



03 a 05 de dezembro de 2025

As Engenharias na Transformação Digital das PMEs

Aplicação do Método RAG para a Otimização do Processo de Avaliação dos Requisitos de Cliente na Cadeia Automotiva

Alan Daniel Bernardino

PPGEP - UFPR

Silvana Pereira Detro

PPGEP - UFPR

Nicolle Christine Sotsek

PPGEP - UFPR

Resumo: A crescente complexidade da cadeia automotiva exige soluções inovadoras para garantir o atendimento aos requisitos específicos dos clientes (CSR). Este estudo investiga a aplicação do método *Retrieval-Augmented Generation* (RAG) na cadeia automotiva para otimizar o processo de avaliação dos requisitos específicos dos clientes, parte essencial para atendimento a norma IATF 16949:2016. A motivação central é a dependência de análises manuais realizada pelos especialistas. O objetivo do artigo é avaliar como a arquitetura RAG pode apoiar a interpretação normativa, reduzir tempo de análise e aumentar a padronização das respostas. O artigo apresenta um framework que integra recuperação semântica dos documentos dos processos e geração de respostas em linguagem natural e a aplicação em um caso simulado. Os resultados indicam que o método RAG apresenta elevado potencial de apoio à tomada de decisão, melhora a abrangência e padronização das informações e contribui para a digitalização do Sistema de Gestão da Qualidade. O estudo reforça a relevância da IA generativa na Indústria 4.0 e abre caminhos para futuras aplicações em ambientes industriais regulados.

Palavras-chave: RAG; Inteligência Artificial Generativa; Requisitos Específicos de Cliente; IATF 16949; Indústria Automotiva.

Application of the RAG Method for Optimizing the Customer Requirements Evaluation Process in the Automotive Supply Chain

Abstract: The increasing complexity of the automotive supply chain demands innovative solutions to ensure compliance with customer-specific requirements (CSR). This study investigates the application of the *Retrieval-Augmented Generation* (RAG) method in the automotive supply chain to optimize the evaluation process of customer-specific requirements, a key component for compliance with the IATF 16949:2016 standard. The central motivation is the reliance on manual analysis performed by specialists. The aim of this paper is to assess how RAG architecture can support normative interpretation, reduce analysis time, and increase response standardization. The paper presents a framework that integrates semantic retrieval of process documents with natural language response generation and the application on a simulated case. The results indicate that the RAG method shows high potential to support decision-making, improve the comprehensiveness and standardization of information, and contribute to the digitalization of the Quality Management System. The study reinforces the relevance of generative AI in Industry 4.0 and opens pathways for future applications in regulated industrial environments.

Keywords: RAG; Generative Artificial Intelligence; Customer-Specific Requirements; IATF 16949; Automotive Industry.

1. Introdução

A crescente complexidade da cadeia de fornecimento global e a necessidade de diferenciação competitiva têm impulsionado as organizações a buscar soluções inovadoras que assegurem qualidade, agilidade e conformidade regulatória. A indústria automotiva, em particular, apresenta um dos cenários mais rigorosos nesse aspecto, devido à combinação de requisitos normativos internacionais e requisitos específicos de clientes (*Customer Specific Requirements* – CSR), estes requisitos são validados e verificados através de auditorias periódicas que demandam elevado controle e rastreabilidade.

Neste contexto, a avaliação de atendimento a requisitos específicos de clientes constitui um processo crítico para os fornecedores automotivos. Tradicionalmente, essa avaliação é realizada de forma manual, com forte dependência da *expertise* de engenheiros e analistas de diferentes áreas da organização. Contudo, esse método apresenta limitações significativas, pois depende do capital intelectual de cada profissional para fazer esta análise, como há uma capacidade limitada para processamento, verificação e processamento dos documentos internos da organização que justificam o atendimento aos requisitos existe uma morosidade no processamento das informações, suscetibilidade a erros interpretativos e a falta de padronização nas respostas, que podem resultar em não conformidades nas auditorias, gerando atividades adicionais para toda a organização, necessidade de novas auditorias e em casos mais críticos podem levar a reprovação da auditoria, perda do certificado e perda de contratos.

O problema central abordado neste artigo é, portanto, a dificuldade de garantir consistência, agilidade e confiabilidade na análise dos requisitos específicos de cliente por meio de processos manuais. Tal cenário justifica a necessidade de investigar alternativas que possam otimizar esse processo, reduzindo o tempo de análise, minimizando erros e assegurando maior padronização nas respostas.

Diante desse desafio, o objetivo deste artigo é analisar a aplicação do método *Retrieval-Augmented Generation* (RAG), aliado a modelos de linguagem de grande escala (LLMs), como solução para otimizar o processo de avaliação de requisitos específicos de cliente na cadeia automotiva. Busca-se verificar se essa arquitetura pode apoiar a tomada de decisão, aumentar a padronização das análises e reduzir a dependência exclusiva do conhecimento humano.

