

ANEXO AO EDITAL Nº 078/2026

Descrição do Projeto, Atividades Complementares e Glossário

LabCLICK — Laboratório de Conformidade, Litigiosidade, Capacidade Institucional e Conhecimento

Centro de Estudos da Ordem Econômica (CEOE) — UNIFESP

1. Contexto Institucional

O LabCLICK — Laboratório de Conformidade, Litigiosidade, Capacidade Institucional e Conhecimento — é um laboratório de pesquisa do Centro de Estudos da Ordem Econômica (CEOE) da Universidade Federal de São Paulo (UNIFESP), com sede em Osasco/SP. O laboratório investiga as condições em que as orientações jurídicas emitidas por órgãos de controle — como procuradorias estaduais e tribunais de contas — são seguidas, contestadas ou ignoradas pelas agências públicas a que se destinam, com foco em métodos empíricos e computacionais. Também investiga aspectos mais gerais de litigiosidade, judiciário, regulação e governança.

O LabCLICK mantém colaborações com instituições brasileiras e internacionais, incluindo a Yale Law School, onde o coordenador do laboratório é pesquisador sênior. As bolsas deste edital integram dois subprojetos em andamento: o Projeto Newton e a Iniciativa NAJA, descritos a seguir.

2. O Projeto Newton

2.1 O problema de pesquisa

A administração pública brasileira é responsável por um volume expressivo de litígios judiciais que poderiam ter sido evitados se as agências envolvidas tivessem acesso claro e confiável às orientações normativas aplicáveis a cada situação. O custo desse fenômeno — chamado no projeto de “litigiosidade evitável” — é suportado tanto pelo erário quanto pelas partes afetadas pelas decisões administrativas contestadas.

O problema tem uma dimensão estrutural, vez que as normas que regem a administração pública brasileira estão distribuídas em múltiplas camadas — a Constituição Federal, as decisões do Tribunal de Contas da União (TCU), leis específicas por área, a lei geral de processo administrativo e as orientações internas das procuradorias — sem uma hierarquia formal clara entre elas. Quando essas camadas entram em conflito, é difícil determinar qual orientação prevalece, e as agências frequentemente tomam decisões com base em uma leitura incompleta do quadro normativo.

2.2 O que o Projeto Newton faz

O Projeto Newton constrói um sistema de inteligência artificial capaz de auxiliar procuradores, controladores e gestores públicos na análise de validade procedimental de atos administrativos. O sistema não substitui o julgamento humano, ele verifica, passo a passo, se cada etapa de um raciocínio jurídico encontra respaldo nas normas aplicáveis — incluindo decisões do TCU e pareceres da Advocacia-Geral da União — e sinaliza quando uma questão é genuinamente controvertida entre os órgãos de controle.

A arquitetura do sistema combina dois tipos de componentes. O primeiro é um modelo de linguagem de grande porte (como os modelos que alimentam assistentes de inteligência artificial amplamente conhecidos), responsável por gerar os passos de raciocínio em linguagem natural. O segundo é um oráculo normativo — um sistema especializado que consulta uma base de conhecimento jurídico estruturada e verifica se cada passo do raciocínio tem fundamento na norma correta, se essa norma estava vigente na data do ato, e se ela é aplicável ao tipo de ato e à instituição em questão. A

interação entre os dois componentes ocorre em tempo real, assim o oráculo pode corrigir o modelo, indicar a norma correta ou sinalizar que a questão é controvertida, antes que o raciocínio incorreto se propague para as etapas seguintes.

O domínio inicial de aplicação é o das contratações públicas diretas sem licitação, disciplinadas pela Lei nº 14.133/2021 (que substituiu a antiga Lei nº 8.666/1993). Esse domínio foi escolhido por três razões: é o setor com maior volume de questionamentos por órgãos de controle; o TCU produziu uma jurisprudência densa e semiestruturada especificamente sobre contratações diretas; e a transição entre as duas leis — com períodos de vigência sobrepostos — gera uma classe bem documentada de erros que os modelos de linguagem cometem quando não têm acesso ao oráculo.

2.3 O benchmark AdmBench-BR

Um dos produtos centrais do projeto é o AdmBench-BR. Trata-se de uma coleção de cenários de avaliação para medir a capacidade de sistemas de inteligência artificial de produzir raciocínio jurídico válido em direito administrativo brasileiro. Cada cenário descreve uma situação administrativa concreta — por exemplo, uma agência que pretende contratar diretamente um escritório de advocacia — e tem uma resposta correta verificável em fontes públicas (decisões do TCU, pareceres da AGU, texto das leis). O benchmark permite comparar sistematicamente o desempenho de diferentes sistemas de IA e identificar onde e como cada um falha.

