

Smart OCR: A Vertical AI Agent for Financial Document Validation — An Exploratory Case Study

Daniel Mitsuaki da Silva Utyiama¹, Oziel Sênior Coelho Moraes¹, Felipe Augusto Souza Guimarães¹, Paolo Bruno Silva Ramos¹, Leonardo Carneiro Marques¹

¹Instituto de Pesquisa e Desenvolvimento SiDi
CEP: 69037-000 – Manaus– AM – Brasil

{d.mitsuaki, oziel.cm, fa.guimaraes, paolo.sr, lc.marques}@sidi.org.br

Abstract. Research Context: The growing complexity of fiscal and financial documents increases the need for intelligent solutions that automate validation. Vertical AI agents are emerging to improve accuracy and efficiency in financial workflows. **Scientific and/or Practical Problem:** Despite their potential, many AI systems face adoption barriers due to usability issues, lack of training, and misalignment with established practices. Understanding user perceptions in early stages is key to sustainable integration. **Proposed Solution and/or Analysis:** This study examines Smart OCR, a vertical AI agent that automates manual validation of fiscal and financial documents and provides a web interface for reviewing discrepancies. We assess perceptions of usefulness, usability, value, satisfaction, and recommendation intent. **Related IS Theory:** The analysis is grounded in the Technology Acceptance Model (TAM) and human–computer interaction literature, which emphasize perceived usefulness and ease of use as determinants of adoption. **Research Method:** The study took place in the financial sector of a research and development institute. Over two weeks, analysts and coordinators used Smart OCR integrated into their workflow. Data were collected through a structured questionnaire combining Likert items and open-ended questions. **Summary of Results:** Findings show high perceived usefulness, expected impact, and value, with strong recommendation intent. Usability and initial satisfaction were moderate, highlighting the need for clearer guidance and interface improvements. **Contributions and Impact to IS area:** The study offers empirical evidence on the acceptance of vertical AI agents in financial contexts and provides insights for IS design, onboarding, and organizational support to foster AI adoption.

1. Introdução

O setor do financeiro, mesmo com os avanços tecnológicos recentes, ainda enfrenta desafios, especialmente no processo de validação de documentos [Szortyka 2024]. Atividades como conferência de notas fiscais de serviço, DANFE, conta de consumo, DARF, nota de débito, FGTS PIX, Impostos Municipais, boletos e outros documentos financeiros consomem tempo, demandam esforço humano e estão sujeitas a erros [Szortyka 2024]. A diversidade de formatos, layouts e estruturas, muitas vezes com dados semiestruturados ou não estruturados [Roy et al. 2025], dificulta a padronização e compromete a detecção de inconsistências [(FSB) 2024], anomalias ou indícios de fraude. Além disso, métodos tradicionais baseados apenas em regras fixas ou soluções de visão computacional

mostram baixa flexibilidade para lidar com novos tipos de documentos ou mudanças nos formatos existentes, limitando a escalabilidade e a capacidade de adaptação do processo [Balch et al. 2023].

Para superar essas limitações, propõe-se o uso de Agentes Inteligentes baseados em Modelos de Linguagem de Grande Escala (LLMs) como núcleo de um sistema de automação financeira [WANG et al. 2024]. Esses agentes podem interpretar, cruzar e validar informações contidas em documentos de diferentes formatos, identificando divergências e alertando para possíveis irregularidades [WANG et al. 2024]. Ao integrar IA e automação inteligente de processos, combinando LLMs, técnicas de automação de processos [Li et al. 2023] e mecanismos de análise contextual [Yang et al. 2024], é possível reduzir o tempo de verificação, aumentar a acurácia e liberar equipes para atividades mais estratégicas [Szortyka 2024, Park 2024]. Entre os ganhos esperados estão a redução de custos operacionais, a aceleração do ciclo de auditoria e a melhoria na qualidade e consistência da validação de dados.

Agentes inteligentes são sistemas capazes de perceber o ambiente, raciocinar e agir de forma autônoma para alcançar objetivos definidos, apoiando-se em conhecimentos e experiências prévias [Szortyka 2024] [Cheng et al. 2024]. Fundamentados em LLMs como GPT, LLaMA e Gemini, esses agentes utilizam arquiteturas *transformer* treinadas em grandes volumes de dados, o que lhes confere elevada capacidade de compreensão e análise da linguagem natural [Li et al. 2024] [Li et al. 2023]. No contexto da validação financeira, podem executar tarefas como extração de campos-chave, comparação automática entre documentos, detecção de inconsistências e identificação de padrões suspeitos [Park 2024] [Yang et al. 2024]. Além disso, frameworks como LangChain, CrewAI e OpenAI Swarm possibilitam a cooperação entre múltiplos agentes especializados, permitindo a decomposição de processos complexos em subtarefas como classificação documental e conferência fiscal [Li et al. 2024].

O sistema proposto utiliza modelos de última geração, incluindo qwen-qwq-32b [Qwen Team 2025], meta-llama/llama-4-scout-17b-16e-instruct [Meta AI 2025], anthropic.claude-3-7-sonnet-20250219-v1:0 [Anthropic 2025] e anthropic.claude-3-5-sonnet-20241022-v2:0 [Anthropic 2024], combinando a precisão e adaptabilidade dos modelos Qwen, LLaMA e Claude. Suas funcionalidades incluem a identificação automática de documentos, extração e validação de dados críticos, associação de irregularidades a datas e entidades específicas e detecção proativa de inconsistências financeiras. Estudos prévios com LLMs indicam potencial para alcançar métricas de acurácia elevadas, com impactos diretos na mitigação de riscos, aumento da transparência e redução de até 60% no tempo total de análise de documentos. A adoção desta solução representa um avanço estratégico para empresas que buscam segurança, eficiência e escalabilidade nos processos de validação documental no setor financeiro.

2. Referencial Teórico e Trabalhos Relacionados

2.1. Conceitos Fundamentais

Os Agentes Inteligentes são sistemas computacionais capazes de perceber o ambiente, processar informações e agir de forma autônoma para atingir objetivos específicos [Russell and Norvig 2021]. Suas principais propriedades incluem a *reatividade*, isto é, a capacidade de responder a estímulos externos; a *proatividade*, relacionada à atuação

orientada a metas; e a *sociabilidade*, que possibilita a interação com outros agentes ou sistemas [Wooldridge 2009]. No setor financeiro, agentes inteligentes têm sido empregados para coordenar processos de extração, validação e auditoria de informações, contribuindo para áreas como conformidade regulatória, análise de relatórios contábeis e auditoria contínua [Nelson et al. 2000, Han et al. 2024]. Arquiteturas multiagentes ampliam esse potencial ao permitir especialização em subtarefas e cooperação entre módulos, tornando os sistemas mais flexíveis e escaláveis [Jennings 2000, Weiss 1999].

O desenvolvimento dos Modelos de Linguagem de Grande Escala (LLMs) expandiu significativamente a atuação dos agentes inteligentes, permitindo que operem não apenas com regras pré-programadas, mas com capacidades linguísticas avançadas [Yin et al. 2024]. Baseados na arquitetura *Transformer* [Vaswani et al. 2017], os LLMs utilizam mecanismos de *self-attention* para capturar dependências de longo alcance no texto, possibilitando avanços expressivos em compreensão e geração de linguagem natural. Além disso, apresentam propriedades emergentes como *generalização zero-shot e few-shot*, capacidade de raciocínio contextual e extração de entidades críticas em textos não estruturados. Estudos recentes apontam seu impacto em setores diversos, incluindo o financeiro [Zhao et al. 2024], mas também destacam riscos como alucinações semânticas, viés nos dados de treinamento e dependência de um *ground truth* confiável para validação [Bommasani et al. 2021].

