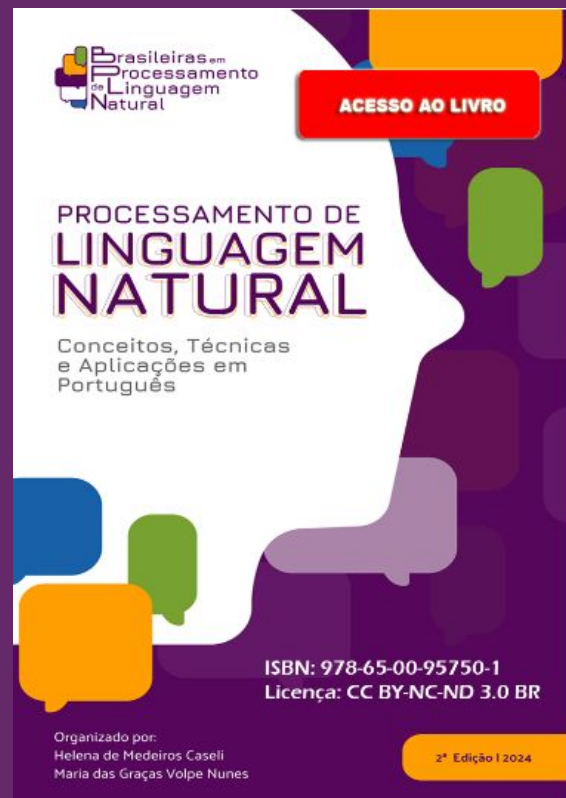




PLN na Saúde

Adriana Pagano - UFMG
Claudia Moro - PUCPR
Elisa Terumi - Celepar / InCor
Lilian Mukai - UEPG
Yohan Gumiel - InCor



<https://brasileiraspln.com/livro-pln>

Sumário

- ❑ Introdução
- ❑ O texto livre em narrativas clínicas
- ❑ Aplicações de PLN na Saúde
- ❑ Recursos de PLN na área clínica
- ❑ Para onde estamos caminhando?

Introdução



Uso da Tecnologia na Área da Saúde

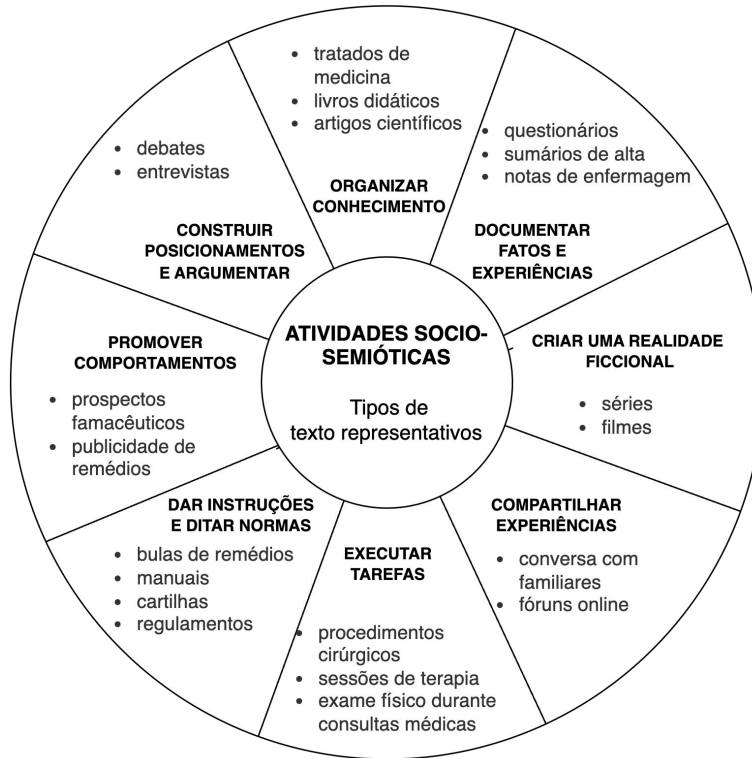
- **Tecnologia na Saúde**

- A área da saúde se beneficia cada vez mais da tecnologia para diagnósticos, tratamentos e gestão de pacientes
- O **Processamento de Linguagem Natural (PLN)** é uma das principais tecnologias, permitindo a análise de grandes volumes de dados clínicos não estruturados.

