

BioAutoML-API: Democratizando o Acesso à Inteligência Artificial na Biologia

Felipe César S. Rampazzo¹,
Octávio Onofre A. Souza¹,
Robson P. Bonidia¹,

¹Faculdade de Tecnologia de Ourinhos (Fatec)
Ourinhos - SP - Brasil

Abstract. *Technology has historically contributed to the development of human life. However, recent crises involving biological and global health issues have underscored the importance of accelerating a crucial process—its democratization. In this article, we will present a case study addressing the democratization of technology, with a special focus on the application of Machine Learning (ML) in biology. Throughout this work, we will highlight the significance of advances achieved through this relationship between technology and biology, emphasizing its impact on scientific research and underscoring the relevance of these advancements for society. Motivated by this, we have developed an API to enhance the pipeline of the BioAutoML application, a tool with end-to-end ML capabilities for the classification of biological sequences. Considering BioAutoML’s ability to optimize scientific research in the microbiology and biotechnology field, we identified the need to build a scalable and robust application capable of supporting the execution of multiple simultaneous processes, contributing to a more user-friendly and, consequently, more democratic environment. Our goal is to facilitate access and usage of tools like BioAutoML, encouraging the responsible development and deployment of ML-based solutions to create a more inclusive and egalitarian technological environment.*

Resumo. *A tecnologia tem contribuído historicamente com o desenvolvimento da vida humana. Contudo, os recentes episódios de crises com problemas biológicos e de saúde global, destacaram a importância de acelerar um processo importante, a sua democratização. Neste artigo, apresentaremos um estudo de caso que aborda a democratização da tecnologia, com foco especial na aplicação de Aprendizado de Máquina (AM) na área da biologia. Ao longo deste trabalho, destacaremos a significância dos avanços obtidos por meio dessa relação entre tecnologia e biologia, o impacto que isso traz à pesquisa científica, ressaltando a relevância destes avanços para a sociedade. Estimulados por isso, desenvolvemos uma API para aprimorar o pipeline da aplicação BioAutoML, uma ferramenta com recursos de AM fim-a-fim para classificação de sequências biológicas. Considerando a capacidade do BioAutoML de otimizar pesquisas científicas na área de microbiologia e biotecnologia, constatamos a necessidade de construir uma aplicação escalável e robusta o suficiente para suportar a execução de diversos processos simultâneos, contribuindo para um ambiente mais amigável e consequentemente mais democrático. Nosso objetivo é facilitar o acesso e o*

uso de ferramentas como o BioAutoML, incentivando o desenvolvimento e a disponibilização responsável de soluções baseadas em AM a fim de criar um ambiente tecnológico mais inclusivo e igualitário.

1. Introdução

Na atualidade, é difícil imaginar a vida sem recursos computacionais, já que a tecnologia é uma aliada importante na resolução de diversos problemas em uma ampla variedade de contextos [Ghazal et al. 2023]. Um exemplo é o uso da Inteligência Artificial (IA), que tem demonstrado ser uma ferramenta valiosa em diversas áreas, incluindo o diagnóstico de doenças como o Alzheimer [Li et al. 2021] e descobertas na área da oncologia, da microbiologia [AlQuraishi 2021, Littmann et al. 2020], epidemiologia [Holzinger et al. 2023, Daniel et al. 2023], dentre outras.

A IA compreende uma ampla variedade de campos, sendo o AM uma de suas subáreas, o qual se concentra na criação de algoritmos e modelos. O AM diz respeito à capacidade de sistemas computacionais processarem dados e aprimorarem seus resultados com base nos dados que foram processados anteriormente, gerando assim o que chamamos de aprendizado [Sharifani and Amini 2023]. Estas técnicas e ferramentas são profundamente presentes no nosso dia a dia, encontrando suas aplicações nos mais diversos cenários, como assistentes de voz e sistemas de controle de tráfego [Pizoñ and Gola 2023].