Uma evolução particularmente relevante é o surgimento dos LLMs multimodais, capazes de processar não apenas texto, mas também imagens e outros tipos de dados. Modelos como CLIP [Radford et al. 2021], Flamingo [Alayrac et al. 2022] e Gemini (Google) integram representações visuais e textuais de forma *end-to-end*. Ao contrário de sistemas que dependem de OCR (Reconhecimento Óptico de Caracteres, do inglês - Optical character recognition) clássico para converter imagens em caracteres [Smith 2007], os modelos multimodais realizam simultaneamente o reconhecimento de texto, a interpretação do layout (como tabelas e colunas) e a compreensão semântica do conteúdo [Liu et al. 2023]. Tal capacidade é especialmente relevante em domínios fiscais e financeiros, em que documentos digitalizados apresentam grande heterogeneidade de qualidade e formato [Wan et al. 2024a].

Em síntese, os Agentes Inteligentes, os Modelos de Linguagem de Grande Escala (LLMs) e os LLMs multimodais representam três pilares tecnológicos interconectados. Esses conceitos, quando considerados em conjunto, oferecem a base tecnológica para novas abordagens de processamento e análise de documentos complexos, incluindo aplicações voltadas à validação fiscal e financeira, que constituem o foco deste trabalho.

2.2. Trabalhos Relacionados

Diversos estudos têm investigado soluções voltadas à validação de documentos fiscais e financeiros. Sistemas baseados em regras fixas, fluxos de trabalho automatizados e técnicas tradicionais de OCR foram amplamente utilizados para conferência de documentos e relatórios financeiros [Nelson et al. 2000, Yimyam et al. 2020, Cho et al. 2023]. Embora esses métodos tenham permitido ganhos iniciais de produtividade, apresentam limitações significativas: são rígidos diante de mudanças de layout, têm dificuldade em lidar com dados semiestruturados e não conseguem capturar contextos complexos em documentos heterogêneos [Roy et al. 2025].

Com o avanço dos Modelos de Linguagem de Grande Escala (LLMs), surgiram aplicações direcionadas ao setor financeiro, incluindo análise automatizada de contratos, suporte ao atendimento ao cliente e auxílio em auditorias [Zhao et al. 2024, WANG et al. 2024]. Esses modelos demonstraram notável capacidade de compreensão da linguagem natural e extração de informações críticas de textos extensos e complexos [Li et al. 2023, Yin et al. 2024]. Contudo, apresentam riscos associados a alucinações semânticas, viés nos dados de treinamento e dependência de bases de conhecimento externas para garantir precisão [Bommasani et al. 2021]. Além disso, muitos trabalhos permanecem restritos ao uso de LLMs unimodais, focados exclusivamente em texto, sem integração com dados visuais ou estruturados.

O surgimento de LLMs multimodais abriu novas possibilidades ao permitir a análise integrada de texto e imagens. Esses modelos são capazes de reconhecer layouts, interpretar tabelas e compreender conteúdos de documentos com maior robustez que abordagens baseadas em OCR ou baseada em regras. Estudos preliminares indicam o potencial dessas arquiteturas para lidar com documentos fiscais e financeiros digitalizados, frequentemente marcados por heterogeneidade de qualidade e formato [Wan et al. 2024a, Yang et al. 2024]. Entretanto, as aplicações ainda são pontuais e carecem de integração com mecanismos de orquestração inteligente capazes de coordenar múltiplas análises [Li et al. 2024, Park 2024].

Em síntese, a literatura evidencia três linhas principais: (i) soluções baseadas em regras, eficazes apenas em contextos estáveis; (ii) LLMs aplicados a tarefas financeiras, ainda limitados a abordagens unimodais; e (iii) modelos multimodais, promissores para lidar com documentos complexos, mas pouco explorados em processos de validação em escala organizacional. Neste trabalho, adotam-se LLMs multimodais com o objetivo de realizar a validação fiscal e financeira em nível organizacional.

3. Smart OCR

Esta seção apresenta a utilização de LLMs para a extração de informações de documentos, com o objetivo de apoiar o trabalho dos revisores fiscais de documentos bancários. Essa atividade é complexa e exige o processamento de grandes volumes de documentos financeiros [Wan et al. 2024b, Yang et al. 2024, Jin 2025], muitas vezes sob prazos apertados, tornando a extração manual demorada e suscetível a erros [Wan et al. 2024b, Yang et al. 2024], especialmente em documentos não estruturados ou semiestruturados, como notas fiscais, boletos, DANFE e DARF. A avaliação precisa e consistente desses documentos é essencial para a identificação de irregularidades [Xu et al. 2024, World Economic Forum 2025, Yeo et al. 2025], uma vez que falhas podem resultar em impactos legais ou financeiros, o que reforça a necessidade de sistemas mais eficientes para o processamento em larga escala.

Para enfrentar esses desafios, propomos agentes inteligentes que utilizam LLMs para a extração automatizada de informações e identificação de irregularidades em documentos financeiros. Modelos avançados de NLP como GPT, Claude e BERT podem processar de maneira eficiente texto estruturado, mitigando o trabalho de leitura e identificação de irregularidades em documentos fiscais.

Conforme ilustrado na Figura 1, a nossa solução consiste de um exame de agentes de consórcio e publicam em filas, esses agentes são condicionados por prompts a esco-

lher a melhor ferramenta para a extração e detecção de irregularidades. Além disso, esses agentes se comunicam por meio de um serviço na figura chamado de Exame de Agentes que orquestra os agentes. Basicamente o fluxo é o lote de documento é enviado para a fila de entrada que é consumida pelos agentes de extração, que fazem a extração de caracteres (OCR) em seguida os agentes de revisão assumem para comparar os dados extraídos com um documento de referência, o resultado é enviado para a fila de saída.

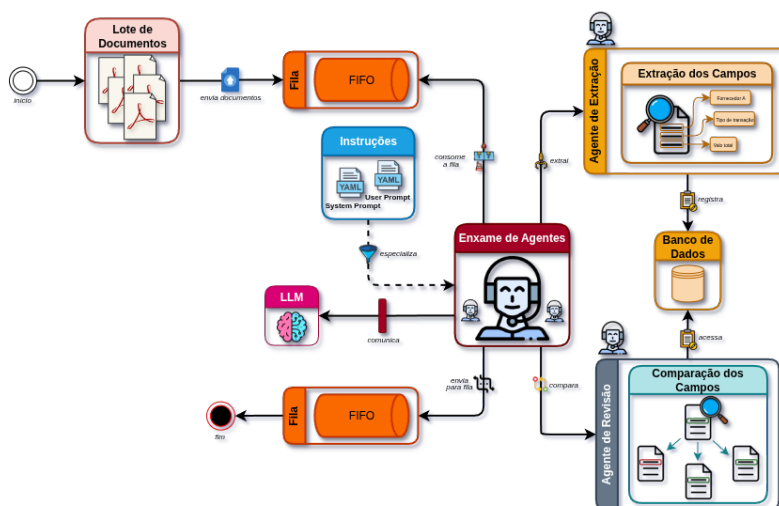


Figura 1. Arquitetura da Aplicação

3.1. Agentes Inteligentes

3.1.1. Engenharia de Prompt

A engenharia de prompt desempenhou papel importante no funcionamento dos agentes de extração e resposta, orientando o modelo de linguagem em tarefas específicas de extração e validação. Os prompts foram projetados para garantir que o modelo compreenda o contexto do documento, identifique os campos relevantes e retorne os dados formatados.

