

SHIFT
FIAP

Ana Raquel



Carreira

- Tecnólogo em banco de dados pela faculdade FIAP.
- MBA em inteligência artificial pela FIAP.
- Mais de 8 anos de experiência como profissional na área de dados tendo atuado em diversos projetos de Banco de Dados, BI, Analytics e Data Science.
- Cientista de dados na FIAP e professora de Machine Learning , Deep Learning, Processamento de Linguagem Natural e Data Viz na FIAP.



Como o computador entende texto?



livro alice no pais das maravilhas



Todas

Shopping

Imagens

Videos

Livros

Mais

Ferramentas

Aproximadamente 606.000 resultados (0,56 segundos)

Alice no País das Maravilhas

Livro por Lewis Carroll

Visão geral

Personagens

Videos

Comprar livro

Avaliações



Anúncios · Comprar livro alice no pais das maravilhas



Alice no País das...

R\$ 46,67
Amazon.co...



Alice no País das...

R\$ 54,90
DarkSide Bo...



Alice: edição bolso de lux...

R\$ 22,82
Amazon.co...



Livro - Alice no País das...

PROMOÇÃO
R\$ 39,01 R\$...
Americanas...



Alice no país das...

R\$ 19,99
Amazon.co...



Visualização do livro

31/143 páginas disponíveis

Visualizar



Sobre

85% gostaram desse livro
Usuários do Google





impala



Mais resultados de [impala.com.br](#) »

Imagens de impala



supernatural



bailarina



wallpaper



perolado



coleção



Feedback

Preciso ser mais específico (a)?



carro impala



Todas

Imagens

Shopping

Notícias

Vídeos

Mais

Ferramentas

Aproximadamente 7.200.000 resultados (0,49 segundos)

Imagens de carro impala



tunado



dentro



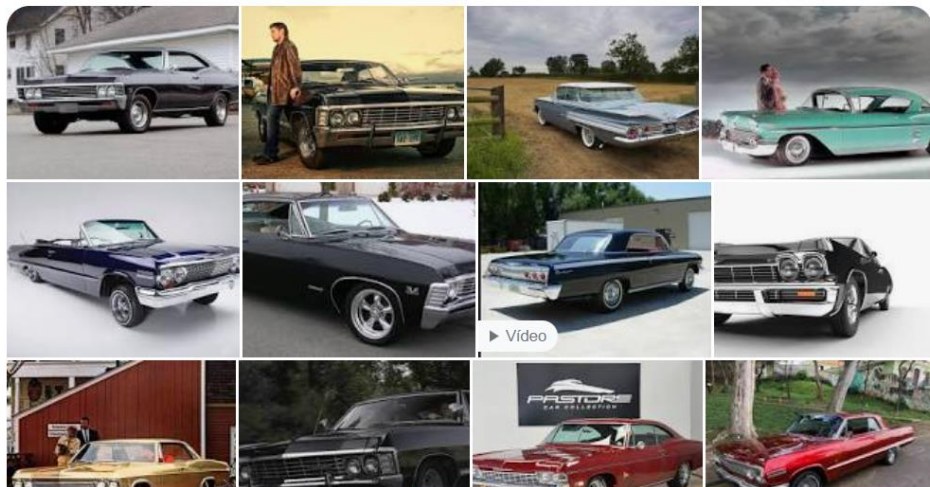
4 porta



brinquedo



roxo



esmalte impala



Todas

Imagens

Shopping

Notícias

Vídeos

Mais

Ferramentas

Aproximadamente 2.550.000 resultados (0,42 segundos)