A metodologia adotada é caracterizada como pesquisa aplicada, de abordagem qualitativa, com condução de um estudo de caso em uma empresa do setor automotivo. O framework foi concebido para integrar documentos dos processos da organização para formar a base vetorial que guiará a recuperação semântica na arquitetura RAG com prompts de comando de acordo com os requisitos de cliente e, a partir disso, gerar respostas em linguagem natural referente ao atendimento à tais requisitos, justificando como a organização atende ao requisito ou informando a não aderência ao requisito, apontando quais são os pontos de não atendimento.

Este artigo está dividido em cinco seções. Após esta introdução, a segunda seção apresenta a metodologia adotada, detalhando as etapas metodológicas e o framework proposto. A terceira seção aborda os conceitos de gestão da qualidade, requisitos específicos de cliente, inteligência artificial e trabalhos correlatos. A quarta seção apresenta a classificação da pesquisa e os resultados encontrados. Por fim, a quinta seção reúne as conclusões, destacando as contribuições, limitações e propostas para estudos futuros.

2. Metodologia

O desenvolvimento deste estudo foi estruturado em diferentes etapas metodológicas. Inicialmente, realizou-se um levantamento de artigos correlatos com o objetivo de identificar o estado da arte sobre a aplicação da arquitetura *Retrieval-Augmented Generation* (RAG) em domínios regulados, especialmente em contextos que demandam rastreabilidade e auditabilidade. Essa revisão permitiu mapear elementos técnicos comumente empregados em pesquisas anteriores, tais como mecanismos de recuperação semântica, granularidade de indexação, estratégias de *prompting* e critérios utilizados para avaliação de respostas. Essas informações forneceram subsídios para a elaboração do framework proposto, adequado às particularidades do setor automotivo.

Na etapa seguinte, procedeu-se ao mapeamento documental da organização estudada, com o intuito de estruturar uma base de informações que serviria de suporte para a construção da base vetorial. Esse mapeamento incluiu documentos que descrevem os processos internos da empresa e instruções operacionais associadas. Os documentos de processos, organizados a partir do modelo de diagrama de tartaruga, continham descrições detalhadas de entradas, saídas, riscos, oportunidades, etapas de execução, recursos envolvidos e indicadores de desempenho.

Para viabilizar a aplicação prática do framework em ambiente controlado, foi criada uma cadeia de suprimentos fictícia, representando de forma simulada a estrutura de uma rede automotiva. Nessa cadeia foram desenvolvidas normas técnicas de fabricação de carros de brinquedo, inspiradas em requisitos normativos automotivos; foram descritos processos de fabricação modelados a partir da estrutura da empresa estudada; e foram definidos requisitos específicos do montador de brinquedos para sua rede de fornecimento, em analogia aos *Customer Specific Requirements* (CSR) das montadoras automotivas. Esse cenário simulado permitiu reproduzir a complexidade documental e normativa de um ambiente real, assegurando condições para a validação da proposta sem expor dados sensíveis da organização analisada.

Na sequência, foi definido um manual contendo os requisitos específicos do montador fictício, selecionado como base para a validação do framework. Esses requisitos foram convertidos em perguntas auditáveis, elaboradas de forma a facilitar a verificação prática de conformidade. Cada questão foi estruturada com critérios de aplicabilidade, fontes de evidência esperadas e notas adicionais sobre exceções ou condições específicas. Essa conversão teve o objetivo de tornar os requisitos mais operacionais e compatíveis com a linguagem utilizada em auditorias e rotinas organizacionais.

Com os documentos internos mapeados e os requisitos transformados em perguntas, o *framework* foi estruturado integrando essas duas bases a um sistema de recuperação semântica e geração de respostas. O módulo de recuperação (*retriever*) foi configurado para localizar trechos relevantes nos documentos indexados, enquanto o módulo de geração (*generator*) teve como função elaborar respostas fundamentadas nos trechos recuperados, incluindo referências documentais que assegurassem rastreabilidade. A arquitetura foi configurada de modo a responder exclusivamente com base em evidências disponíveis, reduzindo a possibilidade de extrapolações não fundamentadas. A arquitetura foi estruturada através de agentes de IA dentro da plataforma do Copilot Studio em conjunto com o LLM GPT4-o.

A etapa final consistiu na validação das respostas produzidas pelo *framework*. Para isso, foram adotados critérios previamente estabelecidos e inspirados no *grading method* de Wang et al. (2024). A análise contemplou três dimensões principais: corretude, relacionada à aderência factual das respostas aos requisitos normativos; completude, referente ao grau em que as respostas abrangiam todos os aspectos relevantes da questão; e honestidade, ligada à transparência em relação a eventuais lacunas ou limitações. Essa avaliação foi

realizada por especialistas da área de qualidade da empresa, que atribuíram notas em escala ordinal a cada resposta, sendo 1 para o não atendimento ao critério, 2 para atendimento parcial e 3 para o atendimento do critério.