A elaboração desses prompts segue princípios de clareza, contexto e estrutura. Primeiramente, são incluídas instruções explícitas sobre o tipo de documento e os campos a serem extraídos, utilizando uma linguagem objetiva e com exemplos quando necessário. Em seguida, são aplicadas restrições quanto ao formato de saída, normalmente JSON, o que garante padronização para integração com as demais camadas do sistema. Além disso, os prompts contêm orientações adicionais para lidar com ambiguidades, como instruções de fallback, por exemplo "caso o campo não seja encontrado, retorne null" e, validações lógicas como, "garanta que valores numéricos venham sem separadores de milhar".

Na prática, o design dos prompts se adapta ao tipo de documento processado. Para uma NFS-e (Nota Fiscal de Serviço Eletrônica), o prompt orienta a extração de informações fiscais obrigatórias, como CNPJ, data de emissão e valor total. Já para boletos bancários, a ênfase recai sobre vencimento, data de vencimento, cedente e linha digitável. Essa personalização é essencial, visto que cada tipo documental possui estrutura e semântica próprias.

Outro ponto relevante é o uso de técnicas de few-shot prompting, nas quais exemplos de entradas e saídas corretas são fornecidos ao modelo, aumentando sua precisão em contextos mais específicos.

Por fim, a engenharia de prompt também contemplou aspectos de controle. Isso inclui a definição de guardrails lógicos, checagem de consistência entre campos extraídos e reforço semântico por meio de descrições adicionais, como "o campo 'valor total' deve corresponder ao somatório de itens mais impostos". Dessa forma, os prompts deixam de ser apenas instruções textuais e passam a atuar como parte integrante da lógica de negócio, influenciando diretamente na qualidade dos dados extraídos e revisados.

3.1.2. Agentes de Extração

A ingestão de dados prioriza a eficiência operacional ao eliminar a necessidade de pré-processamento manual pelo usuário. O sistema executa um fluxo em lote no backend, aceitando arquivos PDF em seu estado original. Nesse contexto, cabe aos agentes de extração a responsabilidade de normalizar os dados e converter o conteúdo visual em representações textuais estruturadas, viabilizando as etapas subsequentes de análise e validação. Os agentes de extração representam uma das camadas mais importantes da arquitetura proposta, sendo responsáveis por interpretar diferentes tipos de documentos financeiros e transformar seu conteúdo não estruturado em dados organizados em um formato JSON. Esses agentes atuam sobre documentos como Nota Fiscal de Serviço eletrônica (NFS-e), boleto bancário, Documento Auxiliar da Nota Fiscal Eletrônica (DANFE) e Documento de Arrecadação de Receitas Federais (DARF), que são comumente encontradas em rotinas fiscais.

Cada agente possui ferramentas especializadas em um tipo de documento, conhecendo seus elementos estruturais e semânticos. Por exemplo, no caso de uma NFS-e, o agente é capaz de identificar e extrair campos como CNPJ do emitente, nome do tomador, valor total, ISS destacado, data de emissão e número da nota. Para boletos bancários, a extração foca em informações como sacado, cedente, valor nominal, vencimento, código de barras e linha digitável. Da mesma forma, agentes voltados para DANFE ou DARF têm suas próprias estruturas e regras de negócio que orientam a extração de dados relevantes.

O funcionamento desses agentes ocorre em etapas bem definidas. Primeiramente, o sistema realiza uma classificação automática do tipo de documento, utilizando a análise de palavras-chave para direcionar o processamento a ferramenta adequada. Em seguida, há uma etapa de pré-processamento, que pode incluir OCR para documentos digitalizados, limpeza de ruídos textuais, e segmentação por blocos (como cabeçalho, corpo e totais).

A extração em si é realizada por meio de modelos de linguagem, utilizando prompts pensados para guiar o modelo na identificação dos campos desejados. Os dados extraídos são retornados no formato JSON, permitindo sua validação. Essa validação envolve verificações via objeto de transferência de dados, que desempenhando o papel de guardrails e, quando necessário, aplicam regras ao tipo de documento.

Os agentes foram projetados com uma arquitetura modular, permitindo que cada um opere de forma independente. Isso garante flexibilidade na manutenção e escalabili-

dade do sistema, uma vez que novos tipos de documentos podem ser incorporados com o mínimo de impacto no restante da aplicação. Também permite a paralelização das tarefas de extração, o que é essencial em ambientes com alto volume de documentos.

Apesar de sua robustez, existem desafios típicos de ambientes reais. Documentos digitalizados de baixa qualidade podem comprometer a precisão do OCR, e a grande diversidade de layouts, especialmente em NFS-e e boletos, que variam conforme a prefeitura ou o banco emissor, exige modelos que consigam se adaptar. Outro ponto crítico são os erros semânticos ocasionais dos modelos de linguagem, que podem inferir incorretamente um valor ou campo quando o contexto está mal definido ou ambíguo. Ainda, no domínio financeiro e fiscal, é comum que existam campos com nomenclatura ou interpretação diferentes entre documentos similares, o que exige atenção especial na elaboração dos prompts e validações.

Para lidar com esses desafios e garantir uma extração cada vez mais confiável, o sistema pode evoluir para incorporar modelos com ajuste fino com exemplos reais, aprendizado contínuo com base em feedback humano e integração de workflows mais complexos via langChain expression language ou LangGraph, possibilitando um controle mais refinado das etapas e decisões durante a extração.

Em resumo, os agentes de extração automatizam a leitura de documentos financeiros, servindo como base para etapas posteriores, como comparação e auditoria. Eles representam um avanço frente às abordagens tradicionais baseadas apenas em regras fixas ou OCR isolado, oferecendo uma solução adaptativa, inteligente e preparada para operar em contextos de maior complexidade e variáveis como o sistema fiscal brasileiro.

3.1.3. Agentes de Revisão

Enquanto os agentes de extração são responsáveis por transformar documentos financeiros em dados estruturados, os agentes de revisão assumem a função de validar a conformidade entre diferentes documentos, com base em um documento de referência, ou ground truth. Esses agentes foram projetados para operar em fluxos de validação, auditoria e conferência automática de documentos, atuando como um segundo nível de inteligência sobre os dados extraídos.

O funcionamento dos agentes de revisão parte da premissa de que um determinado documento, como uma nota fiscal ou um contrato registrado no sistema, é considerado verdadeiro ou esperado, e deve ser utilizado como base para a conferência de outro documento recebido. Por exemplo, um agente pode receber uma NFS-e extraída como referência e verificar se um boleto bancário recebido posteriormente reflete corretamente os dados originais como valores, datas e razão social. Em outro cenário, o agente pode comparar um DANFE com um comprovante de pagamento de DARF, garantindo que os dados de impostos estejam consistentes.

Esses agentes operam utilizando tanto LLMs quanto técnicas tradicionais de comparação semântica, similaridade vetorial e regras lógicas. Ao receber os dados estruturados de dois ou mais documentos, o agente avalia cada campo relevante e identifica divergências, sejam elas numéricas, textuais ou estruturais. Essa análise não se limita a apontar que “os dados são diferentes”; o agente também é capaz de justificar as diferenças, ofe-

recendo explicações contextuais como descontos aplicados, variações de impostos, erros de digitação ou ausência de campos obrigatórios.