Imagens de esmalte impala



perolado



azul



coleção



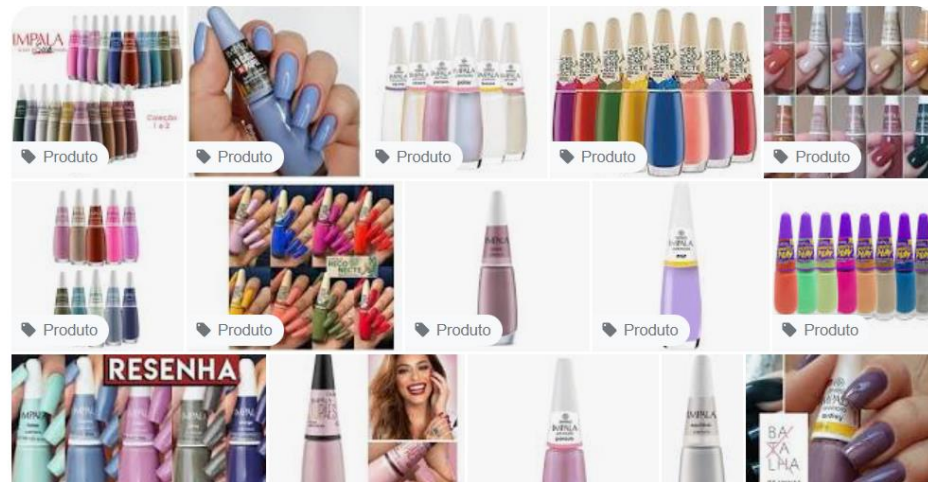
audrey



bailarina



cor





**Como o computador
entende o contexto?**



todas as bandas anos 70 exceto led zeppelin



Todas

Notícias

Shopping

Imagens

Vídeos

Mais

Ferramentas

Aproximadamente 83.500 resultados (0,41 segundos)

<https://www.letras.mus.br> › blog › bandas-de-rock-anos... ▾

Nostalgia: as maiores bandas de rock dos anos 70

5 de jun. de 2020 — Com estilo hippie, cheio de looks coloridos e penteados cheios de estilo, surgia um novo período para o rock. **Banda Led Zeppelin** Integrantes do ...

Não encontrados: ~~exete~~ | Precisa incluir: **exceto**

<https://www.collectorsroom.com.br> › 2019/09 › playlist... ▾

Playlist: Hard 70 - Collectors Room

13 de set. de 2019 — ... a **década de 1970** alcançou status mítico ao dar ao mundo **bandas** lendárias como Black Sabbath, Deep Purple, **Led Zeppelin** e muitas outras.

<https://www.consultoriadorock.com> › 2020/06/03 › mel... ▾

Melhores de Todos os Tempos: Anos 70 - Consultoria do Rock

3 de jun. de 2020 — Trata-se de **Led Zeppelin**. Ao longo dos **anos 70**, a **banda** foi inquestionavelmente a mais importante, a frente inclusive de Stones e Queen, ...

E se eu for muito específico (a)?



imagens de filmes da disney exceto procurando nemo



Todas

Imagens

Shopping

Videos

Notícias

Mais

Ferramentas

Coleções SafeSearch



marlin



dory



amazon



frases



livro



procurando dory

frases marcantes



dvd



peixe nemo



DVD Disney Pixar - Proc...
magazineluiza.com.br



Disney clássicos ilustra...
amazon.com.br · Em estoq...



Procurando nemo filme | Extra
extra.com.br



Procurando Nemo | Obser...
observatoriodocinema.uol.c...



Dvd - Procurando Nemo
magazineluiza.com.br · Em ...





Procurando Nemo esconde segredo sombrio, revela t...
observatoriodocinema.uol.com.br



E o chat GPT?

+ New chat

PLN - Processamento d  

New chat

Numpy Analysis on Test Score


Função DAX DATESYTD



Plotting offline graphs.

Séries Temporais Unimulti

Utilizando NumPy em Séries T

DAX code for calculating stude

 Upgrade to Plus NEW

 Ana Raquel Fernandes Cunha 



Processamento de linguagem natural (PLN) é um campo de estudo da inteligência artificial que se concentra na interação entre humanos e computadores usando a linguagem natural. O objetivo do PLN é permitir que os computadores entendam, interpretem e produzam linguagem humana de maneira que seja natural e significativa para os seres humanos.



O PLN usa algoritmos de aprendizado de máquina para analisar grandes quantidades de dados linguísticos, como textos escritos, transcrições de fala e dados de mídia social. Esses algoritmos permitem que os computadores reconheçam padrões na linguagem, identifiquem tópicos e sentimentos, realizem traduções automáticas e até mesmo gerem respostas em linguagem natural.

Algumas das aplicações práticas do PLN incluem assistentes virtuais, chatbots, análise de sentimento de mídias sociais, corretores ortográficos, tradutores automáticos, entre outros. O PLN está em constante evolução e tem o potencial de transformar a forma como os humanos interagem com os computadores.

 Regenerate response



Send a message...

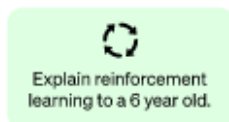


[ChatGPT Mar 23 Version](#). Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts.

Step 1

Collect demonstration data and train a supervised policy.

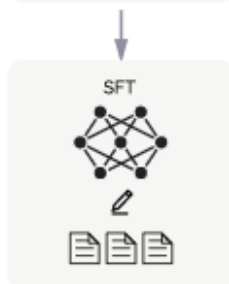
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2

Collect comparison data and train a reward model.