3. Referencial Teórico

O referencial teórico deste artigo concentra-se em dois eixos principais: (i) os sistemas de gestão da qualidade e os requisitos específicos de cliente (CSR), que fundamentam a prática organizacional na indústria automotiva; e (ii) as tecnologias de Inteligência Artificial Generativa, com ênfase na arquitetura *Retrieval-Augmented Generation* (RAG), que é apresentada como solução inovadora para otimização do processo de avaliação dos requisitos específicos dos clientes.

3.1 Os sistemas de gestão da qualidade e os requisitos específicos de cliente

A compreensão de “qualidade” evoluiu acompanhando as necessidades do mercado e a sofisticação dos processos industriais. Autores clássicos consolidaram definições que, em conjunto, formam um núcleo conceitual ainda vigente. Para Juran (1951), qualidade é “adequação ao uso”, enfatizando o atendimento ao propósito do cliente; Crosby (1979) entende como “conformidade aos requisitos”, direcionando o foco a especificações e zero defeitos; Feigenbaum (1983) amplia o escopo ao afirmar que a qualidade é o resultado total das características de um produto/serviço envolvendo marketing, engenharia, produção e manutenção — ou seja, é sistêmica e organizacional, não apenas uma função de inspeção; Ishikawa (1985) destaca o atendimento às necessidades e expectativas do cliente, reforçando a centralidade da voz do cliente; Deming (1986) sintetiza qualidade como previsibilidade consistente do desempenho ao menor custo total possível, conectando-a diretamente à gestão e ao pensamento estatístico.

Essas formulações desembocam, algumas décadas depois, na filosofia da Gestão da Qualidade Total (TQM). A TQM surge como resposta à pressão competitiva global — catalisada pela crise do petróleo e pelo avanço japonês em manufatura — e combina pilares como foco no cliente, melhoria contínua, envolvimento de todos e gestão por processos (Deming, 1986). Um ponto crucial dessa transição é que, conforme o mercado evolui, as organizações deixam de tratar qualidade como “custo” e passam a vê-la como alavanca de valor e meio de reduzir despesas no longo prazo. Em outras palavras, qualidade deixa de ser essencialmente “controle” para se tornar estratégia (Crosby, 1984).

Apesar de seus méritos, a TQM — por ser filosofia — carecia de padronização suficiente que viabilizasse auditorias comparáveis entre empresas e setores. Com a intensificação da globalização, surgiu a necessidade de critérios uniformes que tornassem a qualidade auditável e certificável. É nesse contexto que, em 1987, a ISO 9001 formaliza princípios da TQM em um modelo estruturado, sistematizando requisitos, documentação e práticas de gestão. A partir daí, o conceito de Sistema de Gestão da Qualidade (SGQ) passa a integrar fundamentos da TQM a uma arquitetura de requisitos explícitos, gerando padronização, rastreabilidade, comparabilidade e certificação (ISO, 2015).

Essa institucionalização é vista por autores como um dos grandes marcos da qualidade no último século: ao transformar a filosofia em sistema normativo, a ISO 9001 permitiu ampla difusão de boas práticas e um vocabulário comum para clientes, fornecedores e auditores (Goetsch, 2014).

Na cadeia automotiva, a criticidade de segurança, confiabilidade e rastreabilidade impõe um patamar superior de controle. Processos, produtos e serviços precisam atender a um corpo normativo extenso, composto não apenas pela ISO 9001, mas por requisitos adicionais definidos por montadoras (OEMs) e por auditorias de segunda/terceira parte que

verificam aderência no dia a dia. Desse cenário emerge a IATF 16949:2016, norma setorial que integra a ISO 9001 e acrescenta requisitos específicos voltados à prevenção de falhas, redução de variação, robustez de processos, controle de fornecedores e conformidade documental (TRICKER, 2019).

A IATF 16949 organiza-se em 10 capítulos, dos fundamentos (escopo, referências, termos) à operação, avaliação de desempenho e melhoria. Em comparação com a ISO 9001:2015 (77 requisitos), a IATF 16949:2016 expande a exigência para 106 requisitos, além de centenas de *Customer Specific Requirements* (CSR) emitidos por cada montadora. O incremento quantitativo e qualitativo é expressão do aumento de complexidade regulatória do setor e, ao mesmo tempo, da necessidade de governança robusta para garantir conformidade, fluidez operacional e credibilidade em auditorias (Norma IATF 16949:2016, 2016).

O capítulo 8 (Operação) é o coração operacional do SGQ, cobrindo planejamento/control operacional, requisitos de produtos/serviços, projeto e desenvolvimento, controle de terceirizados, produção/serviço, liberação e controle de saídas não conformes. Na ISO 9001, esses pontos estruturam a gestão do fluxo de valor; na IATF 16949, eles são complementados por requisitos que aprofundam o controle (gestão de mudanças, validação de processo com foco na aprovação do cliente, controle de *software* embarcado quando aplicável, planos de contingência, verificação de *layout* e gestão de retrabalhos/concessões, entre outros).