A comparação é orientada por esquemas pré-definidos, nos quais se define quais campos são considerados críticos, opcionais ou derivados. O agente pode, por exemplo, entender que divergências no campo “valor total” devem ser sinalizadas com maior severidade do que no campo “descrição do serviço”, e que datas com pequenas diferenças podem ser aceitáveis, dependendo da margem de tolerância definida no contexto.

Além disso, os agentes de revisão podem operar com diferentes níveis de sensibilidade. Em contextos contábeis ou fiscais mais rígidos, podem ser configurados para sinalizar qualquer divergência, por menor que seja. Já em contextos operacionais, como conferência de faturas ou cobranças, podem aplicar regras mais flexíveis, considerando variações aceitáveis de valores ou prazos. Essa adaptabilidade torna os agentes de revisão extremamente úteis em ambientes corporativos, onde os documentos passam por múltiplas etapas e sistemas antes de serem conferidos.

Embora a implatação desses agentes trouxe celeridade no processo de validação dos documentos, a operação desses agentes enfrenta desafios importantes. Um deles é a necessidade de um ground truth confiável, se o documento de referência estiver incorreto, toda a cadeia de validação será comprometida. Outro desafio está na interpretação semântica: campos com nomes diferentes, mas conteúdo equivalente, como “valor da nota” e “valor do serviço”, precisam ser mapeados corretamente. Além disso, há casos em que diferenças são esperadas e legítimas, como impostos retidos ou descontos concedidos, exigindo que o agente tenha inteligência contextual e não apenas regras fixas.

Para mitigar esses riscos e aumentar a confiabilidade, os agentes de revisão podem evoluir para incorporar feedback humano supervisionado, aprendizado contínuo com base em erros históricos, e integração com fontes externas de dados fiscais ou contratuais. Também podem ser combinados com sistemas de alerta e dashboards analíticos, permitindo que times financeiros ou jurídicos acompanhem indicadores de conformidade em tempo real.

Assim, os agentes de revisão também desempenham um papel importante na solução ao garantir que os documentos financeiros estejam em conformidade com os registros esperados, automatizando processos que tradicionalmente exigiriam verificação manual, suscetível a erros e demorada. Quando integrados de forma coordenada com os agentes de extração, criam um pipeline robusto e inteligente capaz de transformar rotinas operacionais em processos automatizados com alto valor, auditáveis e também escaláveis.

3.1.4. Modelos de Larga Escala utilizados

A escolha dos LLMs utilizados foi orientada pelas necessidades de cada tipo de agentes, levando em conta critérios como capacidade de compreensão semântica, eficiência no processamento de prompts, custo computacional, estabilidade e segurança das informações financeiras trafegadas entre o agente e o provedor do modelo.

Nos agentes de extração, foi utilizado o modelo `anthropic.claude-3-7-sonnet-20250219-v1:0` [Anthropic 2025]. Esse modelo se mostrou eficiente na interpretação de instruções em prompts e na manipulação de contextos com variabilidade nas estruturas

dos documentos, como ocorre em documentos fiscais do tipo NFS-e, boletos bancários e DANFES. Sua capacidade de extração de campos garantiu resultados mais consistentes mesmo em situações de documentos digitalizados com ruídos ou estruturas não padronizadas. Outro ponto relevante foi a boa relação entre custo e desempenho, permitindo o processamento de grandes volumes de documentos sem comprometer a escalabilidade do sistema.

Já os agentes de revisão foram implementados com o modelo `us.anthropic.claude-3-5-sonnet-20241022-v2:0` [Anthropic 2024], escolhido por sua maior estabilidade em análises comparativas. Esse modelo demonstrou bom desempenho na interpretação de diferenças semânticas, como a equivalência entre campos com nomes diferentes, e na avaliação de discrepâncias numéricas ou textuais de acordo com as regras configuradas.

Em síntese, a combinação do Claude 3.7 Sonnet na etapa de extração e do Claude 3.5 Sonnet na etapa de revisão criou uma complementaridade entre os modelos. Enquanto o primeiro atua como motor principal da interpretação documental, especializado em estruturar informações, o segundo assume o papel de validador inteligente. Adicionalmente, visando atender aos requisitos de privacidade e confidencialidade associados a documentos fiscais e financeiros, o processamento dos dados foi realizado por meio de serviços da Amazon Web Services, os quais mantêm acordo contratual de confidencialidade com o SiDi. Nesse contexto, a comunicação entre os componentes do sistema e os agentes ocorre exclusivamente via canais criptografados. Além disso, a adoção de APIs em modalidade corporativa veta, contratualmente, o armazenamento ou o uso dos dados para retreinamento de modelos. Tais medidas garantem o sigilo das informações e a plena conformidade com a Lei Geral de Proteção de Dados.

3.2. Interface e Fluxo de Uso

A aplicação SmartOCR inicia pela tela de login, onde o usuário informa e-mail e senha. Após a autenticação, a tela Minhas Atividades, Figura 2, funciona como o painel central, reunindo os tickets de validação organizados em itens divergentes, pendentes e aprovados. A interface oferece filtros por unidade organizacional e data de pagamento, além de exibir metadados resumidos e integração com ferramentas externas, como o JIRA.

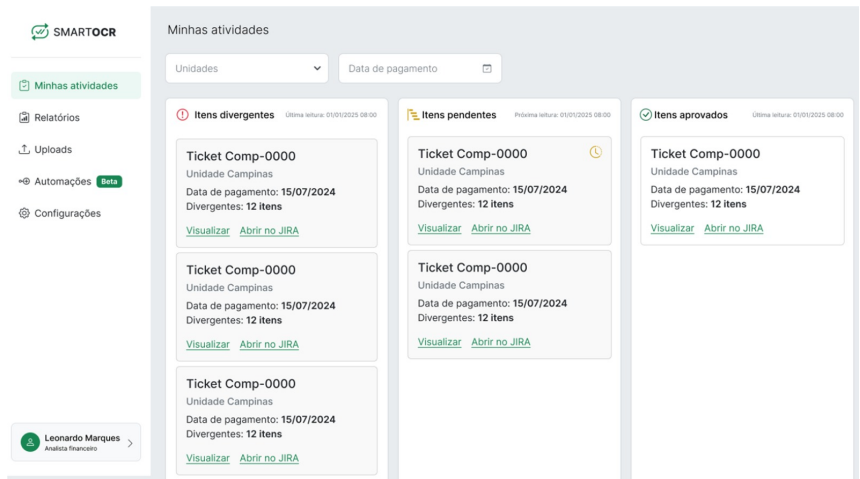


Figura 2. Tela de Minhas Atividades

A validação assistida ocorre na tela de detalhe do Ticket, Figura 3, que materializa a colaboração entre a inteligência artificial e o julgamento humano. Neste ambiente, o sistema apresenta estruturadamente as divergências identificadas pela comparação semântica em campos críticos, como CNPJ e vencimento. O usuário visualiza simultaneamente o documento base e os arquivos comparados para realizar a revisão final. Portanto, essa interface assegura a confiabilidade do processo, permitindo que a IA automatize a extração enquanto o usuário exerce a decisão final sobre as inconsistências.

