

Levantamento de iniciativas de IA no Ministério Público

**Comissão de
Planejamento Estratégico
Estratégia Nacional do MP Digital**

2023



CONSELHO
NACIONAL DO
MINISTÉRIO PÚBLICO

SUMÁRIO

	APRESENTAÇÃO	4
1.	MINISTÉRIO PÚBLICO FEDERAL.....	5
	1.1 TRIA - TRIAGEM EM HABEAS CORPUS.....	5
	1.2 RESULTADOS DE JULGAMENTOS.....	5
	1.3. CLASSIFICAÇÃO PUBLICAÇÕES STJ.	6
	1.4. MANIFESTAÇÃO OUVIDORIA	6
	1.5. ATHOS-MPF (NOME PROVISÓRIO).....	7
	1.6. INTERNALIZAÇÃO GPSMED	7
	1.7. TRIAGEM DE INQUÉRITOS POLICIAIS.	8
2.	MINISTÉRIO PÚBLICO DO TRABALHO	8
	2.1. SARA (SISTEMA ASSISTENTE DE REDAÇÃO E ANÁLISE).....	8
3.	MINISTÉRIO PÚBLICO DO DISTRITO FEDERAL E TERRITÓRIOS.....	9
	3.1. ROBOCON.....	9
	3.2. SISTEMA DE APOIO ÀS PROMOTORIAS DE EXECUÇÃO PENAL (PEPSI).....	9
	3.3. LUMINARIA.....	10
	3.4. SELFIEPONTO – PONTO ELETRÔNICO ATRAVÉS DO RECONHECIMENTO FACIAL. REGISTRA O PONTO ATRAVES DO RECONHECIMENTO FACIAL E LOCALIZAÇÃO GEOGRÁFICA DO SERVIDOR. ...	11
	3.5. RISCO VD	11
	3.6. SISTEMA INTELIGENTE DE INTERPRETAÇÃO DE AUTOS (SINTIA)	12
4.	MINISTÉRIO PÚBLICO DO ESTADO DO ACRE.....	12
	4.1.LLAMA 2 DA EMPRESA META	12
5.	MINISTÉRIO PÚBLICO DO ESTADO DO AMAPÁ.....	13
	5.1 ASSISTENTE CHATBOT COM ASSISTENTE VIRTUAL – PONTO ELETRÔNICO COM RECONHECIMENTO FACIAL.....	13
6.	MINISTÉRIO PÚBLICO DO ESTADO DA BAHIA	14
	6.1. FRATRIA - FERRAMENTA DE AUTOMAÇÃO DE TAREFAS COM RECURSO EM INTELIGÊNCIA ARTIFICIAL DO MPBA.....	14
7.	MINISTÉRIO PÚBLICO DO ESTADO DO CEARÁ.....	14
	7.1. ACORDOS DE NÃO-PERSECUÇÃO PENAL	14
	7.2. CONECTA.CONSUMIDOR	16
	7.3. PORTAL DE EXPLORAÇÃO DE EVIDÊNCIAS DIGITAIS (PEED).....	17
8.	MINISTÉRIO PÚBLICO DO ESTADO DO ESPÍRITO SANTO	21
	8.1. JULIUS - FERRAMENTA DE HIPERAUTOMAÇÃO PARA ANÁLISE DE DADOS DE TRANSPARÊNCIA PÚBLICA, FOCANDO INICIALMENTE NA IDENTIFICAÇÃO DE SERVIDORES PÚBLICOS COM ACÚMULOS DE CARGOS.	21
9.	MINISTÉRIO PÚBLICO DO ESTADO DE GOIÁS.....	21
	9.1. ANÁLISE DE NOTÍCIA DE FATO PARA INFORMAR OUTRAS SIMILARES	21
	9.2. IDENTIFICAÇÃO DE DISTORÇÕES DE PAGAMENTO AOS INTEGRANTES DA INSTITUIÇÃO	22
10.	MINISTÉRIO PÚBLICO DO ESTADO DE MATO GROSSO DO SUL	22
	10.1 PROPOSITOR DE DENÚNCIAS DE VIOLÊNCIA DOMÉSTICA	22
	10.2 RECONHECIMENTO DE ENTIDADES EM PROCESSOS.....	22
11.	MINISTÉRIO PÚBLICO DO ESTADO DE MINAS GERAIS.....	23
	11.2. ARCANJO	24
	11.3. TRENA.....	25
	11.4. CAMPANHA ELEITORAL LEGAL	26
	11.5. TIRA-TEIMA INFO	27
	11.6. QUANTO CUSTA.....	28
	11.7. PAINEL DE TRILHAS DE LICITAÇÕES	29
	11.8. PROCONDATA.....	30
	11.9. PAINEL DE TRILHAS DE SERVIDORES PÚBLICOS DA SAÚDE.....	30
	11.10. ÁDUNA	31
	11.11. LINS	33
12.	MINISTÉRIO PÚBLICO DO ESTADO DA PARAÍBA	34
	12.1. SISTEMA ARIEL.....	34
	12.2. PANDORA SPEECH	34
13.	MINISTÉRIO PÚBLICO DO ESTADO DO PIAUÍ	35

13.1	CHATBOT PARA ABERTURA DE CHAMADOS DA CENTRAL DE SERVIÇOS DE TI.....	35
14.	MINISTÉRIO PÚBLICO DO ESTADO DO RIO GRANDE DO NORTE	35
14.1.	PROJETO GENIUS	35
14.2.	PROJETO CHATMP	36
14.3.	UTILIZAÇÃO DE CHATBOT PARA O SERVIÇO CENTRAL DE INFORMAÇÃO AO CIDADÃO QUE SE ENCONTRA ACESSÍVEL NO PORTAL DO MPRN	36
15.	MINISTÉRIO PÚBLICO DO ESTADO DE SANTA CATARINA.....	37
15.1.	CATARINA	37
15.2.	PROGRAMA CÉOS	38
16.	MINISTÉRIO PÚBLICO DO ESTADO DE SÃO PAULO	39
16.1	ATIVIDADE DE JURIMETRIA E ANÁLISE DE DADOS, UTILIZANDO SOLUÇÃO DE PREDIÇÃO ANÁLISE ESTATÍSTICA E INTELIGÊNCIA ARTIFICIAL.....	39
16.2	DIVERSOS.....	39
16.3	PROCESSAMENTO DE LINGUAGEM NATURAL EM PEÇAS PROCESSUAIS COM REDES NEURAS.	40
17.	MINISTÉRIO PÚBLICO DO ESTADO DE TOCANTINS	41
17.1.	DESENVOLVIMENTO DE SISTEMA E-PONTO DE CONTROLE DE PONTO ELETRÔNICO COM USO DE RECONHECIMENTO FACIAL.....	41
17.2.	SISTEMA RADAR AMBIENTAL.....	41

APRESENTAÇÃO

Introdução

Este relatório apresenta o levantamento de iniciativas de inteligência artificial (IA) desenvolvidas ou em desenvolvimento nos ramos e unidades do Ministério Público, e teve como objetivo principal mapear e entender o estágio atual e as perspectivas futuras da adoção de tecnologias de IA dentro do contexto específico do MP, visando direcionar efetivamente a atuação estratégica do MP Digital no fomento à evolução digital da Instituição.

Contextualização

Em outubro de 2023, a Comissão de Planejamento Estratégico do Conselho Nacional do Ministério Público iniciou um processo de consulta aos ramos e unidades do Ministério Público, por meio do ofício-circular (nº 57/2023/CPE), que solicitou o preenchimento de formulário destinado a coletar informações detalhadas sobre as iniciativas de IA já desenvolvidas ou em desenvolvimento.

Além da descrição da solução, o formulário coletou informações sobre as tecnologias e os bancos de dados (datasets) utilizados para alimentar o modelo de IA, os resultados esperados ou já obtidos, os desafios enfrentados no desenvolvimento, implementação e/ou contratação, entre outras.

Abrangência

Considerando que não existe um consenso sobre o conceito de IA, que varia desde definições amplas, incluindo diversas tecnologias que permitem processamento e análise de dados, até aquelas que se restringem a aplicações ou ações específicas, como a criação de modelos para predição, recomendação e decisões automatizadas ou determinadas técnicas, a exemplo da aprendizagem de máquina (TCU, 2022; Hoekstra et al., 2021; Ubaldi et al., 2019), o levantamento adotou uma perspectiva ampla dessas tecnologias, nos termos da Pesquisa TIC Governo Eletrônico 2021¹, considerando iniciativas como uso de *chatbots* ou assistentes virtuais, sistemas de reconhecimento facial ou de imagem ou ferramentas de predição e análise de dados a partir de aprendizagem de máquina.

Considerações Finais

A iniciativa de levantamento e mapeamento das tecnologias de IA desenvolvidas pelos Ministérios Públicos reflete o compromisso do CNMP com a inovação responsável e orientada para o futuro.

Os resultados deste levantamento servirão como base para estratégias do MP Digital, além de estimular o desenvolvimento cooperativo dessas e de outras soluções, com consequente redução de custos e tempo investidos pelos ramos e unidades na persecução dos mesmos objetivos.

Tendo em vista as rápidas transformações tecnológicas, o presente estudo será anualmente atualizado.

¹Disponível em:

https://cetic.br/media/docs/publicacoes/2/20220725170710/tic_governo_eletronico_2021_livro_eletronico.pdf

1. MINISTÉRIO PÚBLICO FEDERAL

1.1 TRIA - TRIAGEM EM HABEAS CORPUS.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Realiza automaticamente a extração de informações sobre o sentido das decisões proferidas pelo Superior Tribunal de Justiça em Habeas Corpus (HC), comparando-as com os pareceres produzidos pelo MPF no mesmo processo e indicando se a decisão é favorável ou contrária à manifestação do membro, ou mesmo se foi proferida sem manifestação prévia do MPF. A ferramenta facilita a triagem dos HCs nos gabinetes criminais da PGR.

Bancos de Dados Utilizados: Para o modelo que analisa as decisões do STJ, foi montada e rotulada base de treinamento/validação contendo 15 mil decisões/acórdãos proferidos pelo STJ em HC's. Para o treinamento do modelo que analisa os pareceres proferidos pelo MPF, foi montada e rotulada base com cerca de 48000 mil pareceres.

Tecnologias Empregadas: Foi utilizada aprendizagem supervisionada, por meio da aplicação de modelo Random Forest.

Resultados Esperados: Sistema encontra-se em produção, com informações sendo disponibilizadas no sistema de tramitação processual do MPF (Sistema Único). Em 08/2023, foi verificado que a acurácia do Tria em produção é de 94,80%.

Desafios: Necessidade de rotulagem manual de grande quantidade de documentos. Necessidade de construção de infraestrutura de IA para processamento dos dados recebidos do STJ.

Parcerias: Desenvolvido exclusivamente pelo MPF.

Estágio Atual: Concluído com sucesso.

Considerações: --

1.2 RESULTADOS DE JULGAMENTOS.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Extrai, de forma automática, informações das sentenças produzidas em ações penais de modo a ajudar servidores de gabinete da 1ª instância no preenchimento dos dados acerca do resultado do julgamento no sistema Único.

São disponibilizadas as seguintes informações: réu, sentido do resultado do julgamento, fundamentação jurídica, base normativa, pena, multa, data da sentença e juiz.

Bancos de Dados Utilizados: Foi montada e rotulada base contendo 3797 sentenças proferidas em ações penais.

Tecnologias Empregadas: Foi utilizado modelo de Reconhecimento de Entidade Nomeada (NER), através da biblioteca Spacy, com a utilização de tecnologia Transformers.

Resultados Esperados: Sistema encontra-se em produção, com informações sendo disponibilizadas no sistema de tramitação processual do MPF (Sistema Único). Durante o treinamento, a acurácia geral do NER foi de 88,21%. Em produção, tem-se observado acurácias diferentes dependendo de qual entidade está sendo extraída. Algumas entidades são extraídas com acurácias acima de 90%, enquanto outras possuem acurácias consideradas ainda insuficientes (abaixo de 80%).

Desafios: Necessidade de rotulagem manual de sentenças proferidas em ações penais. Falta de padronização nos metadados recebidos da Justiça Federal via MNI, dificultando até mesmo a identificação das sentenças. Modelo atual tem a capacidade de analisar apenas sentenças com um réu e um crime associado. Necessário grande esforço ou utilização de tecnologias mais recentes (modelos LLM) para evolução da solução.

Parcerias: Desenvolvido exclusivamente pelo MPF.

Estágio Atual: Concluído com sucesso.

Considerações: --

1.3. CLASSIFICAÇÃO PUBLICAÇÕES STJ.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Analisa textos de publicações realizadas pelo STJ a fim de auxiliar o usuário na análise da necessidade de requisitar o processo para apreciação do MPF.

Bancos de Dados Utilizados: Foi montada base contendo dados cerca de 920 mil publicações realizadas pelo STJ. Não foi necessário realizar a rotulagem, visto que as ações de requisição ou não requisição já constavam registradas no sistema SUBJUR.

Tecnologias Empregadas: Foi utilizada aprendizagem supervisionada, por meio da aplicação de modelo Random Forest.

Resultados Esperados: Sistema encontra-se em produção, com a disponibilização de sugestão de requisição em sistema próprio para análise das publicações do STJ. Foram treinados modelos por grupos de órgãos julgadores, sendo que alguns grupos tiveram precisões consideradas ótimas (acima de 94%) enquanto em outros grupos o modelo ainda não possui boa performance. Considerando as métricas avaliadas, está prevista a realização de automatização de requisições nas faixas de probabilidade em que o modelo possui precisões próximas a 100%.

Desafios: As publicações de alguns órgãos julgadores ainda não possuem boas métricas de precisão, sendo necessária, portanto, a realização de mais experimentos e estudos para a evolução do sistema.

Parcerias: Desenvolvido exclusivamente pelo MPF.

Estágio Atual: Concluído com sucesso.