O AdmBench-BR será publicado em plataforma de acesso aberto, com código de avaliação e uma tabela pública de classificação de sistemas (“leaderboard”), e destina-se à comunidade internacional de pesquisa em inteligência artificial aplicada ao direito.

3. A Iniciativa NAJA

A Iniciativa NAJA — Núcleo Aberto de Dados Jurisdicionais e Administrativos — é um projeto complementar ao Newton, voltado para a construção e manutenção de um repositório público de dados judiciais e administrativos brasileiros em formato acessível à pesquisa. Dados públicos sobre decisões judiciais, pareceres de procuradorias e atos administrativos existem nos portais das instituições, mas raramente estão disponíveis em formatos estruturados e padronizados que permitam análise quantitativa por pesquisadores independentes. A NAJA endereça essa lacuna.

A iniciativa tem quatro componentes: um corpus de dados estruturados extraídos de fontes públicas (decisões do TCU, pareceres da AGU, legislação federal e seus históricos de alteração, e a base DataJud do Conselho Nacional de Justiça); um Índice de Abertura de Dados, publicado anualmente, que avalia o grau em que cada tribunal disponibiliza seus dados em formato efetivamente utilizável por pesquisadores; um protocolo e uma interface de programação (API) que permitem a outros grupos de pesquisa contribuir e consumir dados do repositório sob padrões comuns; e um Fórum de Advocacia junto ao Conselho Nacional de Justiça (CNJ) e ao Ministério da Justiça, promovendo a melhoria das práticas de abertura de dados do Judiciário brasileiro.

4. Fontes de Dados Utilizadas no Projeto

Todos os dados utilizados no projeto são obtidos de fontes públicas e estão disponíveis sem acesso institucional restrito. As principais fontes são:

- API de dados abertos do TCU: acesso estruturado a acórdãos, súmulas e à Jurisprudência Seleccionada do Tribunal de Contas da União, que consolidam as posições operacionais do TCU sobre compras públicas e outros temas.

- Portal de pareceres da AGU: pareceres vinculantes da Advocacia-Geral da União, que orientam a administração pública federal sobre questões jurídicas relevantes.
- LexML: base de legislação federal em formato XML estruturado, incluindo as leis de compras públicas e seus históricos de alteração.
- DataJud/CNJ: base nacional de dados do Poder Judiciário, mantida pelo Conselho Nacional de Justiça, com metadados de processos judiciais de todos os tribunais brasileiros.
- INEP: base de dados educacionais do Instituto Nacional de Estudos e Pesquisas Educacionais, utilizada na linha de pesquisa sobre políticas públicas e desenvolvimento econômico.

5. Atividades Complementares dos Bolsistas

Além das atribuições específicas de cada função, todos os bolsistas participam, na medida em que suas atribuições o permitam, das atividades institucionais do LabCLICK e do CEOE. Essas atividades são parte integrante da formação dos bolsistas e compreendem:

- Participação em reuniões periódicas de alinhamento da equipe, incluindo apresentação de resultados parciais e discussão de andamento dos projetos.
- Apoio à organização de eventos científicos e institucionais coordenados pelo CEOE de forma geral e pelo LabCLICK em particular, como workshops, seminários e conferências, incluindo o workshop de junho de 2026 (São Paulo), o seminário de agosto de 2026 e a III Conferência Internacional CEOE/LabCLICK prevista para novembro-dezembro de 2026.
- Contribuição à preparação de relatórios técnicos, notas de impacto e materiais institucionais destinados a parceiros, órgãos financiadores e instituições colaboradoras.
- Participação em grupos de leitura e discussão de literatura científica relevante para os projetos, organizados periodicamente pelo coordenador.
- Apoio à preparação de publicações científicas — incluindo levantamentos bibliográficos, organização de dados e revisão de manuscritos — conforme a área de formação do bolsista.
- Colaboração com a equipe de comunicação do projeto na produção de conteúdo para o site, redes institucionais e materiais de divulgação, quando solicitado.

A participação nessas atividades complementares é considerada parte da experiência de formação proporcionada pela bolsa e não implica remuneração adicional. O volume de horas dedicadas a atividades complementares é proporcionalmente reduzido em relação às atribuições principais e não deve comprometer o cumprimento das entregas estabelecidas no plano de trabalho.

6. Glossário de Termos Técnicos

Os termos abaixo aparecem no edital ou nos planos de trabalho das bolsas e são explicados aqui para facilitar a avaliação das candidaturas por candidatos de diferentes áreas de formação.

