

## Ficha de Curso

### Agrupamento de Textos Jurídicos e IA - Parceria CNJ, PNUD e Universidade

### Federal de Goiás - UFG

<b>Curso:</b> Agrupamento de Textos Jurídicos e IA - Parceria CNJ, PNUD e UFG
<b>Modalidade:</b> Ensino a Distância – Autoinstrucional
<b>Público-alvo:</b> Servidores técnicos do Poder Judiciário brasileiro, com atuação na área de Tecnologia da Informação e no desenvolvimento de aplicações para a Plataforma Digital do Poder Judiciário Brasileiro – PDPJ-Br.
<b>Carga horária total:</b> 15 horas

#### 1. Contextualização e Objetivo da Capacitação

O projeto de Agrupamento de Textos Jurídicos e IA, fruto da parceria entre o Conselho Nacional de Justiça (CNJ), o Programa das Nações Unidas para o Desenvolvimento (PNUD) e a Universidade de Goiás (UFG), junto ao Programa Justiça 4.0, teve como objetivo prover uma visão geral da Solução SAS e o conhecimento de cunho prático no formato *hands on* a equipes técnicas das áreas de Inteligência Artificial (IA) e de engenharia de dados do CNJ e dos Tribunais brasileiros.

Como resultado do projeto, foi possível identificar agrupamentos de processos similares por meio da aplicação de técnicas de Processamento de Linguagem Natural (*Natural Language Processing – NLP*) e de Inteligência Artificial.

Para tal, criou-se um *dataset* de referência (coleção dourada), contendo grupos de processos anotados por uma equipe de especialistas do direito, para, em seguida, identificar qual a combinação de modelos alcançaria um resultado mais próximo aos

grupos formados manualmente naquele *dataset*, culminando com a escolha da combinação de modelos vencedora.

No decorrer desta capacitação o/a cursista terá uma visão geral das etapas que compõem a *pipeline* final de processamento utilizada no Projeto SAS, com eventuais ajustes que permitiram a execução de cada etapa ao longo da jornada.

Nesse sentido, o objetivo desta capacitação é: **oferecer subsídios para o desenvolvimento de novas metodologias, estudos e soluções de IA para a promoção de inovação, com foco na efetividade da realização da Justiça.**

## **2. Programa**

O curso está dividido em **10 Módulos**, distribuídos da seguinte maneira:

### **Módulo 1: Introdução**

- Boas-Vindas
- Orientações Gerais do Curso
- Pipeline SAS
- Conhecimentos Gerais em NLP – Parte I
- Conhecimentos Gerais em NLP – Parte II

### **Módulo 2 – Aplicação SAS**

- Visão Geral da Aplicação SAS
- Metadados dos processos
- Uso de filtros
- Processos similares: índices de similaridade
- Estratégias de Análises em um conjunto de dados

- Exportação de dados
- Feedback em relação aos grupos formados

### **Módulo 3 – Modelos de Linguagem**

- Treinamento de um Modelo de Linguagem

### **Módulo 4: Reconhecimento Frases e Entidades**

- Frases - Introdução
- Frases - Requirements
- Frases - Configuração
- Frases - Treino
- Frases - Avaliação
- Frases - Inferência
- NER - Introdução / Requirements
- NER - Dataset Reader
- NER - Model
- NER - Trainer
- NER - Data Loader
- NER - Treino
- NER - Avaliação
- NER - Inferência

### **Módulo 5 - Sintetização**

- Sintetização

### **Módulo 6 – Representação Vetorial**

- Apresentação do SPECTER

- Criação de dataset
- Arquivos de input em formato SPECTER
- Preparação de dados de treinamento
- Treinamento do modelo SPECTER
- Conversão do modelo SPECTER para o formato Transformer da Hugging Face

### **Módulo 7 – Agrupamentos**

- Agrupamentos

### **Módulo 8 – Avaliação de Qualidade**

- Avaliação de Qualidade

### **Módulo 9 - Integração com Sinapses**

- Explicação do pipeline em Python
- Módulos disponíveis do Sinapses
- Customização dos módulos no Sinapses

### **Módulo 10 - Treinamento Mão na Massa**

- Modelo de linguagem RoBERTa Jurídico - Parte I
- Modelo de linguagem RoBERTa Jurídico - Parte II
- Modelo de linguagem RoBERTa Jurídico - Parte III
- Modelo de linguagem RoBERTa Jurídico – Parte IV
- Modelo de linguagem RoBERTa Jurídico - Parte V
- Modelo de linguagem RoBERTa Jurídico - Parte VI
- Sintetização - Parte I
- Sintetização - Parte II

- Modelo Specter - Parte I
- Modelo Specter - Parte II
- Modelo Specter - Parte III
- Modelo Specter - Parte IV
- Modelo Specter - Parte V
- Modelo Specter - Parte VI
- Modelo de Agrupamento - Parte I
- Modelo de Agrupamento - Parte II
- Projeto SAS - pipeline - Parte I
- Projeto SAS - pipeline - Parte II

Destaca-se que o curso sofreu refinamentos durante seu processo de finalização, de forma que podem ocorrer menções de nomenclaturas que foram descontinuadas, como, por exemplo, Projeto Kairos para se referenciar ao Projeto de Agrupamento de Textos Jurídicos e IA, desenvolvido pela UFG. Nesse sentido, pedimos para desconsiderar essas menções.

#### **4. 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);
- Materiais complementares;
- Atividade avaliativa de desempenho para certificação; e
- Avaliação sobre a qualidade do curso.

Destaca-se que o presente curso, no modelo autoinstrucional nesta plataforma, é originário da versão síncrona e online do treinamento previamente ofertado. Nesse sentido, eventuais cortes no áudio ou na imagem dos conteudistas durante as videoaulas podem ser fruto de instabilidade da rede no momento da gravação.

## 5. 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>
Estudo individual de conteúdo pré-aula	0h30	3,33%
Estudo individual em aula assíncrona	11h	73,33%
Estudo individual pós-aula	2h30	16,67%
Atividade Avaliativa e Avaliação do Curso	1h	6,67%
<b>Total</b>	<b>15 horas</b>	<b>100%</b>

## 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 mesmo. Requer-se, no mínimo, 70% de aproveitamento na atividade avaliativa para obtenção de certificado.

## 7. 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.