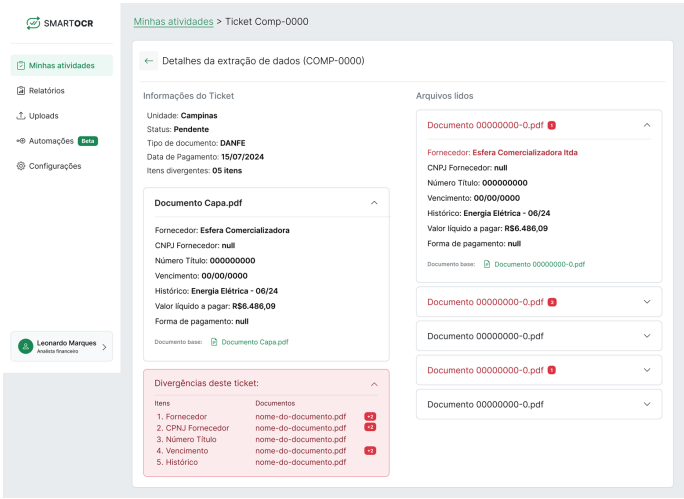


Figura 3. Tela de Detalhamento de Ticket

Em suma, a articulação dessas telas compõe o fluxo integrado que melhora a simples automação de tarefas. Essa organização reflete a fidelidade ao contexto de uso real, suportando desde o acesso seguro até o monitoramento gerencial. Ao alinhar usuários reais a suas tarefas, o SmartOCR consolida-se não apenas como um extrator de dados, mas como uma ferramenta de apoio à decisão na validação de documentos fiscais e financeiros.

4. Estudo

O presente trabalho caracteriza-se como um estudo de caso exploratório, conduzido no setor financeiro de um Instituto de P&D. A estratégia de amostragem adotada foi não probabilística e intencional, contemplando a totalidade da equipe diretamente responsável pelo fluxo de validação fiscal na organização. Dessa forma, a amostra de N=5, embora insuficiente para generalizações estatísticas, apresenta aderência ao contexto real de uso, uma vez que representa todos os especialistas que executam o processo em ambiente real de trabalho. Os resultados devem, portanto, ser interpretados dentro de seus limites exploratórios e contextuais, como evidências iniciais sobre o fenômeno investigado. Além disso, apesar de a pesquisa não ter sido submetida a um comitê de ética oficial, o estudo foi conduzido com a permissão dos participantes, que assinaram um Termo de Consentimento Livre e Esclarecido (TCLE) disponibilizado antes do questionário. O acesso ao questionário só era permitido após a assinatura do TCLE.

O objetivo da investigação foi compreender a percepção dos colaboradores quanto à utilidade, usabilidade e valor percebido da ferramenta, considerando sua aplicação piloto em processos de validação fiscal e financeira. Para isso, o Smart OCR foi incluído diretamente no fluxo de trabalho já existente dos analistas e coordenadores, auxiliando

na análise dos documentos que tradicionalmente eram verificados de forma manual. As etapas de conferência passaram a ser operacionalizadas automaticamente pelo sistema, enquanto as divergências identificadas podiam ser consultadas em uma interface web desenvolvida especificamente para essa finalidade.

A aplicação do sistema ocorreu em uma rodada inicial de uso, ao longo de um período de duas semanas, sem treinamento formal prévio. Os participantes receberam apenas uma explicação inicial de como a ferramenta funcionaria dentro de seu fluxo de trabalho, e em seguida tiveram contato direto com o sistema em um cenário próximo ao real de operação. Essa escolha metodológica permitiu avaliar a experiência genuína de uso, sem vieses de instrução ou suporte estruturado, refletindo as condições práticas de adoção em ambiente corporativo.

Ao término das duas semanas, foi aplicado um questionário estruturado, composto por duas partes: (i) questões demográficas e itens em escala Likert de 5 pontos, avaliando os construtos de utilidade percebida, usabilidade percebida, impacto esperado, valor percebido, satisfação e intenção de recomendação; (ii) questões abertas, que permitiram aos participantes expressar opiniões livres sobre pontos fortes, fragilidades e sugestões de melhoria.

Essa configuração do estudo possibilitou a triangulação entre dados quantitativos e qualitativos, de modo a oferecer uma análise mais abrangente da aceitação inicial do sistema. Na próxima seção são apresentados os resultados obtidos em relação ao objetivo do estudo.

5. Resultados

5.1. Caracterização da amostra

Na Figura 4 é possível analisar visualmente que a maioria dos respondentes tinha entre 25–34 anos (40%) ou 35–44 anos (40%), com um participante na faixa de 45–54 anos (20%). Em relação à função atual, observou-se diversidade de cargos (analistas e coordenadores), sendo que 80% estavam há 1–3 anos na posição atual. Quanto à experiência na área financeira, 60% relataram mais de 6 anos de atuação, evidenciando que a amostra representa profissionais experientes no domínio de validação fiscal e financeira. A qualificação técnica dos respondentes, de maioria sênior, mitiga o viés de aprendizado e confere maior confiabilidade às percepções de utilidade e valor, uma vez que as avaliações provêm de especialistas conhecedores das nuances do problema de negócio, e não de usuários novatos



Figura 4. Dados Demográficos dos Participantes

5.2. Estatísticas descritivas dos construtos

A Tabela 1 apresenta as estatísticas descritivas para os construtos avaliados. Os *scores* foram calculados a partir da média dos itens correspondentes (quando aplicável).

Tabela 1. Estatísticas descritivas dos construtos (escala 1–5). UP = Utilidade Percebida; US = Usabilidade Percebida; IE = Impacto Esperado; VP = Valor Percebido; SAT = Satisfação; REC = Recomendação.

Construto	N	M	DP	Med.	Q1	Q3	IC95% Inf.	IC95% Sup.
UP	5	4,73	0,44	5,00	4,50	5,00	4,34	5,12
US	5	3,90	0,74	4,00	3,50	4,50	3,24	4,56
IE	5	4,60	0,60	4,60	4,20	5,00	4,07	5,13
VP	5	4,60	0,55	5,00	4,20	5,00	4,11	5,09
SAT	5	3,80	0,84	4,00	3,00	4,50	3,05	4,55
REC	5	4,80	0,45	5,00	4,50	5,00	4,40	5,20

As percepções de utilidade (M=4,73), impacto esperado (M=4,60) e valor percebido (M=4,60) foram avaliadas como elevadas, reforçando que o Smart OCR foi percebido pelos participantes do estudo como relevante e benéfico. A intenção de recomendação (M=4,80) foi a mais alta, indicando propensão favorável à adoção plena do sistema. Por outro lado, usabilidade percebida (M=3,90) e satisfação inicial (M=3,80) mostraram valores mais moderados, sugerindo oportunidades de melhoria na experiência de uso.

5.3. Estatísticas descritivas dos itens

A Tabela 2 detalha as médias por item individual, possibilitando identificar quais aspectos específicos foram mais bem avaliados.

Os itens mais bem avaliados concentraram-se em redução de erros e tempo, diminuição de esforço e potencial estratégico (todos com M=4,8). Em contrapartida, os aspectos mais críticos foram simplicidade do teste inicial (M=3,6) e satisfação geral (M=3,8), indicando barreiras ligadas mais à experiência de uso do que ao valor percebido (veja Figura 5).

