

## Ficha de Curso

### Anotação de Dados para Modelos de Inteligência Artificial no Judiciário

|   |                                 |
|---|---------------------------------|
| <b>Curso:</b> Anotação de Dados para Modelos de Inteligência Artificial no Judiciário |                                 |
| <b>Modalidade:</b> Ensino a distância – Autoinstrucional                              |                                 |
| <b>Carga horária total:</b> 10h   | <b>Carga horária semanal:</b> - |

#### 1. Objetivos

O objetivo principal do curso é habilitar o(a) cursista a compreender o conceito de anotação de dados, os métodos de anotação e o papel dessa anotação na geração de modelos de Inteligência Artificial (IA). Ademais, busca-se apresentar e discutir a importância da qualidade dos dados, destacando a necessidade de dados corretos, completos, válidos, relevantes e confiáveis para geração de datasets para algoritmos de *Machine Learning*, bem como para uma gestão comercial orientada a dados.

Um projeto de anotação tem como meta garantir a confiança no dataset anotado, bem como minimizar eventuais vieses. Abrange, entre outras etapas, coletar e reunir dados corretos e relevantes, escolher o método de anotação dos dados, definir os guias de anotação, selecionar anotadores, revisores e ferramentas próprias para o tipo de anotação almejada. Durante o curso, são apresentados exemplos práticos de projetos de tribunais e do CNJ/PNUD em parceria com universidades que realizaram anotação de dados. Ao final, é proporcionado ao(a) cursista informações sobre recursos de anotação da plataforma SINAPSES e novas abordagens de anotação baseadas em LLMs.

## 2. Programa

O curso apresenta como conteúdos programáticos centrais: A importância dos Dados e sua qualidade; Big Data e tipos de Dados; Dados como ativo estratégico para o Poder Judiciário; Dados para Algoritmos de Inteligência Artificial; Etapas de Projeto de Anotação de Dados e suas boas práticas; os Projetos ALICIA, PEDRO, Janus, Sofia e Elis; Anotação de Dados no SINAPSES e suas aplicações, e; Anotação de Dados para modelos de Inteligência Artificial usando LLMs.

## 3. Procedimentos didáticos

Como suporte ao Ensino Remoto será utilizada a infraestrutura do Ambiente Virtual de Aprendizagem (AVA) da plataforma moodle do Centro de Formação e Aperfeiçoamento de Servidores do Poder Judiciário (CEAJUD). Na Plataforma estarão disponíveis os seguintes materiais:

- Ficha do Curso;
- Links para acesso às videoaulas legendadas (módulos gravados);
- Material Complementar em formato de slides;
- Atividade avaliativa de desempenho para certificação; e
- Avaliação sobre a qualidade do curso.

## 4. Carga horária do(a) cursista

A seguir, é apresentada a carga horária total do curso.

| <b>Atividades</b>                                | <b>Carga horária (horas)</b> | <b>Fração (%)</b> |
|--|------------------------------|-------------------|
| <b>Estudo individual de conteúdo pré-aula</b>    | 1h                           | 10%               |
| <b>Estudo individual em aula assíncrona</b>      | 5h                           | 50%               |
| <b>Estudo individual pós-aula</b>                | 2h                           | 20%               |
| <b>Atividade Avaliativa e Avaliação do Curso</b> | 2h                           | 20%               |
| <b>Total</b>                                     | <b>10h</b>                   | <b>100%</b>       |

## **5. Conteúdo programático**

O curso está dividido em 3 módulos, detalhados abaixo.

**Módulo 1:** A importância da qualidade dos dados para o desenvolvimento de sistemas inteligentes

- A Importância dos Dados;
- Big Data e Tipos de Dados;
- Garantindo a Qualidade dos Dados;
- Dados como ativo estratégico para o Poder Judiciário – Parte I;
- Dados como ativo estratégico para o Poder Judiciário – Parte II.

**Módulo 2:** Princípios básicos da Anotação de Dados

- Dados para Algoritmos de Inteligência Artificial;
- Etapas de Projeto de Anotação de Dados – Boas Práticas;
- Projeto ALICIA: Análise Linguística para Classificação Automática: parceria CNJ/PNUD com a Universidade de Fortaleza – Parte 1 Visão Geral do Projeto;
- Projeto ALICIA: Análise Linguística para Classificação Automática: parceria CNJ/PNUD com a Universidade de Fortaleza – Parte 2 Metodologia de Anotação;
- Projeto PEDRO: Plataforma de Extração e Descoberta de Precedentes dos Tribunais: parceria CNJ/PNUD com a Universidade de Brasília - Parte 1 Visão Geral do Projeto;
- Projeto PEDRO: Plataforma de Extração e Descoberta de Precedentes dos Tribunais: parceria CNJ/PNUD com a Universidade de Brasília - Parte 2 - Metodologia de Anotação;

- Projeto PEDRO: Plataforma de Extração e Descoberta de Precedentes dos Tribunais: parceria CNJ/PNUD com a Universidade de Brasília - Parte 3 - Ferramentas de Anotação.