Considerações: --

1.4. MANIFESTAÇÃO OUVIDORIA

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Classifica manifestações recebidas através de canal de atendimento ao cidadão pela Ouvidoria do MPF, indicando se o direcionamento da demanda deveria ter sido feito à Sala de Atendimento ao Cidadão que recebe denúncias e representações que não são objeto de atuação da Ouvidoria.

Bancos de Dados Utilizados: Foi utilizada base de manifestações protocoladas pelo cidadão no Sistema da Ouvidoria e no Sistema de Atendimento ao Cidadão já existente no MPF, indexada no Sistema Aptus.

Tecnologias Empregadas: Foi utilizada aprendizagem supervisionada, por meio da aplicação de modelo Random Forest.

Resultados Esperados: For obtida, em fase de avaliação do modelo, métrica de precisão para a indicação de responsabilidade da Sala de Atendimento ao Cidadão na ordem de 97%.

Desafios: Não houve.

Parcerias: Desenvolvido exclusivamente pelo MPF.

Estágio Atual: Concluído com sucesso.

Considerações: --

1.5. ATHOS-MPF (NOME PROVISÓRIO)

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Ferramenta para agrupamento de autos por meio de análise de similaridade semântica entre documentos.

Inicialmente, a ferramenta está sendo desenvolvida para a análise de manifestações relacionadas a Recursos Especiais e Extraordinários produzidas pelo MPF, de modo a auxiliar setor responsável pela análise da atuação do MPF na aplicação dos institutos da repercussão geral e da súmula vinculante.

Bancos de Dados Utilizados: Estão sendo realizados testes com cerca de 117 mil íntegras de manifestações relacionadas a recursos especiais ou recursos extraordinários. Estão também sendo realizados testes com cerca de 45 mil íntegras de manifestações relacionadas à atuação do MPF como custos legis.

Tecnologias Empregadas: Estão sendo testados modelos Doc2Vec e Transformers para geração de vetores de embeddings. Além disso, para a realização de agrupamentos, estão sendo testados modelos como HDBSCAN e também algoritmo desenvolvido pelo STJ.

Resultados Esperados: Sistema ainda em desenvolvimento. Espera-se auxiliar os membros do MPF fornecendo informações e ferramentas para análise conjunta de feitos.

Desafios: Sistema encontra-se ainda em desenvolvimento, em fase de testes de POC construída para validação das tecnologias pelos usuários.

Parcerias: Termo de Adesão STJ Nº 03/2023 firmado com o STJ para repasse da tecnologia ao MPF.

Estágio Atual: Em desenvolvimento.

Considerações: --

1.6. INTERNALIZAÇÃO GPSMED

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Extração de informações sobre medicamentos, doenças e tratamentos em processos judiciais em trâmites no MPF, organizando as informações em Painel de Business Intelligence.

As informações a serem coletadas pelo sistema têm por objetivo auxiliar nas análises e diagnósticos de demandas na área de saúde pública, de forma a subsidiar o desenvolvimento de políticas públicas voltadas à prevenção ou desjudicialização de litígios em matéria sanitária.

Bancos de Dados Utilizados: Estão sendo coletadas dados e íntegras de petições iniciais, decisões e sentenças de processos da área de saúde que estão tramitando no MPF. Até o momento, foram coletadas cerca de 500 peças processuais.

Tecnologias Empregadas: Inicialmente, estão sendo implementadas técnicas de extração de dados estruturados por meio de expressões regulares, conforme TJ/RN. Pretende-se evoluir para técnicas de processamento de linguagem natural (Reconhecimento de Entidades Nomeadas - NER)..

Resultados Esperados: Espera-se que o sistema forneça dados para auxiliar os membros na atuação do MPF na área de saúde.

Desafios: Sistema encontra-se em desenvolvimento.

Parcerias: Cooperação técnica realizada com o TJ/RN para repasse da tecnologia ao MPF.

Estágio Atual: Em desenvolvimento.

Considerações: --

1.7. TRIAGEM DE INQUÉRITOS POLICIAIS.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Identificação do motivo da entrada de Inquéritos Policiais que chegam ao MPF (dilação de prazo, relatório final ou outro motivo), através da análise dos textos das peças processuais.

Bancos de Dados Utilizados: Estão sendo coletados dados e íntegras de peças processuais de IPLs que chegam ao MPF, para comporem base de treinamento e avaliação. Até o momento, foram coletadas cerca de 12 mil peças.

Tecnologias Empregadas: Pretende-se utilizar técnicas de processamento de linguagem natural para classificação a partir dos textos das peças processuais.

Resultados Esperados: Espera-se auxiliar os gabinetes com atuação criminal no MPF na realização de triagem de Inquéritos Policiais.

Desafios: Sistema encontra-se em desenvolvimento.

Parcerias: Desenvolvido exclusivamente pelo MPF.

Estágio Atual: Em desenvolvimento.

Considerações: --

2. MINISTÉRIO PÚBLICO DO TRABALHO

2.1. SARA (SISTEMA ASSISTENTE DE REDAÇÃO E ANÁLISE)

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Diversas soluções estão sendo desenvolvidas e testadas em regime de piloto pela Secretaria de Pesquisa e Análise de Informações - SPAI/GabPGT. A principal solução sendo trabalhada é denominada SARA (Sistema Assistente de Redação e Análise).

Bancos de Dados Utilizados: O piloto do SARA utiliza as bases transacionais dos sistemas finalísticos do MPT, assim como outras bases de dados, externas, mantidas pela SPAI.

Tecnologias Empregadas: O SARA estava sendo desenvolvido com infraestrutura onpremises, mas o piloto em curso está utilizando as API da OpenAI, com possível expansão futura através da aquisição de créditos AZURE.

Resultados Esperados: Em ambiente de testes, foi possível dialogar com o SARA para obter resumo de processos, proposição de temas ou outras informações presentes em processos finalísticos do MPT.

Desafios: Falta de pessoal e complexidade de contratação/licenciamento de plataformas em nuvem para a disponibilização de LLMs.

Parcerias: O desenvolvimento é totalmente interno.

Estágio Atual: Em desenvolvimento

Considerações:

3. MINISTÉRIO PÚBLICO DO DISTRITO FEDERAL E TERRITÓRIOS

3.1. ROBOCON

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: – Permite a investigação de danos coletivos ao consumidor a partir do monitoramento dos processos autuados nos Juizados Especiais Cíveis.

Bancos de Dados Utilizados: Utiliza o Diário da Justiça Eletrônico para tomar conhecimento dos processos do Juizado Especial Cível, e em seguida, faz a consulta no PJE para obter detalhes do processo e as peças processuais. Além disso, utiliza a base de pessoa jurídica fornecida pela Receita Federal.

Tecnologias Empregadas: Usa modelos de redução de dimensionalidade LSI e NMF, e Tesseract para OCR dos documentos.

Resultados Esperados: A ferramenta foi utilizada com sucesso pela promotoria para proposição de TACs https://www.mpdf.mp.br/portal/images/pdf/tacs/prodecon/Prodecon_TAC_2021_848.pdf e instrução de processos.

Desafios: Os maiores desafios enfrentados referem-se a extração de dados dos autos dos processos oriundos do PJE. Em relação a formação de equipes multidisciplinares, tem-se como dificuldade reunir servidores da área fim e de tecnologia com a devida capacitação para contribuir no mapeamento de necessidades que justificam a investigação de uma solução tecnológica

Parcerias: Estamos em conversas para realização de testes com a MAAT e a Whoopi.

Estágio Atual: Cada iniciativa está em estágio diferente

Considerações: --

3.2. SISTEMA DE APOIO ÀS PROMOTORIAS DE EXECUÇÃO PENAL (PEPSI)

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Provê visualização integrada e especializada de todas as fontes de dados necessárias ao tratamento dos feitos. Além disso, a identificação automatizada dos pedidos e eventos do processo e a sugestões de manifestações emanadas em casos similares.

Bancos de Dados Utilizados: Coleta dados do PJE (SEEU), serviços extras do SEEU. Pretende-se expandir para consumir dados do BNMP e SIAPEN

Tecnologias Empregadas: Usa Support Vector Machine (SVM) e Tesseract.

Resultados Esperados: A ferramenta está em desenvolvimento. Portanto espera-se:

- Redução no tempo necessário para coletar o material de apoio, considerando que hoje é necessário visitar vários sistemas para coleta do material de apoio;

- Redução no tempo de exame do material, considerando que o sistema pode organizar o material de modo a facilitar o exame;

- Redução no tempo de manifestação, ao sugerir pareceres de casos similares.

Desafios: Os maiores desafios enfrentados referem-se a extração de dados dos autos dos processos oriundos do PJE. Em relação a formação de equipes multidisciplinares, tem-se como dificuldade reunir servidores da área fim e de tecnologia com a devida capacitação para contribuir no mapeamento de necessidades que justificam a investigação de uma solução tecnológica

Parcerias: Estamos em conversas para realização de testes com a MAAT e a Whoopi.

Estágio Atual: Cada iniciativa está em estágio diferente

Considerações: --

3.3. LUMINARIA

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Automação do fluxo de feitos do Juizado Especial Criminal, compreendendo a extração automatizada dos dados e fatos do boletim de ocorrência, e encaminhamento automático (se possível) ou sugestão de encaminhamento por meio de IA generativa.

Bancos de Dados Utilizados: Utiliza dados dos processos do Juizado Especial Criminal onde o MP é parte. Os dados são colhidos do PJE.

Tecnologias Empregadas: Pretende-se experimentar o Azure Document Intelligence, OpenAI GPT e LLama2 para fazer a extração de dados dos documentos do processo.

Resultados Esperados: – A ferramenta está em desenvolvimento. Portanto espera-se:

- Redução no tempo de coleta e consolidação dos dados do Feito presentes apenas nos autos.

- Redução no tratamento de casos simples e repetitivos

Desafios: Os maiores desafios enfrentados referem-se à extração de dados dos autos dos processos oriundos do PJE. Em relação a formação de equipes multidisciplinares, tem-se como dificuldade reunir servidores da área fim e de tecnologia com a devida capacitação para contribuir no mapeamento de necessidades que justificam a investigação de uma solução tecnológica

Parcerias: Estamos em conversas para realização de testes com a MAAT e a Whoopi.

Estágio Atual: Cada iniciativa está em estágio diferente

Considerações: --

3.4. SELFIEPONTO – PONTO ELETRÔNICO ATRAVÉS DO RECONHECIMENTO FACIAL. REGISTRA O PONTO ATRAVÉS DO RECONHECIMENTO FACIAL E LOCALIZAÇÃO GEOGRÁFICA DO SERVIDOR.

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: foi uma iniciativa para resolver o problema que estamos com a necessidade de compra de coletora biométrica a cada 4 anos. Estamos testando a possibilidade de substituí-las pelo ponto registrado diretamente do celular do servidor/estagiário. Para isto estamos utilizando 2 tecnologias o reconhecimento facial e o georreferenciamento para validar se o servidor se encontra no perímetro estabelecido das instalações do MPDFT

Bancos de Dados Utilizados: Banco de dados do sistema de gestão de pessoa

Tecnologias Empregadas: Georef- Ferramenta Azure Map e Rec Facial- ferramenta Face API Azure

Resultados Esperados: – Esperamos acabar com as coletoras de biometria existente, gerando uma economia para o órgão.

Desafios: Os maiores desafios enfrentados referem-se a extração de dados dos autos dos processos oriundos do PJE. Em relação a formação de equipes multidisciplinares, tem-se como dificuldade reunir servidores da área fim e de tecnologia com a devida capacitação para contribuir no mapeamento de necessidades que justificam a investigação de uma solução tecnológica

Parcerias: Estamos em conversas para realização de testes com a MAAT e a Whoopi.

Estágio Atual: Cada iniciativa está em estágio diferente

Considerações: --

3.5. RISCO VD

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Análise de risco conforme o questionário de violência doméstica que acompanha os inquéritos policiais envolvendo crimes dessa natureza.

Bancos de Dados Utilizados: Questionário de Risco vindo juntado ao Inquérito Policial vindo do Pje.

Tecnologias Empregadas: Tesseract, LLama2

Resultados Esperados: A ferramenta está em desenvolvimento. Portanto espera-se:

- Redução no tempo necessário para coletar o material de apoio, considerando que hoje é necessário visitar vários sistemas para coleta do material de apoio;
- Redução no tempo de exame do material, considerando que o sistema pode organizar o material de modo a facilitar o exame;
- Redução no tempo de manifestação, ao sugerir pareceres de casos similares.

Desafios: Os maiores desafios enfrentados referem-se a extração de dados dos autos dos processos oriundos do PJE. Em relação a formação de equipes multidisciplinares, tem-se como dificuldade reunir servidores da área fim e de tecnologia com a devida capacitação para contribuir no mapeamento de necessidades que justificam a investigação de uma solução tecnológica

Parcerias: Estamos em conversas para realização de testes com a MAAT e a Whoopi.

Estágio Atual: Cada iniciativa está em estágio diferente

Considerações: --

3.6. SISTEMA INTELIGENTE DE INTERPRETAÇÃO DE AUTOS (SINTIA)

Estrutura de IA: Sim

Uso de Tecnologias de IA: Não

Resumo da Iniciativa: - Uso de inteligência artificial generativa para sumarização de processos judiciais eletrônicos de natureza cível e familiar.

Bancos de Dados Utilizados: Processos Judiciais Cíveis e de Família oriundos do Pje.

Tecnologias Empregadas: Tesseract, OpenAI GPT e LLama2 para fazer a extração de dados dos documentos do processo.

Resultados Esperados: A ferramenta está em desenvolvimento. Portanto espera-se:

- Redução no tempo necessário para coletar o material de apoio, considerando que hoje é necessário visitar vários sistemas para coleta do material de apoio;
- Redução no tempo de exame do material, considerando que o sistema pode organizar o material de modo a facilitar o exame;
- Redução no tempo de manifestação, ao sugerir pareceres de casos similares.

