidades de melhoria nos *prompts*, além de calibrar o nível de confiança que deveria ser atribuído às recomendações automáticas em diferentes situações.

A partir desta onda, a plataforma passou a contemplar também a geração automatizada das fichas de Especificação Técnica (ET) e de Regulamentação de Preços e Critérios de Medição (RPCM) no formato padrão da Sabesp. A equipe de analistas participou ativamente do processo de melhoria desta funcionalidade, indicando de que forma trabalhava, quais informações eram mais relevantes para cada etapa da análise e de que forma gostariam que o conteúdo fosse apresentado. Esse processo colaborativo de desenvolvimento transformou o refinamento dos prompts em uma atividade guiada pela prática real, favorecendo a utilização posterior pela equipe.

Quarta Onda: Sistema GPT-Vizca e Banco de Dados

Esta onda consolidou a infraestrutura de registro e análise dos dados produzidos nas fases anteriores. Desenvolveu-se o sistema GPT-Vizca, com o banco de dados Relats BDP, que passou a armazenar, organizar e disponibilizar os serviços processados. Ao final de fevereiro de 2026, o sistema havia registrado 3.407 serviços processados. A criação de uma interface dedicada para a equipe de analistas, com controles de acesso por perfil de usuário e funcionalidades de edição, validação e acompanhamento, garantiu a rastreabilidade e a integridade dos dados — requisitos fundamentais para a confiabilidade dos resultados e para a auditabilidade do processo. Em paralelo a este trabalho, os serviços foram agrupados de forma a possibilitar ganho de escala considerando que os serviços individuais poderiam compartilhar especificações técnicas e regulamentações de preços e critérios de medição. A Figura 2 a seguir apresenta a estrutura conceitual do agrupamento adotado.

Figura 2 – Estudo de Agrupamento



Fonte: Autores.

A interface de trabalho foi estruturada em dois módulos complementares. O módulo de Referencial Teórico apresenta os resultados da análise realizada pelo LLM em cada base de dados consultada, organizados por tópico: normas aplicáveis, composições similares de bases externas, especificidades técnicas relevantes identificadas nas fontes. O módulo de Base IA (Analistas) oferece um painel de controle operacional com informações de acompanhamento (analista responsável, datas de início, término e revisão), uma imagem ilustrativa do serviço, e os resultados de Especificação Técnica e RPCM pré-preenchidos no formato padrão. A Figura 3, a seguir, apresenta esta interface.

Figura 3 – Interface dos Analistas



Fonte: Autores.

O módulo também disponibiliza, para download, arquivos em PDF e Excel já formatados e preenchidos — a partir dos quais o analista conclui a análise e finaliza o produto, concentrando seu tempo na conferência técnica e nos ajustes necessários. A figura 4 apresenta um exemplo de especificação técnica extraída diretamente do sistema.

Figura 4 – Exemplo de extração de especificação técnica gerada por LLM



ESPECIFICAÇÃO TÉCNICA

Capítulo: CANTEIRO DE OBRAS **Grupo:** CERCAMENTO
Cat 1: **Cat 2:**
Cat 3: **Família:** TAPUME

Objeto

Fornecimento, montagem, manutenção e posterior remoção de tapume contínuo de fechamento provisório do canteiro, executado em chapas metálicas galvanizadas, incluindo estrutura de apoio (caibros de madeira 60×50 mm ou perfis leves, conforme projeto), fixadores, arremates e pintura de proteção/identificação, com sinalização e segurança conforme NR-18.

Materiais

Chapas metálicas galvanizadas (ondulada/lisa), esp. típica 0,430,50 mm (ou conforme projeto), sobreposição mínima longitudinal/transversal definida em projeto.

