

FORMULÁRIO DE DESCRIÇÃO DE VAGA PARA DIVULGAÇÃO E PROCESSO SELETIVO	
Lead Institution: USP at São Paulo	
Address: CEP: - São Paulo - SP	
Supervisor name: Prof ^a Dr ^a Giselle Beiguelman	
Type: (TT-4A) Bolsista TT-4A para Infraestrutura Computacional – REF 23TT4-005	
Project title: Inteligência Artificial e Arquivos: Contribuições à História da Arte e às Humanidades Digitais	
Research Group: C4AI	
<p>Description (Portuguese and English)</p> <p>Cientista de Dados com ênfase em Inteligência Artificial: responsável pelo desenvolvimento dos scripts de programação do projeto e sua articulação às diferentes redes que alimentarão o acervo dinâmico de arte, arquitetura e design, fruto da pesquisa desenvolvida no Projeto Temático Acervos Digitais e Pesquisa: arte, arquitetura e design.</p> <p>No âmbito do Projeto Temático Acervos Digitais e Pesquisa, este bolsista investigará métodos de classificação e catalogação de obras artísticas, arquitetônicas e de design com recursos de Inteligência Artificial, buscando compreender as novas possibilidades de arranjos e organização do conhecimento histórico pelas IAs e contribuindo para a discussão sobre métodos baseados em visão computacional na formação de um discurso crítico no campo das Humanidades Digitais.</p> <p>Plano de atividades:</p> <p>O bolsista terá as seguintes funções no projeto ao longo de 24 meses:</p> <ul style="list-style-type: none"> • Coleta de dados: construir scripts para busca e recuperação da informação em APIs, incluindo processamento de imagem para análise de grandes coleções de vídeos ou imagens estáticas. • REST de mídias sociais com maior ênfase em ambiente Wikidata e Wikimedia Commons; • Tratamento dos dados: construir scripts para normalização, limpeza e transformação dos dados e modelagem de banco de dados para armazenamento dos dados coletados; • Modelos de Aprendizagem de Máquina: modelagem, customização, parametrização e disponibilização em ambiente de produção de modelos de aprendizagem de máquina, com maior ênfase em modelos deep learning e eventualmente LLMs em tecnologia universo Python (TensorFlow e Pytorch). A depender das demandas de experimentação do projeto, modelos mais simples providos por bibliotecas como Sci-kit Learn e Sci-kit Imagem serão utilizados. • Carga de dados: interface com API-REST do software livre WordPress para entrega dos dados processados e consumo por aplicação de gestão da informação de acervos digitais Tainacan (https://wordpress.org/plugins/tainacan/). <p>Data Scientist with a focus on Artificial Intelligence: responsible for developing programming scripts for the project and their integration into the various networks that will feed the dynamic collection of art, architecture, and design, resulting from the research carried out in the Thematic Project Digital Collections and Research: art, architecture, and design.</p> <p>Within the scope of the Thematic Project Digital Collections and Research, this scholar will investigate methods of classification and cataloging of artistic, architectural, and design works using Artificial Intelligence resources, aiming to understand the new possibilities for arrangements and organization of historical knowledge by AIs and contributing to the discussion on computer vision-based methods in the formation of critical discourse in the field of Digital Humanities.</p>	

Activity Plan:

The scholar will have the following responsibilities in the project over a 24-month period:

- Data collection: Building scripts for information retrieval and processing in APIs, including image processing for the analysis of large collections of videos or static images.
- REST of social media with a greater emphasis on the Wikidata and Wikimedia Commons environments.
- Machine Learning Models: modeling, customization, parameterization, and deployment of machine learning models in a production environment, with a greater emphasis on deep learning models and potentially LLMs in the Python technology universe (TensorFlow and Pytorch). Depending on the project's experimentation demands, simpler models provided by libraries like Sci-kit Learn and Sci-kit Image will be used.
- Data loading: interface with the open-source WordPress REST API for delivering processed data and consumption by the information management application of digital collection Tainacan (<https://wordpress.org/plugins/tainacan/>).

Requirements to fill the position. (Ex: specific experience, minimum or maximum years after concluding the course) (Portuguese and English)

Esta bolsa se destina a candidatos graduados, com no mínimo quatro anos de experiência em pesquisa ou experiência profissional após a graduação, em área relacionada ao Plano de Atividades proposto;

O perfil do bolsista envolve alguém com formação de graduação nas áreas de Ciência da Computação, Engenharia da Computação, Engenharia Elétrica, Engenharia de Software, Análise e Desenvolvimento de Sistemas ou Sistemas de Informação. Deve ter experiência profissional em projetos de Ciência de Dados, envolvendo as etapas de coleta, tratamento e análise de dados. É fundamental conhecimento técnico e experiência em APIs REST para coleta de dados, técnicas de pré-processamento de dados, tais como normalização, transformação e limpeza de dados e experiência em treinamento e produção com algoritmos de aprendizagem de máquina.

Exigido:

- Github, Gitlab, Python (nível avançado), SQL, MySQL/MariaDB, Pandas, Sci-kit Learn
- Familiaridade com Visão Computacional aplicada a cervos de vídeos e imagens

Desejável:

- Sci-kit Image, TensorFlow, Keras, Pytorch e ElasticSearch

This scholarship is intended for candidates with a bachelor's degree and a minimum of four years of research or professional experience after graduation in a field related to the proposed Activity Plan.

The ideal candidate profile includes someone with an undergraduate degree in Computer Science, Computer Engineering, Electrical Engineering, Software Engineering, Systems Analysis and Development, or Information Systems. They should have professional experience in Data Science projects, involving data collection, preprocessing, and analysis stages. Technical knowledge and experience in REST APIs for data collection, data preprocessing techniques such as normalization, transformation, and cleaning, as well as experience in training and production with machine learning algorithms, are essential.

Required:

- Github, Gitlab, Python (nível avançado), SQL, MySQL/MariaDB, Pandas, Sci-kit Learn
- Familiarity with Computer Vision applied to video and image datasets

Desirable:

- Sci-kit Image, TensorFlow, Keras, Pytorch e ElasticSearch

Informações sobre a bolsa / local de trabalho

A bolsa TT-4A será concedida pela Fapesp por um período de 24 meses.

Local de trabalho: Universidade de São Paulo Campus da Capital (FAUUSP ou no INOVA-USP)

Características e valores: Presencial, 40h semanais, 24 meses, R\$ 6.242,40

Endereço: Av. Prof. Lucio Martins Rodrigues 370 – Cidade Universitária, São Paulo, SP

Scholarship/place of work information

The TT-4A scholarship will be awarded by Fapesp for a period of 24 months.

Workplace: University of São Paulo Campus da Capital (FAUUSP or INOVA-USP)

Characteristics and values: On-site, 40 hours per week, 24 months, R\$ 6,242.40

Address: Av. Prof. Lucio Martins Rodrigues 370 - Cidade Universitária, São Paulo, SP