Dentro do capítulo 8 destaca-se o item 8.2.3.1.1 – Requisitos Específicos dos Clientes (CSR). A norma determina que cada organização deve identificar, compreender, documentar e aplicar os CSR de cada montadora. Esses requisitos complementam as normas gerais e devem ser incorporados de forma sistêmica aos processos internos, com evidências acessíveis às partes interessadas, além de monitoramento contínuo (por auditorias internas, indicadores e *feedback* de clientes). Na prática, os CSR são o ponto de contato normativo mais direto entre o processo do fornecedor e as expectativas do cliente automotivo — e, por isso, um dos elementos mais sensíveis em auditorias (Norma IATF 16949:2016, 2016).

Levantamentos recentes indicam que, apenas para a IATF 16949:2016, há centenas de CSR vigentes entre grandes grupos (Stellantis, Volkswagen, Mercedes-Benz, GM, Ford, BMW, Renault, Volvo, Iveco, Geely, PSA, etc.). O volume e a heterogeneidade desses documentos explicam por que a avaliação de conformidade tende a ser manual, demorada e suscetível a erros interpretativos, especialmente diante de revisões frequentes e linguagem técnica específica (Norma IATF 16949:2016, 2016).

3.2 Inteligência Artificial

A Inteligência Artificial (IA) vem sendo aplicada à qualidade desde os anos 1990 (redes neurais, árvores de decisão, regressões) para controle estatístico de processos, manutenção preditiva, inspeção automática e detecção de padrões (Zhou et al., 1996; Kim et al., 2003). Estudos reportam ganhos relevantes de eficiência e precisão quando a IA é bem integrada a dados e rotinas operacionais. Contudo, persistia uma lacuna: documentos textuais normativos e requisitos de clientes são difíceis para modelos clássicos, porque exigem compreensão semântica, contexto e rastreabilidade.

Essa barreira começou a ceder com a arquitetura Transformers (Vaswani et al., 2017), que substituiu a dependência de processamento sequencial por mecanismos de auto-atenção. O resultado foi uma revolução no Processamento de Linguagem Natural (PLN): modelos como BERT e GPT passaram a capturar contexto de forma muito mais eficaz, permitindo sumarização, classificação, resposta a perguntas e geração de texto com coerência antes

inalcançável. A família de LLMs (*Large Language Models*) elevou o patamar do que é possível em linguagem (Radford et al., 2019; Brown et al., 2020).

Ainda assim, LLMs padrões enfrentam limitações graves para o mundo regulado: (i) conhecimento estático (corte temporal do treinamento), (ii) risco de alucinações (respostas plausíveis, porém incorretas) e (iii) falta de referências explícitas às fontes utilizadas. Em um ambiente como o automotivo, onde evidência documental e auditoria são essenciais, isso é um grande impeditivo.

Para superar essas limitações, surgiu a arquitetura *Retrieval-Augmented Generation* (RAG), em português, geração aumentada via recuperação, proposta por Lewis et al. (2020). O RAG combina dois blocos: um *retriever*, que busca trechos/documentos relevantes em uma base indexada semântica por *embeddings*, e um gerador, que elabora a resposta ancorada nos trechos recuperados. Em vez de “inventar” com base apenas no que o modelo aprendeu no pré-treino, o RAG consulta um repositório atualizado e específico e cita as evidências.

Na prática organizacional, a implantação de RAG exige: (i) Documentação e fonte de dados organizado, (ii) pré-processamento e indexação, (iii) um mecanismo de recuperação eficiente, (iv) um LLM para redigir e gerar respostas. O desempenho do RAG é tão bom quanto a qualidade da documentação utilizada para a preparação da base vetorial e a calibração do *retriever* (Lewis et al., 2020).

A vantagem central do RAG é tornar a resposta auditável: o usuário pode ver de onde veio a informação. Em setores regulados, isso viabiliza governança, compliance e confiabilidade, características alinhadas aos princípios da IATF 16949 e à realidade dos CSR.

