

# Assistente Virtual Inteligente para Instituições Educacionais: Aplicação da Técnica RAG no IFBot

Wesley Folly Volotão de Souza<sup>1</sup>, Ianne Lima Nogueira<sup>1</sup>, Anderson Veiga de Souza<sup>1</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia Fluminense – campus Bom Jesus do Itabapoana (IFF)

CEP 28360-000 – Bom Jesus do Itabapoana – RJ – Brasil

wesley.souza@iff.edu.br, ianne.nogueira@iff.edu.br,  
anderson.silva@iff.edu.br

**Abstract.** *This paper presents the IFBot, an artificial intelligence chatbot developed to optimize access to institutional information in an educational institution. The solution uses the RAG (Retrieval-Augmented Generation) approach, combining web scraping and analysis of public documents to provide automated and reliable responses. As a preliminary result, the tool is expected to reduce face-to-face service demands, with planned evaluations to measure response accuracy and user satisfaction. Additionally, social and cultural impacts, such as digital inclusion and ethical data handling, are discussed.*

**Resumo.** *Este trabalho apresenta o IFBot, um chatbot de inteligência artificial desenvolvido para otimizar o acesso a informações institucionais em uma instituição de ensino. A solução utiliza a abordagem RAG (Retrieval-Augmented Generation), combinando web scraping e análise de documentos públicos para fornecer respostas automatizadas e confiáveis. Como resultado preliminar, espera-se que a ferramenta reduza demandas presenciais, com testes previstos para avaliar a taxa de resposta correta e o grau de satisfação dos usuários. Além disso, são discutidos impactos sociais e culturais, como inclusão digital e tratamento ético de dados sensíveis.*

## 1. Introdução

O avanço da inteligência artificial (IA) tem proporcionado novas soluções para otimizar o acesso à informação e aprimorar processos administrativos em diversas instituições. No contexto educacional, a aplicação de IA vem sendo explorada para melhorar a eficiência no atendimento a alunos e servidores, automatizando tarefas repetitivas e reduzindo a carga administrativa (McAfee & Brynjolfsson, 2017). A agilidade e precisão na obtenção de informações são fundamentais para o bom funcionamento da comunidade acadêmica, garantindo que os usuários tenham acesso rápido a dados institucionais relevantes. Nesse contexto, a inteligência artificial tem se mostrado uma ferramenta poderosa, pois, como destacam Chen, Chen e Lin (2020), as tecnologias de IA têm sido amplamente adotadas em instituições educacionais para automatizar tarefas administrativas, melhorar a eficiência e permitir maior foco dos educadores na interação com os alunos.

Diante desse cenário, este trabalho apresenta o IFBot, um assistente virtual inteligente desenvolvido para facilitar o acesso a dados institucionais de forma

automatizada e confiável. O IFBot utiliza a abordagem *Retrieval-Augmented Generation* (RAG), combinando *web scraping* e análise de documentos públicos para fornecer respostas automatizadas e contextuais. Essa abordagem é baseada na combinação de recuperação de informações e geração de texto, permitindo que o assistente recupere informações relevantes e as apresente de maneira clara e precisa, conforme proposto por Lewis et al. (2020).

Além de otimizar o atendimento administrativo, o IFBot promove a inclusão digital e a democratização da informação, garantindo que alunos e servidores obtenham respostas rápidas e precisas sobre processos institucionais. De acordo com Winkler & Söllner (2018), ferramentas baseadas em IA, como o IFBot, têm demonstrado grande potencial para tornar o acesso à informação mais equitativo, especialmente em ambientes educacionais. No âmbito social e sustentável, a solução contribui para a modernização do serviço público, otimizando o tempo e os recursos humanos da instituição, reduzindo a necessidade de papel e proporcionando um modelo de atendimento escalável e eficiente.

Outro aspecto relevante é o tratamento ético de dados sensíveis. O IFBot foi desenvolvido em conformidade com a Lei Geral de Proteção de Dados (LGPD), garantindo que as informações dos usuários sejam protegidas e que o uso da ferramenta esteja alinhado a práticas responsáveis. O controle de acesso e a anonimização de interações são medidas adotadas para assegurar a privacidade e segurança dos usuários.

Dessa forma, o IFBot se destaca como uma solução tecnológica inovadora para melhorar a gestão educacional, promovendo inclusão, acessibilidade e sustentabilidade no ambiente acadêmico, enquanto respeita os princípios éticos e legais relacionados ao uso de dados.

