

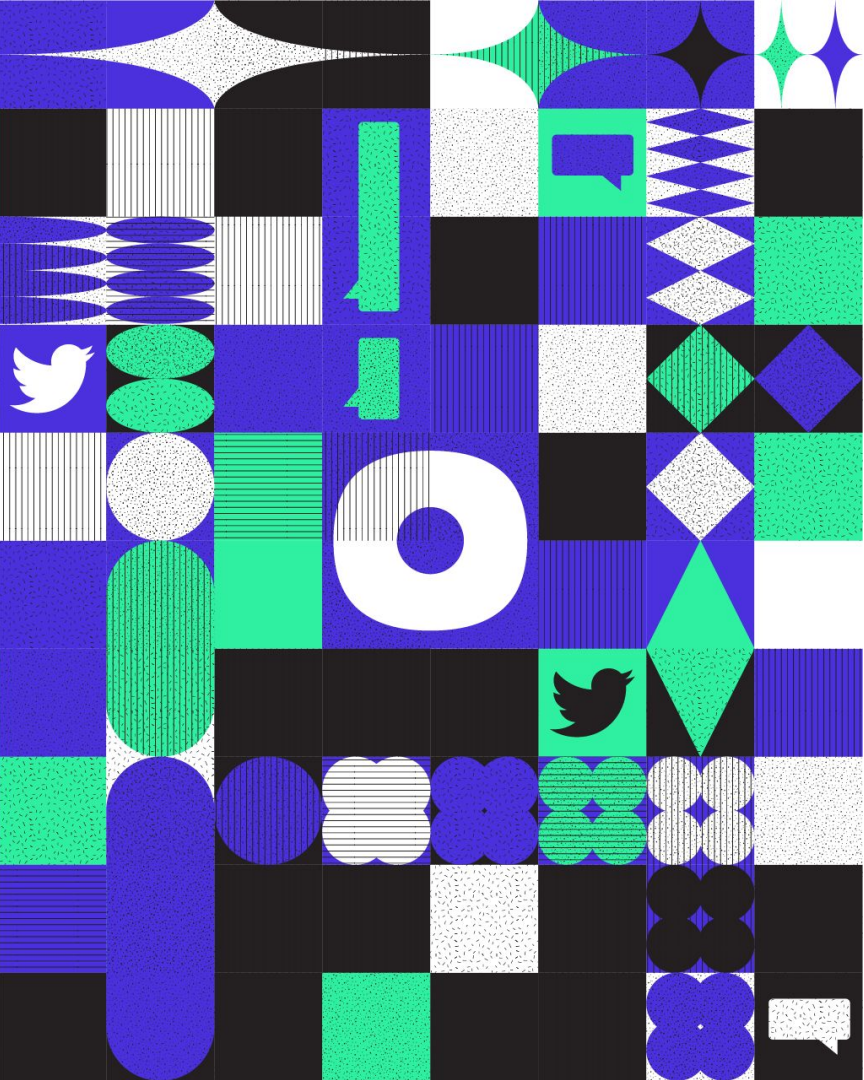
Hands On

<https://nucle.ooo/coda-colab>

<https://nucle.ooo/coda-slides>

<https://nucle.ooo/coda-dados>





Por dentro da caixa-preta: **Uma abordagem prática de IA e NLP.**

Henrique Rieger e Michel Gomes



Henrique Rieger

É estudante de Ciência da Computação na UFPR e é apaixonado por Matemática e Paleontologia.

Foi diretor de comunicação da Ecomp (empresa júnior de Computação da UFPR) em 2021 e desde 2022 é estagiário de tecnologia no Núcleo Jornalismo, lidando com infraestrutura e raspagens de dados.



Michel Gomes

Formado em Ciência da Computação pela UTFPR, mestrando em Sistemas Inteligentes na mesma instituição.

Atua como programador no Núcleo Jornalismo desde 2022.

Tem profundo interesse em Inteligência Artificial, raspagem de dados e eletrônica básica, e é apaixonado por explorar a intersecção entre esses campos.