4. Resultados

4.1 Classificação do artigo

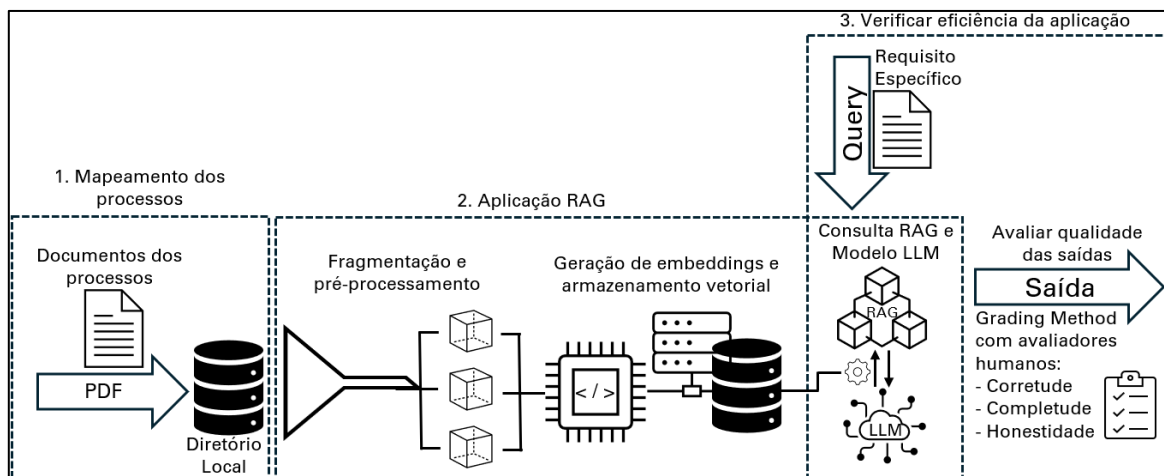
A pesquisa desenvolvida caracteriza-se como aplicada, de abordagem qualitativa e conduzida por meio de um estudo de caso em uma empresa do setor automotivo. O objetivo central consistiu em propor e avaliar um framework baseado na arquitetura *Retrieval-Augmented Generation* (RAG), com apoio de modelos de linguagem de grande escala (LLMs), para otimizar o processo de avaliação de requisitos específicos de clientes (*Customer Specific Requirements – CSR*). A escolha desse delineamento deve-se à natureza prática do problema enfrentado, à necessidade de compreender a aplicabilidade de uma solução inovadora em um contexto real e à possibilidade de interpretar os resultados de maneira descritiva e analítica.

4.2 Framework proposto

O framework proposto neste estudo foi desenvolvido com o objetivo de estruturar a aplicação da arquitetura *Retrieval-Augmented Generation* (RAG) no contexto da avaliação de requisitos específicos de clientes (Figura 1).

Sua concepção buscou alinhar aspectos técnicos da Inteligência Artificial Generativa com as necessidades práticas de rastreabilidade e padronização exigidas pelos Sistemas de Gestão da Qualidade. Para isso, o modelo foi organizado em três macro etapas interdependentes: mapeamento dos processos, aplicação do RAG e verificação da eficiência da aplicação.

Figura 1 - Framework proposto para avaliação CSR

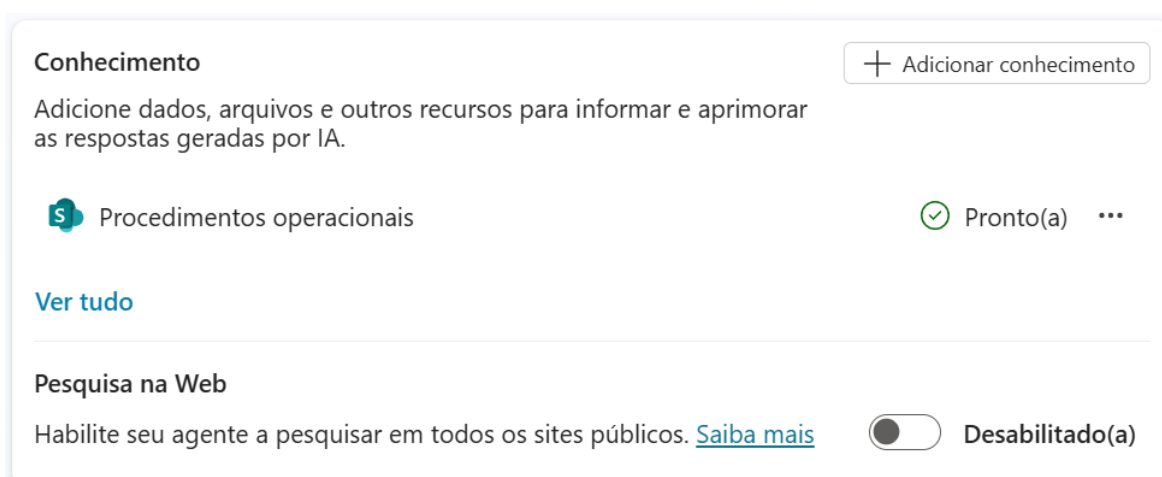


Fonte: O Autor (2025)

A primeira etapa consistiu no mapeamento e organização dos documentos que compuseram a base de conhecimento utilizada no experimento. Foram incluídos documentos fictícios criados para representar uma cadeia de suprimentos simulada, composta por normas técnicas de fabricação de carros de brinquedo, instruções operacionais inspiradas em processos reais da empresa estudada e requisitos específicos definidos por um “montador de brinquedos” para sua rede de fornecedores. Esses documentos foram reunidos em formato digital (PDF) e armazenados em um diretório na nuvem (sharepoint), em uma pasta chamada “Procedimentos operacionais”, garantindo controle de versão e rastreabilidade. O mapeamento seguiu a lógica do diagrama de tartaruga, contemplando informações como entradas, saídas, recursos, etapas de execução, riscos, oportunidades e indicadores, assegurando assim a construção de um repositório representativo para a posterior aplicação da arquitetura RAG.