### **Módulo 3:** Plataforma SINAPSES e Anotação de Dados

- Introdução ao SINAPSES;
- Anotação de Dados no SINAPSES e Aplicações;
- Projeto Janus – TRE-BA;
- Projeto Sofia – TJBA;
- Projeto Elis – TJPE;
- Anotação de Dados para modelos de Inteligência Artificial usando LLMs.

## **6. Avaliação de desempenho**

O desempenho no curso será determinado por nota na atividade avaliativa realizada ao final do curso e pelo preenchimento da avaliação sobre a qualidade do curso. Requer-se, no mínimo, 70% de aproveitamento na atividade avaliativa para obtenção de certificado.

## **7. Informações sobre os conteudistas**

Conselho Nacional de Justiça (CNJ)

- Dr. Rafael Leite Paulo - Juiz Auxiliar da Presidência do CNJ.

Programa das Nações Unidas para o Desenvolvimento (PNUD)

- Sílvia Maria Wanderley Moraes - Professora da Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), desde agosto de 1997. Professora colaboradora do Grupo de Pesquisa em Linguagem Natural da PUCRS, do Grupo de Pesquisa de Inteligência Artificial na Saúde e

mais recentemente no Centro de Pesquisa em Ciência de Dados. Obteve título de doutora em Ciência da Computação PPGCC / PUCRS em 2012. Anteriormente, cursou mestrado em Ciência da Computação no PPGCC / UFRGS de 1994 a 1997. Obteve o título de bacharel em Informática também na FACIN / PUCRS em 1992. Atualmente, também é Analista de Inteligência Artificial do PNUD.

- José Avelino Placca - Possui graduação em Engenharia de Computação pelo Instituto Militar de Engenharia (1989), mestrado em Computação pela Universidade Federal Fluminense (1999) e doutorado em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro (2007). Atualmente é Analista de Inteligência Artificial do PNUD, avaliador do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira e consultor ad-hoc da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco e professor da Universidade Estadual do Mato Grosso do Sul.

#### Tribunal de Justiça da Bahia (TJBA)

- Leonardo Ribeiro de Oliveira – Assessor da Coordenação dos Juizados Especiais do TJBA.
- Jonathas Costas de Carvalho - Assessor da Coordenação dos Juizados Especiais do TJBA.

#### Tribunal de Justiça de Pernambuco (TJPE)

- Hadautho Roberto Barros da Silva - Analista de Inteligência Artificial e desenvolvedor do ELIS.

#### Tribunal Regional Eleitoral da Bahia (TRE-BA)

- Desembargador Roberto Frank – Presidente do TRE-BA e idealizador do projeto Janus.
- André Cavalcante - Secretário de Tecnologia da Informação e gerente do projeto Janus.
- Hercília Barros - Secretária Judiciária do 1º grau de jurisdição e integrante da equipe gestora do Janus.

#### Universidade de Brasília (UnB)

- Nilton Correia da Silva – Professor Pesquisador em Inteligência Artificial e Coordenador do projeto PEDRO.
- Débora Bonat – Professora Pesquisadora em Direito do projeto PEDRO.
- Fabício Braz - Professor Pesquisador em Inteligência Artificial do projeto PEDRO.

#### Universidade de Fortaleza (UNIFOR)

- João José Peixoto Vasco Furtado – Professor Pesquisador em Inteligência Artificial e Coordenador do Projeto ALICIA.
- João Araújo Monteiro Neto – Professor Pesquisador em Direito do Projeto ALICIA.

## **8. Requisitos Técnicos**

Para cursar a presente capacitação, faz-se necessário possuir:

- Computador e/ou tablet capazes de acessar a plataforma moodle do CEAJUD;
- Versões mais atualizadas de Windows, Linux ou Mac;
- Conexão de Internet.

## 9. Referências bibliográficas

- He, X., Lin, Z., Gong, Y., Jin, A., Zhang, H., Lin, C., Jiao, J., Yiu, S.M., Duan, N., & Chen, W. (2023). AnnoLLM: Making Large Language Models to Be Better Crowdsourced Annotators. ArXiv, abs/2303.16854.
- Tseng, Tina & Stent, Amanda & Maida, Domenic. (2020). Best Practices for Managing Data Annotation Projects.
- Y. Roh, G. Heo and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328-1347, 1 April 2021.
- Youdi Gong, Guangzhen Liu, Yunzhi Xue, Rui Li, Lingzhong Meng, A survey on dataset quality in machine learning, *Information and Software Technology*, Volume 162, 2023.

## 10. Créditos e Agradecimentos

- Programa Justiça 4.0: Inovação e Efetividade na Realização da Justiça para Todos;
- Programa das Nações Unidas para o Desenvolvimento – PNUD;
- Conselho da Justiça Federal – CJF;
- Conselho Nacional de Justiça – CNJ;
- Buffalo Digital LTDA;
- Tribunal de Justiça da Bahia – TJBA;
- Tribunal de Justiça de Pernambuco – TJPE;
- Tribunal Regional Eleitoral da Bahia – TRE-BA;
- Universidade de Brasília – UnB;
- Universidade de Fortaleza – UNIFOR.