Termo	Explicação
AdmBench-BR	Coleção de cenários de avaliação produzida pelo projeto para medir a capacidade de sistemas de inteligência artificial de raciocinar corretamente sobre direito administrativo brasileiro. O nome vem de “Administrative Benchmark — Brazil”. Cada cenário tem uma resposta correta verificável em fontes públicas (decisões do TCU, pareceres da AGU, texto das leis).
Anotação (corpus)	Processo de classificar e rotular documentos ou trechos de texto segundo categorias predefinidas. No projeto, bolsistas de direito aplicam um manual de anotação para identificar, em decisões do TCU e pareceres da AGU, as normas citadas, os tipos de ato administrativo analisados e as conclusões jurídicas. Esse trabalho é necessário para que o sistema de inteligência artificial aprenda a reconhecer padrões normativos e para que os cenários do AdmBench-BR tenham respostas corretas verificáveis.
Anotação gold standard	Conjunto de documentos anotados com alta confiança, após revisão e reconciliação de discrepâncias entre anotadores, que serve como referência definitiva para treinar e avaliar o sistema. O termo “gold standard” é usado na área de processamento de linguagem natural para designar dados de alta qualidade garantida. No projeto, a produção do gold standard envolve pelo menos dois anotadores, medição de concordância e revisão pelo coordenador.
API (interface de programação)	Canal padronizado pelo qual dois sistemas de software trocam informações automaticamente. No projeto, a API do TCU, a API do DataJud/CNJ e a API do LexML permitem que os programas do projeto extraiam dados dessas fontes sem navegar manualmente pelos portais. O oráculo normativo do sistema Newton também é exposto como uma API, o que permite que o modelo de linguagem consulte a base normativa durante o raciocínio.
ASP / Clingo	“Answer Set Programming” é uma forma de lógica computacional declarativa usada para raciocinar sobre regras com exceções — exatamente o tipo de estrutura que caracteriza sistemas normativos jurídicos, nos quais uma norma mais específica pode afastar uma norma mais geral. O Clingo é o principal programa (solver) usado para executar regras escritas em ASP. No projeto, as regras de resolução de conflitos entre camadas normativas são codificadas em ASP.
ASPIC+	Framework formal de argumentação jurídica, desenvolvido na área de inteligência artificial e direito, que representa argumentos em conflito e as relações de ataque entre eles. No projeto, é usado para representar as questões normativas genuinamente contestadas — aquelas em que TCU e AGU têm posições divergentes sem resolução formal — de forma que o sistema possa sinalizar a controvérsia ao invés de fornecer uma resposta incorretamente definitiva.
CATE (Efeito Médio Condicional)	Método estatístico que estima o impacto de uma intervenção não como um único número médio, mas como um valor que varia conforme as características dos indivíduos ou unidades analisadas. No projeto, é usado para medir se certos tipos de cenário (por camada

Termo	Explicação
de Tratamento)	normativa, tipo de ato, etc.) produzem mais erros em determinados modelos de linguagem do que em outros — o que ajuda a identificar onde o oráculo é mais necessário.
Compras públicas	Processo pelo qual órgãos públicos adquirem bens, serviços e obras. No contexto do projeto, o foco está especificamente nas “contratações diretas” — casos em que a lei autoriza o órgão a contratar sem realizar licitação, desde que cumpridos requisitos específicos. Esse é o domínio normativo inicial do sistema Newton.
Corpus	Conjunto de documentos reunidos sistematicamente para fins de pesquisa ou treinamento de sistemas computacionais. No projeto, o corpus normativo é composto por decisões do TCU, pareceres da AGU, textos legais e outros documentos jurídicos, obtidos de fontes públicas e tratados de acordo com um protocolo de coleta e organização padronizado.
DataJud / CNJ	DataJud é a base nacional de dados do Poder Judiciário brasileiro, mantida pelo Conselho Nacional de Justiça (CNJ). Contém metadados de todos os processos judiciais em tramitação ou encerrados, com atualização mensal. No projeto, é usada para vincular processos administrativos contestados a seus desfechos judiciais, permitindo medir empiricamente a litigiosidade decorrente de diferentes tipos de decisão administrativa.
Dispensa de licitação	Hipótese legal em que a administração pública está autorizada a contratar diretamente, sem realizar processo licitatório, desde que preenchidos os requisitos da lei. A Lei nº 14.133/2021 (nova Lei de Licitações) organiza as hipóteses de dispensa nos artigos 72 a 79, substituindo os artigos 24 e 25 da Lei nº 8.666/1993. A transição entre as duas leis é o principal domínio normativo do sistema Newton na fase inicial.
DSPy	Framework de software que permite otimizar automaticamente as instruções enviadas a modelos de linguagem, em vez de escrevê-las manualmente por tentativa e erro. No projeto, é usado para ajustar as instruções do sistema à medida que ele é expandido para novos domínios normativos, garantindo que os resultados sejam reproduzíveis e comparáveis entre diferentes versões.
Grafo de conhecimento normativo	Representação estruturada das normas jurídicas e das relações entre elas em formato computacionalmente consultável. No projeto, cada norma é um nó do grafo, com informações sobre sua instituição de origem, data de publicação e período de vigência. As relações entre normas (superveniência, revogação, escopo de aplicação) são codificadas como arestas. O oráculo normativo do sistema Newton consulta esse grafo para verificar a validade dos passos de raciocínio do modelo de linguagem.
GRF (Causal Forest Generalizado)	Método de machine learning desenvolvido para estimar efeitos heterogêneos de tratamento em dados experimentais ou quasi-experimentais. É uma extensão dos modelos de “florestas aleatórias” (random forests) que incorpora garantias de inferência causal. No