Apesar do fato de que plataformas que permitem acesso aos recursos de IA sejam populares atualmente, o conceito de democratização ainda é vago para muitos, pois a maioria das pessoas sequer sabe por onde começar a estudar sobre IA [Dwivedi et al. 2023]. Embora o assunto esteja muito fomentado, a especulação sobre este mercado persiste. Algumas opiniões dizem que os governos deveriam regulamentar o desenvolvimento de IA [Chamberlain 2023, Koniakou 2023], enquanto outros argumentam sobre uma abordagem global [Järvelä et al. 2023]. É exatamente neste ponto que chegamos às questões éticas do desenvolvimento deste tipo de ferramenta. De acordo com [Himmelreich 2023], *“The question is not so much whether AI should be democratized but how!”*.

Podemos especular sobre o potencial uso indevido destes recursos, mas de certa forma, esta preocupação geralmente se volta a fatores humanos que podem ser imprevisíveis e subjetivos [Matyashova 2023], quando na realidade, existem questões que necessitam ser discutidas antes mesmo das ferramentas baseadas em AM serem disponibilizadas [Balasubramaniam et al. 2022]. O crescimento exponencial no desenvolvimento de ferramentas baseadas em AM e IA fomenta o debate sobre as problemáticas a respeito do avanço desta área [Parmar et al. 2023]. Por trás de toda a tecnologia que vemos, é importante salientar que essas ferramentas são fundamentadas em modelos estatísticos e matemáticos aplicados usando recursos computacionais para gerar percepções utilizando técnicas específicas [He 2023].

Um algoritmo necessita de dados para otimizar seus resultados. A seleção de parâmetros e conjuntos de dados que a máquina irá interpretar não são apenas uma etapa, mas sim o produto de uma pré-análise minuciosa da necessidade de determinar características que a máquina utilizará para tomar decisões posteriores e se aprimorar [Letzgus et al. 2022]. Vale salientar que, muitas vezes, podemos nos deparar com

desafios éticos [Varma et al. 2023]. Por exemplo, em um algoritmo de seleção de currículos, a inclusão de atributos como gênero ou raça de um candidato pode levar a interpretações que segregariam candidatos com base em tais características [Medeiros et al. 2023]. Mesmo que a máquina consiga gerar resultados a partir destes dados, não significa que é ético ou justo fazê-lo [John-Mathews et al. 2022]. Portanto, é fundamental que nossa abordagem seja justa e totalmente orientada pelos dados [Gyory et al. 2022], ou seja, "*Data-Driven*".

A ascensão da área de AM e processamento de Big Data é recente, por mais que seu crescimento seja exponencial. Mesmo assim, o mercado da tecnologia é um dos mais fomentados e movimentados bilhões [Chowdhury et al. 2023]. As empresas que lideram este cenário estão sempre em busca de otimizar seus processos e resultados. A realidade é que muitas vezes os recursos computacionais necessários para estas aplicações [Enholm et al. 2022], assim como todo tipo de tecnologia emergente, são caros, além da necessidade de grande especialização técnica, o que tornou esse tipo de ferramenta menos acessível. Tais fatores fizeram com que, por muito tempo, o uso de recursos de ML não fosse tão popular quanto poderia, algo como um processo de "de-democratização".

O cenário resultante disso é o que temos hoje, grandes empresas que dominam o mercado de IA, muitas pesquisas em andamento utilizando estes recursos, e o aspecto que mais nos chama atenção, a especulação ao redor deste mercado [Bozkurt et al. 2023]. Por trás das especulações, pode-se notar a importância do debate sobre IA, que se dá, em partes, pela falta de informação, como é comum com qualquer coisa nova. No entanto, ao invés de considerar a desaceleração do processo de evolução tecnológica [Hupont et al. 2023], poderíamos fomentar cada vez mais o estudo do campo. Assim, com o auxílio da tecnologia, pode-se desenvolver soluções mais justas, éticas e responsáveis, e que também, colaborem para a criação de um ambiente tecnológico cada vez mais democrático e seguro [Prem 2023].