- **Diversidade Textual na Medicina**

- A medicina envolve diversas atividades produtivas de significado, chamadas de atividades sócio-semióticas
- Estudos identificam **oito** tipos de atividades, como procedimentos cirúrgicos e registros clínicos falados e escritos

Tipos de texto no domínio da medicina



Oito tipos de **atividades sócio-semióticas** e os tipos de texto mais representativos de cada um deles no domínio da medicina.

Documentação de Fatos e Experiências

- **Narrativas Clínicas na Medicina**

- A documentação de fatos e experiências é uma atividade sócio-semiótica essencial
- Inclui questionários, relatórios clínicos, notas de enfermagem, sumários de alta e boletins médicos

- **Valor das Narrativas Clínicas**

- Textos não estruturados que revelam a história do paciente
- Mineração de dados identifica padrões, aprimorando a compreensão das condições e criando modelos preditivos

O texto livre em narrativas clínicas



A Importância do Registro Eletrônico de Saúde

- **Impacto do Prontuário Eletrônico no Brasil**

- O RES (EHR) ampliou os dados sobre a atenção aos pacientes
- Contém dados estruturados, semiestruturados e não estruturados
- A mineração desses dados identifica tendências, diagnósticos e tratamentos, aprimorando a gestão do cuidado ao paciente.



Desafios das Narrativas Clínicas em Texto Livre

- **Desafios ao Processar Narrativas Clínicas**

- As narrativas clínicas são textos livres que geralmente possuem jargões, abreviaturas e variações de linguagem
- Erros de digitação e ortografia, que dificultam a interpretação
- Algoritmos de IA precisam de dados anotados
- A anotação manual é essencial, mas demanda tempo e recursos, limitando a criação de grandes *datasets*.

Exemplo de Narrativa Clínica

- **Categorização das Narrativas Clínicas**

- Classificadas por tipos, como notas de enfermagem e sumários de alta
- Cada tipo de texto apresenta desafios únicos
- Dependendo da especialidade mudam totalmente os termos utilizados
- Notas de ambulatório, por exemplo, tendem a ser mais informais e fragmentadas.

Exemplo de Narrativa Clínica

AP: HAS . Obesidade . DM. Tabagista.	Refere HAS desde 21 anos de idade, irregular controle da pressão arterial.
Ao exame: BEG, CHAAA, eupneico. ACV: BRNF, sem sopro. PA 159X100 mmHg, FC 89. AR: MV+ s/ RA. ABD : flácido , indolor a palpação , sem visceromegalias . MMII : edema bilateral ++/+4 .	
USG rins normal . Doppler MMII normais . ECG : BAV 1º G , BRD . TE Externo : baixa resposta cronotrópica , EEVV polimórficas frequentes , comportamento normal de PA . ECO TT Ext 19/06/20: AE 28 Septo 12 Parede Post 12 Massa VE 218 FEVE 70 % VE 48x29 Hipertrofia simétrica discreta do VE . Alteração do relaxamento do VE .	
Laboratório : 22/09/21: Hb:14,3 Leuco:4.960 PLaq:204mil Colesterol:169 HDL:49 LDL:102 Triglic:91 Glicose:96 Ur:40 Cr:0,93 Na:142 K:4,3 TSH:2,08	
Oriento MEV , Restrição Hidrossalina . Oriento importância da adesão medicamentosa . Solicito ECO TT .	
Prescrevi : sinvastatina 40 mg comp . (1 comp vo 1xd) ; ácido acetilsalicílico 100 mg comp (1 comp vo 1xd) ; omeprazol 20 mg cáps; losartana 50 mg comp (1 comp vo 12/12h) ;	
Solicitei : Ecocardiograma com doppler colorido adquiridas ; hemoglobina glicada ; ácido úrico ; colesterol (sangue) ; CPK; creatinina / soro ; glicose / soro ; TGO	