## **2. Justificativa**

A crescente demanda por informações institucionais e a necessidade de otimização dos serviços administrativos tornam essencial a adoção de soluções tecnológicas que aprimorem esses processos. Atualmente, grande parte das dúvidas e solicitações acadêmicas e administrativas exige a consulta manual a documentos institucionais ou a interação direta com setores específicos, o que gera sobrecarga nos atendimentos e dificulta o acesso ágil à informação. Para McAfee & Brynjolfsson (2017), no contexto da transformação digital, tecnologias baseadas em inteligência artificial e plataformas digitais têm o potencial de automatizar processos e reduzir essa sobrecarga. Ferramentas como o IFBot são exemplos práticos desse potencial, proporcionando atendimento ágil e equitativo aos usuários e otimizando o tempo dos servidores (Winkler & Söllner, 2018).

Diante desse cenário, a implementação do IFBot, um assistente virtual baseado na abordagem RAG, alimentado por *web scraping* e documentos públicos, se justifica como uma solução inovadora para otimizar o acesso à informação. De acordo com Lewis et al. (2020), a abordagem RAG combina recuperação de informações e geração de texto, permitindo que o IFBot forneça respostas rápidas e confiáveis, garantindo maior acessibilidade e escalabilidade dos serviços.

Ao automatizar a recuperação de informações institucionais, o IFBot reduz a carga de trabalho dos setores administrativos, permitindo que os profissionais se concentrem em tarefas mais estratégicas. Segundo Zhang, Følstad e Bjørkli (2021), a implementação bem-sucedida de chatbots depende de fatores organizacionais, como

gestão de recursos, capacitação da equipe e estabelecimento de métricas de desempenho. Além disso, o uso do IFBot promove inclusão digital, proporcionando uma interface acessível que atende a diferentes perfis de usuários, independentemente de suas habilidades tecnológicas.

No contexto educacional, o IFBot democratiza o acesso à informação e moderniza a gestão acadêmica, garantindo que alunos e servidores obtenham respostas rápidas e precisas. Outro aspecto fundamental é o compromisso com o tratamento ético de dados sensíveis, em conformidade com a LGPD. O IFBot foi projetado para garantir a privacidade dos usuários, utilizando técnicas de anonimização e controle de acesso, assegurando que o uso da ferramenta esteja alinhado às melhores práticas de segurança e proteção de dados.

Por fim, o desenvolvimento do IFBot oferece uma oportunidade educacional significativa para os alunos envolvidos no projeto. Ao participarem da criação e implementação de uma solução tecnológica avançada, os alunos são inseridos na vanguarda do mercado de trabalho, adquirindo competências essenciais em inteligência artificial, desenvolvimento de software e ética digital. Dessa forma, o IFBot não apenas atende a uma demanda institucional, mas também capacita alunos e servidores no desenvolvimento de tecnologias emergentes, preparando-os para os desafios da transformação digital.

### **3. Fundamentação Teórica**

O desenvolvimento de assistentes virtuais baseados em inteligência artificial (IA) tem avançado significativamente nos últimos anos, impulsionado pelo aprimoramento de modelos de Processamento de Linguagem Natural (PLN) e pela introdução de arquiteturas híbridas, como a RAG. Segundo Lewis et al. (2020), a abordagem RAG combina recuperação de informações e geração de texto, permitindo que os assistentes virtuais acessem dados relevantes e os transformem em respostas contextuais. No contexto deste trabalho, o IFBot é um exemplo prático dessa tecnologia, utilizando técnicas de recuperação de informações e geração de linguagem natural para fornecer respostas rápidas, precisas e contextualizadas.

Além disso, os avanços em modelos de linguagem baseados em redes neurais profundas, como o GPT-3, ampliam significativamente a capacidade de compreensão e interação dos assistentes virtuais (Brown et al., 2020). Esses modelos são capazes de interpretar o contexto, adaptar respostas e fornecer informações com alto nível de precisão, tornando o IFBot uma ferramenta mais eficaz e interativa.

No contexto educacional e administrativo, a automação proporcionada pelo IFBot está alinhada às tendências globais de transformação digital. De acordo com Russell e Norvig (2020), os algoritmos inteligentes desempenham um papel fundamental na otimização de processos, permitindo que tarefas repetitivas sejam automatizadas, reduzindo a carga de trabalho humano e garantindo acesso ágil e preciso às informações. Ao adotar essa abordagem, o IFBot não apenas melhora a eficiência operacional, mas também promove uma experiência mais acessível e dinâmica para alunos e servidores.