Considerando isso, este artigo visou democratizar o acesso a uma ferramenta criada e auditada em ambiente acadêmico, supervisionada pelo público interessado em utilizá-la na área da biologia, chamada BioAutoML [Bonidia et al. 2022]. O BioAutoML é uma ferramenta que realiza recomendação de pipelines robustos para representar sequências biológicas, automação da seleção de características e ajuste de hiperparâmetros. Contudo, por mais que seu uso não dependa de amplo conhecimento na área de AM, ainda são necessários conhecimentos computacionais avançados para executar os códigos-fonte. Por isso, a migração de uma ferramenta local para uma API (do inglês, *Application Programming Interface*) de fácil acesso e amplamente disponível aumenta o impacto deste tipo de tecnologia, acelerando a pesquisa, conseqüentemente, na descoberta de percepções científicas, contribuindo assim, para uma maior compreensão da área biológica. Ao mesmo tempo, nos desafia a considerar implicações e desafios éticos destas aplicações na sociedade.

2. Trabalhos Relacionados

Ao realizar uma revisão da literatura, constatamos que já existem algumas ferramentas relacionadas que propõem soluções semelhantes para a classificação de sequências biológicas utilizando recursos de AM. No entanto, a Figura 1 classifica

quatro ferramentas relacionadas com base na utilização de engenharia de recursos, algoritmos de AM, ajuste de hiperparâmetros e disponibilização de API.

FERRAMENTAS	ENGENHARIA DE RECURSOS	ALGORITMO DE AM	AJUSTE DE HIPERPARÂMETROS	FLUXO DE DADOS FIM A FIM	API
iLearn	-	V	V	-	-
iLearnPlus	-	V	V	-	-
autoBioSeqpy	-	V	V	-	-
AutoGenome	-	V	V	-	-
BioAutoML	V	V	V	V	V

Figura 1. Trabalhos Relacionados

Além de ser uma das primeiras ferramentas deste campo com um pipeline fim-a-fim, BioAutoML API obtém vantagem sobre as outras ferramentas. BioAutoML é uma ferramenta totalmente democrática, pois, ao usuário acessar a API e ler sua documentação, ele é capaz de executar seu processo com poucos cliques, ao contrário de ferramentas que necessitam ser utilizadas a partir do código fonte que requerem conhecimento de diversos recursos computacionais. Outro fator importante de ser citado é que ao disponibilizar uma API, o BioAutoML passa a ser mais acessível para um único usuário e também pode ser implementado a outras aplicações, podendo assim, ser utilizado em ambientes empresariais e acadêmicos da maneira com que venha a ser mais conveniente.

3. Materiais e Métodos

Conforme citado anteriormente, para o desenvolvimento da aplicação *BioAutoML-API*, consideramos a capacidade de processamento que o sistema na totalidade teria de oferecer. Fatores como segurança, simplicidade no uso, escalabilidade nortearam o processo de desenvolvimento. Conhecendo as características citadas acima, é possível diminuir a curva de aprendizado de uma pessoa para utilizar uma ferramenta de AM em seu computador local via terminal de comando, para a experiência de uma aplicação escalável online, capaz de suportar centenas de processos simultâneos. A Figura 2 apresenta o fluxo de trabalho do **BioAutoML-API**. Todos os processos detalhados na Figura são detalhados nas próximas seções.

3.1. Tecnologias Utilizadas

Para disponibilizar um ambiente robusto e assíncrono, optamos por utilizar uma estrutura de microsserviços para construir as aplicações que orquestram o pipeline de dados do BioAutoML. Um ambiente orientado a microsserviços nada mais é que uma grande aplicação subdividida em pequenas aplicações, cada uma com seu conjunto de funcionalidades específicas.