Legenda

Antecedentes pessoais	Exames laboratoriais
História pregressa	Orientações
Exame físico e sinais vitais	Medicamentos prescritos
Exames realizados	Planejamento

7

Exemplo de uma **narrativa clínica** ilustrada, destacando as categorias de informações que podem ser encontradas nesses textos.



Anotação de Narrativas Clínicas

- **Anotação Manual em PLN na Saúde**
 - Essencial para refinar modelos de PLN
 - Anotação com revisão manual para maior precisão
 - Permite a caracterização de entidades nomeadas.
- **Construção de *Corpora* de Narrativas Clínicas**
 - O *fine-tuning* de modelos genéricos com dados clínicos melhora a precisão
 - Vários exemplos destacam a importância de modelos ajustados ao domínio médico
 - Sujeita a restrições técnicas e regulatórias (ex privacidade dos dados).

Aplicações de PLN na Saúde



Predição

- **A predição em PLN na saúde abrange:**
 - Diagnóstico
 - Tratamento
 - Evolução clínica
 - Alta médica
 - Detecção de quedas e depressão
- **Importância:**
 - Classificação de dados clínicos (ex: prontuários eletrônicos)
 - Identificação precoce de doenças
 - Personalização do tratamento

Desidentificação (anonimização)

- **Desidentificação**
 - Remoção de informações identificáveis dos pacientes (ex: nome, endereço).
 - Garantir privacidade e cumprir regulamentações (LGPD e GDPR).
 - Permite uso de dados clínicos para pesquisa sem comprometer a privacidade
- **Técnicas de Desidentificação:**
 - Substituição de nomes por símbolos ou pseudônimos
 - Remoção de informações geográficas e de contato
 - Uso de modelos de linguagem para anonimização

Extração de Conceitos Clínicos

- **Identificação de entidades relevantes**
 - NER (Reconhecimento de Entidades Nomeadas)
 - Extração de sintomas, diagnósticos, medicamentos
 - Detecção de negação
- **Mapeamento de terminologia**
 - CID, SNOMED CT, UMLS
 - Análise de dados clínicos em larga escala e suporte à decisão médica.

Relações Temporais

- **Organização cronológica das informações clínicas de um paciente**
 - Extração de relações temporais em dados longitudinais
 - Análise da evolução clínica
 - Identificação de padrões e tendências
 - Integração com sistemas de suporte à decisão.

Sumarização

- **Extração de informações relevantes**
 - Criar resumos legíveis de dados clínicos
 - Permite análise rápida de informações essenciais
 - Sumarização extrativa: seleção de frases importantes
 - Sumarização abstrativa: síntese das informações