Adicionalmente, o uso de técnicas de web *scraping* para extração automática de informações públicas é uma característica essencial do IFBot. Essa abordagem permite que o chatbot mantenha sua base de conhecimento sempre atualizada, garantindo

respostas precisas e relevantes. Conforme destacado por Khder (2021), o uso de *web scraping* deve ser conduzido de forma ética e responsável, garantindo que apenas informações públicas sejam acessadas e que a privacidade dos usuários seja preservada, em conformidade com a LGPD.

A ética no tratamento de dados é um pilar fundamental do IFBot, que adota práticas de anonimização e controle de acesso para proteger as informações dos usuários. Essas medidas asseguram que o sistema opere de forma segura e transparente, garantindo que o uso da tecnologia esteja alinhado às melhores práticas de segurança da informação.

Dessa forma, a fundamentação teórica deste projeto se baseia na interseção entre PLN, recuperação de informações, geração de respostas e ética no tratamento de dados. O IFBot se apresenta como uma solução inovadora e escalável, que não apenas otimiza o acesso a dados institucionais, mas também promove uma experiência de usuário segura, eficiente e em conformidade com as melhores práticas de proteção de dados.

#### **4. Objetivos**

O objetivo deste trabalho é descrever o desenvolvimento do IFBot, um assistente virtual inteligente baseado em técnicas de PLN e RAG. O IFBot é projetado para fornecer respostas automatizadas e confiáveis a alunos e servidores de uma instituição de ensino, otimizando o acesso à informação e reduzindo a sobrecarga administrativa.

Para isso, o trabalho contempla o processo de coleta e estruturação de dados institucionais por meio de técnicas de *web scraping* e análise de documentos públicos, garantindo que a base de dados utilizada seja confiável e atualizada. A implementação da abordagem RAG permite que o IFBot combine recuperação de informações e geração de respostas contextuais, assegurando precisão e coerência no atendimento aos usuários. Adicionalmente, é desenvolvido um ambiente de interação acessível, permitindo que alunos e servidores utilizem o IFBot de forma intuitiva em múltiplos canais de comunicação, como plataformas web e aplicativos de mensagens.

Outro aspecto fundamental é o compromisso com a segurança e privacidade dos dados, em conformidade com a LGPD. O IFBot é projetado para garantir que as informações dos usuários sejam protegidas, com medidas de anonimização e controle de acesso.

Ao apresentar o desenvolvimento do IFBot, este trabalho visa discutir os benefícios esperados da ferramenta, como inclusão digital, democratização da informação e otimização dos serviços administrativos na instituição de ensino.

#### **5. Metodologia**

A metodologia adotada para o desenvolvimento do IFBot segue uma abordagem sistemática, estruturada em etapas que garantem a precisão e eficiência na criação do assistente virtual. O IFBot é projetado para otimizar o acesso a informações institucionais, combinando técnicas de PLN e RAG.

##### **5.1. Materiais e Tecnologias**

A execução do projeto utiliza o uso de diversas ferramentas e tecnologias, incluindo:

- Linguagens de Programação: Python, que será utilizado para a manipulação de dados, treinamento da IA e desenvolvimento da interface.
- Bibliotecas e Frameworks:
  - *Scrapy* e *BeautifulSoup* para web *scraping* e extração de dados de documentos públicos.
  - *NLTK* e *spaCy* para o Processamento de Linguagem Natural (PLN).
  - *Hugging Face Transformers* para implementação do modelo de RAG (*Retrieval-Augmented Generation*).
  - *FAISS* (Facebook AI Similarity Search) para indexação vetorial e recuperação eficiente de informações.
  - *Flask* ou *FastAPI* para criação da API de interação do chatbot.
- Banco de Dados: *PostgreSQL* ou *MongoDB* para o armazenamento e recuperação dos documentos processados.
- Plataformas de Comunicação: Web, Telegram e WhatsApp para a interação dos usuários com o chatbot.

## 5.2. Etapas da Metodologia

O desenvolvimento do IFBot será conduzido através de seis etapas principais, conforme descrito a seguir:

### 1. Coleta e Estruturação de Dados

A primeira etapa envolve a coleta de dados institucionais. A extração de documentos públicos será realizada utilizando técnicas de web *scraping*, com o auxílio das bibliotecas *Scrapy* e *BeautifulSoup*. Além disso, arquivos institucionais serão coletados diretamente de fontes oficiais. Os dados extraídos serão estruturados em formatos padronizados (JSON, CSV) e armazenados em um banco de dados adequado (*PostgreSQL* ou *MongoDB*), o que permitirá uma recuperação eficiente posterior.