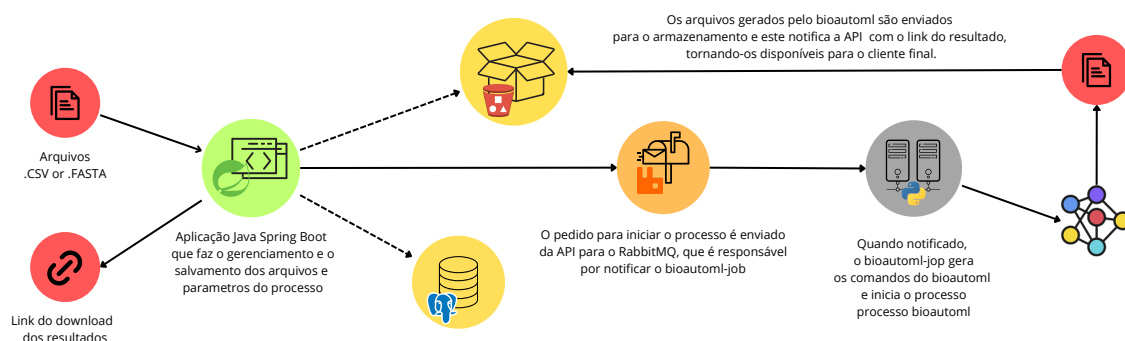


Figura 2. Fluxograma de trabalho do *BioAutoML-API*

- **Java Spring Boot:** O Spring Boot é um projeto feito para simplificar e acelerar o processo de desenvolvimento de aplicativos Java. Ele oferece uma plataforma para criar aplicativos Java robustos, incluindo aplicações Web. Está entre os *frameworks* mais usados do mercado e é um dos principais componentes do nosso sistema [Reddy 2017].
- **Cloud AMQP:** O CloudAMQP é um serviço de Mensageria (do inglês, *Messaging*) em nuvem que fornece mensagens em tempo real e em lote para aplicativos. Ele é construído em cima da tecnologia RabbitMQ. O CloudAMQP facilita a criação, o gerenciamento e a escalabilidade de filas de mensagens, o que é fundamental para a comunicação entre componentes de aplicativos distribuídos. É ele quem gerencia a fila de processos do *BioAutoML-API* [Johansson and Dossot 2020].
- **Rclone:** No geral, o Rclone é uma ferramenta de linha de comando multi-plataforma de código aberto, que permite a sincronização e transferência de dados entre sistemas de armazenamento em nuvem e sistemas locais. Ele é amplamente utilizado para gerenciar arquivos e diretórios em serviços de armazenamento em nuvem, como Google Drive, Dropbox, Amazon S3. No caso, ele é responsável por transferir os arquivos resultantes dos processos para a pasta de resultados [Padhy 2021].
- **PostgreSQL:** PostgreSQL é um sistema de gerenciamento de banco de dados relacional de código aberto. É uma solução muito utilizada para aplicações que exigem um banco de dados relacional robusto, seguro e altamente personalizável. Este, armazena os dados de cada processo realizado no sistema [Smith 2010].
- **ObjectStorage:** Um *Object Storage* é um tipo de serviço de armazenamento em nuvem que difere dos sistemas de armazenamento de arquivos tradicionais. Em vez de armazenar dados em uma estrutura hierárquica de pastas, o *Object Storage* organiza os dados em objetos individuais, identificados por chaves exclusivas. Sua flexibilidade, escalabilidade e durabilidade o tornam uma escolha popular para armazenar, acessar e distribuir grandes volumes de dados de forma confiável e eficiente [Odun-Ayo et al. 2017].

3.2. Sistema Híbrido

Como o BioAutoML foi desenvolvido em ambiente acadêmico, com um suporte computacional disponibilizado para o funcionamento do *Core* da ferramenta, optamos

por construir o sistema em um ambiente híbrido, que consiste em um sistema com componentes funcionando tanto em um sistema *cloud*, quanto componentes funcionando em um sistema físico. Vale ressaltar que é possível centralizar todo o sistema em um único ambiente *cloud* ou em um ambiente multi-*cloud*. Contudo, a disponibilidade de um ambiente físico por parte do BioAutoML com bons recursos computacionais ajuda na redução de custos da aplicação [Fisher et al. 2018]. Por exemplo:

- **Ambiente *cloud*:** Um ambiente *Cloud*, também conhecido como "*Cloud Computing*", é uma infraestrutura que fornece serviços, recursos computacionais e armazenamento de dados através da internet. Em vez de depender de servidores e bancos de dados locais, estes ambientes possibilitam provisionar e acessar recursos de computação por meio de provedores de serviços em nuvem, ou seja, na rede. Neste ambiente, o uso de um *framework* como Java Spring Boot fornece uma gama de ferramentas ótimas para construir serviços assíncronos independentes, como o principal microserviço da aplicação *BioAutoML-API*, a qual é responsável tanto pelas etapas de autenticação, envio de mensagens para a fila de processamento, quanto pelo armazenamento dos arquivos fornecidos pelo usuário no *object storage*. O *design-pattern* utilizado nessa aplicação é o *OutBox*. Basicamente, ele recebe os objetos do formulário do usuário, os dados de identificação do usuário, realiza um enriquecimento com os dados dos arquivos fasta ou CSV inseridos para que ele possa ter todas as informações e converter estes objetos em mensagens de texto base x64 antes de inseri-las na fila assíncrona.
- **Ambiente *On Premise*:** Um ambiente *On Premise* funciona como servidor físico local, ou data center local, onde se hospedam aplicações diversas. A disponibilidade de usar este ambiente potencializou a escalabilidade do sistema. Dentro deste ambiente temos um serviço Python conectado à fila de mensagens assíncronas. Este serviço é responsável por orquestrar quantas instâncias da *IA* serão executadas no ambiente, quantos processos serão executados por vez e qual o limite de processos nas filas. Ao receber novos processos nas filas, o orquestrador decodifica a mensagem enviada pela *OutBox* direciona o processo para uma das instâncias. Ao finalizar um dos processos, o orquestrador também é responsável por ser um *listener* da pasta de resultados de cada um dos processos e por fim, como há persistência no banco de dados *PostgreSQL* do ambiente *cloud*, é possível identificar cada um dos processos pelo seu id, atualizar seu *status* e enviar os arquivos resultantes para a pasta de resultados no *Object Storage* com o envio massivo de arquivos do *Rclone*.

4. Resultados

A eficiência da aplicação de AM na análise e processamento de dados biológicos é inegável, tendo em vista os avanços significativos ocorridos nos últimos anos. No entanto, a forma como o usuário pode acessar esse tipo de ferramenta desempenha um papel crucial na acessibilidade e aplicabilidade prática. No contexto da biologia computacional, a disponibilização do BioAutoML como uma API apresenta vantagens sobre a maioria das ferramentas que necessitam ser utilizadas a partir de seus códigos-fonte, o que dificulta o uso deste tipo de recurso por conta de demandar grande

capacidade de processamento e expertise computacional por parte do usuário final. Além disso, vale ressaltar que a aplicação online do BioAutoML garante um fluxo de dados fim-a-fim. Isso garante muitas vantagens para consistência, segurança de dados e também confere uma possibilidade de rastrear possíveis falhas com mais facilidade nos *logs* da aplicação. Para entender a importância da democratização dessa ferramenta, dois estudos de caso foram realizados, comparando o resultado do BioAutoML-API a outras ferramentas aplicadas para a classificação de peptídeos anticancerígenos [Yu et al. 2020] e RNAs não-codificadores longos [Han et al. 2019].

- **Estudo de caso 1:** Os peptídeos têm um importante papel em uma variedade de funções biológicas. Eles são conjuntos de aminoácidos que podem se agrupar de diversas formas. Existem peptídeos encontrados no sistema nervoso central dos organismos, desempenhando um papel importante na transmissão de sinais neurais. Com isso, a identificação deste tipo de estrutura nos permite compreender o comportamento, emoções e funções orgânicas de diversas formas de vida. Outros peptídeos são relacionados ao combate de doenças metabólicas como hipertireoidismo ou até mesmo doenças cardiovasculares. As aplicações dos estudos dessa área avançam em busca de soluções para problemas biológicos complexos e já obtiveram avanços significativos em doenças infecciosas como COVID e doenças neurodegenerativas como Alzheimer e Parkinson. Por mais que existam diversas classificações e tipos de peptídeos, neste estudo abordaremos um tipo específico: Peptídeos Anticancerígenos (ACPs). A identificação deste tipo de peptídeo nos possibilita atuar na inibição do crescimento e proliferação de células cancerígenas. A principal proposta da área é trazer novos tratamentos menos tóxicos e mais eficazes para o combate dessa doença.
- **Estudo de caso 2:** O segundo estudo de caso abordado é a respeito do uso de ML para identificação de RNAs não codificadores longos (lncRNAs). Os lncRNAs, são uma classe de moléculas de ácido ribonucleico (RNA) que não codificam proteínas, ao contrário dos RNAs mensageiros (mRNAs). Essas moléculas desempenham uma variedade de funções cruciais nas células e estão envolvidas em processos regulatórios complexos dentro do genoma. Apesar de sua incapacidade para serem traduzidos em proteínas, os lncRNAs são essenciais para a regulação da expressão gênica e para a modulação de diversos processos celulares. Além disso, eles são encontrados em diversas regiões do genoma. Mesmo que seja uma frente de pesquisa relativamente nova, cientistas descobriram a relação deste tipo de estrutura genética em funções como diferenciação celular, processos metabólicos e dinâmica cromossômica. As evidências dos estudos sugerem uma relação direta entre alguns tipos de lncRNA e doenças humanas complexas como doença de Alzheimer, doenças cardiovasculares, dentre outras. O estudo dos lncRNA têm um futuro promissor para compreender melhor o funcionamento de diversos processos bioquímicos e com isso trazer soluções biológicas mais eficazes para diversos problemas genéticos.