Na segunda etapa foi utilizado a construção de um agente através do Copilot Studio para a construção da solução utilizando a arquitetura RAG, os documentos foram referenciados pelo agente (Figura 2), desabilitando a busca na web para reduzir o risco de alucinações ao buscar respostas em fontes não rastreáveis.

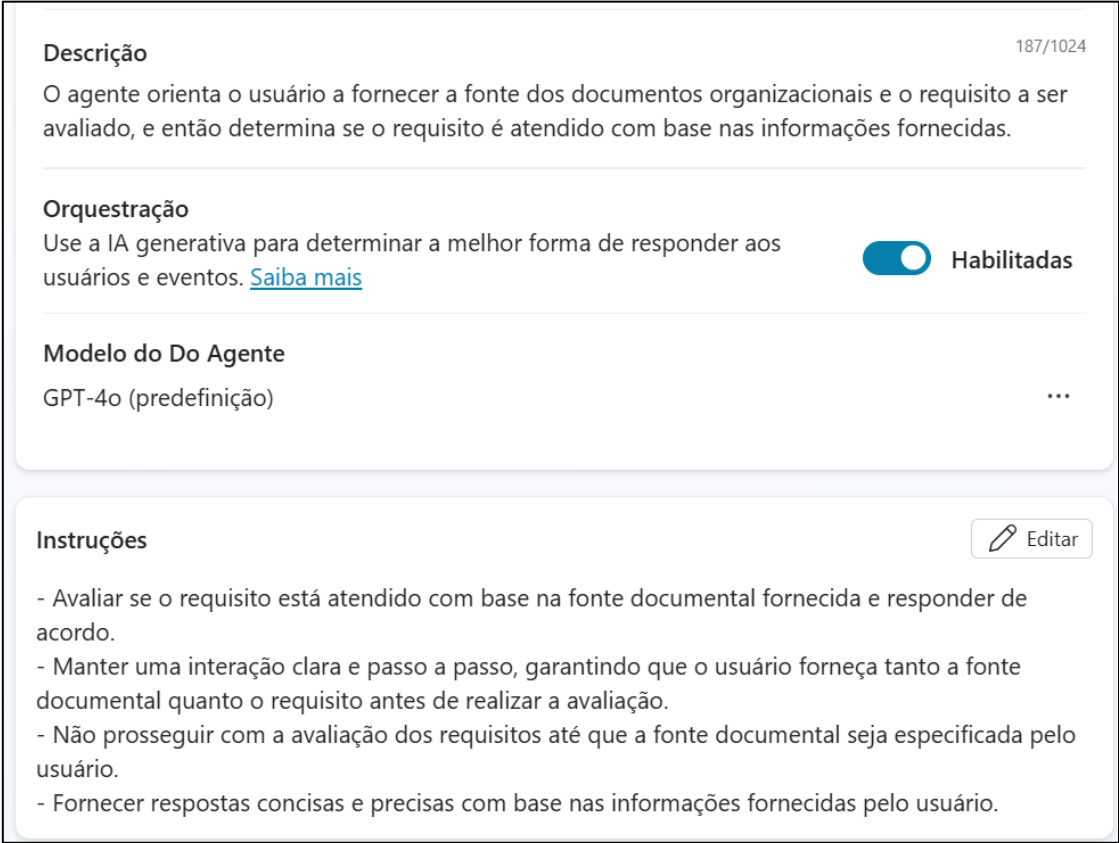
Figura 2: Definição da base de conhecimento para o agente



Fonte: O Autor (2025)

A construção das respostas foi realizada através do módulo de geração (*generator*) considerando o GPT 4-o. deste modo a arquitetura foi configurada para que o sistema respondesse apenas com base em evidências documentais disponíveis, incluindo referências aos documentos que embasaram a resposta, assegurando assim rastreabilidade e transparência, a Figura 3 apresenta as configurações de descrição, orquestração, modelo do agente e instruções adotadas.

Figura 3: Configurações do agente no Copilot Studio



Fonte: O Autor (2025)

Na Tabela 1 estão apresentados os requisitos que foram simulados para validar o framework para a avaliação do atendimento ao CSR e as respostas esperadas.

Tabela 1: Requisitos para validação do framework

n	Requisito	Resposta Esperada
1	A empresa segue a certificação NM 300-1?	Atende
2	O teste de queda considera a altura de 1,0 metro?	Atende
3	As amostras iniciais contêm os relatórios dimensionais de desempenho e laudo de toxidade?	Atende
4	As alterações são avaliadas por uma equipe multifuncional?	Atende
5	É aplicado cera de abelha para fazer o arredondamento dos cantos?	Não Atende

Fonte: O Autor (2025)

As saídas geradas pelo sistema seguem o padrão da Figura 4. Esse processo de validação permitiu verificar a qualidade das respostas, mantendo a rastreabilidade das fontes consultadas e aproximando o sistema das práticas utilizadas em auditorias de clientes.