Hands On

<https://nucle.ooo/coda-colab>

<https://nucle.ooo/coda-slides>

<https://nucle.ooo/coda-dados>



Redes Neurais

Ideias e fundamentos

Inteligência Artificial e Aprendizado de Máquina

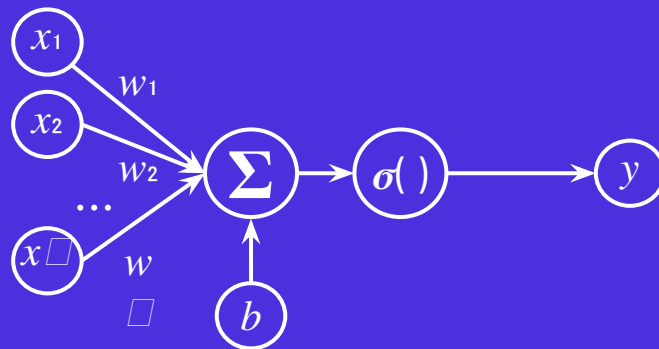
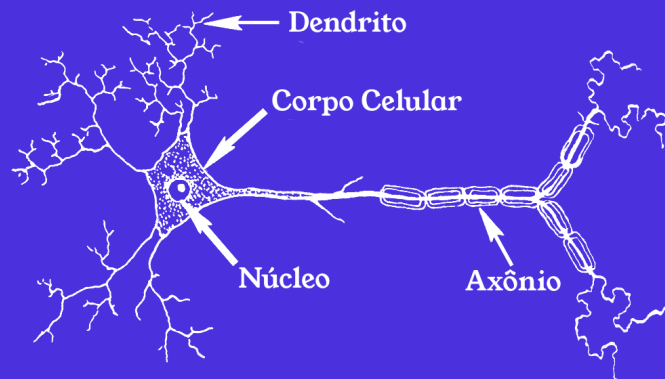
- Inteligência Artificial:

“Computação usando agentes inteligentes, que são capazes de tomar decisões sem programação explícita por um humano.”

- Aprendizado de Máquina:

“Área de IA que trata da ciência e da arte de fazer com que máquinas aprendam comportamentos através de dados.”

PERCEPTRON: O Neurônio Artificial



PERCEPTRON: O Neurônio Artificial

$$y = \sigma(\vec{w}^T \cdot \vec{x} + b)$$

- Multiplica as **entradas** pelos **pesos**;
- Soma tudo + **bias**;
- Passa por uma **função de ativação**;

Como definir \vec{w} e b ?

Atualização de Pesos

TREINAMENTO:

1. Inicie os pesos e *bias* com valores aleatórios;
2. Avalie um exemplo de teste com o valor atual;
3. Calcule o erro;
4. Atualize os pesos e *bias* conforme o erro;
5. Repita o passo 2 para o próximo exemplo;

Repita esse algoritmo até que a base se esgote, e depois comece de novo.

Validação

- Quando eu paro?
 - Quando um conjunto de **validação** atingir o erro mínimo;
- Usamos um conjunto diferente do treinamento para avaliar a qualidade da resposta;
- Evita *overfitting*;

Teste

- Outro conjunto de dados é usado como teste;
- Avalia a qualidade final do classificador;
- Mede:
 - ACURÁCIA;
 - Precisão;
 - Recall;
 - F1-Score;
 - Matriz de confusão;
 - ...