Desafios: Os maiores desafios enfrentados referem-se a extração de dados dos autos dos processos oriundos do PJE. Em relação a formação de equipes multidisciplinares, tem-se como dificuldade reunir servidores da área fim e de tecnologia com a devida capacitação para contribuir no mapeamento de necessidades que justificam a investigação de uma solução tecnológica

Parcerias: Estamos em conversas para realização de testes com a MAAT e a Whoopi.

Estágio Atual: Cada iniciativa está em estágio diferente

Considerações: --

4. MINISTÉRIO PÚBLICO DO ESTADO DO ACRE

4.1. LLAMA 2 DA EMPRESA META

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Um LLM (Large Language Model), modelos de linguagem grandes em português, para automatizar tarefas repetitivas que levariam horas e concluí-las em apenas alguns minutos blip image captioning large (em testes)

É um modelo de linguagem pré treinado para entender e descrever o conteúdo de uma imagem ChatPDF.

Modelo de linguagem que permite a leitura e compreensão de arquivos PDF para gerar resumos e/ou responder dúvidas relacionadas ao conteúdo'-

O objetivo do uso do Blip é fazer uma varredura por várias imagens e buscar por palavras chaves, como "drogas, armas e tatuagens" para melhorar o resultado de nossas aplicações.

O objetivo do ChatPDF é poder cruzar informações de vários arquivos PDFs e extrair informações que levariam muito tempo (ou passaria despercebido) se feito por humanos.

Bancos de Dados Utilizados: Ainda não trabalhamos com aprendizado de máquina.

Tecnologias Empregadas: Llama2, OpenAI, gpt3, blip, chatpdf.

Resultados Esperados: Resultados obtidos, mas ainda em desenvolvimento:

- Conversação com AI para tirar dúvidas sobre tópicos diversos, interpretação de imagens e descrevendo o que tem na imagem, autocompletar para frases não terminadas
- Usar AI para ler e compreender arquivos PDF para cruzar informações, gerar resumos e responder dúvidas sobre o conteúdo interpretado
- Reconhecimento facial.

Desafios: Infraestrutura talvez não seja o suficiente para tarefas mais pesadas por falta de GPU nos servidores
- Curva de aprendizado longa - Falta de equipe para dedicar-se exclusivamente ao desenvolvimento de soluções com A.I.

Parcerias: Não.

Estágio Atual: Recém iniciado (menos de 3 meses). Em testes

Considerações:

5. MINISTÉRIO PÚBLICO DO ESTADO DO AMAPÁ

5.1 ASSISTENTE CHATBOT COM ASSISTENTE VIRTUAL – PONTO ELETRÔNICO COM RECONHECIMENTO FACIAL

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Sistema da ouvidoria

Tecnologias Empregadas: Azure

Resultados Esperados: Melhorar a qualidade do atendimento robotizado dos cidadãos através do assistente virtual

Desafios: O refinamento entre as ferramentas envolvidas, é algo que ainda está sendo melhorado

Parcerias: Não

Estágio Atual: Planejamento

Considerações: --

6. MINISTÉRIO PÚBLICO DO ESTADO DA BAHIA

6.1. FRATRIA - FERRAMENTA DE AUTOMAÇÃO DE TAREFAS COM RECURSO EM INTELIGÊNCIA ARTIFICIAL DO MPBA

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Esta tecnologia desenvolvida internamente pela equipe do MPBA tem por objetivo receber procedimentos extrajudiciais, iniciando por Inquéritos Policiais, para, após a leitura de todo o procedimento, usando LLM, sugerir qual decisão deve ser tomada pelo membro (ex: Denúncia, diligência, arquivamento, cumulando ou não) e, ao fim, preminutar a peça processual sugerida, já com os dados necessários, tudo sob supervisão humana. Também usamos como análise preditiva o chat do TEAMS.

Bancos de Dados Utilizados: A ferramenta FRATRIA utiliza um conjunto de tecnologias, entre as quais OCR, scraping, LLM, etc., com intuito de melhor atingir seu objetivo, mas com redução de custos e maior eficiência. Ela pode se adaptar a qualquer modelo de IA, mas neste momento está acoplada a nuvem da AZURE, usando o GPT. Tecnologias Empregadas: GPT - OpenAI no Microsoft Azure, R Project, Python, IA Builder e SQL Server

Tecnologias Empregadas: Conforme explicação do item anterior

Resultados Esperados A iniciativa se encontra em fase de testes e refinamento, já em vias de utilização em promotoria piloto.

Desafios: Inúmeros desafios foram encontrados para essa iniciativa: inquéritos desorganizados, com arquivos de imagens incompatíveis com leitura OCR, dificuldade de contratação respeitando a lei de licitações e a LGPD, do modelo de linguagem, elaboração e refino da engenharia de Prompts, limitação de tokens, custo, entre outros

Parcerias: Esta iniciativa, em específico, está sendo desenvolvida totalmente com equipe própria do MPBA.

Estágio Atual: Em desenvolvimento

Considerações: --

7. MINISTÉRIO PÚBLICO DO ESTADO DO CEARÁ

7.1. ACORDOS DE NÃO-PERSECUÇÃO PENAL

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Desenvolver ferramenta organizacional, dotada de mecanismos de automação e inteligência artificial, que possibilite agilizar a análise de inquéritos policiais passíveis de acordos de não-persecução penal. Realiza-se uma seleção dos processos na fila da promotoria que são passíveis de aplicação de um Acordo de Não-Persecução Penal - ANPP. A partir da seleção inicial, permite-se ao usuário escolher uma “parte investigada” de um processo e gerar um relatório socioeconômico e histórico processual criminal, aplicando mecanismos de interoperabilidade entre diversos sistemas internos e externos ao MPCE para pesquisa de informações sobre a parte escolhida.

Pretende-se agilizar a proposição de acordo de não-persecução penal, pois a análise desses tipos penais incorre em entrar em diversos sistemas, internos e externos ao MPCE e, em seguida, organizar e montar um perfil resumido com relação ao autor, a fim de visualizar informações econômicas, sociais e criminais, vitais para elaboração das penas a serem aplicadas nos acordos.

A aplicação filtrará os processos passíveis de acordo na promotoria e pesquisará as informações relativas aos autores nas bases do SIMPCE, SAJMP, TJCE e outras, objetivando apresentar dados diversos e úteis em um único relatório. Quando da elaboração documental de uma proposta de acordo, o promotor redige o texto com base em seu conhecimento tácito, com a aplicação o promotor poderá tomar decisões com base em dados concretos, diminuindo a diferenciação inerente a entrevista em tempo de audiência, na qual pode ocorrer visão parcial dos fatos.

Com os resultados desta aplicação, futuramente, o MPCE poderá guardar uma maior quantidade de dados sobre os acordos firmados e desenvolver painéis gráficos capazes de indicar tendências de perfis criminais por regiões do estado, embasar a criação de políticas públicas, priorizar ações de fiscalização junto a órgãos estaduais, elaborar planos de atuação específicos para realidade de cada município, fomentar ações de prevenção a incidentes criminais. A gestão baseada em informações e algoritmos de inteligência objetiva alcançar um maior nível de maturidade organizacional na previsão do acontecimento de crimes, e a respeito do retorno social e econômico que o desempenho direcionado do MPCE promoverá para a sociedade.

Bancos de Dados Utilizados:

MPCE – Processos do Ministério Público do Estado do Ceará

RECEITA FEDERAL: Secretaria da Receita Federal do Brasil

RAIS/CAGED: Relação Anual de Informações Sociais/ Cadastro Geral de Empregados e Desempregados

CDL: Câmara de Dirigentes Lojistas

TJCE: Processos do Tribunal de Justiça do Estado do Ceará

SINDIÔNIBUS: Sindicato das Empresas de Transporte de Passageiros do Estado do Ceará

NOTAIS FISCAIS: Documentos Fiscais e Notas Fiscais

SISOBI: Sistema de Observação Aérea da Secretaria de Segurança Pública

CAGECE: Companhia de Água e Esgoto do Ceará

DETRAN/CE: Departamento Estadual de Trânsito do Ceará

TCE/CE: Tribunal de Contas do Estado do Ceará

AERONAVES (RAB): Registro Aeronáutico Brasileiro

EMBARCAÇÕES (MARINHA): Registro de Embarcações na Marinha do Brasil

SAP/SISPEN/SISGEPEN (SEJUS): Sistema Penitenciário / Sistema de Gestão Penitenciária da Secretaria de Justiça do Estado

BOLSA FAMÍLIA / AUXÍLIO EMERGENCIAL: Programa Bolsa Família / Auxílio Emergencial do Governo Federal

Tecnologias Empregadas: GPT Azure, OCR

Resultados Esperados:

Melhor aproveitamento da força de trabalho em atividades não repetitivas.

Diminuição do tempo de proposição de transação penal e acordos de não-persecução penal.

Diminuição do tempo dispendido em coletar e organizar as informações sobre os autores.

Diminuição do tempo para se encontrar as partes do processo para intimação de audiência.

Desafios: O principal desafio tem sido a integração da solução ANPP com a ferramenta SAJMP, devido a concorrência do projeto ANPP com a fila de outras demandas de melhorias deste sistema.

Parcerias: A solução foi construída no MPCE, em parceria do LINO-Laboratório de Inovação do MPCE com a equipe de Soluções da Área-Fim (SAF) da Secretaria de Tecnologia da Informação do MPCE (SETIN). A solução construída está sendo integrada ao sistema processual do MPCE, o SAJMP, que é uma ferramenta contratada pelo MPCE da empresa privada Softplan.

Estágio Atual: Em desenvolvimento

Considerações: --

7.2. CONECTA.CONSUMIDOR

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Desenvolver uma solução tecnológica dotada de mecanismos de automação que possibilitem agilizar o atendimento ao consumidor no DECON – Programa Estadual de Proteção e Defesa do Consumidor, bem como fornecer painéis gerenciais dotados de gráficos e indicadores das relações consumeristas no estado do Ceará.

Atualmente, é imprescindível estabelecer uma conexão direta entre o DECON e o consumidor de todo o estado do Ceará, utilizando as ferramentas de comunicação mais difundidas no mercado, a fim de proporcionar um atendimento ágil, eficiente e tecnologicamente avançado. Nesse sentido, o canal de atendimento pelo WhatsApp tem se destacado como uma das principais formas de comunicação entre o setor de atendimento do DECON e os consumidores que buscam seus serviços. Cerca de 85% das interações realizadas pelo setor são feitas através desse canal, atendendo diariamente de 190 a 215 pessoas, abrangendo uma ampla variedade de demandas.

No entanto, é importante salientar que as demandas recebidas não se limitam apenas a reclamações e dúvidas típicas do setor de atendimento, que representam apenas cerca de 30% do total. Além disso, existe uma parcela significativa, correspondendo a 70% das demandas, que são represadas fora do horário de atendimento. Esse cenário evidencia a necessidade de otimizar os recursos do DECON, considerando ferramentas que permitam a triagem adequada das demandas, filtrando aquelas pertinentes ao DECON e redirecionando outras para os setores responsáveis correspondentes. Além disso, é importante implementar mecanismos que limitem o recebimento e o tratamento das demandas nos horários adequados de atendimento.

Com o intuito de aprimorar o serviço prestado pelo DECON, é fundamental adotar uma abordagem mais eficiente na gestão das demandas, priorizando aquelas relacionadas diretamente ao papel e competências do órgão. Dessa forma, é possível direcionar os recursos disponíveis de forma mais estratégica, maximizando o potencial de atendimento e incrementando a qualidade dos serviços prestados aos consumidores.

Em resumo, a implementação de uma automatização no atendimento ao consumidor no DECON baseia-se na necessidade de ampliar a conexão direta com os cidadãos, através do aumento do número de atendimento, possibilitado pela aplicação de ferramentas de automação amplamente utilizadas, como o chatbot para WhatsApp. Além disso, a visualização e análise de painéis de Business Intelligence (BI) poderá otimizar os recursos do DECON, por meio da triagem adequada das demandas e do redirecionamento para os setores responsáveis, permitirá um crescimento do atendimento e uma gestão mais eficiente das demandas, superando os desafios relacionados ao volume de solicitações e aos horários de atendimento.

Bancos de Dados Utilizados: Dados estruturados: Consumidor.gov; Proconsumidor

Tecnologias Empregadas: Chatbot

Resultados Esperados:

- Atendimento com horário delimitado pelo MPCE: o chatbot estará disponível para atendimento aos consumidores em tempo limitado, incluindo horários do expediente de trabalho, proporcionando uma resposta imediata e evitando o represamento de demandas.

- Triage eficiente de demandas: o chatbot será capaz de realizar uma triagem inicial das demandas recebidas, direcionando-as aos setores responsáveis dentro do DECON de acordo com a sua natureza, garantindo um encaminhamento mais rápido e assertivo. Deixando para o atendimento humano as demandas mais complexas.
- Foco em demandas complexas: com o chatbot lidando com demandas mais simples e frequentes, os atendentes humanos poderão se dedicar a casos mais complexos, que exigem análise aprofundada e intervenção especializada, resultando em um atendimento mais qualificado.
- Agilidade no atendimento: o chatbot será capaz de fornecer respostas instantâneas a perguntas frequentes e demandas mais simples, reduzindo o tempo de espera dos consumidores e agilizando o fluxo de atendimento.
- Resolução de problemas comuns: o chatbot poderá oferecer soluções imediatas para problemas e dúvidas comuns dos consumidores, evitando a necessidade de intervenção humana e reduzindo a carga de trabalho dos atendentes.
- Redução de custos operacionais: com a automatização dos atendimentos, haverá uma redução significativa nos custos operacionais, pois um chatbot pode lidar com um grande volume de demandas sem exigir recursos humanos adicionais.
- Melhoria na experiência do consumidor: ao proporcionar um atendimento ágil e eficiente, o projeto de automação com chatbot contribuirá para uma experiência positiva por parte dos consumidores, aumentando sua satisfação com o serviço prestado pelo DECON.
- Monitoramento e análise de dados: o chatbot poderá coletar e analisar dados sobre as demandas dos consumidores, fornecendo informações valiosas para a identificação de tendências, necessidades específicas e oportunidades de melhoria no atendimento.
- Aumento da eficiência operacional: a automação dos atendimentos permitirá que o DECON melhore sua eficiência operacional, otimizando o fluxo de trabalho, reduzindo gargalos e proporcionando uma alocação mais eficaz dos recursos disponíveis.
- Fortalecimento da imagem institucional: a implementação de um projeto de automação com chatbot demonstra o compromisso do DECON em adotar tecnologias inovadoras para melhor atender aos consumidores, fortalecendo sua imagem como um órgão moderno, acessível e eficiente.
- Reuso da solução em outros setores: o case com o DECON poderá servir de exemplo para utilização do chatbot em outros setores que necessitam de atendimento ao público, podendo ser uma solução facilmente replicável.