Recursos de PLN na área clínica

Recursos de PLN na Área Clínica

- **Corpora em Português no Contexto Clínico**

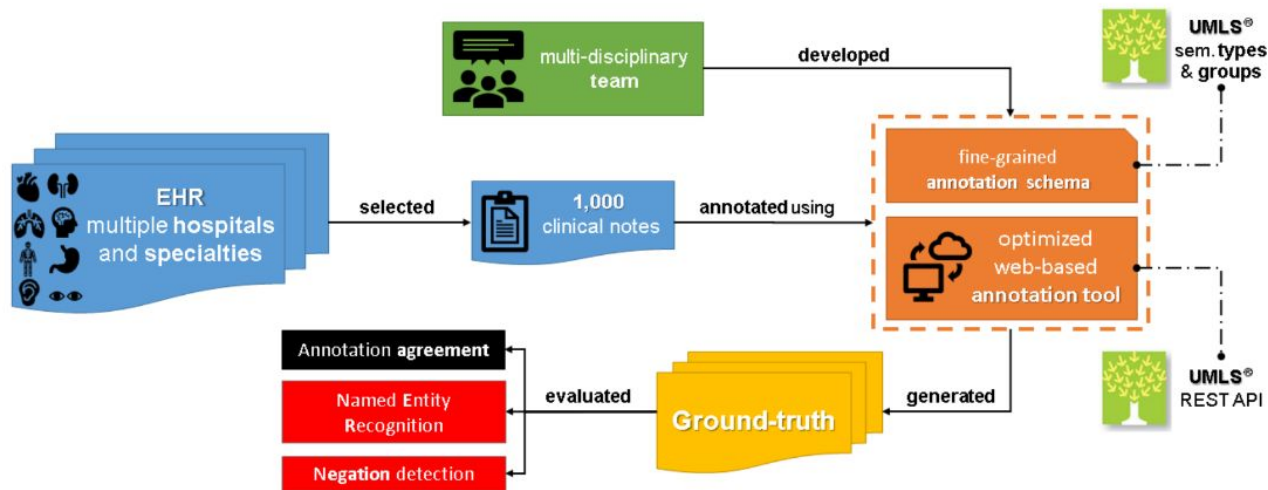
- Conjuntos de dados textuais organizados e anotados
- Treinar, avaliar e testar modelos de PLN

- **Modelos de PLN**

- Modelos de inteligência artificial treinados para entender, gerar e interpretar textos
- Tarefas como tradução, resumo, análise de sentimentos, e classificação de informações textuais.

Corpus SemClinBr

Um *corpus* semanticamente anotado multi-institucional e multiespecializado para tarefas clínicas de **PLN** em **português**.



Corpus SemClinBr

- Contém **1.000** notas clínicas
- **65.117** entidades e **11.263** relações anotadas
 - Aspectos clínicos: condições médicas, medicamentos e procedimentos
 - Dicionários de negação
 - Abreviações médicas
- Disponível para uso acadêmico em: <https://github.com/HAILab-PUCPR/SemClinBr>

Corpus Detecção de Quedas

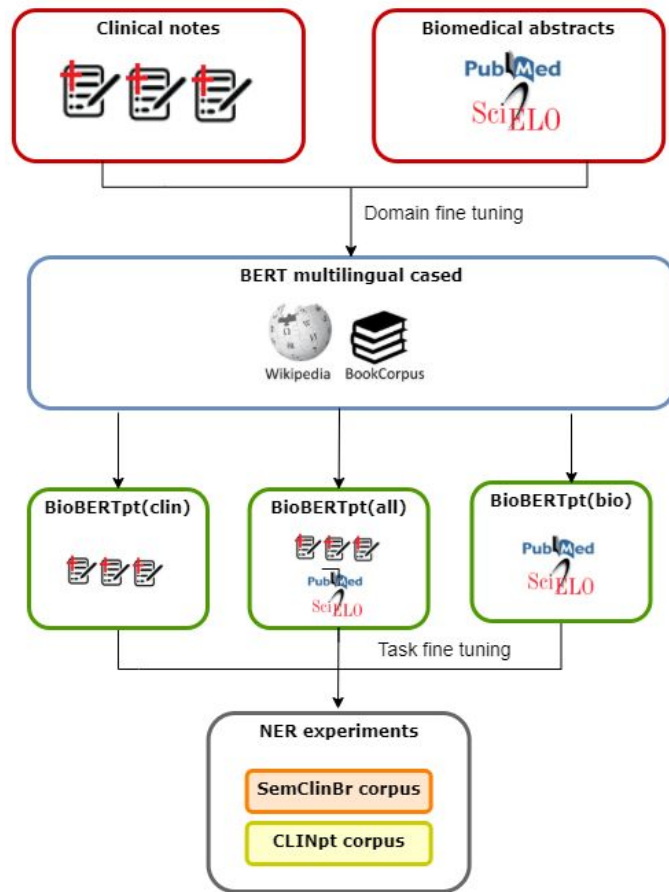
- Foco na **detecção de incidentes de quedas**
 - Eventos adversos em hospitais
 - Melhorar a compreensão dos incidentes
 - Qualidade do cuidado ao paciente
- Contém **1.078** notas de progresso desidentificadas
 - Registros clínicos sobre os pacientes durante sua estadia no hospital
- Disponível em <https://github.com/nlp-pucrs/fall-detection>