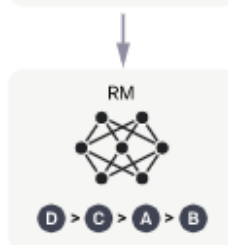
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.



The PPO model is initialized from the supervised policy.



The policy generates an output.



The reward model calculates a reward for the output.



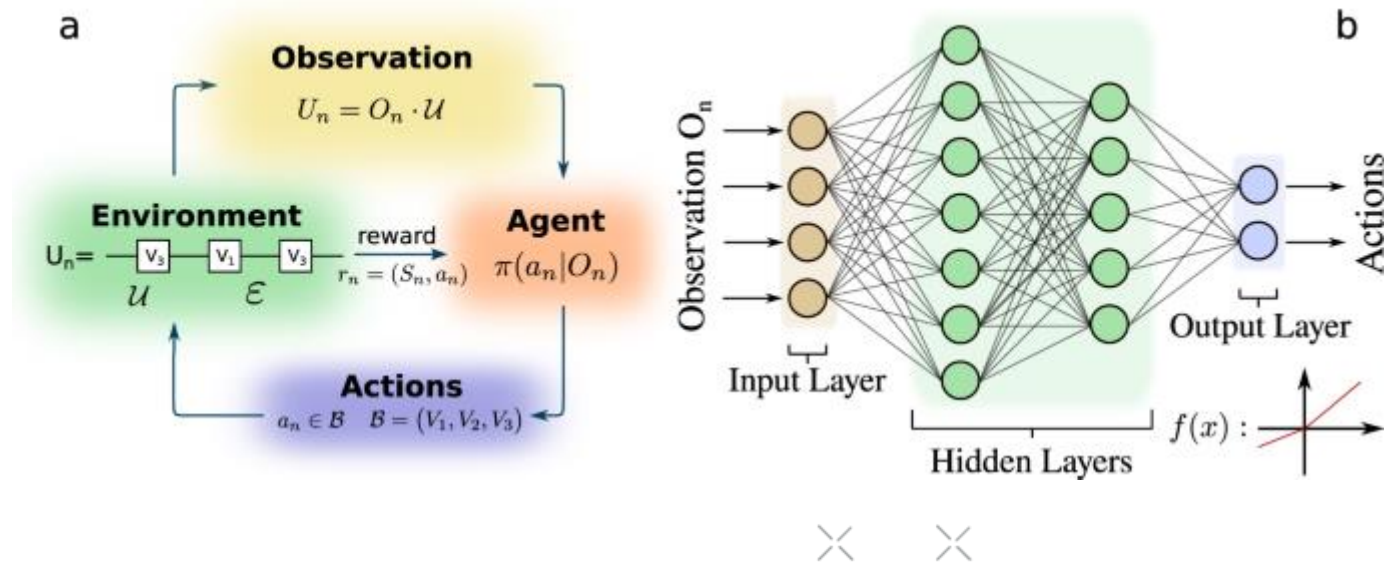
The reward is used to update the policy using PPO.



E como funciona o modelo?

O modelo foi **treinado** usando bancos de dados de texto da internet. Isso incluiu incríveis 570 GB de dados obtidos de livros, textos da web, Wikipedia, artigos e outros textos na Internet. Para ser ainda mais exato, 300 bilhões de palavras foram inseridas no sistema.

Como um modelo de linguagem, ele trabalha com probabilidade, capaz de adivinhar qual deve ser a próxima palavra em uma frase. O modelo utilizado é do tipo **Reinforcement Learning from Human Feedback (RLHF)**.



Aprendizagem por reforço:

Basicamente funciona como tentativa e erro, que podem ser recompensadas por alguma recompensa quando o modelo de fato gera um acerto, e punição quando o modelo erra a previsão.



Punição



Recompensa

Mas o chat GPT possui limitações?

Claro, como todo modelo de machine learning e deep learning! :)

Não é capaz de responder às perguntas “atuais”. O modelo foi treinado até 2021.

Mas o chat GPT possui algum problema?

Como o sistema é treinado em grande parte usando palavras da internet, ele pode captar preconceitos, estereótipos e opiniões gerais da internet.



O que é processamento de linguagem natural?

Essa tecnologia é uma vertente da inteligência artificial utilizada para **traduzir a linguagem humana** para uma linguagem que os dispositivos tecnológicos **consigam entender**.