NLP

Ideias e fundamentos

Processamento de Linguagem Natural

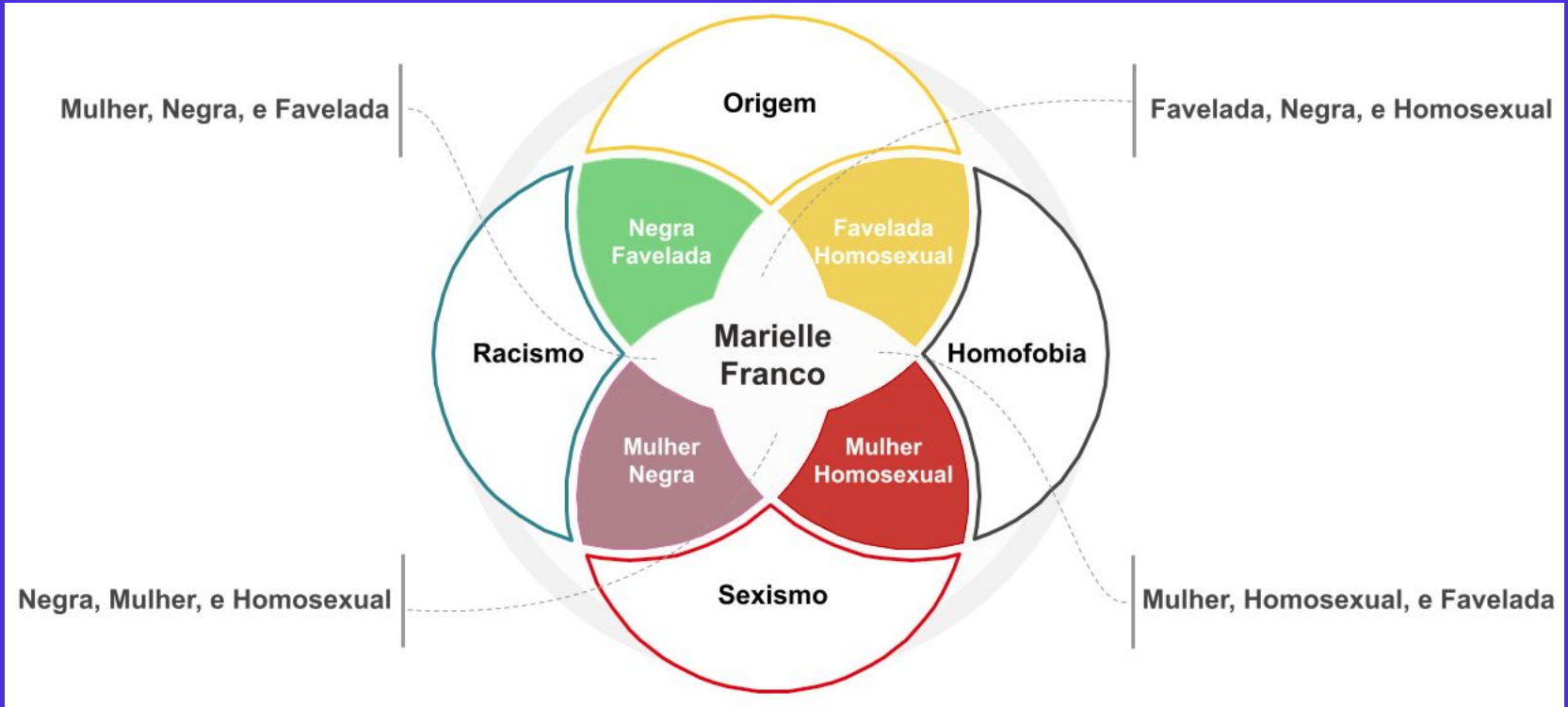
Manipulação/Interpretação seja de texto ou fala da linguagem Humana para processamento computacional;



A word cloud graphic on a white background. The central text reads "Transform Data Into Insights". Surrounding this central text are various other words in different colors and sizes, including: "options", "import", "count", "frequency", "font", "visualize", "free", "generator", "sentiment", "cloud", "print", "color", "download", "word", and "customize".

- Análise de sentimentos;
- Autocompletar;
- Filtro de spam;
- Chatbots.

Exemplo concreto de NLP / PLN



Exemplo concreto de NLP / PLN

Walter Casagrande PER criticou no Fim de Papo MISC as falas do presidente do Corinthians ORG ,
Duilio Monteiro Alves PER , durante uma sabatina para fazer o balanço de sua gestão. Para o colunista do UOL MISC ,
Duilio PER mentiu sobre seu período à frente do clube como quem vive numa realidade paralela.