### 2. Processamento e Indexação das Informações

Após a coleta, será aplicado PLN para a segmentação e remoção de ruídos, bem como para a análise semântica dos textos. Para garantir uma recuperação eficiente das informações com base na similaridade semântica, será implementada uma indexação vetorial utilizando *FAISS*. Este processo assegura que as informações recuperadas sejam relevantes para as consultas realizadas pelos usuários.

### 3. Treinamento do Chatbot com RAG

O modelo de RAG será treinado para otimizar a geração de respostas a partir da recuperação de documentos relevantes. Esse modelo combinará a recuperação de informações com a geração de respostas em linguagem natural, aumentando a precisão e contextualização das respostas fornecidas. Haverá também a realização de testes de refinamento, com ajustes nos hiperparâmetros do modelo, visando aprimorar sua performance.

#### 4. Desenvolvimento da Interface e Integração

Será desenvolvida uma interface interativa que permita aos usuários interagir com o IFBot de maneira intuitiva. A comunicação entre o chatbot e os usuários será viabilizada por meio da criação de uma API utilizando *FastAPI* ou *Flask*. O IFBot será integrado a diversas plataformas de comunicação, como Telegram, WhatsApp e o portal web da instituição, ampliando a acessibilidade ao serviço e atendendo a diferentes preferências de uso dos usuários.

#### 5. Testes e Validação

Após o desenvolvimento, o IFBot passará por uma série de testes unitários e testes de desempenho para garantir a precisão das respostas fornecidas. A validação será realizada com um grupo de usuários internos da instituição, que fornecerão feedbacks importantes para ajustes e melhorias. Um mecanismo de feedback contínuo também será implementado, permitindo que os usuários reportem inconsistências ou sugiram melhorias nas respostas geradas pelo chatbot.

#### 6. Implantação e Monitoramento Contínuo

Após a fase de validação, o IFBot será disponibilizado oficialmente para toda a comunidade acadêmica. Para garantir o bom desempenho contínuo do sistema, será implementado um sistema de monitoramento contínuo, que avalia métricas como a taxa de uso, precisão das respostas e satisfação dos usuários. Com base nos dados coletados, serão realizados aprimoramentos periódicos no sistema para otimizar sua eficiência e atender de forma mais eficaz às necessidades dos usuários.

### 5.3. Compromisso com a Segurança e Privacidade

O desenvolvimento do IFBot é realizado em conformidade com a LGPD. As informações dos usuários são protegidas por meio de técnicas de anonimização e controle de acesso, garantindo que o uso da ferramenta esteja alinhado às melhores práticas de proteção de dados.

### 6. Considerações Sociais, Éticas e Culturais

O desenvolvimento e uso do IFBot têm implicações que vão além da otimização administrativa. Em termos sociais, o IFBot pode promover inclusão digital, permitindo que alunos e servidores com diferentes níveis de familiaridade tecnológica acessem informações institucionais de maneira prática. No entanto, é necessário garantir que a ferramenta seja clara e acessível, evitando exclusão digital.

No aspecto ético, o IFBot será desenvolvido em conformidade com a LGPD, garantindo que apenas informações públicas sejam utilizadas como base de conhecimento e que nenhuma interação dos usuários seja armazenada de forma identificável. As medidas adotadas incluem anonimização de dados e controle de acesso, assegurando que a privacidade dos usuários seja protegida.

Culturalmente, o IFBot tem o potencial de transformar a forma como alunos e servidores interagem com o ambiente acadêmico, promovendo uma gestão mais

moderna e acessível. No entanto, é essencial garantir que a automação não substitua totalmente o contato humano em situações que exijam suporte mais pessoal.

Esses cuidados reforçam o compromisso do projeto com a inclusão, segurança e ética na aplicação da inteligência artificial.

## **7. Resultados Esperados**

Com a implementação do projeto, espera-se que alunos e servidores tenham acesso rápido e preciso às informações institucionais por meio de interações naturais e intuitivas. O modelo de RAG garantirá que as respostas sejam baseadas em documentos oficiais, assegurando confiabilidade e atualização contínua. A integração do IFBot a múltiplos canais, como plataformas web e aplicativos de mensagens, permitirá uma ampla acessibilidade à ferramenta, facilitando o acesso à informação. Esse projeto não se limita a um único sistema de automação, mas também pode abrir portas para outras inovações tecnológicas, posicionando a instituição de maneira proativa na adoção de novas tecnologias.