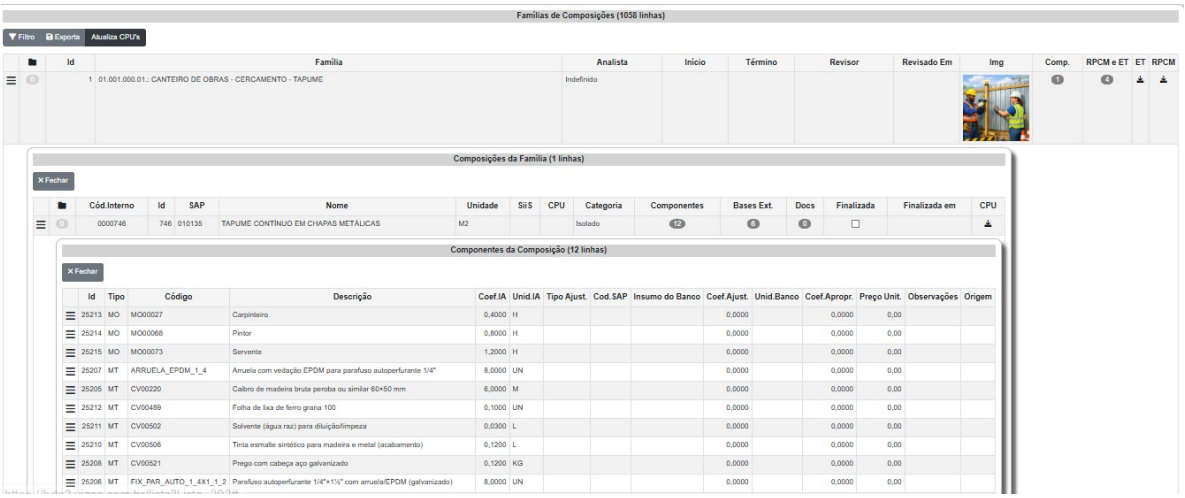
Estrutura de apoio: caibros de madeira 60×50 mm (ou equivalente), tratada/preservada quando aplicável, dimensionada como peça secundária segundo NBR 7190; admitir perfis metálicos equivalentes se previsto.

Fixação: parafusos autoperfurantes com arruela/vedação, complementados por pregos nas peças de

Fonte: Autores.

A Composição de Preços Unitários (CPU) é formada pelos insumos, serviços auxiliares, coeficientes de produtividade de cada um, unidades de medida e demais parâmetros necessários ao cálculo do custo unitário do serviço. O sistema GPT-Vizca apresenta uma base estruturada que organiza as informações de referência — insumos identificados nas bases externas, coeficientes comparativos, unidades de medida adotadas em cada fonte — facilitando e agilizando o trabalho do analista na etapa de análise e validação da composição. O sistema apresenta uma base que facilita o trabalho do analista. A figura 5 a seguir ilustra as informações disponibilizadas e os campos para preenchimento.

Figura 5 – Exemplo de Base para Análise da Composição de Preços



Fonte: Autores.

Quinta Onda: Aprimoramento Contínuo, Métodos Não Destrutivos e Chatbot com RAG

A quinta onda concentrou-se em dois eixos complementares e simultaneamente em desenvolvimento. O primeiro foi o aprimoramento sistemático dos instrumentos da pesquisa: refinamento contínuo dos *prompts* com base no feedback acumulado dos analistas ao longo das ondas anteriores, criação de *checklist* e curva ABC para priorização dos itens com maior impacto nos orçamentos, e desenvolvimento de referencial teórico específico para Métodos Não Destrutivos (MND) — categoria de serviços não contemplada no banco original.

Os Métodos Não Destrutivos representaram um caso especial que demandou uma abordagem adaptada — denominada, neste trabalho, de "onda paralela". Por se tratar de serviços com pouca ou nenhuma representação em bases de preços externas, a estratégia convencional de comparação com bancos de referência era insuficiente e produzia resultados de baixa utilidade. A solução desenvolvida foi reorientar o foco da consulta da IA: em vez de buscar referenciais de composição de preços, o sistema passou a consultar referenciais teóricos — normas técnicas específicas, publicações científicas, documentações de fabricantes de equipamentos e relatórios técnicos setoriais. Essa adaptação demonstrou a flexibilidade da arquitetura desenvolvida e consolidou um princípio metodológico importante: a estratégia de consulta da IA deve ser adaptada à natureza e à disponibilidade de referências de cada categoria de serviço, não aplicada de forma uniforme a todos os casos. Nesta etapa, foram elaboradas as bases teóricas para 583 itens.

O segundo eixo da quinta onda foi o desenvolvimento de um chatbot baseado em RAG, configurado com fontes autorizadas da Sabesp — Normas Técnicas (NTS), procedimentos operacionais, catálogos de fornecedores e especificações técnicas. A arquitetura RAG foi adotada por duas razões convergentes e igualmente relevantes: a qualidade das respostas, garantida pelo ancoramento exclusivo em documentação técnica predefinida e validada; e a segurança da informação, assegurada pela restrição do acesso do modelo aos documentos explicitamente autorizados, sem acesso a dados sensíveis ou estratégicos da companhia. A preocupação com a exposição de informações confidenciais norteou as decisões de design do sistema desde sua concepção.