Desafios: O principal desafio será a contratação de ferramenta de chatbot, devido a limitações de formas de pagamentos a empresas privadas que operam este tipo de tecnologia: restrita a cartão de crédito, pagamento por tarifação de usuários, conexão de pagamento ao WhatsApp, dentre outras.

Parcerias: instituições públicas, privadas ou da Academia (Identifique) *A solução está sendo construída no MPCE, em parceria do LINO-Laboratório de Inovação do MPCE com as equipes do DECON.

Estágio Atual: Em desenvolvimento

Considerações: --

7.3. PORTAL DE EXPLORAÇÃO DE EVIDÊNCIAS DIGITAIS (PEED)

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: O PEED é uma ferramenta em desenvolvimento no LINO-Laboratório de Inovação do MPCE que tem como principal objetivo automatizar a identificação e análise de informações provenientes de

diversas fontes, sejam elas estruturadas ou não. Utilizando técnicas de ciência de dados e inteligência artificial, o PEED é uma colaboração entre o LINO, o Ministério Público de Minas Gerais e a Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap), que conecta as entidades de pesquisa e desenvolvimento da Universidade de Fortaleza (Unifor) e do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE).

As principais funcionalidades do PEED incluem:

1. **Reconhecimento Facial:** funcionalidade que analisa rostos em imagens para identificar relacionamentos entre pessoas ou mesmo indicar possível identidade. Aplicação: Imagine que um investigador tenha várias fotos de indivíduos relacionados a um caso criminal. O PEED pode ser usado para identificar automaticamente e relacionar esses rostos a registros de suspeitos, testemunhas ou vítimas, auxiliando na identificação de pessoas-chave em investigações.
2. **Análise de Contexto de Conversas:** algoritmo capaz de extrair significado por trás das interações verbais. Aplicação: Considere um grande volume de mensagens de texto ou e-mails relacionados a uma investigação. O PEED pode analisar o contexto e o tom das conversas para detectar ameaças, conspirações ou informações relevantes, permitindo uma compreensão mais profunda das comunicações.
3. **Elaboração de Trilhas de Pesquisa de Dados Estruturados:** Auxílio na organização e análise de informações estruturadas. Aplicação: Em um processo judicial complexo, com inúmeras partes envolvidas, o usuário do PEED pode criar trilhas de pesquisa para encontrar rapidamente informações específicas em diversas bases de dados, como datas, veículos, nomes de pessoas e empresas ou números de referência. Isso economiza tempo na análise de diversas bases de dados.
4. **Busca com Ferramentas Elastic Search em Peças Processuais:** Facilidade de pesquisa em documentos processuais gerados no MPCE. Aplicação: Em um contexto jurídico, quando se tem um grande volume de documentos processuais criados ou manipulados, o usuário do PEED pode pesquisar livremente (como uma pesquisa em buscadores da internet) com base no Elastic Search para localizar jurisprudência relevante, leis, precedentes, partes citadas em processos, endereços, telefones em processos judiciais passados, agilizando a busca no acervo processual.
5. **Emissão de Relatórios Socioeconômicos e Histórico Processual:** Apoio na tomada de decisões por meio de informações detalhadas sobre as partes investigadas. Aplicação: Suponha que um membro precise tomar uma decisão sobre a multa de um réu. Este usuário do PEED poderá gerar relatórios que incluem informações sobre o histórico criminal da pessoa, bem como dados socioeconômicos que auxiliam na determinação apropriada da multa a ser aplicada.

Portanto, o PEED representa uma ferramenta aliada na busca por evidências digitais e no apoio à tomada de decisões.

Bancos de Dados Utilizados:

Dados estruturados:

- MPCE – Processos do Ministério Público do Estado do Ceará
- RECEITA FEDERAL: Secretaria da Receita Federal do Brasil
- RAIS/CAGED: Relação Anual de Informações Sociais/ Cadastro Geral de Empregados e Desempregados
- CDL: Câmara de Dirigentes Lojistas
- TJCE: Processos do Tribunal de Justiça do Estado do Ceará
- SINDIÔNIBUS: Sindicato das Empresas de Transporte de Passageiros do Estado do Ceará
- NOTAIS FISCAIS: Documentos Fiscais e Notas Fiscais
- SISOBI: Sistema de Observação Aérea da Secretaria de Segurança Pública
- CAGECE: Companhia de Água e Esgoto do Ceará

- DETRAN/CE: Departamento Estadual de Trânsito do Ceará
- TCE/CE: Tribunal de Contas do Estado do Ceará
- AERONAVES (RAB): Registro Aeronáutico Brasileiro
- EMBARCAÇÕES (MARINHA): Registro de Embarcações na Marinha do Brasil
- SAP/SISPEN/SISGEPEN (SEJUS): Sistema Penitenciário / Sistema de Gestão Penitenciária da Secretaria de Justiça do Estado
- BOLSA FAMÍLIA / AUXÍLIO EMERGENCIAL: Programa Bolsa Família / Auxílio Emergencial do Governo Federal
- Planilhas CSV que o usuário poderá inserir na ferramenta

Dados não-estruturados:

- Extrações de equipamentos apreendidos em operações
- Conversas de WhatsApp
- Fontes de imagens que o usuário do PEED poderá inserir na ferramenta

Tecnologias Empregadas: Modelos de linguagem (BERTimbau), OCR, algoritmos de aprendizagem de máquina, reconhecimento de padrões (imagem e áudio), algoritmos de reconhecimento de entidades nomeadas (REN).

Resultados Esperados:

- Eficiência Aprimorada: A ferramenta PEED tem agilizado a identificação e análise de fontes não-estruturadas, permitindo que os órgãos de investigação do MPCE otimizem seus processos e economizem tempo.
- Precisão Aprimorada: A utilização de técnicas de inteligência artificial tem aprimorado a precisão na detecção de informações relevantes em dados estruturados e não-estruturados, ajudando a identificar pessoas, por exemplo.
- Expansão para Centros de Apoio: A decisão de expandir o uso da ferramenta PEED para os Centros de Apoio tem o potencial de enriquecer as bases de dados disponíveis. Isso permitirá uma análise mais abrangente, fortalecendo as capacidades de investigação e tomada de decisão em todo o MPCE.
- Aproximação com outros Ministérios Públicos e o meio acadêmico: A parceria com a o MPMG e a FUNCAP tem proporcionado trazer tecnologia de ponta para o MP, bem como o reuso de soluções já adotadas com sucesso em outros MPs.
- Mudança cultural: A introdução de ferramentas como o PEED em uma instituição jurídica representa uma mudança cultural significativa que afeta tanto os procedimentos quanto a mentalidade dos profissionais envolvidos. Essas mudanças podem ser percebidas da seguinte forma:
 - Modernização da Prática Jurídica: A adoção de ferramentas tecnológicas como o PEED representa a modernização da prática jurídica. Isso incentiva os profissionais do setor a se adaptarem a um ambiente digital e a abraçarem novas tecnologias como parte integrante de seu trabalho cotidiano.
 - Foco na Eficiência e Produtividade: As funcionalidades do PEED foram projetadas para melhorar a eficiência e produtividade. Isso motiva os profissionais a buscar maneiras mais ágeis de realizar tarefas, economizar tempo e recursos, e reduzir a burocracia.
 - Ênfase na Colaboração e Aprendizado Contínuo: A adoção do PEED, por utilizar computação em nuvem, IA, e outras ferramentas integradas, incentiva a colaboração entre diferentes profissionais e setores dentro do MPCE. Isso cria uma cultura de aprendizado contínuo, onde as equipes compartilham conhecimentos e experiências para melhorar o uso das tecnologias.

- **Mudança na Abordagem Analítica:** Ferramentas como o PEED utilizam análise de dados e inteligência artificial. Isso muda a forma como os membros e outros profissionais abordam a pesquisa, análise de casos e tomada de decisões, incorporando uma abordagem mais orientada por dados.

- **Ênfase na Segurança de Dados:** A mudança cultural inclui a conscientização sobre a importância da segurança de dados, especialmente quando se lida com informações sensíveis relacionadas dados pessoais. Os profissionais devem estar cientes das práticas de segurança e regulamentos de proteção de dados.

- **Resistência Inicial e Capacitação Necessária:** A introdução de novas ferramentas pode encontrar resistência inicial, pois os profissionais podem se sentir desconfortáveis com a mudança. No entanto, programas de capacitação eficazes podem ajudar a superar essa resistência, capacitando as pessoas a utilizar eficazmente as novas tecnologias.

Desafios: A construção do PEED enfrentou alguns desafios, com dois dos principais sendo a contratação de infraestrutura na nuvem e a rotatividade de profissionais de TI. Abaixo, seguem esses desafios em detalhes:

(1) **Contratação de Infraestrutura na Nuvem:**

- **HUB de Tecnologia da Informação e Comunicação:** a empresa pública estadual ETICE possui um HUB de aquisições e contratações de TIC que agiliza o processo de contratação de outros entes do Governo do Estado. O MPCE seguiu a estratégia de contratar por meio desse HUB, dada a inovação desse tipo de contratação e suas peculiaridades no setor público.
- **Dimensionamento Adequado:** Garantir que a infraestrutura na nuvem seja dimensionada adequadamente para atender às demandas do PEED foi um fator crítico. Foram diversas rodadas de análise com a equipe de infraestrutura de TI do MPCE, arquitetos em nuvem e a equipe de pesquisadores da FUNCAP.
- **Segurança e Conformidade:** A segurança dos dados e a conformidade com regulamentações sempre foram preocupações básicas. Garantir que os dados armazenados na nuvem estivessem protegidos e em conformidade com as leis e regulamentações aplicáveis foi um desafio constante.
- **Alta Performance:** Garantir alta performance dos serviços na nuvem foi crucial para evitar interrupções e baixo desempenho no funcionamento do PEED.

(2) **Rotatividade de Profissionais de TI:**

- **Aquecimento do mercado de TI na pandemia COVID-19:** Houve alta rotatividade de profissionais de TI resultante do aquecimento desse nicho de mercado durante a pandemia. Foi desafiador garantir que os profissionais especializados estivessem devidamente capacitados e treinados para construir o PEED.
- **Continuidade Operacional:** Essas mudanças de mercado afetaram a forma de se projetar a continuidade operacional do PEED no MPCE. Foi importante investir em documentação adequada e planos de contingência a fim de minimizar interrupções no desenvolvimento e nas operações do PEED.
- **Gestão do Conhecimento:** A transferência de conhecimento de profissionais que deixam a equipe para aqueles que entram é fundamental. A criação de um sistema eficaz de gestão do conhecimento foi desafiadora e ainda se encontra em aprimoramento. Para superar esses desafios, o MPCE contou com um planejamento cuidadoso, investimento em capacitação da equipe e parcerias sólidas internas (SETIN, NUGEP, NATI, Órgãos de Investigação) e externas, além de uma abordagem proativa para lidar com a rotatividade de profissionais de TI.

Parcerias: O PEED é resultado de uma colaboração entre o LINO – Laboratório de Inovação do MPCE, o Ministério Público de Minas Gerais e a Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap), que conectou as entidades de pesquisa e desenvolvimento da Universidade de Fortaleza (Unifor) e do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE).

Estágio Atual: Em operação assistida

Considerações: --

8. MINISTÉRIO PÚBLICO DO ESTADO DO ESPÍRITO SANTO

8.1. JULIUS - FERRAMENTA DE HIPERAUTOMAÇÃO PARA ANÁLISE DE DADOS DE TRANSPARÊNCIA PÚBLICA, FOCANDO INICIALMENTE NA IDENTIFICAÇÃO DE SERVIDORES PÚBLICOS COM ACÚMULOS DE CARGOS.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Bot que extrai dados estruturados e não estruturados de todos portais da transparência dos Municípios do ES, consolida em uma base e aplica ferramentas de análise. Solução toda em código aberto, containerizada e disponibilizada no git: <https://github.com/mpes-uis/Julius>

Bancos de Dados Utilizados: Banco gerado diretamente pelo Bot com todos os dados de todos portais da transparência do ES.

Tecnologias Empregadas: OCR, PLN, APIs, Data Mining

Resultados Esperados Consolidação e estruturação de base de dados públicas com informações de contratos, servidores, licitações, compras, e outros elementos, de todos municípios do estado.

Desafios: Servidores que fornecem as APIs públicas suportarem a quantidade de requisições, priorização de tarefas.

Parcerias: Não. Iniciativa 100% realizada pelos servidores do MPES.

Estágio Atual: Em desenvolvimento

Considerações: --

9. MINISTÉRIO PÚBLICO DO ESTADO DE GOIÁS

9.1. ANÁLISE DE NOTÍCIA DE FATO PARA INFORMAR OUTRAS SIMILARES

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Banco de dados de autos digitais

Tecnologias Empregadas: OCR

Resultados Esperados: análise de identidade com novas notícias de fato.