Termo	Explicação
	projeto, é usado para analisar como os erros dos modelos de linguagem variam conforme as características dos cenários normativos.
Inferência causal por machine learning	Área metodológica que combina técnicas de aprendizado de máquina com princípios de identificação causal para estimar efeitos de intervenções em dados observacionais ou experimentais. Inclui métodos como causal forests, double-selection LASSO e estimadores duplamente robustos. No projeto, é aplicada à análise de dados de processos judiciais e administrativos.
Leaderboard	Tabela pública de classificação de sistemas de inteligência artificial em um benchmark específico, atualizada à medida que novos sistemas são avaliados. Serve como instrumento de transparência e estimula a melhoria progressiva dos sistemas. O AdmBench-BR terá um leaderboard público hospedado na plataforma HuggingFace.
LLM (Modelo de Linguagem de Grande Porte)	Sistema de inteligência artificial treinado em grandes volumes de texto para gerar e compreender linguagem natural. Exemplos conhecidos incluem o ChatGPT (OpenAI), o Gemini (Google) e o Claude (Anthropic). No projeto Newton, LLMs são usados como geradores de raciocínio jurídico, e o oráculo normativo verifica e corrige esse raciocínio em tempo real.
MCP (Model Context Protocol)	Protocolo de comunicação aberto, publicado pela Anthropic em novembro de 2024, que define uma interface padronizada entre um modelo de linguagem e servidores de ferramentas externas. No projeto, o servidor MCP do Newton é o componente que recebe consultas do modelo de linguagem e as encaminha ao grafo de conhecimento normativo, retornando respostas tipadas (aceite, rejeição, incerteza).
NAJA	Núcleo Aberto de Dados Jurisdicionais e Administrativos. Iniciativa do LabCLICK voltada para a construção de um repositório público de dados judiciais e administrativos brasileiros em formato estruturado e acessível à pesquisa.
Newton (Projeto)	Projeto de pesquisa do LabCLICK que constrói um sistema de inteligência artificial para verificar a validade procedimental de atos administrativos no âmbito do direito administrativo brasileiro. O nome homenageia Isaac Newton como símbolo da articulação entre raciocínio formal e verificação empírica.
Oráculo normativo	Componente do sistema Newton responsável por verificar, em tempo real, se cada passo do raciocínio jurídico gerado pelo modelo de linguagem tem fundamento na norma correta, se essa norma estava vigente na data relevante e se é aplicável ao tipo de ato em questão. O oráculo responde com três tipos de sinal: aceite (o passo está correto), rejeição (o passo está errado, com indicação da norma correta) ou incerteza (a questão é genuinamente controvertida entre órgãos de controle).