As métricas usadas para comparação dos resultados são:

- **Acurácia:** Representa a proporção de instâncias corretamente classificadas em relação ao total de instâncias.

- **Precisão:** Representa a proporção de instâncias positivas identificadas corretamente em relação ao total de instâncias classificadas como positivas pelo modelo.
- **Recall:** Indica a proporção de instâncias positivas que foram corretamente identificadas pelo modelo em relação ao total real de instâncias positivas.
- **Valor F:** É uma métrica que combina precisão e recall em uma única medida.

Como podemos observar nas Figuras 3 e 4, as métricas resultantes dos processos realizados pelo BioAutoML assemelham-se com as métricas das demais ferramentas do mercado em ambos os estudos de caso. Este fato por si só já faria do BioAutoML uma ferramenta competitiva no mercado, mas vale ressaltar que o BioAutoML além de classificar os dados e entregar as métricas, ele também entrega o resultado do seu módulo de engenharia de recursos, matriz de confusão, dentre outros dados que são de grande valia para o usuário final.

FERRAMENTA	ACC (%)	PRECISION (%)	RECALL (%)	FV(%)
iACP	74.4	0.778	68.3	72.2
CNN	78.6	0.790	78.0	78.5
ACPred-FL	88.4	0.957	80.5	87.4
DeepACP	82.7	0.657	80.4	82.3
BioAutoML	81.7	0.900	71.9	79.7

Figura 3. Comparação de ferramentas para classificação de ACPs

É importante dizer que ao realizar uma revisão da literatura, constatamos que grande parte das ferramentas relacionadas ao BioAutoML não disponibilizam um pipeline de dados fim-a-fim, sendo, geralmente, necessário executar os códigos fonte para utilizar as ferramentas em seu computador local. O **BioAutoML-API** destaca-se por toda sua estrutura computacional, desde o algoritmo de AM com seu módulo de engenharia de recursos e auto-aprendizado até a disponibilização destes recursos em uma aplicação *web* hospedada em um ambiente híbrido que funciona de maneira assíncrona e que é orquestrada por um serviço que pode gerar múltiplas instâncias do *core* da ferramenta e escalar sua capacidade de executar processos simultâneos. Também é importante citar que a estrutura da API armazena os metadados resultantes dos processos executados, informando ao usuário o status de processamento e informando caso haja algum problema durante a execução. Todos estes componentes unidos nos permitiram criar uma das ferramentas mais democráticas do mercado. Ainda que hajam outras ferramentas que tenham um fluxo

FERRAMENTA	ACC (%)	PRECISION (%)	RECALL (%)	FV(%)
CPAT (Default Model)	0.9600	NA	NA	0.9592
CNCI	0.9450	NA	NA	0.9465
PLEK (Default Model)	0.9410	NA	NA	0.9441
LncFinder	0.9728	NA	NA	0.9726
BioAutoML	0.9708	0.9594	0.9832	0.9711

Figura 4. Comparação de ferramentas para classificação de lncRNAs

de dados fim-a-fim ou que usem engenharia automatizada de recursos, não existem ferramentas que agrupam todas estas funcionalidades e vantagens tecnológicas em uma só aplicação que seja disponibilizada de forma democrática, ética e gratuita. Isso nos mostra que o **BioAutoML-API** tem total condições de cumprir o papel para o qual foi desenvolvido: democratizar o conhecimento para acelerar o desenvolvimento de soluções biológicas para diversos tipos de problemas.