'Ele vive no universo paralelo': "É um universo paralelo, ele vive num universo paralelo. Ele faz uma narrativa que eu não sei se ele mente ou se ele acredita plenamente naquilo que ele fala, porque quem está olhando, está ouvindo, fala: 'o que esse cara está falando? Onde o Corinthians ORG é mais organizado que Palmeiras ORG e Flamengo ORG ? Onde o Corinthians ORG é maior que todo mundo no Brasil LOC ?'.

Fonte: <https://demos.explosion.ai/displacy-ent>

Embeddings



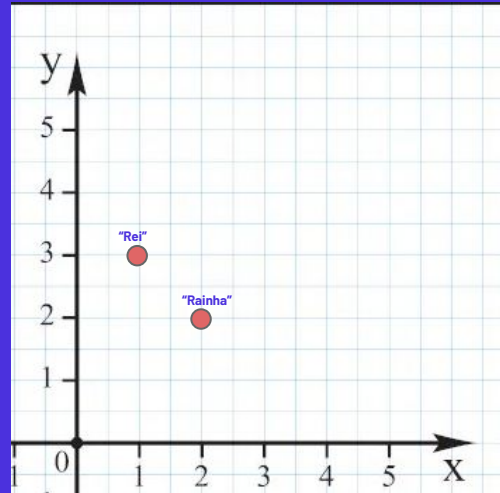
Embeddings

Necessidade de representação (números)



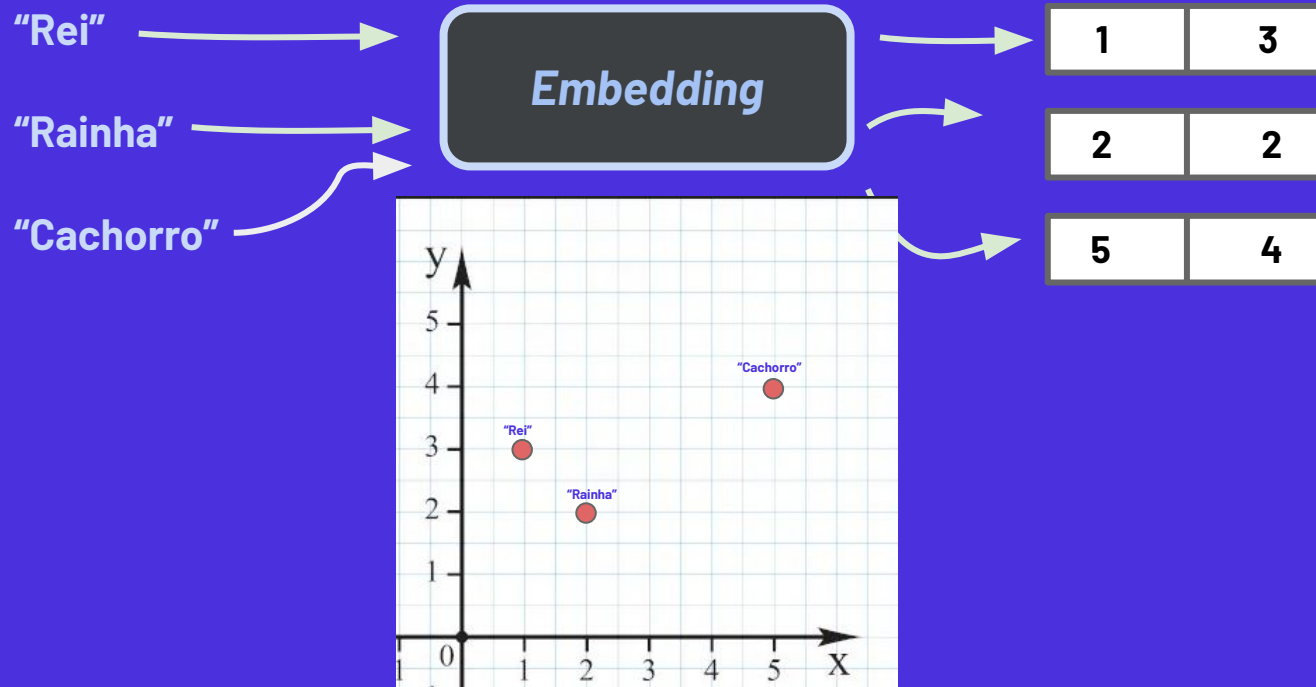
Embeddings

Necessidade de representação (números)