Corpus Casos Clínicos - Neurologia

- Contém **281** textos de casos clínicos em **neurologia**
 - Números 1 e 2 do volume 17 da revista clínica Sinapse
 - Sociedade Portuguesa de Neurologia
- Pré-processamento com **NLPPort**
- Anotação manual de entidades clínicas
 - Ex. Caracterização, evolução, genética, anatomia, resultado, negação, etc
- Disponível para pesquisa acadêmica em <https://github.com/fabioacl/PortugueseClinicalNER>

Modelo BioBERTpt

- Modelo **BERT** adaptado para o português
 - BERT multilíngue como base
 - Textos clínicos e biomédicos
- Família de 3 modelos:
 - BioBERTpt(clin)
 - BioBERTpt(bio)
 - BioBERTpt(all)
- Melhoria de **2,72%** em F1-score no SemClinBr
- Disponível em <https://github.com/HAILab-PUCPR/BioBERTpt> e <https://huggingface.co/pucpr>



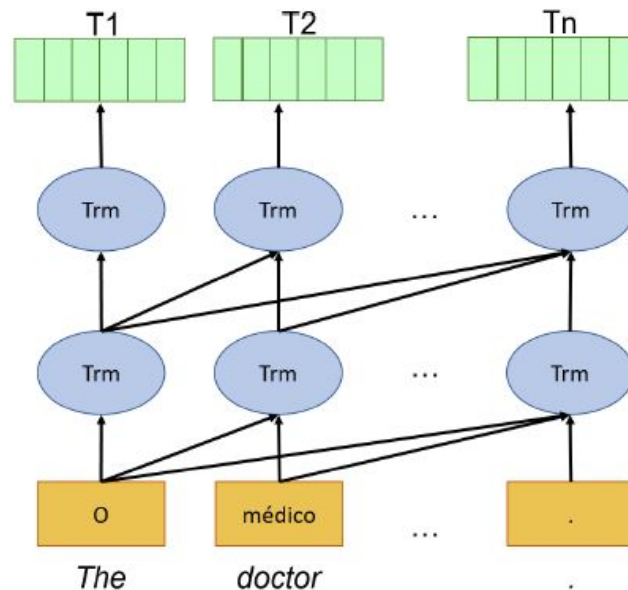
Modelo BioBERTptRT

- Versão do **BioBERTpt** para **classificação de sentenças clínicas**
 - Padrão **SOAP** (Subjetivo, Objetivo, Avaliação e Plano)
 - Estruturar e organizar as notas clínicas
- Treinado (ajustado) com **234.673** notas clínicas
 - Mais de um milhão de sentenças únicas
 - 100.000 sentenças rotuladas manualmente
- Melhor compreensão do contexto e da terminologia médica em português.

Modelo GPT2-Bio-Pt

- Modelo baseado em **GPT-2** especializado para o domínio biomédico em português
 - Apoiar tarefas de PLN
 - Geração de texto sintético
 - Classificação
- Desempenho superior na detecção de quedas
- Disponível em <https://github.com/HAILab-PUCPR/gpt2-bio-pt> e <https://huggingface.co/pucpr/gpt2-bio-pt>