5. Conclusão

O desenvolvimento do **BioAutoML-API** desempenha um importante papel social ao proporcionar acesso a recursos de AM na área da biologia. Essa solução, além de otimizar uma série de processos e reduzir a complexidade de uso, também amplia a escalabilidade de aplicação da ferramenta, diminuindo a distância entre a aplicação e o usuário final. Vale ressaltar que o fato da aplicação estar hospedada em um ambiente híbrido, nos permite manter os dados em nuvem com alta disponibilidade. Essa característica, aliada a assincronicidade na coleta dos resultados dos processos, uma dinâmica melhor para o dia-a-dia dos usuários e mais conforto. No entanto, é importante examinar aspectos relacionados à privacidade e segurança de dados, além de abordar a responsabilidade científica, viés algorítmico e os demais desafios éticos envolvidos para garantir que essa ferramenta impacte a comunidade científica de forma responsável e eficaz. A disponibilização destes recursos não só representa um passo significativo em direção a um ambiente científico mais tecnológico e inovador, mas também cria um ambiente mais democrático, no qual pesquisadores com níveis de experiência interesses distintos podem usufruir dos mesmos recursos.

Referências

AlQuraishi, M. (2021). Machine learning in protein structure prediction. *Current opinion in chemical biology*, 65:1–8.

- Balasubramaniam, N., Kauppinen, M., Hiekkänen, K., and Kujala, S. (2022). Transparency and explainability of ai systems: ethical guidelines in practice. In *International Working Conference on Requirements Engineering: Foundation for Software Quality*, pages 3–18. Springer.
- Bonidia, R. P., Santos, A. P. A., de Almeida, B. L. S., Stadler, P. F., da Rocha, U. N., Sanches, D. S., and de Carvalho, A. C. P. L. F. (2022). BioAutoML: automated feature engineering and metalearning to predict noncoding RNAs in bacteria. *Briefings in Bioinformatics*, 23(4):bbac218.
- Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., et al. (2023). Speculative futures on chatgpt and generative artificial intelligence (ai): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, 18(1).
- Chamberlain, J. (2023). The risk-based approach of the european union’s proposed artificial intelligence regulation: Some comments from a tort law perspective. *European Journal of Risk Regulation*, 14(1):1–13.
- Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., and Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through ai capability framework. *Human Resource Management Review*, 33(1):100899.
- Daniel, Cenggoro, T. W., and Pardamean, B. (2023). A systematic literature review of machine learning application in covid-19 medical image classification. *Procedia Computer Science*, 216:749–756. 7th International Conference on Computer Science and Computational Intelligence 2022.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koochang, A., Raghavan, V., Ahuja, M., et al. (2023). “so what if chatgpt wrote it?” multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management*, 71:102642.
- Enhölm, I. M., Papagiannidis, E., Mikalef, P., and Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, 24(5):1709–1734.
- Fisher, C. et al. (2018). Cloud versus on-premise computing. *American Journal of Industrial and Business Management*, 8(09):1991.
- Ghazal, T. M., Hasan, M. K., Ahmad, M., Alzoubi, H. M., and Alshurideh, M. (2023). Machine learning approaches for sustainable cities using internet of things. In *The Effect of Information Technology on Business and Marketing Intelligence Systems*, pages 1969–1986. Springer.
- Gyory, J. T., Soria Zurita, N. F., Martin, J., Balon, C., McComb, C., Kotovsky, K., and Cagan, J. (2022). Human versus artificial intelligence: A data-driven approach to real-time process management during complex engineering design. *Journal of Mechanical Design*, 144(2):021405.
- Han, S., Liang, Y., Ma, Q., Xu, Y., Zhang, Y., Du, W., Wang, C., and Li, Y. (2019). Lncfinder: an integrated platform for long non-coding rna identification utili-