O chatbot passou a ser utilizado tanto pela equipe de campo (apropriadores) quanto pelo backoffice, para finalidades diversas: tirar dúvidas técnicas sobre procedimentos executivos, compreender detalhes de normas específicas, localizar requisitos aplicáveis a determinados serviços, identificar critérios de medição, consultar especificações de materiais e equipamentos, entre outros usos. A adoção espontânea e multifuncional da ferramenta — para além do escopo originalmente previsto — foi um indicativo relevante de que o sistema respondia a necessidades reais e latentes da organização, não atendidas pelos canais de informação existentes. A Figura 6 apresenta a interface inicial do Chatbot.

Figura 6 – Interface do Chatbot com RAG



Fonte: Autores.

Procedimentos Éticos e de Segurança

Em conformidade com as boas práticas de pesquisa e desenvolvimento em ambiente corporativo, foram implementados desde o início do projeto os seguintes controles: controles de acesso baseados em perfis de usuário, garantindo que cada membro da equipe acesse apenas as informações pertinentes à sua função; registro de todas as operações realizadas na plataforma, assegurando rastreabilidade e auditabilidade; anonimização de dados sensíveis nos contextos em que dados estratégicos precisavam ser utilizados para calibração; e definição de fontes autorizadas como única base de conhecimento dos sistemas de IA, tanto para o sistema de revisão quanto para o chatbot. Esses cuidados visaram preservar a confidencialidade e a integridade das informações estratégicas da organização ao longo de todo o processo.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise dos resultados desta pesquisa deve ser contextualizada à luz de suas condições metodológicas específicas. Trata-se de uma pesquisa-ação conduzida em ambiente corporativo real, com múltiplas variáveis concomitantes que não puderam ser controladas: crescimento da maturidade técnica da equipe ao longo do tempo, ajustes na composição da equipe com entrada e saída de membros, demandas extraordinárias surgidas no decorrer do processo, melhorias em processos paralelos ao de revisão e variações na complexidade dos serviços processados em cada período. Esse conjunto de variáveis torna impraticável o isolamento de um resultado quantitativo único e inequívoco para o ganho de produtividade atribuível exclusivamente à IA.

Essa condição, longe de representar uma limitação do estudo, é característica inerente e esperada de pesquisas-ação conduzidas em organizações reais. Thiollent (2011) é explícito sobre essa natureza: a pesquisa-ação não busca o controle experimental das variáveis, mas a compreensão profunda dos fenômenos em seu contexto natural, com ênfase na utilidade prática do conhecimento gerado. O valor científico deste trabalho reside, portanto, na consistência da trajetória de evolução documentada, na riqueza das observações qualitativas e na capacidade de generalização metodológica dos princípios identificados.

Redefinição do papel do analista: de executor a revisor

Um dos resultados mais significativos observados ao longo do projeto foi a transformação qualitativa do trabalho dos analistas — uma mudança que pode ser descrita como a passagem do papel de executor para o papel de revisor. No estágio anterior ao uso da IA, a maior parte do tempo especializado era consumida em atividades de busca e compilação: localizar o serviço em bases externas com nomenclaturas não padronizadas, pesquisar normas técnicas aplicáveis, reunir composições de referência e consolidar anotações em formatos

heterogêneos. Esse conjunto de atividades, embora necessário, não constituía o núcleo do valor técnico que os especialistas poderiam agregar.

Com a implementação progressiva do sistema GPT-Vizca, esse conjunto de atividades passou a ser executado pela IA de forma sistemática, rápida e padronizada. O analista passou a receber, já estruturado e formatado, o referencial teórico necessário para iniciar sua análise: normas aplicáveis organizadas por relevância, composições similares de bases externas com indicação das diferenças em relação à composição vigente, especificação técnica e regulamentação em formato padrão pré-preenchido. Sua atividade passou a se concentrar no que genuinamente demanda conhecimento especializado: validar a pertinência das referências apresentadas, identificar discrepâncias técnicas e suas causas, ajustar composições com base no conhecimento das condições locais de execução e atestar a conformidade do produto final.