Avaliações dos Usuários por Critério

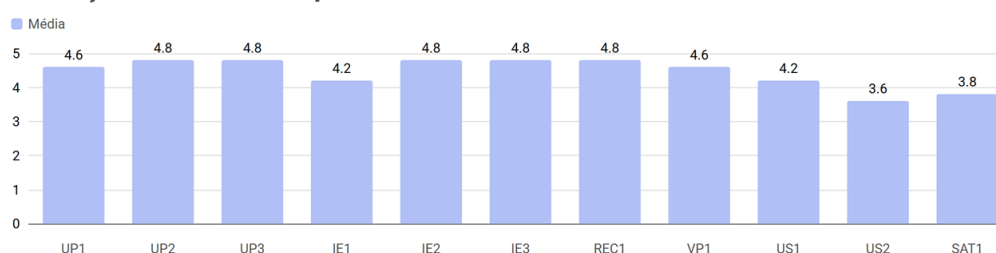


Figura 5. Avaliações dos Usuários por Critério

Tabela 2. Estatísticas descritivas por item (escala 1–5).

Item	M	DP
UP1 — O Smart OCR aparenta ser útil para apoiar meu trabalho.	4,60	0,55
UP2 — O Smart OCR tem potencial para reduzir erros na validação fiscal e financeira.	4,80	0,45
UP3 — O Smart OCR pode reduzir o tempo necessário para concluir as tarefas de validação de documentos.	4,80	0,45
IE1 — Mesmo após uso inicial, pode melhorar a qualidade das entregas.	4,20	0,84
IE2 — Pode contribuir para diminuir o esforço ou estresse na validação de documentos.	4,80	0,45
IE3 — Pode liberar tempo para atividades mais analíticas ou estratégicas.	4,80	0,45
REC1 — Recomendo que o Smart OCR seja integrado definitivamente ao processo financeiro.	4,80	0,45
VP1 — O Smart OCR agrega valor ao processo de gestão financeira da empresa.	4,60	0,55
US1 — A interface do Smart OCR é fácil de entender e utilizar.	4,20	0,45
US2 — Foi simples realizar o teste inicial com o Smart OCR.	3,60	0,89
SAT1 — Estou satisfeito(a) com minha experiência inicial.	3,80	0,84

5.4. Consistência interna das escalas

A Tabela 3 apresenta os valores de alfa de Cronbach para os construtos avaliados com múltiplos itens.

Tabela 3. Confiabilidade interna (alfa de Cronbach).

Construto	k (itens)	α
UP (Utilidade percebida)	3	0,882
US (Usabilidade percebida)	2	0,636
IE (Impacto esperado)	3	0,750

Os valores indicam boa consistência interna para utilidade percebida e impacto esperado, e consistência aceitável para usabilidade, considerando o número reduzido de itens ($k=2$). Esses resultados conferem validade interna às escalas aplicadas e sustentam a interpretação dos *scores* médios apresentados.

5.5. Percepções qualitativas dos participantes

Além das escalas fechadas, o questionário incluiu questões abertas para que os participantes expressassem suas percepções sobre o uso inicial do Smart OCR. A análise das respostas permitiu identificar três categorias principais: (i) *pontos positivos*, (ii) *sugestões de melhoria*, e (iii) *observações gerais*. Esses achados qualitativos complementam as estatísticas descritivas apresentadas anteriormente, conferindo maior profundidade à interpretação dos resultados.

5.5.1. Pontos positivos

Os participantes destacaram ganhos relacionados à eficiência e ao potencial estratégico do sistema:

“O sistema facilitou muito a conferência dos documentos, evitando retrabalho e agilizando o processo.” (P01)

“Vejo que pode liberar tempo para análises mais profundas em vez de ficar só na validação operacional.” (P02)

Esses comentários complementam os *escores* em utilidade percebida ($M=4,73$) e impacto esperado ($M=4,60$). A percepção de que o Smart OCR contribui para reduzir erros e tempo reflete também os resultados por item (UP2, UP3, IE2, IE3), todos com média 4,8. Dessa forma, a análise qualitativa confirma que os usuários enxergaram o sistema como um facilitador operacional e um potencializador estratégico, convergindo com a intenção de recomendação elevada ($M=4,80$).

5.5.2. Sugestões de melhoria

As principais críticas concentraram-se na usabilidade inicial e na necessidade de apoio ao usuário:

“Precisei de ajuda para entender como inserir os documentos, a interface não foi totalmente intuitiva.” (P03)

“Acho que com treinamento o uso será mais fácil.” (P04)

Essas falas corroboram com os resultados quantitativos mais baixos de usabilidade percebida ($M=3,90$) e satisfação inicial ($M=3,80$). Enquanto os participantes reconheceram o valor do sistema, identificaram dificuldades práticas em sua utilização inicial. Isso sugere que barreiras de adoção não residem na percepção de valor do sistema pelos usuários, mas sim na experiência de interação. Esse padrão é comum em estudos de aceitação tecnológica quando não há treinamento prévio, reforçando a necessidade de estratégias de *onboarding* e suporte ao usuário [Dulebohn et al. 2005, Karunarathne et al. 2020, Langendorf et al. 2025].

5.5.3. Observações gerais

Por fim, alguns respondentes destacaram o caráter promissor do sistema, mas ponderaram que ajustes são necessários para sua adoção plena:

“Gostei do conceito, mas acho que precisa de mais explicações ou tutoriais.” (P05)

“O potencial é claro, mas acredito que a interface precisa ser mais simples para todos do time.” (P03)

“Vejo futuro na ferramenta, mas seria importante ter materiais de apoio para que todos entendam bem o processo.” (P04)

Essas observações reforçam a visão de que, embora o Smart OCR seja percebido como relevante e com potencial de impacto, ainda existe uma lacuna entre a *aceitação cognitiva* (reconhecimento do valor) e a *adoção efetiva* (uso fluido e satisfatório). Isso explica por que construtos como utilidade e valor percebido tiveram médias altas (acima de 4,6), mas satisfação e usabilidade ficaram em torno de 3,8 e 3,9. Assim, os comentários qualitativos fornecem indicativos importantes, especialmente para a empresa X: os usuários estão abertos à adoção, mas condicionam uma aceitação maior a melhorias na clareza, documentação e suporte inicial.

6. Discussão

Os resultados desta avaliação inicial do Smart OCR indicam que, mesmo em uma rodada experimental de uso, o sistema foi percebido como útil e com impacto positivo esperado no processo de validação fiscal e financeira. Tais achados são coerentes com modelos clássicos de aceitação tecnológica, como o Technology Acceptance Model (TAM) [Davis 1989], segundo o qual a utilidade percebida é um dos principais determinantes da intenção de uso. O *score* médio de utilidade ($M=4,73$; $\alpha = 0,882$) e a intenção de recomendação ($M=4,80$) sugerem uma predisposição favorável à adoção, alinhando-se a evidências de que a percepção de ganhos concretos influencia fortemente a aceitação de novas tecnologias em ambientes organizacionais [Venkatesh and Davis 2000, Venkatesh et al. 2003].

O construto de impacto esperado também apresentou valores elevados ($M=4,60$; $\alpha = 0,750$), indicando que os usuários identificam benefícios potenciais como redução de esforço, economia de tempo e liberação de recursos para tarefas analíticas. Tais percepções dialogam com estudos recentes sobre a adoção de sistemas de inteligência artificial no setor financeiro, nos quais a eficiência operacional e a redução de erros são apontadas como principais motivadores de aceitação [Davenport and Ronanki 2018, Syed et al. 2020].