Desafios: Análise dos dados

Parcerias: Desenvolvimento com equipe própria da instituição.

Estágio Atual: em desenvolvimento

Considerações: --

9.2. IDENTIFICAÇÃO DE DISTORÇÕES DE PAGAMENTO AOS INTEGRANTES DA INSTITUIÇÃO

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Banco de dados de autos digitais e folha de pagamento.

Tecnologias Empregadas: OCR

Resultados Esperados: otimização da auditoria em folha de pagamento, permitindo realizar direcionamento do acompanhamento de situações que variam em relação à normalidade.

Desafios: Análise dos dados

Parcerias: Desenvolvimento com equipe própria da instituição.

Estágio Atual: Concluído com sucesso.

Considerações: --

10. MINISTÉRIO PÚBLICO DO ESTADO DE MATO GROSSO DO SUL

10.1 PROPOSITOR DE DENÚNCIAS DE VIOLÊNCIA DOMÉSTICA

Estrutura dedicada para projetos com IA: Sim.

Uso de Tecnologias de IA: Sim.

Resumo da Iniciativa: O projeto visa desenvolver um sistema capaz de propor denúncias de violência doméstica de forma automatizada, melhorando a eficiência e a rapidez no atendimento desses casos.

Bancos de Dados Utilizados: Banco de Dados do SAJMP.

Tecnologias Empregadas: OCR, GPT - Azure, LLama2.

Resultados Esperados: Identificação assertiva de casos de violência doméstica, agilizando o processo de denúncia.

Desafios: Limitação da janela de contexto, fine-tuning dos modelos para contexto legal específico.

Parcerias: Desenvolvimento interno, sem parcerias externas.

Estágio Atual: Em desenvolvimento.

Considerações: --

10.2 RECONHECIMENTO DE ENTIDADES EM PROCESSOS

Estrutura de IA: Sim.

Uso de Tecnologias de IA: Sim.

Resumo da Iniciativa: Este projeto foca no reconhecimento automático de entidades legais em processos judiciais, visando melhorar a precisão e a velocidade na análise de documentos.

Bancos de Dados Utilizados: Banco de Dados do SAJMP.

Tecnologias Empregadas: GPT - Azure, BERT.

Resultados Esperados: Capacidade aprimorada de identificar entidades legais em textos processuais.

Desafios: Limitação da janela de contexto, ajuste fino dos modelos para contexto legal específico.

Parcerias: Desenvolvimento interno, sem parcerias externas.

Estágio Atual: Em desenvolvimento.

Considerações: --

11. MINISTÉRIO PÚBLICO DO ESTADO DE MINAS GERAIS

11.1 ÁQUILA.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: sistema de reconhecimento de pessoas por meio de faces e aspectos biométricos (tatuagens) e para detecção de objetos de interesse em imagens e vídeos.

O ÁQUILA tem como objetivo auxiliar o MPMG no reconhecimento de pessoas e detecção de objetos de interesse utilizando imagens e vídeos, tanto de domínio aberto quanto geradas por sistemas de outras repartições e órgãos públicos. Do ponto de vista computacional, a ferramenta realiza as operações de detecção e reconhecimento de pessoas e objetos usando estratégias e conceitos ligados a diferentes tarefas de Visão Computacional, como classificação, detecção e recuperação de informação. Do ponto de vista prático, módulos de reconhecimento e detecção foram implementados com o intuito de “ensinar” à máquina os padrões existentes nas imagens/vídeos de forma que, posteriormente, o sistema seja capaz de realizar tais padrões de forma automática, retornando ao usuário, “quem” (pessoa) e/ou “o quê” (objeto) está presente na imagem ou vídeo.

Bancos de Dados Utilizados: Para a parte de reconhecimento de pessoas, dataset públicos foram utilizados, como CASIA-WebFace, LFW e YaleB. Para a parte de reconhecimento de pessoas por aspectos biométricos (ou seja, tatuagens) foi encontrado somente um dataset público que pôde ser utilizado para esta tarefa: o DeMSI. Diante disso, foi necessário criar um dataset para essa tarefa, usando imagens de tatuagens públicas recuperadas em diferentes websites (como Flickr, Google, Reddit, etc.). Após a coleta, dataset ficou com a seguinte distribuição de classes: Águia, Coringa, Escorpião, Índia, Tigre, Teia de aranha, Lágrima, Caveira, Palhaço, Cobra, Arame farpado, Demônio-da-Tasmânia, Chucky, Polvo, Duende, Saci pererê, Morte com fuzil, Mago, Nossa Senhora Aparecida, Diabo, Carpa, Coração transfixado, Aranha, Cruz, Jesus Crucificado, Jesus. Para a parte de detecção de objetos, diversos dataset públicos, de propósito geral, puderam ser utilizados. Entretanto, dada a definição das classes de interesse, isto é, veículos (bicicleta, carro, moto, ônibus, caminhão, etc) e armas (pistola, metralhadora, canivete, facão, faca, etc), não foi possível encontrar um dataset com essas classes em número significativo. Dessa forma, um novo conjunto de dados foi criado. O primeiro passo foi buscar, em conjuntos de dados públicos (como no VOC Pascal 2012, MS COCO, e OD-Weapon), imagens que continham essas classes. Em seguida, para complementar o dataset, foi realizada uma busca na internet (usando crawlers que coletaram imagens do Google, Bing e Naver). O dataset ficou com a seguinte distribuição de classes: pistola, faca, bicicleta, carro, ônibus, moto, caminhão, facão, faquinha, machado, rifle/sniper, shotgun, metralhadora, submetralhadora.

Tecnologias Empregadas: WSO2 (autenticação, gerenciamento de identidade, segurança), JavaScript, CSS, HTML, React (para desenvolvimento da Interface Web), FastAPI e Python (Backend), Redes neurais artificiais.

Resultados Esperados: Do ponto de vista negocial, a ferramenta encontra-se em fase de testes de homologação por usuários internos, sendo esperado que a ferramenta, ao realizar a identificação de pessoas relacionadas a atividades criminosas, auxilie sobremaneira os órgãos de execução nos processos investigativos.

Do ponto de vista tecnológico, os resultados alcançados nas diferentes áreas de reconhecimento e detecção são expressos em termos de mean Average Precision (mAP). Para o reconhecimento de pessoas usando faces, o método CurricularFace foi escolhido como o mais eficaz após testar vários métodos e pré-processamentos. No caso do reconhecimento de pessoas através de tatuagens, o método CosFace se destacou como a escolha mais adequada. Para o módulo de detecção de objetos, o método YoloV5 foi selecionado com base nos resultados obtidos. Essas escolhas são baseadas em métricas de desempenho e indicam a eficácia dos métodos escolhidos para cada tarefa, com mAPs de aproximadamente 95%, 65% e 79%, respectivamente, demonstrando o sucesso das abordagens selecionadas. Como resultado dessas conquistas, o sistema Águila foi desenvolvido e está disponível como uma interface web de fácil usabilidade. Além disso, a plataforma permite a incorporação de novos repositórios de dados, ampliando ainda mais sua funcionalidade e utilidade.

Desafios: Com o objetivo e o escopo definidos, metas foram traçadas para desenvolver o sistema. O principal desafio foi, dentro dos limites e fronteiras estabelecidos, utilizar ferramentas que o tornassem mais robusto, tendo em vista as situações reais às quais o sistema será utilizado. Como solução para esse problema, o sistema foi preparado para receber novos dados constantemente e, dessa forma, aprender novas características, sempre que possível. Com isso, o aprendizado é atualizado continuamente.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações: --

11.2. ARCANJO

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: sistema para identificação de exploração sexual infantil em imagens e vídeos.

O ARCANJO tem como objetivo uma solução baseada em aprendizado de máquina para identificar a presença de exploração sexual infantil em imagens e vídeos. O sistema é dotado de um método combinado para detectar exploração sexual infantil, usando um módulo de detecção de face infantil e um módulo detector de pornografia. O método proposto classifica as imagens como relacionadas à exploração sexual infantil apenas se as contiverem pornografia e se for detectado um rosto infantil. O módulo de detecção de rosto infantil utilizado no método proposto foi construído através da integração de um detector de rosto e um método de estimativa de idade.

Bancos de Dados Utilizados: Por se tratar de um problema cujos dados possuem severas restrições legais, ou seja, sua posse ou acesso configuram uma ação criminosa, é preciso contar com bases de dados auxiliares. Essas bases se referem aos seguintes subproblemas da detecção de exploração sexual infantil: (1) detecção de faces, (2) estimativa de idade dos indivíduos presentes na imagem e (3) detecção de pornografia, em conformidade com a lei. Esses são problemas cuja quantidade de amostras disponíveis é abundante, possibilitando um treinamento mais robusto de um modelo de aprendizado. Foram utilizados os respectivos conjuntos de dados auxiliares:

WIDER FACE [1], Adience [2] e NSFW [3].

Para mitigar a evidente escassez de bases de dados que retratam a exploração sexual infantil, a Polícia Federal de Minas Gerais construiu o RCPD [4] (Region-based annotated Child Pornography Dataset). O dataset é constituído por 2.138 imagens, com 1.630 contendo pelo menos uma pessoa presente e 508 não contendo pessoas, possibilitando, assim, a verificação da taxa de falsos positivos.

[1] Yang, Shuo, et al. "Wider face: A face detection benchmark." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] Eidinger, Eran, Roei Enbar, and Tal Hassner. "Age and gender estimation of unfiltered faces." IEEE Transactions on Information Forensics and Security 9.12 (2014): 2170-2179.

[3] Mahadeokar, Jay, and Gerry Pesavento. "Open sourcing a deep learning solution for detecting NSFW images." Retrieved August 24 (2016): 2018.

[4] Macedo, João, Filipe Costa, and Jefersson A. dos Santos. "A benchmark methodology for child pornography detection." 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 2018..

Tecnologias Empregadas: Redes neurais artificiais, JavaScript, CSS, HTML, React (para desenvolvimento da Interface Web), FastAPI e Python (Backend).

Resultados Esperados: A abordagem proposta foi avaliada no conjunto de dados RCPD. A acurácia na detecção de exploração sexual infantil, no conjunto de dados RCPD foi de 79,84%. Os modelos auxiliares de classificação utilizados no Arcanjo tiveram uma acurácia de 82,55% para a tarefa de detecção de crianças e 85,78% para a etapa de classificação de pornografia (nu ou sexo). O sistema Arcanjo foi preparado para realizar uma classificação mais flexível, possibilitando, também, identificar conteúdo seminu, além de conteúdos com nudez e sexo, obtendo uma acurácia de 81,15%. A ferramenta ARCANJO foi desenvolvida e está disponível em produção como uma interface web de fácil usabilidade.

Desafios: Como anteriormente exposto (item 7), por se tratar de um problema cujos dados possuem severas restrições legais, ou seja, sua posse ou acesso configuram uma ação criminosa, foi preciso contar com bases de dados auxiliares para o treinamento dos algoritmos.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações: --

11.3. TRENA

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: plataforma para perícia digital e perícia cidadã em obras públicas. O TRENA é uma plataforma para monitoramento e acompanhamento de obras públicas. No dashboard, o analista do MPMG pode monitorar as obras, seus estágios de execução e vistorias realizadas sobre ela. O dashboard permite ainda a atribuição de vistorias aos parceiros técnicos e a geração de relatórios sobre essas vistorias, que podem ser usados para compor processos. Já no APLICATIVO TRENA, os técnicos parceiros do MPMG podem realizar o envio de vistorias das obras de acordo com as solicitações recebidas. As vistorias são acompanhadas de metadados de localização e horário de onde foram feitas as fotos e os vídeos. Cidadãos comuns também podem visualizar as vistorias das obras (desde que não sejam sigilosas) e contribuir com envios de vistorias. Nesse caso, seus envios devem ser validados pelo analista do MPMG no dashboard. A plataforma conta ainda com um modelo de inteligência artificial, treinado para identificar automaticamente em qual estágio uma obra está de acordo com as fotos recebidas. O modelo de análise computacional automatizada está preparado para os tipos de obra Escola, Creche e Pavimentação.

Bancos de Dados Utilizados: Foram utilizadas imagens de obras presentes no SIMEC e Geoobras, do TCEMG, além de um repositório de dados construído a partir de imagens coletadas do Google e do Bing.

Tecnologias Empregadas: WSO2 (autenticação, gerenciamento de identidade, segurança), JavaScript, CSS, HTML, React (para desenvolvimento da Interface Web), FastAPI e Python (Backend), Android, iOS, Redes neurais artificiais.

Resultados Esperados: A disponibilização de um aplicativo móvel para plataformas Android e iOS, voltado para a realização de perícias digitais em obras de natureza diversa, constitui um resultado significativo deste

projeto. Adicionalmente, foi desenvolvido um Dashboard destinado ao monitoramento e à gestão das vitorias técnicas e cidadãs. Este Dashboard incorpora funcionalidades variadas, incluindo a automação na geração de relatórios, contribuindo para a eficiência, economicidade e praticidade do processo.

Desafios: Acesso a bancos de dados de obras públicas com informações atualizadas e de qualidade para treinamento do modelo/algoritmo.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações: --

11.4. CAMPANHA ELEITORAL LEGAL

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: é uma solução para identificação de campanhas de propaganda eleitoral antecipada por meio da análise de dados do Twitter ("X"), o desenvolvimento e a avaliação de uma estratégia para essa identificação, bem como a integração do módulo desenvolvido à infraestrutura do MPMG, considerando tanto mensagens compartilhadas individualmente quanto grupos de mensagens compartilhadas em momentos distintos.