Esse ponto de estabilização — no qual a IA fornece a base e o analista realiza a conferência técnica e os ajustes necessários — representa o equilíbrio funcional desejado pelo projeto. A atividade do analista não é eliminada; ela é qualificada. O tempo antes gasto em tarefas repetitivas e de baixo valor agregado é liberado para o exercício do julgamento crítico, que é exatamente onde o conhecimento especializado faz diferença no resultado final.

Democratização do conhecimento técnico e resiliência do processo

Um efeito colateral relevante da implementação do sistema, não previsto explicitamente nos objetivos iniciais, foi a democratização do conhecimento técnico dentro da equipe. Analistas com formação generalista, sem especialização prévia em determinadas categorias de serviço, passaram a conseguir conduzir revisões com qualidade a partir do referencial estruturado fornecido pela IA. O sistema funcionou como nivelador de competências: ao apresentar de forma organizada os normativos aplicáveis, as composições de referência e o contexto técnico do serviço, reduziu a dependência de conhecimento tácito individual — aquele que estava "na cabeça" de especialistas específicos — e aumentou a resiliência do processo frente a variações na composição da equipe.

Limites do sistema e estratégias adaptativas

O desempenho do sistema mostrou-se diretamente proporcional à disponibilidade e à qualidade das fontes de referência externas. Para serviços com ampla representação em bases públicas e privadas, os resultados foram consistentemente bons. Para serviços com poucas referências externas, o desempenho foi mais limitado — condição que, importa notar, igualmente dificultaria o trabalho de um analista humano: a escassez de referências é uma característica do domínio do problema, não uma limitação específica da IA. Nesses casos, contudo, o sistema ainda agregou valor ao confirmar rapidamente a escassez de referências, poupando ao analista o tempo que seria gasto em pesquisas infrutíferas, ou redirecionando pesquisas específicas para outras bases não contempladas na IA.

O caso dos Métodos Não Destrutivos (MND) ilustra com clareza a capacidade de adaptação da arquitetura desenvolvida. Diante da ausência de referenciais em bancos de preços convencionais, a estratégia foi reorientada para a consulta de referenciais teóricos — normas técnicas, publicações científicas e documentações de fabricantes. Essa adaptação, desenvolvida de forma paralela às ondas principais, demonstra um princípio metodológico importante: o sucesso da aplicação de IA em domínios técnicos especializados depende da capacidade de identificar e adaptar a estratégia de consulta à natureza específica de cada categoria de problema, e não da aplicação uniforme de uma abordagem genérica.

Segurança da informação como premissa de arquitetura

A decisão de adotar a arquitetura RAG foi motivada de forma convergente por duas razões: a qualidade das respostas e a segurança da informação. Ao restringir o conhecimento do sistema exclusivamente a fontes autorizadas e predefinidas, o RAG oferece uma solução técnica para dois problemas distintos: o viés de alucinação inerente aos LLMs de uso geral — que, sem ancoramento documental, podem gerar informações

plausíveis, mas factualmente incorretas — e o risco de exposição de dados estratégicos da organização a sistemas externos não controlados.

A preocupação com a confidencialidade norteou toda a arquitetura da solução desde sua concepção — não como requisito adicional a ser atendido depois, mas como premissa inegociável de design. Essa abordagem tem uma implicação direta para a replicação do método: qualquer implementação análoga em outro contexto corporativo precisará enfrentar o mesmo desafio. A qualidade e a segurança do sistema são diretamente proporcionais à qualidade, à organização e ao controle da base documental que o alimenta. Estruturar e curar essa base — reunindo, organizando, atualizando e controlando o acesso às fontes autorizadas — constitui um investimento crítico e frequentemente subestimado em projetos desta natureza.

Engajamento da equipe e adoção da ferramenta

Um fator determinante para o sucesso do projeto foi a ausência de resistência por parte da equipe técnica. Ao contrário do que frequentemente se observa em processos de introdução de novas tecnologias em organizações — onde a resistência à mudança é um obstáculo comum e documentado —, os analistas manifestaram desde o início interesse ativo e demanda por ampliação da ferramenta para outras atividades que consumiam seu tempo. A solicitação de suporte para novas atividades partiu da própria equipe, não de uma imposição hierárquica.
Esse engajamento facilitou decisivamente o processo de refinamento: os próprios analistas contribuíram com informações detalhadas sobre seus fluxos de trabalho reais, sobre as etapas que mais demandavam tempo e sobre o formato ideal de apresentação das informações para cada etapa da análise. O desenvolvimento dos prompts tornou-se, assim, um processo genuinamente colaborativo — ancorado na prática real dos especialistas, e não em suposições sobre o que seria útil.