Por outro lado, a usabilidade percebida ($M=3,90$; $\alpha = 0,636$) e a satisfação inicial ($M=3,80$) mostraram-se moderadas, apontando para barreiras de experiência do usuário. Essa discrepância entre percepção de valor/utilidade e menor percepção de usabilidade é consistente com a literatura de Interação Humano-Computar (IHC), que destaca que a adoção sustentada depende não apenas da percepção de benefícios, mas também da facilidade de uso e qualidade da experiência interativa [Venkatesh et al. 2012, Nielsen 2012]. Em contextos de sistemas de suporte financeiro, fatores como onboarding adequado, clareza na apresentação de resultados e integração fluida ao fluxo de trabalho são determinantes para converter a aceitação inicial em uso contínuo [Gimpel et al. 2018]. Além disso, a relação entre usabilidade e satisfação corroboram os resultados do estudo de Klock [Klock et al. 2016], apresentado no Simpósio Brasileiro de Sistema de Informação (SBSI), que também destacou a diferenciação entre satisfação do usuário e usabilidade em sistemas de gerenciamento de referências. Assim como no Smart OCR, onde a usabilidade moderada influenciou a satisfação inicial, o estudo do SBSI evidenciou que a usabilidade impacta diretamente a adoção e aceitação das ferramentas digitais, mesmo quando sua utilidade é reconhecida. Esse paralelo reforça a necessidade de investir aprimoramentos na usabilidade para transformar percepções positivas de valor e utilidade em satisfação e engajamento efetivos, garantindo a plena adoção dos sistemas pelos usuários.

É importante considerar que os participantes não receberam treinamento estruturado e tiveram contato limitado ao sistema. Estudos prévios demonstram que a percepção de usabilidade tende a melhorar com maior familiaridade e com processos de adaptação organizacional [Gefen and Straub 2000, Holden and Karsh 2010]. Assim, é factível que, em fases posteriores de implantação, o Smart OCR apresente maiores níveis de satisfação e experiência positiva, desde que acompanhado de práticas que visem melhorar a usabilidade de experiência dos usuários (UX).

Em termos de implicações práticas, os resultados sugerem que o Smart OCR possui forte potencial de aceitação organizacional, dado que os usuários já reconhecem seu valor e estão dispostos a recomendá-lo. Contudo, para garantir adoção sustentada, este estudo inicial indicou que é necessário investir em: (i) treinamentos e guias de uso que reduzam a curva de aprendizado; (ii) melhorias na interface visando maior clareza e explicabilidade dos resultados; e (iii) estratégias de integração que reduzam a percepção de esforço na interação com o sistema.

Entre as limitações deste estudo, destaca-se o pequeno tamanho da amostra (N=5) e o caráter exploratório da coleta, restrita a uma única rodada experimental de uso. Essas limitações restringem a generalização dos achados, mas não invalidam o caráter diagnóstico inicial da pesquisa. Além disso, é importante ressaltar que os resultados quantitativos apresentados não visam a inferência estatística populacional, mas sim a generalização analítica para contextos organizacionais similares. Estudos futuros devem ampliar o número de participantes, considerar comparações entre perfis de usuários (ex.: analistas vs. gestores) e aplicar modelos quantitativos mais robustos (ex.: regressões, modelagem de equações estruturais) para investigar relações entre utilidade, usabilidade, satisfação e recomendação.

De forma geral, este estudo reforça evidências da literatura de que a utilidade percebida e o impacto esperado são preditores-chave da aceitação de sistemas baseados em IA em organizações, enquanto a usabilidade representa um fator crítico para consolidação da adoção. Nesse sentido, o Smart OCR demonstra potencial para contribuir à modernização dos processos financeiros, desde que acompanhado de ajustes no design de interação e estratégias de suporte aos usuários.

7. Conclusões e Trabalhos Futuros

Este artigo apresentou uma avaliação inicial do Smart OCR, um agente inteligente de apoio às atividades financeiras, cuja função central é a extração de informações de notas fiscais e a identificação de inconsistências contábeis, com possibilidade de atuação corretiva no fluxo de trabalho. A investigação foi conduzida de forma exploratória, contando com a participação de cinco analistas e coordenadores financeiros da empresa X, em caráter de estudo de viabilidade para compreender a percepção de utilidade, usabilidade, impacto esperado e satisfação inicial com o sistema.

Os achados indicaram que o Smart OCR foi percebido como útil e capaz de gerar impacto esperado positivo no processo de validação fiscal e financeira, corroborando modelos clássicos de aceitação tecnológica. A intenção de recomendação por parte dos participantes sugere que o sistema possui potencial de aceitação organizacional, especialmente pelo reconhecimento de benefícios como redução de erros, economia de tempo e liberação de recursos para análises de maior valor agregado. Por outro lado, a usabilidade

e a satisfação inicial apresentaram valores moderados, revelando barreiras relacionadas à experiência do usuário que podem comprometer a adoção do agente no fluxo de trabalho caso não sejam adequadamente endereçadas. Esses resultados reforçam que, embora a percepção de utilidade seja determinante na intenção de uso, a qualidade da interação e o suporte ao usuário são fatores críticos para consolidar a adoção contínua de tecnologias baseadas em IA.

Entre as contribuições deste trabalho, destaca-se o caráter exploratório da avaliação, em ambiente real, que permite compreender dimensões-chave de aceitação tecnológica aplicadas ao contexto financeiro-organizacional. Os resultados sugerem implicações práticas para futuras implantações, como a necessidade de investimentos em treinamento, clareza na apresentação dos resultados e integração fluida aos processos de trabalho.

Como trabalhos futuros, planeja-se a execução de um novo estudo com a ampliação da amostra tanto quantitativamente quanto em termos de diversidade, para incluir diferentes perfis de usuários (por exemplo, analistas, gestores e profissionais de auditoria), o que possibilitaria análises comparativas sobre percepções e expectativas em relação ao sistema. Estudos longitudinais também poderão investigar a evolução da usabilidade e da satisfação à medida que os usuários ganham familiaridade com a ferramenta e que melhorias de interface sejam implementadas. Adicionalmente, o uso de métodos quantitativos mais robustos, como regressões ou modelagem de equações estruturais, poderá contribuir para explorar relações entre os construtos analisados (utilidade, usabilidade, satisfação e recomendação), permitindo maior generalização e rigor estatístico. De maneira geral, os resultados apresentados constituem evidências iniciais de viabilidade para o uso organizacional do Smart OCR e apontam caminhos para aprimorar tanto o potencial tecnológico quanto a experiência de adoção, consolidando o papel de sistemas de IA no apoio aos processos financeiros.

Referências

- Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., et al. (2022). Flamingo: a visual language model for few-shot learning. In *Advances in Neural Information Processing Systems*.
- Anthropic (2024). Claude 3.5 sonnet v2: Multimodal reasoning model. Disponível em: <https://aws.amazon.com/bedrock/claude/>. Acesso em: 8 fev. 2026.
- Anthropic (2025). Claude 3.7 sonnet: Extended thinking model. Disponível em: <https://build.nvidia.com/anthropic/claude-3-7-sonnet-20250219-v1:0>. Acesso em: 8 fev. 2026.
- Balch, T., Veloso, M., et al. (2023). Flowmind: Automatic workflow generation in financial services leveraging large language models. In *ICAIF '23: Proceedings of the 4th ACM International Conference on AI in Finance*, page 1–11, Brooklyn, NY, USA.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Cheng, Y., Zhang, C., Zhang, Z., Meng, X., Hong, S., Li, W., Wang, Z., Wang, Z., Yin, F., Zhao, J., and He, X. (2024). Exploring large language model based intelligent agents: Definitions, methods, and prospects. *arXiv preprint*.