O principal objetivo deste projeto foi investigar estratégias para identificar campanhas de disseminação de propagandas eleitorais antecipadas em dados públicos previamente coletados de redes sociais. Em outras palavras, buscou-se generalizar as análises e estratégias exploradas anteriormente, expandindo o foco para identificar evidências da ocorrência de ações coordenadas, de um ou mais usuários, para promover a disseminação de conteúdo com propaganda eleitoral antecipada. O foco em campanhas se justifica, pois elas fornecem evidências mais fortes de ações criminosas em comparação com mensagens individuais, que podem ter sido compartilhadas sem intenção ou consciência de que tal ação consiste em um crime. Além disso, tais campanhas podem estar por trás de outros padrões possivelmente relevantes, como ação coordenada e disseminação em massa. O escopo do projeto envolveu três etapas principais. Inicialmente, realizou-se uma análise de dados do Twitter, utilizando uma grande amostra de informações, com o objetivo de identificar padrões relacionados a campanhas de propaganda eleitoral antecipada. Essa análise foi fundamental, já que não havia registros prévios na literatura sobre a identificação desse tipo específico de campanha. Com base nos conhecimentos adquiridos na etapa anterior, o projeto prosseguiu com o desenvolvimento e a avaliação de uma estratégia para a identificação de campanhas de propaganda eleitoral antecipada. Essa estratégia foi criada com o intuito de identificar grupos de usuários que apresentassem comportamentos anômalos na disseminação desse tipo de conteúdo em qualquer plataforma. A avaliação dessa estratégia, realizada com dados provenientes do Twitter, evidenciou indícios significativos do aumento do uso dessa plataforma para a disseminação em massa de conteúdo por grupos de usuários. Na terceira etapa do projeto, o foco direcionou-se para a investigação da integração do módulo desenvolvido anteriormente à infraestrutura já existente no MPMG, a qual se destinava à coleta e detecção de propaganda eleitoral, preservando sua utilidade e agregando valor à equipe do MPMG.

Bancos de Dados Utilizados: Banco de dados criado a partir de coletores de mensagens postadas na plataforma Twitter (atualmente conhecida como "X") durante período pré-eleitoral.

Tecnologias Empregadas: WSO2 (autenticação, gerenciamento de identidade, segurança), JavaScript, CSS, HTML, React (para desenvolvimento da Interface Web), FastAPI e Python (Backend, módulo para identificação de ação coordenada, etc), TensorFlow (modelo para classificação de mensagens com conteúdo político), MongoDB (armazenamento dos dados).

Resultados Esperados: O Campanha Eleitoral Legal foi concluído e disponibilizado para utilização pela equipe do MPMG. Espera-se que o sistema proposto possa auxiliar a equipe de especialistas na identificação de

campanhas eleitorais antecipadas realizadas em plataformas digitais de forma isolada ou coordenada, prática esta que é proibida pela legislação atual (Lei Eleitoral nº 13.488, de 6 de outubro de 2017).

Desafios: Um dos principais desafios foi a integração complexa entre as diferentes camadas e componentes do sistema. No entanto, através de uma análise cuidadosa e da adoção de boas práticas de arquitetura, foi possível implementar as integrações de forma eficiente e garantir a comunicação adequada entre os módulos. Além disso, a implementação da camada de segurança exigiu uma compreensão aprofundada das políticas de segurança e autenticação, o que foi superado com sucesso pela equipe. Atualmente, o principal desafio que se impõe é alteração no modelo de negócio adotado pelo twitter / 'X', que passou a cobrar pelas coletas realizadas, mesmo por órgãos públicos.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações:

11.5. TIRA-TEIMA INFO

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: sistema para a identificação de websites que produzem e disseminam desinformação na Internet, com foco específico em plataformas digitais.

Em uma primeira etapa do projeto, foi construída uma base de dados e realizado um estudo de pequeno porte para entendimento de características desses websites que espalham desinformação no contexto brasileiro. Em seguida, a partir da base de dados (estática) construída e caracterizada, investigou-se alternativas para facilitar a interação do usuário final (especialista) no processo de identificação de websites dedicados ao compartilhamento de desinformação no contexto brasileiro. Mais especificamente, foi desenvolvido um sistema, chamado "Tira-Teima", que permite ao usuário final (especialista) visualizar e analisar potenciais websites dedicados a compartilhar desinformação, bem como características relevantes desses websites – que podem fornecer insumos valiosos para dar suporte ao veredito de um especialista. Por fim, foi proposta uma abordagem que possa auxiliar, dinamicamente, o processo de identificação de websites que compartilham desinformação, revelando evidências úteis com potencial de agregar valor à equipe do MPMG. A abordagem proposta foi aplicada no Twitter, plataforma definida como alvo durante o projeto, e os resultados preliminares obtidos ressaltam o seu potencial para identificação de websites com foco na produção e disseminação de conteúdo cuja veracidade seja contestável.

Bancos de Dados Utilizados: Banco de dados de domínios rotulados como alta/baixa credibilidade, e mensagens postadas na plataforma Twitter ("X").

Tecnologias Empregadas: WSO2 (autenticação, gerenciamento de identidade, segurança), JavaScript, CSS, HTML, React (para desenvolvimento da Interface Web), FastAPI e Python (Backend, módulo de extração de atributos de domínios, identificação de Websites, etc), MongoDB (armazenamento dos dados).

Resultados Esperados: O Tira-Teima foi concluído e disponibilizado para testes por equipe de colaboradores do MPMG. Espera-se que o sistema proposto possa auxiliar a equipe de especialistas na identificação de websites (ou domínios) dedicados à produção e espalhamento de desinformação em plataformas digitais.

Desafios: Um dos principais desafios foi a integração complexa entre as diferentes camadas e componentes do sistema. No entanto, através de uma análise cuidadosa e da adoção de boas práticas de arquitetura, foi possível implementar as integrações de forma eficiente e garantir a comunicação adequada entre os módulos. Além disso, a implementação da camada de segurança exigiu uma compreensão aprofundada das políticas de segurança e autenticação, o que foi superado com sucesso pela equipe.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações:

11.6. QUANTO CUSTA

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: sistema que permite consulta de preços de produtos e serviços praticados pela administração pública dos municípios e do estado de Minas Gerais. Além da consulta, o sistema permite a detecção de casos de sobrepreço.

O Quanto Custa é uma solução prática e eficiente para a consulta de preços de itens de licitações em Minas Gerais. Com o objetivo de oferecer informações precisas e atualizadas, tanto para produtos quanto para serviços adquiridos pelo estado e pelos municípios, o sistema se destaca por suas duas funcionalidades principais: permitir que os usuários pesquisem um item específico para obter um preço de referência praticado para o mesmo e detecção de sobrepreço do item pesquisado.

Bancos de Dados Utilizados:

Licitações dos municípios de MG: sistema Informatizado de Contas dos Municípios (SICOM) do TCE-MG

Licitações estaduais de MG: Portal da Transparência do TCE-MG

Dados empresariais: Receita Federal / Serviço Federal de Processamento de Dados (SERPRO).

Tecnologias Empregadas:

Algoritmo de redução de representação textual por vetor - Word2Vec

Algoritmo de redução de dimensionalidade - UMAP

Algoritmo de clusterização - HDBScan

Algoritmo de similaridade de texto - FuzzyWuzzy

Detecção de anomalia - IQR

Análises estatísticas

ElasticSearch

Python.

Resultados Esperados: O Quanto Custa desempenha um papel fundamental na identificação de licitações que apresentaram sobrepreço para itens procurados. O sobrepreço é uma prática fraudulenta em licitações, e o sistema oferece uma ferramenta essencial para a formulação de novas licitações, ao disponibilizar a consulta do preço de referência, e também para a auditoria desses processos, por meio da análise de sobrepreço.

Desafios: Entendimento das regras de negócio para detecção do sobrepreço; disponibilidade e qualidade dos dados necessários ao treinamento do modelo, inclusive no que tange à identificação e tratamento de itens similares (características e medidas).

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações

11.7. PAINEL DE TRILHAS DE LICITAÇÕES

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: destinado à exibição de resultados das trilhas de auditoria e das análises de irregularidades entre vínculos de licitantes.

O Painel de Licitações tem como objetivo disponibilizar visualizações e mecanismos de exploração de dados de trilhas de auditoria através da análise de licitações e de vínculos entre licitantes, envolvidos nos respectivos processos licitatórios, levando em consideração os lotes e itens aos quais concorreram.

Bancos de Dados Utilizados:

Licitações dos municípios de MG: Sistema Informatizado de Contas dos Municípios (SICOM) do TCE-MG

Licitações estaduais de MG: Portal da Transparência do TCE-MG

Dados empresariais: Receita Federal / Serviço Federal de Processamento de Dados (SERPRO)

Dados de sanções e restrições: Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS); Cadastro de Fornecedores Impedidos de licitar e Contratar com a Administração Pública Estadual (CAFIMP)

Dados de ligação de energia elétrica: Companhia Energética de Minas Gerais (CEMIG)

Dados de servidores públicos: Instituto de Previdência dos Servidores do Estado de Minas Gerais (IPSEMG)

Dados de política: Tribunal Superior Eleitoral (TSE) - Eleições 2016, 2018, 2020;

Tecnologias Empregadas:

Modelo de detecção de anomalia

Heurísticas envolvendo processamento de texto

Algoritmos de teoria dos grafos

Python

Banco de dados PostgreSQL

Spark / Zeppelin

Plataforma Lemonade

Resultados Esperados: A partir de metáforas visuais, cruzamentos de dados e mecanismos de inteligência artificial empregados no desenvolvimento do PAINEL DE TRILHAS DE LICITAÇÕES, é possível que auditores e promotores públicos possam identificar e investigar possíveis ilicitudes nos processos licitatórios, sob a ótica de análise da licitação e dos licitantes e suas relações (possivelmente) espúrias que prejudicam a lisura e a livre concorrência, premissa de processos licitatórios.

Desafios: Entendimento das regras de negócio; disponibilidade e qualidade dos dados necessários.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações:

11.8. PROCONDATA

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: assistente inteligente desenvolvido para apoiar as atividades do PROCON-MG.

O PROCONDATA é um sistema de busca desenvolvido para apoiar as atividades do PROCON-MG, visando aprimorar o atendimento ao público e o apoio jurídico. Essa plataforma integrada permite que a equipe do PROCON-MG realize buscas avançadas em uma base de conhecimento que incorpora manifestações e reclamações de consumidores de diversas fontes.

Bancos de Dados Utilizados:

Manifestações e bases de reclamações fundamentadas de consumidores de diversas fontes, incluindo PROCON-MG e Ouvidoria, ambos do MPMG, Consumidor.gov, ReclameAqui e SINDEC.

Tecnologias Empregadas:

Foram implementadas no PROCONDATA funcionalidades avançadas de busca e integração de dados, bem como o uso de modelos de aprendizado de máquina para classificar as manifestações de consumidores das distintas bases de dados, detecção automática de entidades e agrupamento (clusterização) de manifestações. Além dessas tecnologias, foram utilizadas: WSO2 (autenticação, gerenciamento de identidade, segurança), JavaScript, CSS, HTML, React (para desenvolvimento da Interface Web), FastAPI e Python (Backend), PostgreSQL, Elasticsearch.

Resultados Esperados: O PROCONDATA foi disponibilizado em uma interface web simples e de fácil usabilidade, tendo como principal meta melhorar a eficiência das atividades do PROCON-MG, abrangendo o atendimento ao público, o apoio jurídico e o fortalecimento da proteção aos direitos dos consumidores. Para isso, foram implementadas funcionalidades avançadas de busca e integração de dados, bem como o uso de modelos de aprendizado de máquina para classificar as manifestações de consumidores das distintas bases de dados, além da geração de relatórios. Com a disponibilização do sistema, a equipe do PROCON-MG agora tem acesso a recursos avançados e uma plataforma de busca integrada, capaz de pesquisar em várias bases de dados de reclamações, contribuindo para a quantificação e identificação de violações de direitos coletivos e imprimindo substancial eficiência aos processos de trabalho.

Desafios: O principal desafio encontrado ao longo do projeto foi a obtenção de bases de dados para a disponibilização do sistema aos usuários finais (Procon-MG). Obtivemos acesso a amostras de dados de manifestações de cada fonte, o que possibilitou o treinamento do modelo e desenvolvimento das funcionalidades do PROCONDATA. Neste novo ciclo, os dados completos estão em fase de integração ao sistema.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Em desenvolvimento.

Considerações

11.9. PAINEL DE TRILHAS DE SERVIDORES PÚBLICOS DA SAÚDE

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: sistema para a exibição dos resultados das análises de possíveis irregularidades envolvendo os vínculos de trabalho dos servidores públicos da área da saúde no estado de Minas Gerais.

O Painel de Trilhas de Servidores da Saúde engloba as visualizações dos resultados das trilhas de auditoria de dados dos servidores públicos da área da saúde no estado de Minas Gerais. A ferramenta fornece uma visão abrangente das análises exploratórias dos dados, encontrando padrões e auxiliando na detecção de possíveis inconsistências ou irregularidades, mormente a existência de vínculos incompatíveis, consideradas a natureza dos vínculos, as jornadas de trabalho contratadas e a distância entre municípios contratantes ou locais de prestação dos serviços. A ferramenta permite explorar visualmente os padrões e as anomalias encontradas, verificar irregularidades por ocupações, bem como realizar pesquisas por servidor, por município e por tipo de vínculo (público/privado), além de disponibilizar um mapa interativo das possíveis irregularidades encontradas.

Bancos de Dados Utilizados:

Cadastro Nacional de Estabelecimento de Saúde (CNES) - Datasus.

Dados de cadastro de PF e PJ: Receita Federal / Serviço Federal de Processamento de Dados (SERPRO).

Tecnologias Empregadas:

Análises estatísticas: Boxplot e Histograma

Heurísticas envolvendo as regras de negócio

Python

Banco de dados PostgreSQL

Spark / Zeppelin.