A adoção do chatbot com RAG pelos apropriadores em campo, de forma espontânea e para finalidades além das originalmente previstas, é um indicador qualitativo robusto de utilidade percebida. No contexto de trabalho de um apropriador em campo — que precisa de respostas rápidas e confiáveis sobre procedimentos e especificações técnicas, muitas vezes em condições adversas de conectividade e tempo —, uma ferramenta de consulta que responde de forma precisa e ancorada em documentação oficial representa um ganho operacional concreto e imediato.

CONCLUSÕES E RECOMENDAÇÕES

Contribuições metodológicas para a área de engenharia de custos

Além dos resultados operacionais obtidos no contexto específico da Sabesp, a experiência relatada neste trabalho contribui para a literatura de engenharia de custos em infraestrutura sob pelo menos três perspectivas metodológicas.

A primeira contribuição é a demonstração da viabilidade técnica do uso de LLMs para interpretação de linguagem técnica especializada no domínio do saneamento. A hipótese de que modelos de linguagem de propósito geral — sem especialização prévia no domínio — seriam capazes de compreender nomenclaturas, identificar equivalências semânticas entre serviços e estruturar informações técnicas de forma consistente foi confirmada empiricamente ao longo do estudo. Esse resultado não era evidente a priori: o domínio do saneamento possui uma linguagem técnica específica, com terminologias que variam entre bases de dados, empresas e regiões, e que não necessariamente encontra representação proporcional nos corpora de treinamento dos modelos.

A segunda contribuição é metodológica: a estratégia de calibração de *prompts* por comparação com referência conhecida — em que itens validados por especialistas servem como padrão-ouro para o refinamento iterativo do sistema — oferece um procedimento replicável para a adequação de LLMs a domínios técnicos corporativos. Esse procedimento não depende de retreinamento do modelo nem de recursos computacionais especializados: requer apenas a disponibilidade de uma amostra de referência de qualidade e o engajamento

dos especialistas do domínio no processo de avaliação. Sua simplicidade operacional é, paradoxalmente, um de seus maiores ativos em contextos corporativos, onde recursos e tempo são limitados.

A terceira contribuição é a demonstração de que a arquitetura RAG, originalmente desenvolvida para tarefas de recuperação de informação em processamento de linguagem natural, mostra-se eficaz também em aplicações de suporte ao trabalho especializado em engenharia. A ancoragem das respostas em bases documentais controladas — normas técnicas, composições de referência, procedimentos internos — não apenas melhora a qualidade técnica das informações fornecidas, mas também cria um mecanismo natural de rastreabilidade: para cada afirmação do sistema, é possível identificar a fonte documental que a fundamenta, o que é essencial em contextos de engenharia de custos, onde a justificativa técnica das composições precisa ser auditável.

Perspectivas de replicação e generalização do método

O uso de Inteligência Artificial Generativa na revisão do banco de preços de saneamento da Sabesp demonstrou ser uma solução viável, escalável e de alto impacto. Em aproximadamente dez meses de desenvolvimento e implantação progressiva, o sistema GPT-Vizca processou 3.407 serviços, com adoção plena pela equipe e percepção coletivamente avaliada como excelente. O projeto superou suas metas originais e gerou resultados — como a democratização do conhecimento técnico e a adoção espontânea do chatbot para usos não previstos — que evidenciam o potencial transformador da abordagem adotada.

As principais conclusões do estudo podem ser sintetizadas nos seguintes pontos:

- **A IA atuou como copiloto técnico, não como substituto do especialista.** O valor central da solução reside na redistribuição do esforço cognitivo: o analista passou de executor de tarefas de busca e compilação para revisor e validador do referencial gerado pelo sistema, concentrando seu tempo no julgamento técnico de alto valor.
- **A estratégia de evolução por ondas mostrou-se metodologicamente adequada.** A progressão incremental permitiu aprender com baixo custo, validar hipóteses antes de avançar e construir confiança institucional de forma gradual — características essenciais em projetos de implantação de IA em ambientes corporativos complexos.
- **O refinamento de prompts por comparação com referência conhecida é um método robusto.** Calibrar o sistema a partir de resultados já validados por especialistas, em vez de critérios abstratos, conferiu objetividade e rastreabilidade ao processo de melhoria contínua e é aplicável a outros contextos que exijam adequação a linguagem técnica especializada.
- **A arquitetura RAG é recomendada para aplicações corporativas que exijam qualidade e segurança.** O ancoramento das respostas em bases documentais controladas mitiga o risco de alucinações, garante rastreabilidade das informações e permite o controle de acesso — três requisitos críticos em contextos de engenharia de custos com dados estratégicos.
- **O engajamento da equipe técnica é fator crítico de sucesso.** O desenvolvimento colaborativo dos prompts, guiado pelos fluxos de trabalho reais dos analistas, foi determinante para a qualidade do sistema. A ausência de resistência à adoção da tecnologia, substituída por demanda ativa de expansão, acelerou o processo e ampliou os resultados alcançados.
- **O método é transferível para outros domínios.** A condição fundamental não é tecnológica, mas metodológica: a compreensão profunda do processo manual, da linha de raciocínio dos especialistas e do resultado esperado é o insumo indispensável para qualquer replicação bem-sucedida.

É importante registrar que o sucesso alcançado não decorreu apenas de escolhas tecnológicas acertadas, mas também de um conjunto de condições organizacionais favoráveis: a disposição da gestão em apoiar e acompanhar o desenvolvimento desde as etapas iniciais; a abertura da equipe técnica para colaborar ativamente no processo de refinamento; e a existência de uma base documental estruturada que pudesse ser mobilizada como fonte de conhecimento para os sistemas de IA. Essas condições não devem ser tomadas como garantidas em qualquer contexto de replicação — sua presença ou ausência será determinante para o sucesso de iniciativas semelhantes.

O projeto também demonstrou que a introdução de IA em processos de trabalho especializados não é um evento, mas um processo. A maturidade tecnológica e operacional foi construída ao longo de dez meses, em ciclos progressivos de aprendizado, ajuste e consolidação. Essa perspectiva temporal é frequentemente subestimada em projetos de transformação digital, que tendem a superestimar resultados de curto prazo e subestimar os de médio e longo prazo. A estratégia de ondas evolutivas adotada neste trabalho oferece um modelo prático para gerenciar essa progressão de forma estruturada e com controle de riscos — aplicável a outros contextos que envolvam a introdução de IA em domínios técnicos complexos.

Por fim, cabe uma reflexão sobre o papel da IA na engenharia de infraestrutura. A experiência relatada neste artigo sugere que o maior valor da IA nesse contexto não está na automação de decisões técnicas — que continuam a exigir o julgamento de especialistas —, mas na eliminação das barreiras de acesso à informação que limitam a qualidade e a velocidade com que esse julgamento pode ser exercido. Uma ferramenta de IA que coloca o conhecimento técnico relevante na mão certa, no momento certo e na forma adequada não substitui o engenheiro: amplifica sua capacidade de fazer o que somente ele pode fazer.

Recomendações para Implementações Futuras

Com base na experiência acumulada ao longo do projeto, as seguintes recomendações são dirigidas a organizações que considerem implementar abordagens semelhantes:

- **Mapear profundamente o processo manual antes de qualquer decisão tecnológica.** O mapeamento detalhado das etapas, dos artefatos produzidos, dos critérios de qualidade e dos pontos de dificuldade do processo existente é o insumo essencial para o design do sistema de IA. Atalhos nessa etapa podem comprometer a qualidade do resultado de forma difícil de reverter nas fases seguintes.
- **Envolver os especialistas no desenvolvimento dos prompts desde o início.** A participação ativa da equipe técnica não é apenas desejável — é metodologicamente necessária. Os especialistas são os únicos que conhecem o resultado correto; sem essa referência concreta, o processo de calibração não tem âncora objetiva.
- **Adotar uma arquitetura de fontes autorizadas (RAG) em vez de modelos de propósito geral.** Em domínios técnicos corporativos, a qualidade e a segurança das respostas são dimensões tão importantes quanto a velocidade de processamento. O RAG oferece o melhor equilíbrio entre essas dimensões para o perfil de aplicação aqui descrito.
- **Planejar para o médio prazo.** A maturidade de um sistema de IA aplicado a domínios técnicos especializados não é alcançada em poucas semanas. O planejamento deve contemplar ciclos sucessivos de refinamento, com marcos claros de avaliação de qualidade e critérios explícitos de avanço para a fase seguinte.
- **Tratar a base documental como um ativo estratégico a ser construído e mantido.** A qualidade do sistema de IA é diretamente proporcional à qualidade, à organização e à atualização das fontes que o alimentam. O investimento na curadoria documental é tão estratégico quanto o investimento tecnológico — e frequentemente mais negligenciado.