OpenAI GPT Pre-training Architecture

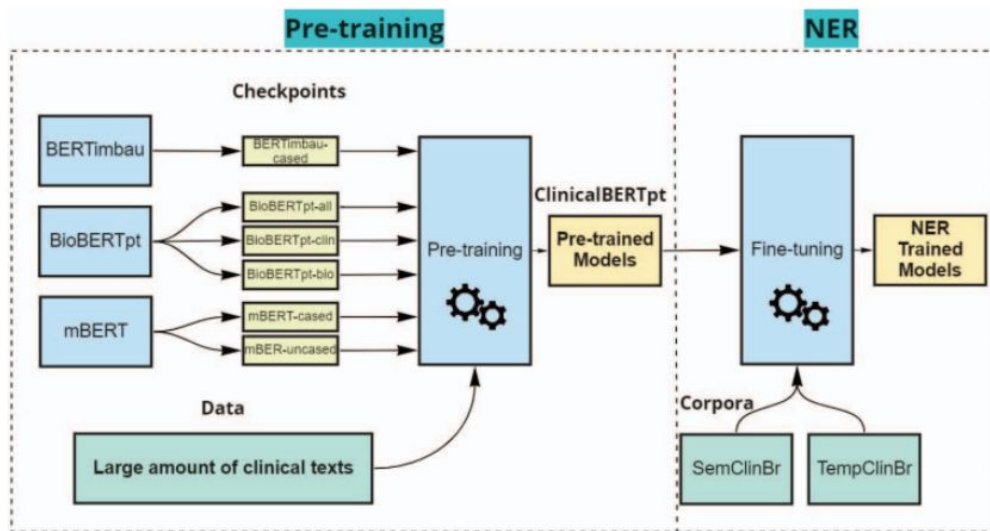


Modelo BERT para Classificação CID-10

- Atribuição de códigos **CID-10** (Classificação Internacional de Doenças) a causas de morte usando descrições em texto livre
 - Certificados de óbito
 - Relatórios de autópsia
 - Boletins clínicos fornecidos pelo Ministério da Saúde de Portugal
- **BERT** adaptado para a tarefa particular de codificação CID-10
- Pré-treinamento que incorpora conhecimento clínico.

Modelo CardioBERTpt

- Modelo **BERT** especializado para **cardiologia**
 - Nuances e terminologias específicas do domínio cardiológico
- Treinado com dados de saúde de um hospital especializado
- Disponível em <https://github.com/HAILab-PUCPR/CardioBERTpt> e <https://huggingface.co/pucpr-br/cardiobertpt>



Para onde
estamos
caminhando?

Para onde estamos caminhando?

- **A evolução da IA e do PLN na saúde**

- Avanços com Grandes Modelos de Linguagem (LLMs) e IA Generativa
- Oportunidades promissoras
 - Eficiência dos sistemas de saúde
 - Otimizar processos de diagnóstico e tratamento
 - Análise de dados não estruturados
 - Desenvolvimento de tratamentos personalizados
 - Pesquisa biomédica
 - Gestão de grandes volumes de dados clínicos
- Desafios para garantir uma aplicação eficaz e ética.

Desafios - PLN na Saúde

- **Qualidade dos dados clínicos**
 - Anonimização eficaz
 - Padronização dos termos utilizados
 - Redução de viés
 - Garantia de ética e privacidade
- **Capacidade de lidar com textos complexos e heterogêneos**
 - Notas de enfermagem, laudos e textos de pacientes
- **Integração com sistemas de saúde**
 - Interoperabilidade
 - Segurança dos dados
- **Aceitação pelos profissionais de saúde**
 - Demonstração de utilidade
 - Validação da eficácia na prática clínica.



Importância do julgamento clínico

- **O papel das ferramentas de PLN na Saúde**

- Auxiliam na análise e interpretação de dados clínicos
- Complementam o cuidado ao paciente, sem substituir o conhecimento dos profissionais
- Auxiliam na tomada de decisões
- Foco na melhoria do cuidado, com ética e privacidade dos dados

Não substitui experiência e o conhecimento de profissionais da saúde!





Obrigada!