Humano = linguagem natural

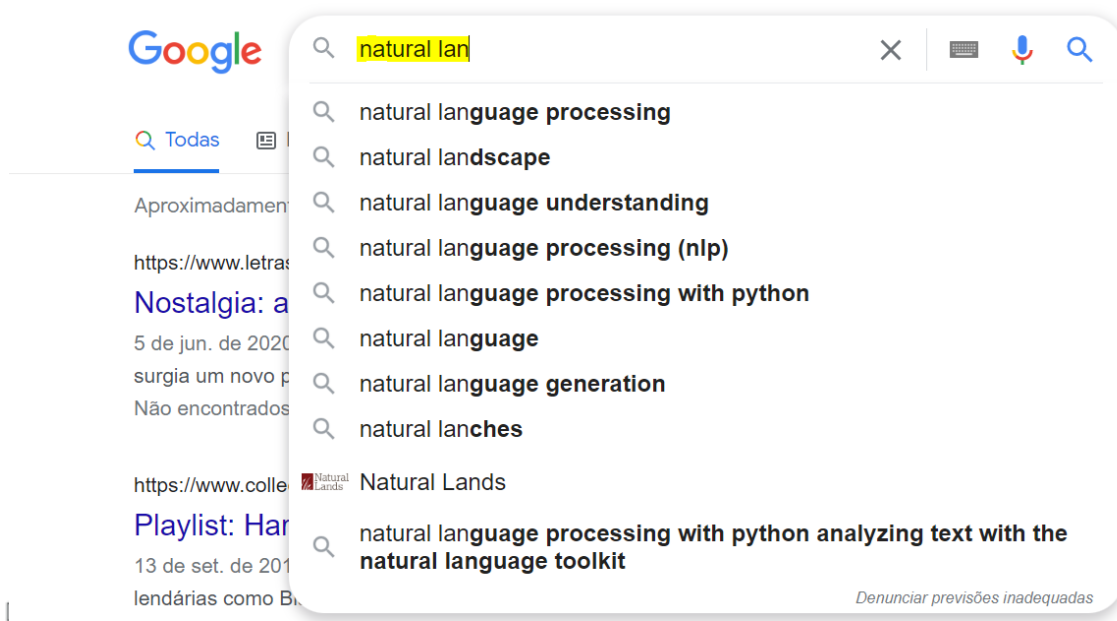
Máquina = linguagem de programação

Como observamos nos slides anteriores, a tarefa de ensinar a máquina compreender contexto não se torna tão simples.

Nesse modulo, vamos aprender a ensinar a máquina entender a linguagem natural humana!

Aonde se utiliza NLP? Por que um cientista de dados precisa aprender?

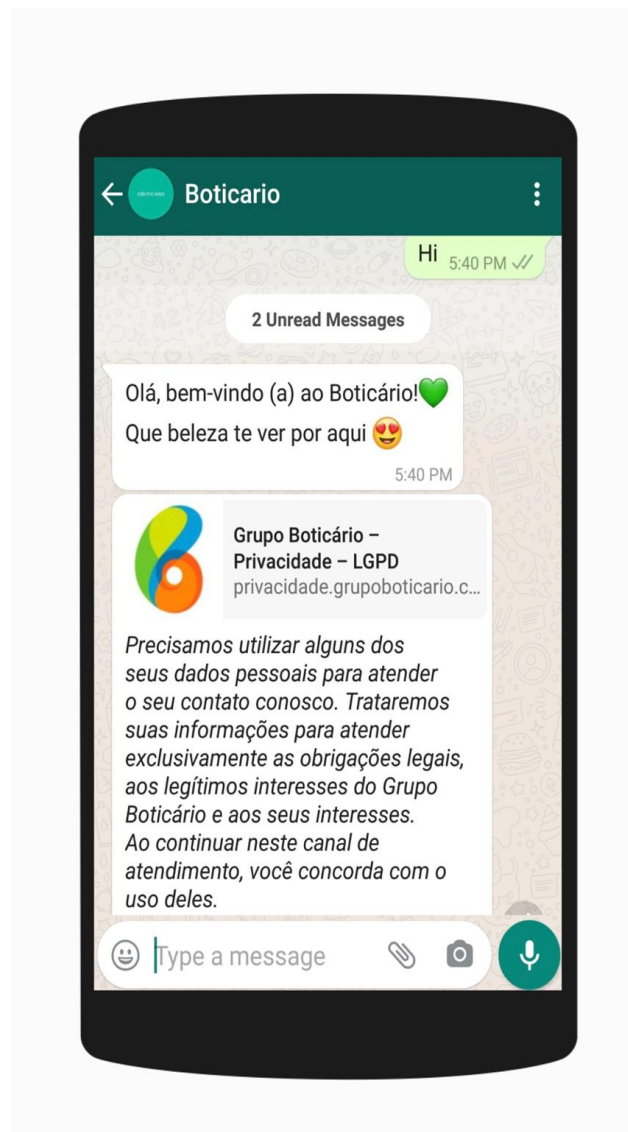
Ferramentas de busca de informações / previsão de palavras