Além de melhorar a eficiência administrativa, o IFBot promoverá benefícios significativos para a comunidade acadêmica. Alunos e servidores frequentemente enfrentam dificuldades para acessar documentos, regulamentos e orientações sobre procedimentos administrativos, resultando em retrabalho e sobrecarga de atendimentos presenciais e por e-mail. A automação proporcionada pelo IFBot contribuirá para a democratização da informação, promovendo a inclusão digital e facilitando o acesso ágil às diretrizes institucionais. Ao incentivar o uso de informações digitais por meio de IA, o projeto também reduzirá o desperdício de recursos físicos, como papel.

Outro benefício relevante será a acessibilidade para pessoas com deficiência visual. O IFBot poderá ser integrado a assistentes de voz e leitores de tela, garantindo um atendimento mais inclusivo e permitindo que todas as partes da comunidade acadêmica possam usufruir de seus serviços.

O IFBot será desenvolvido como um software de código aberto, permitindo que outras instituições possam adaptar e reutilizar a tecnologia conforme suas necessidades. O código-fonte será disponibilizado sob uma licença GPL (*General Public License*) ou MIT, o que possibilitará o aprimoramento contínuo por outras equipes. A base de conhecimento utilizada será composta exclusivamente por documentos públicos, em conformidade com as normas de transparência e garantindo a segurança em relação a dados sensíveis ou protegidos por sigilo institucional.

## **8. Conclusão**

Este trabalho apresentou o desenvolvimento inicial do IFBot, um assistente virtual inteligente baseado em inteligência artificial e na abordagem de RAG. O IFBot é projetado para otimizar o acesso a informações institucionais, promovendo inclusão digital e democratização da informação. Ao utilizar técnicas de web *scraping* e análise de documentos públicos, o assistente virtual oferece respostas automatizadas e confiáveis a alunos e servidores.

Os resultados esperados incluem o aprimoramento da eficiência administrativa, a redução da sobrecarga de atendimentos presenciais e a promoção da acessibilidade digital. O compromisso com a segurança e privacidade dos dados é assegurado pela

conformidade com a LGPD, garantindo que apenas informações públicas sejam utilizadas como base de conhecimento.

Além de sua aplicação imediata no atendimento institucional, o IFBot destaca-se como uma solução escalável, que pode ser adaptada para outras instituições educacionais, contribuindo para a modernização dos serviços acadêmicos e administrativos. Como software de código aberto, o IFBot poderá ser aprimorado continuamente por outras equipes, fortalecendo sua base tecnológica e garantindo a evolução da ferramenta.

Os próximos passos incluem o aperfeiçoamento do modelo de RAG, a expansão da base de conhecimento e o desenvolvimento de mecanismos avançados de segurança e acessibilidade. A validação com usuários reais também será fundamental para identificar melhorias e garantir que o IFBot atenda às necessidades da comunidade acadêmica.

Dessa forma, o IFBot tem o potencial de se consolidar como uma solução tecnológica inovadora para o atendimento acadêmico e administrativo, promovendo inclusão, segurança e eficiência no acesso à informação.

## Referências

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... and Amodei, D. (2020) “Language Models are Few-Shot Learners”. In *Advances in Neural Information Processing Systems*, 33, pages 1877–1901.
- Chen, L., Chen, P., and Lin, Z. (2020) “Artificial Intelligence in Education: A Review”. In *IEEE Access*, 8, pages 75264–75278.
- Khder, M. A. (2021) “Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application”. In *International Journal of Advances in Soft Computing and its Applications*, 13(3), pages 145–168.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020) “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. In *Advances in Neural Information Processing Systems*, 33, pages 9459–9474.
- McAfee, A., and Brynjolfsson, E. (2017) *Machine, Platform, Crowd: Harnessing Our Digital Future*, W. W. Norton & Company.
- Russell, S. J., and Norvig, P. (2020) *Artificial Intelligence: A Modern Approach (4th edition)*, Pearson.
- Winkler, R., and Söllner, M. (2018) “Unleashing the Potential of Chatbots in Education: A State-of-the-Art Analysis”, In *Academy of Management Proceedings*, 2018(1), page 15903. Briarcliff Manor, NY 10510: Academy of Management.
- Zhang, J. J. Y., Følstad, A., and Bjørkli, C. A. (2021) “Organizational Factors Affecting Successful Implementation of Chatbots for Customer Service”. In *Journal of Internet Commerce*, 22(7), pages 1–35.