


| | | |
|--|---|---------------------------|
|  | Fundação de Apoio à Universidade de São Paulo Av. Afrânio Peixoto, 14 –Butantã CEP - 05507-000–São Paulo –SP Tels. 55 11 3035-0550 / 3091-4289 e Fax: 55 11 3035-0580 | Projeto nº 3268 |
| | | Data 10/01/2022 |

Relatório de Atividades de Bolsista

| | |
|----------------|--------------|
| Parcial | Final |
|----------------|--------------|

| | |
|-------------------|--------------------------|
| Título do Projeto | Chatbot Q&A multi-agente |
|-------------------|--------------------------|

| | |
|--|----------------------|
| Nome do Bolsista <i>(sem abreviaturas)</i> | Enzo Bustos da Silva |
|--|----------------------|

| | | | |
|---------------------------|-------|-------------------|--------------------|
| Modalidade da Bolsa | Nível | Duração da Bolsa | |
| Iniciação Científica - IC | I | 1 (um) ano | |
| | | Início 01/02/2021 | Término 01/02/2022 |

| | |
|----------|--|
| Unidade: | Universidade de São Paulo - Escola Politécnica |
|----------|--|

| | |
|--------------|-----|
| Departamento | PCS |
|--------------|-----|

| | |
|---|-----|
| Laboratório/Núcleo <i>(se for o caso)</i> | C2D |
|---|-----|

| | |
|-------------------------------------|-------------------------------------|
| Nome do Coordenador do Projeto | Nome do Orientador |
| PROFA. DRA. ANNA HELENA REALI COSTA | PROFA. DRA. ANNA HELENA REALI COSTA |

Período das atividades desenvolvidas: de 01/02/2021 a 01/02/2022

1 Principais objetivos iniciais do Plano de Pesquisa

Dentre os objetivos iniciais que estavam no Plano de pesquisa, pode-se listar:

- Estudos da literatura e papers de Inteligência Artificial, de forma mais geral
- Aprendizado das técnicas específicas utilizadas em Processamento de Linguagem Natural (PLN)
- Uso das principais técnicas de PLN reportadas como estado-da-arte (Redes Transformers)
- Desenvolvimento de um agente conversacional do tipo Repórter, sobre a base de dados do Diário da Assembleia da República Portuguesa (DAR)*

*Observação: Inicialmente o projeto utilizaria dados referentes a Amazônia Azul, porém a dificuldade de extração de textos de PDF's sobre esse assunto fizeram com que o projeto migrasse para o DAR que tinha arquivos em formato txt

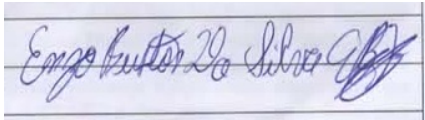


2 Objetivos alcançados pela pesquisa até a presente data

Até a presente data, grande parte dos objetivos foi cumprida. Segundo o Calendário das Etapas e atividades a revisão bibliográfica foi feita ao longo dos 2 primeiros meses, com leituras complementares ao longo de toda a duração da bolsa. Principalmente no que diz respeito a artigos voltados para a área de PLN, que foi o foco do projeto.

Quanto ao uso das técnicas do estado-da-arte, foram implementados diversos módulos para extrair informações das falas transcritas nas atas do DAR e, para isso foram utilizados modelos que utilizam a arquitetura de transformers como o BERT (e seu análogo para português, BERTimbau, ambos com conhecimento de língua), M2M100 (tradutor multilíngue), PEGASUS (sumarizador automático), BERTopic (extração de Tópicos abstratos) e roBERTa (classificação Zero-Shot em temas). Todos estes modelos consistem de redes neurais pré-treinadas que utilizam arquitetura de Transformers.

Na presente data, o Agente conversacional está na fase final de construção, que será desenvolvido utilizando-se a plataforma DiagFlow e terá a intenção de ser um FAQ sobre as Atas do DAR.

3 Principais resultados alcançados pelas atividades desenvolvidas durante a pesquisa

| | | |
|--|--|---|
| (relatórios, publicações, notas, artigos, eventos, etc) | | |
| <p>Quanto aos principais resultados alcançados vale pontuar:</p> <ul style="list-style-type: none"> - Total de 12 relatórios mensais para o Centro de Ciência de Dados para acompanhamento da Bolsa - Participação no 29º Simpósio de Iniciação Científica da USP: Apresentação - Publicação de artigo aceita no XVIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2021): Link - Publicação de artigo aceita na Conferência Internacional de Processamento Computacional da Língua Portuguesa (PROPOR 2022): Link - Publicação de artigo aceita no 1º Seminário Internacional de Humanidades - Artificial Intelligence: Democracy And Social Impacts (C4AI 2021): Ainda não publicado | | |
| 4 Impacto do projeto junto à área acadêmica a qual o bolsista está engajado | | |
| <p>Como impactos acadêmicos pode-se citar a profundidade dos temas de Inteligência Artificial, especificamente em PLN, que foram estudados durante o ano que complementam a formação na ênfase em Computação do bolsista, em especial as técnicas de Sumarização Automática e Modelagem de Tópicos, que exigiram um conhecimento prévio da área para serem utilizadas de forma eficiente.</p> <p>Além disso, como contribuição para a área de Engenharia e Ciências da Computação estão os três artigos aceitos.</p> | | |
| 5 Outras metas a serem atingidas até o término da bolsa <i>(preencher este campo somente se tratar de Relatório Parcial)</i> | | |
| ----- Não aplicável ----- | | |
| 6 Relacionar os documentos que fundamentam a pesquisa realizada e anexá-los a este relatório <i>(se houver)</i> | | |
| ----- Não aplicável ----- | | |
| Outras informações relevantes | | |
| O bolsista irá continuar e expandir essa Iniciação Científica em 2022. | | |
| Assinatura do Bolsista | Assinatura do Orientador | Assinatura do Coordenador |
|  |  |  |
| Setor de Gestão de Pessoas – FUSP | | |
| Nome do Conferente: _____ data: __09__ / __01__ / 2022__ | | |