Resultados Esperados: O objetivo central deste projeto é fornecer uma ferramenta eficiente e com navegação acessível e intuitiva para os integrantes do MPMG, mormente os PJ com atuação na defesa da saúde e na defesa do patrimônio público, para exploração visual das informações relevantes acerca dos múltiplos vínculos dos servidores públicos da área da saúde no Minas Gerais. Com base nas trilhas de auditoria, o painel oportunizou uma análise abrangente e detalhada sobre aspectos essenciais, auxiliando na identificação de eventuais inconsistências ou irregularidades relacionadas, a partir de dados oficiais (CNES) – Datasus, o que reveste de precisão e confiabilidade as análises realizadas pelo painel e, ao final, garante celeridade e acurácia dos resultados alcançados pelo parquet mineiro nesta área de atuação.

Desafios Entendimento das regras de negócio e acesso às bases de dados necessárias.

Parcerias: O projeto foi desenvolvido por meio da parceria com o Departamento de Ciência da Computação da Universidade Federal de Minas Gerais - DCC/UFMG.

Estágio Atual: Concluído com sucesso.

Considerações:

11.10. ÁDUNA

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: O ÁDUNA 2.0 é uma plataforma tecnológica para pesquisa de dados de pessoas físicas e jurídicas.

O ÁDUNA 2.0 é um sistema de busca desenvolvido para apoiar os integrantes do MPMG na localização de pessoas e nas atividades de investigação e inteligência. Reúne em uma única interface, de forma amigável e navegação intuitiva, o cruzamento de dados oriundos de diversas bases nacionais e locais, abertas e fechadas. Possui uma funcionalidade para análise de vínculos de forma gráfica entre entidades (pessoas físicas, empresas, endereços, telefones, e-mails, aeronaves e outros), semelhante ao I2 da IBM, porém, desenvolvido pelo MPMG em software livre (D3). Construído a partir do conceito de Progressive Web App, o ÁDUNA 2.0

está acessível via web, por meio de qualquer dispositivo eletrônico (PC, tablets ou celulares), sem a necessidade de instalação de qualquer aplicativo ou software, descentralizando e facilitando o uso da tecnologia. A ferramenta está parametrizada para perfis diversos de usuários, permitindo que o grupo de usuários vinculado a cada tipo de perfil tenha acesso ao conjunto de dados e funcionalidades compatível com as atribuições de seu cargo/função/lotação. Além disso, o acesso está suportado por autenticação em 2 fatores e todos os logins de acessos e pesquisas são registrados e armazenados, garantindo a auditabilidade da ferramenta e robustecendo a segurança da informação.

Bancos de Dados Utilizados:

CECAD (Cadastro Único)

SISOBI

Eleitores TREMG

Colaboradores MPMG

RFB / SERPRO CNPJ / CPF

PCMG/FAC, PCMG/SIDS_CONDUCTORES, PCMG/SIDS_VEICULOS

REDS (Registros de Eventos de Defesa Social)

CEMIG – Companhia de Energia Elétrica de MG

Credilink.

Tecnologias Empregadas:

Análises estatísticas: Boxplot e Histograma

Heurísticas envolvendo as regras de negócio

Python

Banco de dados PostgreSQL

Spark / Zeppelin.

Resultados Esperados: O ÁDUNA foi disponibilizado em uma interface web simples e de fácil usabilidade, tendo como principal meta melhorar a eficiência e efetividade das atividades do MPMG. Para isso, foram implementadas funcionalidades avançadas de busca e integração de dados, bem como o uso de modelos de aprendizado de máquina para identificar, rotular e apresentar os vínculos entre entidades das distintas bases de dados, além da geração de relatórios de inteligência. Com a disponibilização do sistema, os integrantes do MPMG e de órgãos parceiros agora tem acesso a recursos avançados e uma plataforma de busca integrada, capaz de facilitar a pesquisa de informações de PF e PJ, de modo descentralizado em cada Unidade ministerial, trazendo celeridade e efetividade aos processos de trabalho. Como exemplo, temos que o número de solicitação de pesquisas de dados dirigidos ao GSI, que no ano de 2022 bateu o pico de 1730 em um único mês, caiu ao patamar de 73 solicitações no mês com maior número de pedidos no ano de 2023, após a disponibilização do ÁDUNA 2.0.

Desafios: Partindo do zero, realizar a modelagem e implementação do ambiente de Big Data, ativo essencial que suporta todas as ferramentas desenvolvidas no âmbito do Programa Capacidades Analíticas, bem como a contratação de empresas especialistas, cuja expertise em cada um dos segmentos demandados para levar a cabo essa empreitada foi, sem dúvidas, um projeto desafiador, mas também, na mesma proporção, um sucesso que elevou a capacidade de processamento e análise de dados no MPMG a um nível superior. Vencida a etapa inicial de criação do ambiente de Big Data e início das atividades de pesquisa e desenvolvimento de inovação tecnológica, o maior desafio encontrado ao longo do projeto foi a obtenção de bases de dados para a disponibilização do sistema aos usuários finais.

Parcerias: O projeto vem sendo desenvolvido pela equipe interna da Diretoria de Análises e Tecnologia de Inteligência, com apoio da empresa 4MTI Soluções Ltda.

Estágio Atual: Desenvolvido com sucesso. Sistema disponibilizado e em Fase evolutiva de desenvolvimento de novas funcionalidades.

Considerações:

11.11. LINS

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: O LINS é uma ferramenta para exploração e análise das receitas, despesas e contratações dos municípios mineiros e do estado de MG. O ÁDUNA 2.0 é um sistema de busca desenvolvido para apoiar os integrantes do MPMG na localização de pessoas e nas atividades de investigação e inteligência. Reúne em uma única interface, de forma amigável e navegação intuitiva, o cruzamento de dados oriundos de diversas bases nacionais e locais, abertas e fechadas. Possui uma funcionalidade para análise de vínculos de forma gráfica entre entidades (pessoas físicas, empresas, endereços, telefones, e-mails, aeronaves e outros), semelhante ao I2 da IBM, porém, desenvolvido pelo MPMG em software livre (D3). Construído a partir do conceito de Progressive Web App, o ÁDUNA 2.0 está acessível via web, por meio de qualquer dispositivo eletrônico (PC, tablets ou celulares), sem a necessidade de instalação de qualquer aplicativo ou software, descentralizando e facilitando o uso da tecnologia. A ferramenta está parametrizada para perfis diversos de usuários, permitindo que o grupo de usuários vinculado a cada tipo de perfil tenha acesso ao conjunto de dados e funcionalidades compatível com as atribuições de seu cargo/função/lotação. Além disso, o acesso está suportado por autenticação em 2 fatores e todos os logins de acessos e pesquisas são registrados e armazenados, garantindo a auditabilidade da ferramenta e robustecendo a segurança da informação.

Bancos de Dados Utilizados:

PORTAL_MG LICITAÇÕES

SICOM RECEITAS

SICOM DESPESAS

SICOM LICITAÇÕES

SICOM NOTAS FISCAIS

SICOM CONTRATOS

TCESP RECEITAS

TCESP DESPESAS

RFB/SERPRO CNPJ

RFB/SERPRO CPF.

Tecnologias Empregadas:

Classificação baseada em regras¹ - Rule-based classifiers, Hadoop, Spark e D3.

Resultados Esperados: O LINS foi disponibilizado em uma interface web simples e de fácil usabilidade, tendo como principal meta melhorar a eficiência e efetividade das atividades de investigação na temática defesa do patrimônio público no âmbito do MPMG. Para isso, foram implementadas funcionalidades avançadas de busca e integração de dados. Com a disponibilização do sistema, os integrantes do MPMG e de órgãos parceiros agora tem acesso a recursos avançados e uma plataforma de busca integrada, capaz de facilitar a

pesquisa de modo descentralizado em cada Unidade ministerial, facilitando os processos de trabalho investigativos, norteados a produção de provas e a requisição de documentos e, ao final, alavancando os resultados alcançados.

Desafios: Partindo do zero, realizar a modelagem e implementação do ambiente de Big Data, ativo essencial que suporta todas as ferramentas desenvolvidas no âmbito do Programa Capacidades Analíticas, bem como a contratação de empresas especialistas, cuja expertise em cada um dos segmentos demandados para levar a cabo essa empreitada foi, sem dúvidas, um projeto desafiador, mas também, na mesma proporção, um sucesso que elevou a capacidade de processamento e análise de dados no MPMG a um nível superior. Vencida a etapa inicial de criação do ambiente de Big Data e início das atividades de pesquisa e desenvolvimento de inovação tecnológica, o maior desafio encontrado ao longo do projeto foi a obtenção das bases de dados para a disponibilização do sistema aos usuários finais.

Parcerias: O projeto vem sendo desenvolvido pela equipe interna da Diretoria de Análises e Tecnologia de Inteligência, com apoio da empresa 4MTI Soluções Ltda.

Estágio Atual: Desenvolvido com sucesso. Sistema disponibilizado e em Fase evolutiva de desenvolvimento de novas funcionalidades.

Considerações: --

12. MINISTÉRIO PÚBLICO DO ESTADO DA PARAÍBA

12.1. SISTEMA ARIEL

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Banco de dados de imagens coletadas através de sistema do IPC da Paraíba.

Tecnologias Empregadas:

Biblioteca utilizada face recognition

https://github.com/ageitgey/face_recognition

Resultados Esperados: Os resultados ainda estão sendo avaliados, mas acredita-se que a ferramenta chamada Pandora Speech tem o potencial de automatização do processo de transcrição de áudio, processo bastante utilizado nos núcleos de inteligência.

Desafios: Estudo e domínio de novas tecnologias. Essas ferramentas foram desenvolvidas com recursos próprios da instituição, sendo assim, sem auxílio externo o que tem a vantagem de fazer com que o conhecimento fique dentro da instituição, porém fazendo com que o desenvolvimento desse tipo de tecnologia esteja sujeito a uma curva de aprendizado da tecnologia utilizada.

Parcerias: As duas tecnologias mencionadas foram desenvolvidas sem ajuda de parcerias.

Estágio Atual: Em desenvolvimento.

Considerações:

12.2. PANDORA SPEECH

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Ferramenta desenvolvida para a transcrição de áudios e vídeos (speech-to-text ou STT).

Evidências: <https://pandoraspeech.mppb.mp.br/login>

<https://www.mppb.mp.br/index.php/42-noticias/gaeco/24563-ngc-prepara-versao-beta-de-ferramenta-de-transcricao-de-audio-para-integrantes-do-mppb>

Bancos de Dados Utilizados: Banco de dados de imagens coletadas através de sistema do IPC da Paraíba.

Tecnologias Empregadas: E para o Pandora Speech, o modelo Whisper da OpenAI.

Resultados Esperados: Os resultados ainda estão sendo avaliados, mas acredita-se que a ferramenta chamada Pandora Speech tem o potencial de automatização do processo de transcrição de áudio, processo bastante utilizado nos núcleos de inteligência.

Desafios: Estudo e domínio de novas tecnologias. Essas ferramentas foram desenvolvidas com recursos próprios da instituição, sendo assim, sem auxílio externo o que tem a vantagem de fazer com que o conhecimento fique dentro da instituição, porém fazendo com que o desenvolvimento desse tipo de tecnologia esteja sujeito a uma curva de aprendizado da tecnologia utilizada.

Parcerias: As duas tecnologias mencionadas foram desenvolvidas sem ajuda de parcerias.

Estágio Atual: Em desenvolvimento.

Considerações:

13. MINISTÉRIO PÚBLICO DO ESTADO DO PIAUÍ

13.1 CHATBOT PARA ABERTURA DE CHAMADOS DA CENTRAL DE SERVIÇOS DE TI

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Base de conhecimento da ferramenta ITSM

Tecnologias Empregadas: A empresa terceirizada entregou a solução como serviço de melhoria da ferramenta

Resultados Esperados: Redução de chamados abertos, facilidade na solução dos chamados pelos usuários.

Desafios: Como foi uma entrega da empresa terceirizada, não enfrentamos os desafios.

Parcerias: A iniciativa já foi concluída e entregue

Estágio Atual: Recentemente iniciado o uso (3 meses)

Considerações: --

14. MINISTÉRIO PÚBLICO DO ESTADO DO RIO GRANDE DO NORTE

14.1. PROJETO GENIUS

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: O principal objetivo se dá em agilizar os processos de trabalho, reduzir erros e fornecer insights valiosos para a tomada de decisões com mais embasadas, desempenhando um papel crucial na análise de políticas públicas e na previsão de tendências

A iniciativa está recém iniciada e em fase de projeto.

Bancos de Dados Utilizados: Utilizamos no MPRN principalmente os Bancos SQL SERVER e NO-SQL.

Tecnologias Empregadas: Estamos utilizando nos testes o GPT - OpenAI.

Resultados Esperados: O Projeto Central de Informações ao Cidadão possui 95% de aprovação e 85% tiveram o objetivo do atendimento contemplado.

Desafios: Os principais desafios são o de projetar uma estruturação adequada. Isso envolve a criação de uma infraestrutura de dados sólida, a formação de pessoal qualificado e a implementação de políticas de governança de dados rigorosas para garantir a privacidade e a segurança das informações.

Parcerias: O MPRN está formalizado um convênio com a UFRN- Universidade Federal do RN para o desenvolvimento de uma ferramenta de IA proprietária.

Estágio Atual: em fase inicial.

Considerações:

14.2. PROJETO CHATMP

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: O principal objetivo deste projeto é a melhoria da eficiência na elaboração de peças/documentos, promovendo uma redução de custos, padronização, além de permitir uma minimização de erros e, conseqüentemente, proporcionar agilidade na prestação do serviço por meio de automação dos processos para elaboração das peças/documentos, utilizando-se o sistema e-MP.

A iniciativa está recém iniciada e em fase de projeto.

Bancos de Dados Utilizados: Utilizamos no MPRN principalmente os Bancos SQL SERVER e NO-SQL.

Tecnologias Empregadas: Estamos utilizando nos testes o GPT - OpenAI.

Resultados Esperados: O Projeto Central de Informações ao Cidadão possui 95% de aprovação e 85% tiveram o objetivo do atendimento contemplado.