- zing sequence intrinsic composition, structural information and physicochemical property. *Briefings in bioinformatics*, 20(6):2009–2027.
- He, Y.-H. (2023). Machine-learning mathematical structures. *International Journal of Data Science in the Mathematical Sciences*, 1(01):23–47.
- Himmelreich, J. (2023). Against “democratizing ai”. *AI & SOCIETY*, 38(4):1333–1346.
- Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., and Müller, H. (2023). Ai for life: Trends in artificial intelligence for biotechnology. *New Biotechnology*, 74:16–24.
- Hupont, I., Micheli, M., Delipetrev, B., Gómez, E., and Garrido, J. S. (2023). Documenting high-risk ai: A european regulatory perspective. *Computer*, 56(5):18–27.
- Järvelä, S., Nguyen, A., and Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*.
- Johansson, L. and Dossot, D. (2020). *RabbitMQ Essentials: Build distributed and scalable applications with message queuing using RabbitMQ*. Packt Publishing Ltd.
- John-Mathews, J.-M., Cardon, D., and Balagué, C. (2022). From reality to world. a critical perspective on ai fairness. *Journal of Business Ethics*, 178(4):945–959.
- Koniakou, V. (2023). From the “rush to ethics” to the “race for governance” in artificial intelligence. *Information Systems Frontiers*, 25(1):71–102.
- Letzgus, S., Wagner, P., Lederer, J., Samek, W., Müller, K.-R., and Montavon, G. (2022). Toward explainable artificial intelligence for regression models: A methodological perspective. *IEEE Signal Processing Magazine*, 39(4):40–58.
- Li, Z., Jiang, X., Wang, Y., and Kim, Y. (2021). Applied machine learning in alzheimer’s disease research: omics, imaging, and clinical data. *Emerging topics in life sciences*, 5(6):765–777.
- Littmann, M., Selig, K., Cohen-Lavi, L., Frank, Y., Hönigschmid, P., Kataka, E., Mösch, A., Qian, K., Ron, A., Schmid, S., Sorbie, A., Szlak, L., Dagan-Wiener, A., Ben-Tal, N., Niv, M. Y., Razansky, D., Schuller, B. W., Ankerst, D., Hertz, T., and Rost, B. (2020). Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nature Machine Intelligence*, 2(1):18–24.
- Matyashova, D. (2023). Germany: Rising sociopolitical controversies and threats to psychological security from the malicious use of artificial intelligence. In *The Palgrave Handbook of Malicious Use of AI and Psychological Security*, pages 487–505. Springer.
- Medeiros, A. C. B., Franciscan, J. F. V., Longo, J. B. P., de Moraes, M. V. R., Macedo, M. d. O., and Bonidia, R. P. (2023). Grace: Sistema de recomendação de currículos com inteligência artificial. In *Anais Estendidos do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 8–14. SBC.

- Odun-Ayo, I., Ajayi, O., Akanle, B., and Ahuja, R. (2017). An overview of data storage in cloud computing. In *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, pages 29–34. IEEE.
- Padhy, N. (2021). An automation api to optimize the rate of transmission using rclone from local system to cloud storage environment. *Materials Today: Proceedings*, 37:2462–2466. International Conference on Newer Trends and Innovation in Mechanical Engineering: Materials Science.
- Parmar, J., Chouhan, S., Raychoudhury, V., and Rathore, S. (2023). Open-world machine learning: Applications, challenges, and opportunities. *ACM Comput. Surv.*, 55(10).
- Pizoń, J. and Gola, A. (2023). Human–machine relationship—perspective and future roadmap for industry 5.0 solutions. *Machines*, 11(2):203.
- Prem, E. (2023). From ethical ai frameworks to tools: a review of approaches. *AI and Ethics*, pages 1–18.
- Reddy, K. S. P. (2017). *Beginning Spring Boot 2: Applications and microservices with the Spring framework*. Apress.
- Sharifani, K. and Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, 10(07):3897–3904.
- Smith, G. (2010). *PostgreSQL 9.0: High Performance*. Packt Publishing Ltd.
- Varma, A., Dawkins, C., and Chaudhuri, K. (2023). Artificial intelligence and people management: A critical assessment through the ethical lens. *Human Resource Management Review*, 33(1):100923.
- Yu, L., Jing, R., Liu, F., Luo, J., and Li, Y. (2020). Deepacp: A novel computational approach for accurate identification of anticancer peptides by deep learning algorithm. *Molecular Therapy - Nucleic Acids*, 22:862–870.