Análise de sentimento / Pesquisas de satisfação



Chatbots



Assistentes virtuais



Processos jurídicos e contratos



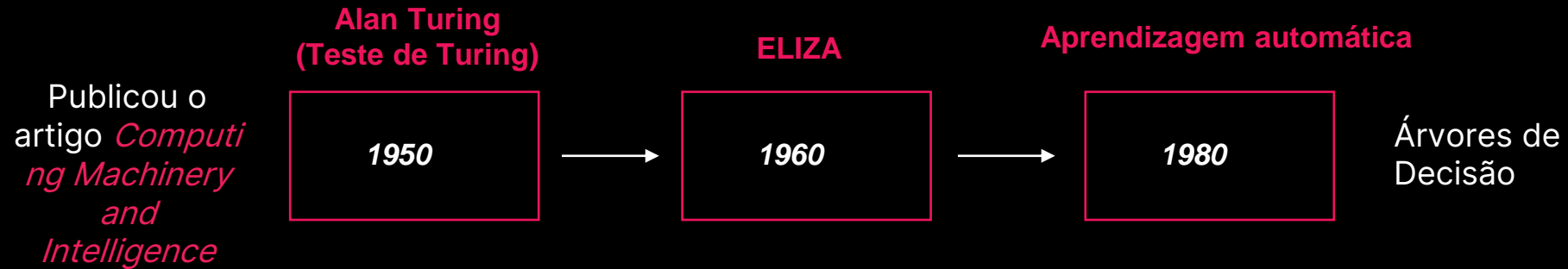


De onde veio a **NLP**?

Origem

Vamos voltar um pouco no tempo...

Simula uma
psicoterapeuta.

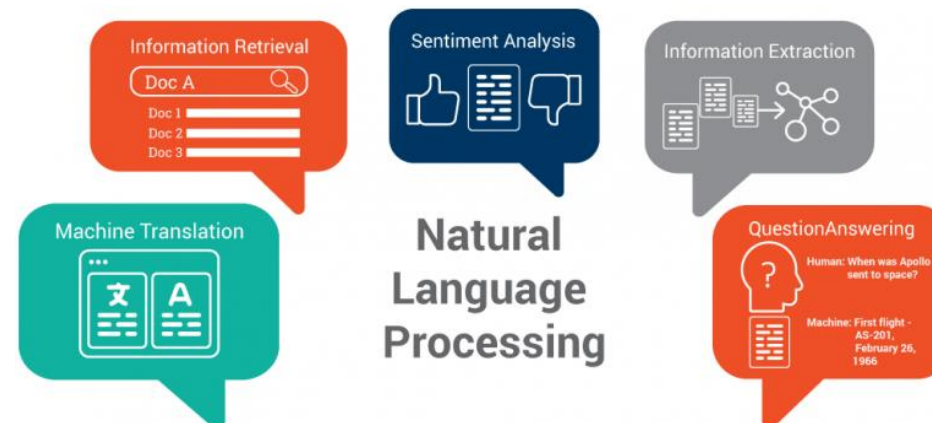


A partir de modelos linguísticos como a teoria da gramática generativa de Noam Chomsky e **modelos estatístico** o NLP processa, analisa e gera valor tendo como input dados em linguagem natural, sejam **textos ou áudios**.

O que vamos aprender nesse módulo?

Trabalhar com texto não se torna tão simples como trabalhar com dados estruturados, porém existem algumas técnicas e processamentos que possibilitam a criação de **algoritmos capazes de interpretar a linguagem humana** e ser a solução de muitos problemas de negócio.

- Manipulação de texto:
 - Tokenização e N-grama
 - Regex
 - Stop-words
 - Normalização de Texto
 - Stemmer o Lemmatizer
- POS-Tagger
- Biblioteca NLTK (Natural Language Toolkit)
- Algoritmos de classificação de texto e LDA (Latent Dirichlet Allocation)
- Análise de sentimento
- Skip-gram
- Word2Vec



Obrigada!

Ana Raquel



[linkedin.com/ana-raquel-fernandes-cunha](https://www.linkedin.com/ana-raquel-fernandes-cunha)

Copyright © 2023 | Ana Raquel Fernandes Cunha

Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem consentimento formal, por escrito, do professor/autor.