- Cho, S., Moon, J., Bae, J., Kang, J., and Lee, S. (2023). A framework for understanding unstructured financial documents using rpa and multimodal approach. *Electronics*, 12(4):939.
- Davenport, T. H. and Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1):108–116.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340.
- Dulebohn, J. H., Murray, B., Sun, M., Rao, V., and Wilkie, D. (2005). Computer training and technology acceptance by end-users: testing a causal model. *Journal of Organizational Behavior*, 26(3):329–347.
- (FSB), F. (2024). Documento ou relatório da fsb. Contém Sumário Executivo e Introdução; Copyright © 2024 Financial Stability Board.
- Gefen, D. and Straub, D. W. (2000). The relative importance of perceived ease of use in is adoption: A study of e-commerce adoption. *Journal of the Association for Information Systems*, 1(1):1–30.
- Gimpel, H., Hosseini, S., Huber, R., Probst, L., and Röglinger, M. (2018). Structuring digital transformation: A framework of action fields and its application at zeiss. *Journal of Information Technology Theory and Application*, 19(1):31–54.
- Han, S., Kang, H., Jin, B., Liu, X. Y., and Yang, S. Y. (2024). Xbrl agent: Leveraging large language models for financial report analysis. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 856–864. ACM.
- Holden, R. J. and Karsh, B.-T. (2010). The technology acceptance model: Its past and its future in health care. *Journal of Biomedical Informatics*, 43(1):159–172.
- Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, 117(2):277–296.
- Jin, J. (2025). Designing the next generation of intelligent manufacturing with generative ai. <https://www.fujitsu.com/global/about/resources/publications/insight/>. Fujitsu Insight Paper. Acessado em 27 ago. 2025.
- Karunaratne, P., Madushani, D., and Perera, S. (2020). Role of employee training and experience on the adoption of digital technologies. *Economic Horizons*, 22(2):123–138.
- Klock, A. C. T., Gasparini, I., and Pimenta, M. S. (2016). Avaliação de usabilidade de sistemas de gerenciamento de referências: um estudo com endnote, mendeley e zotero. In *Anais do XXI Simpósio Brasileiro sobre Sistemas de Informação (SBSI)*, Brasil. Sociedade Brasileira de Computação (SBC).
- Langendorf, J., Meier, C., and Schuster, F. (2025). Systematic literature review on usability and training in digital systems. *Journal of Usability Studies*, 20(1):45–68.
- Li, X., Wang, S., Zeng, S., Wu, Y., and Yang, Y. (2024). A survey on llm-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth*, 1(9):1–43.

- Li, Y., Wang, S., Ding, H., and Chen, H. (2023). Large language models in finance: A survey. In *ICAIF '23: Proceedings of the 4th ACM International Conference on AI in Finance*, Brooklyn, NY, USA.
- Liu, Z., Li, J., Wu, Z., Zhang, Y., Zhou, Y., Lin, K., Yang, Y., Li, J., Hu, Z., Dai, X., et al. (2023). A survey on multimodal large language models. *arXiv preprint arXiv:2306.13549*.
- Meta AI (2025). Llama 4 Scout 17B 16E Instruct: Natively multimodal MoE model. Disponível em: <https://huggingface.co/meta-llama/Llama-4-Scout-17B-16E-Instruct>. Acesso em: 8 fev. 2026.
- Nelson, K. M., Kogan, A., Srivastava, R. P., Vasarhelyi, M. A., and Lu, H. (2000). Virtual auditing agents: the edgar agent challenge. *Decision Support Systems*, 28(3):241–253.
- Nielsen, J. (2012). Usability 101: Introduction to usability. <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>. Acessado em 27 ago. 2025.
- Park, T. (2024). Enhancing anomaly detection in financial markets with an llm-based multi-agent framework. *arXiv preprint*.
- Qwen Team (2025). QwQ-32B: Reasoning model series developed by Qwen. Disponível em: <https://huggingface.co/Qwen/QwQ-32B>. Acesso em: 8 fev. 2026.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR.
- Roy, P., Ghose, B., Singh, P. K., Tyagi, P. K., and Vasudevan, A. (2025). Artificial intelligence and finance: A bibliometric review on the trends, influences, and research directions [version 1; peer review: 2 approved]. *F1000Research*, 14(122).
- Russell, S. and Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson, 4 edition.
- Smith, R. (2007). An overview of the tesseract ocr engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 629–633. IEEE.
- Syed, R., Blome, C., Papadopoulos, T., and Childe, S. J. (2020). Artificial intelligence in operations and supply chain management: A systematic review. *Supply Chain Management*, 25(6):985–1009.
- Szortyka, K. (2024). Inteligentna automatyzacja procesów finansowo-księgowych w centrach outsourcingu procesów biznesowych w polsce. *Stowarzyszenie Księgowych w Polsce*, 48(3):155–175.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.
- Venkatesh, V. and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204.

- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.
- Venkatesh, V., Thong, J. Y. L., and Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1):157–178.
- Wan, X., Deng, H., Zou, K., and Xu, S. (2024a). Enhancing the efficiency and accuracy of underlying asset reviews in structured finance: The application of multi-agent framework. *arXiv preprint arXiv:2405.04294*.
- Wan, X., Deng, H., Zou, K., and Xu, S. (2024b). Enhancing the efficiency and accuracy of underlying asset reviews in structured finance: The application of multi-agent framework. <https://arxiv.org/abs/2405.04294>. arXiv preprint arXiv:2405.04294. Acessado em 27 ago. 2025.
- WANG, L., MA, C., FENG, X., ZHANG, Z., YANG, H., ZHANG, J., CHEN, Z., TANG, J., CHEN, X., LIN, Y., ZHAO, W. X., WEI, Z., and WEN, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):1–26.
- Weiss, G. (1999). *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press.
- Wooldridge, M. (2009). *An Introduction to Multi-Agent Systems*. John Wiley & Sons, 2 edition.
- World Economic Forum (2025). Artificial intelligence in financial services 2025. <https://www.weforum.org/>. World Economic Forum Report. Acessado em 27 ago. 2025.
- Xu, F. F., Song, Y., Li, B., et al. (2024). Theagentcompany: Benchmarking llm agents on consequential real-world tasks. Disponível em: <https://arxiv.org/abs/2412.14161>. Acesso em: 27 ago. 2025.
- Yang, H., Zhang, B., Wang, N., et al. (2024). Finrobot: An open-source ai agent platform for financial applications using large language models. Disponível em: <https://arxiv.org/abs/2405.14767>. Acesso em: 27 ago. 2025.
- Yeo, W. J., Heever, W. V. D., Mao, R., Cambria, E., Satapathy, R., and Mengaldo, G. (2025). A comprehensive review on financial explainable ai. <https://doi.org/10.1007/s10462-024-11077-7>. Published in *Artificial Intelligence Review*, 58:189. Acessado em 27 ago. 2025.
- Yimyam, W., Ketcham, M., Jensuttiwetchakult, T., Hiranchan, S., Pramkeaw, P., and Chumuang, N. (2020). Enhancing and evaluating an impact of ocr and ontology on financial document checking process. In *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pages 1–6. IEEE.
- Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., and Chen, E. (2024). A survey on multimodal large language models. *National Science Review*, 11(12):nwae403.
- Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., Xu, S., Dai, H., Zhao, L., Mai, G., et al. (2024). Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*.