Figura 4: Padrão de respostas geradas



Fonte: O Autor (2025)

Todas as cinco respostas do atendimento aos requisitos estão apresentadas na Tabela 2 e foram avaliadas por meio do *grading method*, segundo três critérios principais: corretude, completude e honestidade com notas de 1 a 3, sendo 1 para não atende ao critério, 2 atende ao critério parcialmente e 3 atende ao critério.

Tabela 2: Requisitos para validação do framework

n	Resposta gerada	Corretude	Completude	Honestidade
1	Sim, a empresa segue a certificação ABNT NBR NM 300-1 – Segurança de brinquedos. Esta norma é mencionada em diversos Procedimentos Operacionais Padrão (POPs) da Toy Parts Corporation	3	3	3
2	Sim, o teste de queda considera a altura de 1,0 metro. O procedimento de teste de queda para rodas de carrinhos de brinquedo inclui a queda livre das rodas de uma altura de 1,0 metro, seguido de uma inspeção visual e teste funcionais para garantir a integridade das rodas após a queda	3	3	3
3	Sim, as amostras iniciais devem incluir: Relatórios dimensionais de desempenho: Realizar medições	3	3	3

	conforme ficha técnica, incluindo dimensões, peso e densidade; Laudo de toxidade: Solicitar e registrar laudos de segurança, incluindo toxidade, conforme necessário			
4	Sim, as alterações são avaliadas por uma equipe multifuncional. O procedimento de controle de mudanças envolve a participação de várias áreas, incluindo Qualidade, Engenharia, Produção, Compras e Logísticas. Cada área tem responsabilidades específicas na solicitação, avaliação, aprovação, implementação e verificação das mudanças	3	3	3
5	Não há menção específica ao uso de cera de abelha para o arredondamento dos cantos nos procedimentos descritivos. No entanto, o processo de acabamento inicial para carrocerias de madeira inclui lixamento leve das superfícies cortadas e remoção de rebarbas no plástico	3	3	3

Fonte: O Autor (2025)

A aplicação do framework demonstrou que o sistema foi capaz de gerar respostas fundamentadas, claras e aderentes aos requisitos simulados, apresentando em todos os casos níveis máximos de corretude, completude e honestidade segundo a avaliação dos especialistas. As respostas seguiram uma estrutura padronizada de acordo com as orientações do agente, contendo tanto a confirmação ou negação do atendimento quanto as referências documentais que sustentaram a saída, o que reforça a capacidade da arquitetura em assegurar rastreabilidade. Essa padronização aproximou o processo da lógica utilizada em auditorias, em que a clareza e a disponibilidade de evidências são fundamentais para validar a conformidade. Os resultados podem ser considerados positivos, ainda que tenham sido obtidos em condições controladas e com um conjunto de requisitos que não corresponde integralmente à complexidade dos manuais reais utilizados na indústria automotiva.

Adicionalmente, destaca-se que a simulação foi conduzida a partir de uma base fictícia e com uma amostragem pequena, o que limita a generalização dos resultados. A conversão dos requisitos específicos em perguntas auditáveis mostrou-se adequada para integrar o modelo às práticas de avaliação da qualidade, permitindo que o sistema operasse de maneira alinhada às necessidades de auditoria documental. A utilização da cadeia de suprimentos fictícia, composta por normas técnicas e processos simulados, possibilitou testar a eficiência do framework em um ambiente seguro, sem comprometer informações sensíveis, mas não reflete a totalidade dos desafios práticos que seriam encontrados na aplicação sobre bases documentais reais. Ainda assim, os testes realizados evidenciam o potencial da abordagem baseada em RAG para apoiar a análise de CSR em cadeias de fornecimento automotivas, especialmente em cenários que demandam padronização, consistência e transparência.

5. Conclusões

Este artigo abordou a aplicação da arquitetura *Retrieval-Augmented Generation* (RAG) no contexto da avaliação de requisitos específicos de clientes (*Customer Specific Requirements – CSR*) em sistemas de gestão da qualidade na cadeia automotiva. O estudo concentrou-se na proposição de um framework capaz de integrar documentos normativos e requisitos do cliente a um sistema de recuperação semântica e geração de respostas em

linguagem natural, com foco em padronização, rastreabilidade e suporte à tomada de decisão.

O objetivo central foi demonstrar como técnicas de Inteligência Artificial Generativa, aliadas à arquitetura RAG, podem contribuir para otimizar o processo de interpretação e análise de requisitos específicos, tradicionalmente realizado de forma manual e sujeito a variações interpretativas.