Para trabalhos futuros, recomenda-se: a avaliação de indicadores quantitativos de produtividade em fases subsequentes, quando as variáveis concomitantes do processo de implantação estiverem estabilizadas e for possível realizar comparações com maior controle metodológico; a expansão do método para outros grupos do banco de preços da Sabesp e para outros tipos de infraestrutura pública e privada; a investigação da aplicação de técnicas de avaliação automatizada da qualidade de outputs de LLMs em contextos de engenharia de custos; e o desenvolvimento de métricas padronizadas para comparação de sistemas de apoio à revisão de preços — lacuna atualmente existente na literatura da área e que limitou a comparação dos resultados deste trabalho com experiências análogas.

REFERÊNCIAS BIBLIOGRÁFICAS

1. **AGÊNCIA BRASIL.** Governo de SP conclui privatização da Sabesp ao vender 32% dos papéis. *Agência Brasil*, Brasília, 23 jul. 2024. Disponível em: https://agenciabrasil.ebc.com.br/geral/noticia/2024-07/governo-de-sp-conclui-privatizacao-da-sabesp-ao-vender-32-dos-papeis. Acesso em: jun. 2026.

2. **BRASIL.** Lei nº 14.026, de 15 de julho de 2020. Atualiza o marco legal do saneamento básico e altera a Lei nº 11.445, de 5 de janeiro de 2007. *Diário Oficial da União*, Brasília, DF, 16 jul. 2020. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l14026.htm. Acesso em: jun. 2026.

3. **BROWN, T. B. et al.** *Language Models are Few-Shot Learners.* arXiv:2005.14165, 2020. Disponível em: https://arxiv.org/abs/2005.14165. Acesso em: maio 2026.

4. **GIL, A. C.** *Métodos e Técnicas de Pesquisa Social.* 6. ed. São Paulo: Atlas, 2008.

5. **INFOMONEY.** Sabesp apresenta nova fase pós-privatização no Investor Day 2026; veja os destaques. *InfoMoney*, São Paulo, 10 abr. 2026. Disponível em: https://www.infomoney.com.br/mercados/sabesp-apresenta-nova-fase-pos-privatizacao-no-investor-day-2026-veja-os-destaques/. Acesso em: jun. 2026.

6. **INSTITUTO TRATA BRASIL.** ***Distante do necessário: investimento em saneamento precisa mais do que dobrar para universalização.*** São Paulo: Instituto Trata Brasil, 2025. Disponível em: https://tratabrasil.org.br/distante-investimento-saneamento-mais-dobrar/. Acesso em: jun. 2026.

7. **LEWIS, P. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.** ***Advances in Neural Information Processing Systems*** **(NeurIPS), v. 33, p. 9459-9474, 2020. Disponível em: https://arxiv.org/abs/2005.11401. Acesso em: jun. 2026.**

8. **MARCONI, M. A.; LAKATOS, E. M.** *Fundamentos de Metodologia Científica.* 9. ed. São Paulo: Atlas, 2021.

9. **SÃO PAULO.** URAE-1 — Unidade Regional de Serviços de Abastecimento de Água Potável e Esgotamento Sanitário Sudeste. *Contrato de Concessão nº 01/2024* [...]. São Paulo, 24 maio 2024. Disponível em: https://semil.sp.gov.br/urae1/[...]. Acesso em: jun. 2026.

10. SABESP — Companhia de Saneamento Básico do Estado de São Paulo. *Sobre a Sabesp.* São Paulo: Sabesp, 2025. Disponível em: https://www.sabesp.com.br/. Acesso em: jun. 2026.

11. **SABESP — Companhia de Saneamento Básico do Estado de São Paulo.** *Relatório de Resultados 2025.* São Paulo: Sabesp, 2026. Disponível em: https://ri.sabesp.com.br/informacoes-financeiras/central-de-resultados/. Acesso em: jun. 2026.

12. **THIOLLENT, M.** *Metodologia da Pesquisa-Ação.* 19. ed. São Paulo: Cortez, 2025.

13. **VASWANI, A. et al. Attention is All You Need.** ***Advances in Neural Information Processing Systems*** (NeurIPS), v. 30, 2017. Disponível em: https://arxiv.org/abs/1706.03762. Acesso em: jun. 2026.