Termo	Explicação
Pipeline de dados / de processamento	Sequência automatizada de etapas de processamento de dados: coleta, limpeza, transformação, armazenamento e disponibilização. O pipeline de ingestão do corpus coleta documentos das APIs públicas, extrai seu conteúdo, identifica referências normativas e os armazena em formato estruturado para preencher o grafo de conhecimento.
PoC (Prova de Conceito)	Segunda fase de desenvolvimento do sistema Newton, que se segue à fase de demonstração inicial (“spike”). Na fase de PoC, o sistema é expandido para cobrir pelo menos dois domínios normativos além das contratações diretas, e os resultados são avaliados sistematicamente com o AdmBench-BR. O objetivo é demonstrar que a arquitetura funciona em mais de um domínio antes de escalar para a versão completa do sistema.
Process tracing	Método qualitativo de análise causal que reconstrói a cadeia de eventos e mecanismos que produziram um resultado específico, a partir do exame detalhado de um caso ou de um conjunto reduzido de casos. No projeto, é usado para analisar casos documentados de litígios evitáveis, identificando em que etapa do processo administrativo a divergência normativa surgiu e como se propagou até a judicialização.
RDF / SPARQL	RDF (Resource Description Framework) é um padrão do World Wide Web Consortium para representar informações como trincas (sujeito — predicado — objeto), formando um grafo de conhecimento. SPARQL é a linguagem de consulta para esse tipo de grafo, análoga ao SQL para bancos de dados relacionais. No projeto, o grafo normativo é armazenado em RDF e o oráculo o consulta usando SPARQL.
Revisão realista (realist review)	Metodologia de síntese de evidências que investiga não apenas se uma intervenção funciona, mas por que funciona, para quem e em que contexto. Ao contrário de revisões sistemáticas convencionais, que buscam uma estimativa média de efeito, a revisão realista busca identificar os mecanismos causais subjacentes e as condições contextuais que os ativam. No projeto, é aplicada ao mapeamento da evidência sobre efetividade de mecanismos de controle administrativo.
Revisão sistemática	Método de síntese da literatura científica que segue um protocolo predefinido para buscar, selecionar, avaliar e sintetizar estudos sobre uma questão específica, de forma reproduzível e comparável. É amplamente usada em políticas públicas, saúde e avaliação de programas. No projeto, serve para mapear o que a literatura sabe sobre as condições em que órgãos de controle influenciam o comportamento de agências públicas.
SEI	Sistema Eletrônico de Informações: plataforma de gestão de documentos e processos adotada pelo governo federal e pela maioria dos governos estaduais brasileiros. Registra cada ato administrativo — criação de documentos, assinaturas, encaminhamentos, decisões — como um evento num fluxo de trabalho. Os registros do SEI são

Termo	Explicação
	uma das fontes potenciais para análise da gramática procedimental real da administração pública.
Spike	Primeira fase de desenvolvimento do sistema Newton, correspondente a um demonstrador funcional de escopo mínimo. O spike cobre apenas o domínio das contratações diretas sem licitação (artigos 72 a 79 da Lei nº 14.133/2021) e tem como objetivo demonstrar que a arquitetura completa — modelo de linguagem + oráculo normativo + grafo de conhecimento — funciona de ponta a ponta antes de ser expandida. O termo é emprestado da engenharia de software, onde designa um experimento exploratório de escopo reduzido.
Superveniência normativa (validity by supersession)	Mecanismo pelo qual uma posição do TCU perde vigência não por revogação formal, mas pela publicação de um acórdão posterior que trata do mesmo assunto com orientação divergente. Ao contrário da revogação legislativa, a superveniência normativa não é sempre explícita, e exige que o sistema identifique cronologicamente qual é a posição mais recente sobre cada questão. Esse é um dos principais desafios técnicos do grafo normativo do projeto Newton.
TCU	Tribunal de Contas da União: órgão de controle externo do governo federal brasileiro, responsável por fiscalizar a aplicação dos recursos públicos. Suas decisões (acórdãos) e súmulas têm natureza quase-normativa: embora não constituam precedente vinculante em sentido estrito, são amplamente seguidas pelas agências públicas como orientação operacional. A Jurisprudência Selecionada do TCU é uma das principais fontes do grafo normativo do projeto Newton.
ToT (Tree of Thoughts)	Estratégia de raciocínio para modelos de linguagem que, em vez de gerar uma única cadeia linear de passos, explora múltiplos caminhos de raciocínio simultaneamente — semelhante à forma como um jogador de xadrez considera várias sequências de jogadas antes de decidir. No projeto Newton, essa estratégia permite que o sistema explore ramos de raciocínio jurídico em paralelo e descarte os que não passam pela verificação do oráculo normativo antes de continuar.
vLLM / Ollama	Programas de código aberto usados para executar modelos de linguagem de grande porte em servidores locais. O vLLM é otimizado para alto volume de requisições simultâneas e é utilizado no servidor de colocação do projeto para executar os modelos de 32 a 70 bilhões de parâmetros. O Ollama é mais simples e é usado durante o desenvolvimento interativo com modelos menores.

Dúvidas sobre o projeto e as atividades podem ser encaminhadas para recrutamento@ceoe.org.br