Os resultados mostraram que o framework foi capaz de produzir respostas consistentes, fundamentadas e transparentes, assegurando rastreabilidade às fontes documentais. Os testes realizados indicaram eficiência na geração de respostas auditáveis, com avaliações positivas em todos os critérios utilizados. No entanto, os experimentos foram conduzidos com um número limitado de requisitos e utilizando uma base fictícia, o que restringe a abrangência dos achados e não reflete integralmente a complexidade dos requisitos reais encontrados na indústria automotiva.

A principal contribuição do trabalho consiste em apresentar uma proposta metodológica que une avanços recentes em Inteligência Artificial Generativa às práticas consolidadas de gestão da qualidade, abrindo caminho para aplicações em auditorias, validações documentais e processos de conformidade. Ao propor um framework auditável, rastreável e alinhado à lógica das normas ISO 9001 e IATF 16949, o estudo oferece uma alternativa inovadora para apoiar a transformação digital na gestão da qualidade.

Como sugestão para trabalhos futuros, recomenda-se ampliar o estudo para múltiplos conjuntos de CSR reais de diferentes montadoras, avaliar a escalabilidade do framework em bases documentais mais extensas, explorar métricas automáticas de avaliação em conjunto com o *grading method* humano e integrar a solução a sistemas corporativos de gestão documental. Essas iniciativas podem fortalecer a robustez da proposta e consolidar seu uso em ambientes produtivos, aproximando ainda mais a tecnologia da realidade das auditorias na indústria automotiva.

Referências

BROWN, T. et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, v. 33, p. 1877–1901, 2020.

CHATTERJEE, A. et al. Artificial intelligence in quality management: predictive models for corrective actions. *Journal of Manufacturing Systems*, v. 59, p. 118–127, 2021.

CHEN, D. et al. Reading Wikipedia to answer open-domain questions. *Association for Computational Linguistics (ACL)*, p. 1870–1879, 2017.

CROSBY, P. B. *Quality is free: the art of making quality certain*. New York: McGraw-Hill, 1979.

DEMING, W. E. *Out of the crisis*. Cambridge: MIT Press, 1986.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, p. 4171–4186, 2019.

FEIGENBAUM, A. V. *Total quality control*. 3. ed. New York: McGraw-Hill, 1983.

GAO, C. et al. A review of artificial intelligence development from symbolic AI to generative AI. *Artificial Intelligence Review*, v. 57, p. 1123–1159, 2024.

GOETSCH, D. L.; DAVIS, S. B. *Quality management for organizational excellence: introduction to total quality*. 8. ed. New York: Pearson, 2014.

ISHIKAWA, K. What is total quality control? The Japanese way. Englewood Cliffs: Prentice-Hall, 1985.

ISO. ISO 9001:2015 – Quality management systems – Requirements. Geneva: International Organization for Standardization, 2015.

IATF. IATF 16949:2016 – Quality management system requirements for automotive production and relevant service parts organizations. International Automotive Task Force, 2016.

JURAN, J. M. Juran's quality control handbook. New York: McGraw-Hill, 1951.

JURAN, J. M.; GODFREY, A. B. Juran's quality handbook. 5. ed. New York: McGraw-Hill, 1999.

KARPUKHIN, V. et al. Dense passage retrieval for open-domain question answering. EMNLP 2020, p. 6769–6781, 2020.

KIM, K. et al. Failure prediction using machine learning in manufacturing systems. Computers & Industrial Engineering, v. 44, n. 2, p. 289–299, 2003.

LIAO, Y. et al. Applications of artificial intelligence in quality inspection: a review. Procedia CIRP, v. 72, p. 1033–1038, 2018.

LEWIS, P. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, v. 33, p. 9459–9474, 2020.

MURPHY, K. P. Machine learning: a probabilistic perspective. Cambridge: MIT Press, 2012.

OHNO, T. Toyota production system: beyond large-scale production. Portland: Productivity Press, 1988.

RADFORD, A. et al. Language models are unsupervised multitask learners. OpenAI Technical Report, 2019.

RAMESH, A. et al. Hierarchical text-conditional image generation with CLIP latents. arXiv preprint, arXiv:2204.06125, 2022.

ROBITAILLE, D. The ISO 9001:2015 implementation handbook: using the process approach to build a quality management system. Milwaukee: ASQ Quality Press, 2017.

VASWANI, A. et al. Attention is all you need. Advances in Neural Information Processing Systems, v. 30, p. 5998–6008, 2017.

VLASIS, A.; SUN, L. Generative AI: opportunities and challenges. AI & Society, v. 38, p. 721–735, 2023.

WANG, B. et al. Evaluating quality of answers for retrieval-augmented generation: a strong LLM is all you need. arXiv preprint, arXiv:2407.12345, 2024.

YIN, W. et al. DocNLI: A large-scale dataset for document-level natural language inference. ACL 2021, p. 4913–4925, 2021.

ZHOU, Z.; SUN, J.; FAN, Z. Machine learning approaches for quality prediction in manufacturing processes. IEEE Transactions on Systems, Man, and Cybernetics, v. 26, n. 5, p. 561–569, 1996.