Desafios: Os principais desafios são o de projetar uma estruturação adequada. Isso envolve a criação de uma infraestrutura de dados sólida, a formação de pessoal qualificado e a implementação de políticas de governança de dados rigorosas para garantir a privacidade e a segurança das informações.

Parcerias: O MPRN está formalizado um convênio com a UFRN- Universidade Federal do RN para o desenvolvimento de uma ferramenta de IA proprietária.

Estágio Atual: Em fase inicial.

Considerações:

14.3. UTILIZAÇÃO DE CHATBOT PARA O SERVIÇO CENTRAL DE INFORMAÇÃO AO CIDADÃO QUE SE ENCONTRA ACESSÍVEL NO PORTAL DO MPRN

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: A Iniciativa se encontra em produção e a tecnologia disponibiliza diversos serviços como por exemplo:

- Consultar processos e Documentos eletrônicos
- Contatos da Procuradoria-Geral de Justiça, Procuradorias e Promotorias
- Atribuições e plantões Ministeriais
- Solicitar informações e orientações específicas
- Realizar denúncias para Ouvidoria do MPRN ou ao GAECO

Bancos de Dados Utilizados: Utilizamos no MPRN principalmente os Bancos SQL SERVER e NO-SQL.

Tecnologias Empregadas: Estamos utilizando nos testes o GPT - OpenAI.

Resultados Esperados: O Projeto Central de Informações ao Cidadão possui 95% de aprovação e 85% tiveram o objetivo do atendimento contemplado.

Desafios: Os principais desafios são o de projetar uma estruturação adequada. Isso envolve a criação de uma infraestrutura de dados sólida, a formação de pessoal qualificado e a implementação de políticas de governança de dados rigorosas para garantir a privacidade e a segurança das informações.

Parcerias: O MPRN está formalizado um convênio com a UFRN- Universidade Federal do RN para o desenvolvimento de uma ferramenta de IA proprietária.

Estágio Atual: em produção e com alto grau de aceitação.

Considerações: ---

15. MINISTÉRIO PÚBLICO DO ESTADO DE SANTA CATARINA

15.1. CATARINA

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: A Catarina funciona como um robô de autoatendimento, que fornece respostas com base em fluxos pré-definidos. Considerando que boa parte das demandas recebidas pelos canais de atendimento podem ser atendidas com as informações já disponibilizadas na página do MPSC, a ferramenta ajuda o cidadão a encontrar rapidamente o que deseja. A assistente virtual Catarina ocupa o canto inferior direito da tela no portal da instituição e facilita o acesso a informações no site do Ministério Público. Trata-se de chatbot totalmente desenvolvido pelo MPSC e pode ser usado tanto pelo público externo, quanto para o público interno (disponível na intranet).

Bancos de Dados Utilizados: Base de dados de endereços de Promotorias de Justiça e base de dados de CEP dos Correios.

Tecnologias Empregadas: Botpress.

Resultados Esperados: Desde o início o chatbot Catarina fez 166 mil atendimentos com 846 mil mensagens.

Desafios: As maiores dificuldades enfrentadas foram a falta de conhecimento sobre chatbots e o fato de ter sido desenvolvida por apenas uma pessoa. Esse fato, também está impossibilitando a evolução e melhoria do projeto.

Parcerias: Não, somente desenvolvimento interno.

Estágio Atual: Concluído com sucesso.

Considerações:

15.2. PROGRAMA CÉOS

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Programa de Inovação Aberta com a UFSC – conta com participação do Núcleo de Inovação da Subprocuradoria Geral de Justiça para Assuntos de Planejamento e Inovação (SubPlan), da Coordenadoria de Tecnologia da Informação (COTEC) e do Laboratório de Inovação do MPSC (iMPulsoLAB) e objetiva o desenvolvimento de projeto em inteligência artificial para auxílio à tomada de decisão inteligente em domínios complexos do setor público, em parceria com a Universidade Federal de Santa Catarina (UFSC). O Programa vem na busca de desenvolver a inovação no Ministério Público de Santa Catarina (MPSC), já que a inovação é um dos valores compreendidos nos propósitos do MP catarinense, fazendo parte da sua estratégia de atuação a busca por soluções tecnológicas integradas e inovadoras, segundo divulgado no Planejamento Estratégico do Ministério Público de Santa Catarina para o período de 2022-2029.

O Programa Céos, uma iniciativa de Inovação Aberta em parceria com a Universidade Federal de Santa Catarina (UFSC), busca impulsionar a inovação no Ministério Público de Santa Catarina (MPSC). Com a crescente quantidade de dados digitais no setor público, especialmente devido ao avanço do Governo Eletrônico, surge a necessidade de desenvolver tecnologias avançadas para gerenciar e extrair inteligência desses dados.

A parceria entre o MPSC e a UFSC, por meio do Programa Céos, visa desenvolver projetos em inteligência artificial para facilitar a tomada de decisões em setores complexos do serviço público. O foco está em aprimorar a eficiência e eficácia das instituições controladoras, reguladoras e defensoras, como o Ministério Público.

O programa se concentra em três linhas de Pesquisa e Desenvolvimento (P&D): detecção de irregularidades em compras e licitações públicas, análise de padrões de ocupação de leitos de UTI, e extração de informações de textos jurídicos. O objetivo é desenvolver técnicas e metodologias de ciência de dados para enfrentar desafios específicos, como o combate ao desvio de recursos públicos.

Além disso, o Programa Céos visa gerar conhecimentos, formar recursos humanos qualificados, produzir materiais de divulgação científica e aumentar a eficiência das atividades públicas, contribuindo assim para benefícios econômicos diretos e uma melhoria na qualidade de vida da população brasileira. A iniciativa também destaca a importância da produção científica, disseminação de conhecimento e formação de profissionais capazes de lidar com a coleta e análise de grandes volumes de dados em diversos domínios de interesse social.

Bancos de Dados Utilizados: Ainda não há definição sobre quais bancos de dados serão usados na pesquisa.

Tecnologias Empregadas: Ainda não há definição sobre quais tecnologias serão utilizadas na pesquisa.

Resultados Esperados: Esperamos que seja descoberta uma técnica moderna para detecção de possíveis irregularidades em licitações, bem como predição de ocupação de leitos e extração de textos jurídicos.

Desafios: O acesso aos dados de licitação para alimentação do modelo de IA parece ser um desafio administrativo.

Parcerias: Com a Universidade Federal de Santa Catarina - UFSC.

Estágio Atual: Em desenvolvimento.

Considerações:

16. MINISTÉRIO PÚBLICO DO ESTADO DE SÃO PAULO

16.1 ATIVIDADE DE JURIMETRIA E ANÁLISE DE DADOS, UTILIZANDO SOLUÇÃO DE PREDIÇÃO ANÁLISE ESTATÍSTICA E INTELIGÊNCIA ARTIFICIAL

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Metadados e Peças Processuais extraídas dos sistemas dos Tribunais - TJSP, STJ e STF

Tecnologias Empregadas: GPT - OpenAI no Microsoft Azure, R Project, Python, IA Builder e SQL Server

Resultados Esperados: Relatórios que apontam tendências nos Tribunais e Atuação do MPSP

Desafios: Dificuldade em contratar profissionais especializados, serviço de integração para buscar as informações e arquitetura corporativa da solução.

Parcerias: MSP e ABJ

Estágio Atual: Em desenvolvimento

Considerações: --

16.2 DIVERSOS

A unidade usou tanto modelos tradicionais de Machine Learning, como regressão logística, regressão linear, boosting, quanto modelos baseados em redes neurais (transformers). A unidade usa boosting (bag of words) para classificação de decisões judiciais conforme o desfecho processual. A partir de uma classificação baseada em regex e validada manualmente a partir de uma amostra, segue-se a uma nova etapa para treinar um modelo de bag of words para classificar especialmente decisões judiciais em procedente e improcedente ou outro desfecho. Além desses modelos, utilizam-se modelos derivados de Llama e Bert para extrair informações de peças processuais, especialmente documento desestruturados.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: Desde o início de 2022, a unidade usa também a API da OPENAI para resumir e extrair informações de peças processuais em massa. No entanto, a unidade, até o momento, não conta com créditos da OPENAI para realizar chamadas à API.

Bancos de Dados Utilizados: Banco de sentenças | Banco de acórdãos | Banco de petições iniciais | Banco de boletins de ocorrência.

Tecnologias Empregadas: PT (API da OPENAI) | Llama2 (Falcon e Mistral) | Bard (fizemos apenas testes) | BOW (boosting, regressão logística e SVM) | Modelagem tópica | Regex | Banco de vetores para armazenar embeddings: pg_vector.

Resultados Esperados: Extração de informações da vítima e agressores, crimes de milhares de B.Os | Resumo de sentenças judiciais de 1º grau do TJ | Resumo de decisões relativas a intimações provenientes do STJ | Extração de informações de decisões e acórdãos em processos que envolvem improbidade administrativa (12.000 processos) | Extração de informações relevantes de processos de idosos (50.000 processos) em crédito consignado | Classificações de decisões judiciais.

Desafios: Falta de unidades de processamento (GPUs) apropriadas para LLMs | Ausência de créditos para chamadas à API da OPENAI | Ausência de banco de vetores para armazenar embeddings.

Parcerias: A iniciativa jurimétrica é desenvolvida de forma sinérgica entre o Núcleo de Inteligência e Gestão de Conhecimento - NIGC, o CAOCrim e o CAOCível do Ministério Público do Estado de São Paulo - MPSP.

Estágio Atual: Em desenvolvimento

Considerações: --

16.3 PROCESSAMENTO DE LINGUAGEM NATURAL EM PEÇAS PROCESSUAIS COM REDES NEURAIAS.

Estrutura de IA: Sim

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: A unidade usou tanto modelos tradicionais de Machine Learning, como regressão logística, regressão linear, boosting, quanto modelos baseados em redes neurais (transformers).

A unidade usa boosting (bag of words) para classificação de decisões judiciais conforme o desfecho processual. A partir de uma classificação baseada em regex e validada manualmente a partir de uma amostra, segue-se a uma nova etapa para treinar um modelo de bag of words para classificar especialmente decisões judiciais em procedente e improcedente ou outro desfecho.

Além desses modelos, usamos modelos derivados de Llama e Bert para extrair informações de peças processuais, especialmente documento desestruturados.

Desde o início de 2022, a unidade usa também a API da OPENAI para resumir e extrair informações de peças processuais em massa.

No entanto, a unidade, até o momento, não conta com créditos da OPENAI para realizar chamadas à API. Os próprios servidores custeiam as chamadas.

Bancos de Dados Utilizados:

Banco de sentenças.

Banco de acórdãos.

Banco de petições iniciais.

Banco de boletins de ocorrência.

Tecnologias Empregadas:

GPT (API da OPENAI).

Llama2 (Falcon e Mistral).

Bard (fizemos apenas testes).

BOW (boosting, regressão logística e SVM).

Modelagem tópica.

Regex.

Banco de vetores para armazenar embeddings: pg_vector.

Resultados Esperados: Obtidos:

Extração de informações da vítima e agressores, crimes, endereço etc de milhares de boletins de ocorrência.

Resumo de sentenças judiciais de primeiro grau do TJSP.

Resumo de decisões relativas a intimações provenientes do STJ.

Extração de informações de decisões e acórdãos em processos que envolvem improbidade administrativa (+- 12.000 processos).

Extração de informações relevantes de processos de idosos (+- 50.000 processos) em crédito consignado.

Classificações de decisões judiciais.

Desafios: Falta de unidades de processamento (GPUs) apropriadas para LLMs. Falta de créditos para chamadas à API da OPENAI. Falta de um banco de vetores para armazenar embeddings provido pelo CTIC.

Parcerias: A iniciativa é desenvolvida apenas pelo setor de jurimetria (NIGC-CAOCRIM-CAOCÍVEL) do MPSP.

Estágio Atual: Em desenvolvimento.

Considerações:

17. MINISTÉRIO PÚBLICO DO ESTADO DE TOCANTINS

17.1. DESENVOLVIMENTO DE SISTEMA E-PONTO DE CONTROLE DE PONTO ELETRÔNICO COM USO DE RECONHECIMENTO FACIAL.

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa:

Bancos de Dados Utilizados: Postgres

Tecnologias Empregadas: LLM, Learn Facial Recognition with Python and OpenCV

Resultados Esperados:

Economia na rotina de aprovação e gerenciamento de folhas de frequência e redução da burocracia e da geração de papel.

Desafios: O principal desafio foi a calibragem das minúcias faciais dos integrantes.

Parcerias: Ambos com desenvolvimento interno

Estágio Atual: Concluído com sucesso.

Considerações: --

17.2. SISTEMA RADAR AMBIENTAL

Estrutura de IA: Não

Uso de Tecnologias de IA: Sim

Resumo da Iniciativa: sistema de reconhecimento e alertas de queimadas e desmatamentos com uso de linguagem natural para tratamento das cicatrizes de queimadas e desmatamentos e geração de alertas.

Bancos de Dados Utilizados: Postgres

Tecnologias Empregadas: LLM, Learn Facial Recognition with Python and OpenCV

Resultados Esperados: Precisão nas informações de queimadas e desmatamentos gerou ações assertivas e que evitaram maiores danos ambientais, também já tendo sido reconhecido e classificado para a final do Prêmio Judiciário Exponencial.

Desafios: Os shapes não tratados das bases georreferenciadas governamentais dificultaram o cruzamento de dados dessas bases.

Parcerias: Ambos com desenvolvimento interno

Estágio Atual: Concluído com sucesso.

Considerações: --



CONSELHO
NACIONAL DO
MINISTÉRIO PÚBLICO

Acesse nosso **portal**:



Siga o **CNMP** nas redes sociais:

 [cnmpoficial](#)

 [cnmpoficial](#)

 [@cnmp_oficial](#)

 [conselhodomp](#)

 [conselhodomp](#)

 [cnmp.mp.br](#)