Embeddings

Necessidade de representação (números)



No mundo real!

```
Rei: [-0.460492 -0.470812  0.670537 -2.76186  -0.622286  1.987087  1.865384  
-1.438193 -0.232919 -0.033758  0.244596 -0.893089 -1.188472  0.114873  
-0.199623  0.418087  0.631197 -0.822691 -1.069761  0.610933 -0.567392  
-0.664472  0.835684 -0.140434  0.027169 -0.944068 -0.05241  -0.849329  
-0.905386  0.610521  0.578016 -0.113413  0.044567  1.453539 -0.101304  
-0.494967  0.561068 -0.923575 -0.339458  0.70646  0.253763 -0.176569  
-0.316958  0.045273  0.17138  0.595086  0.007987  1.160012  0.234011  
-0.480946]
```

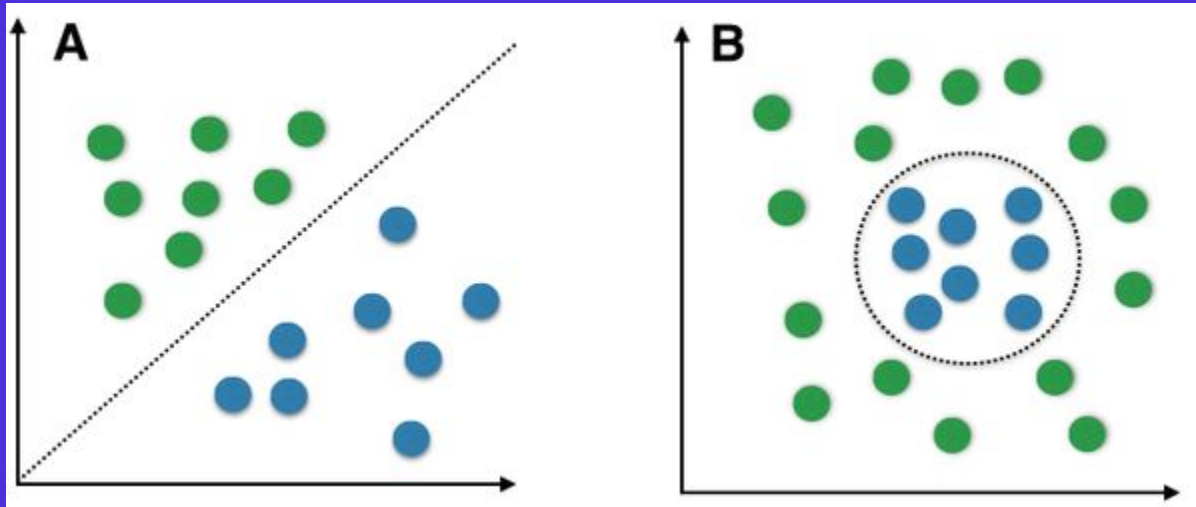
```
Rainha: [-6.809640e-01 -5.037400e-01  4.477700e-01 -2.508161e+00 -4.318160e-01  
 6.174360e-01  1.279938e+00 -1.127896e+00  3.594800e-01  3.583520e-01  
 3.696110e-01 -8.266820e-01 -2.231450e-01 -1.217260e+00  7.343600e-01  
 3.738890e-01  5.589870e-01 -3.882920e-01 -1.257843e+00  4.778760e-01  
-1.454820e-01 -6.047250e-01  1.103206e+00  3.580000e-04 -9.925100e-02  
-5.752430e-01 -2.464990e-01 -7.437500e-01 -1.197329e+00  8.515760e-01  
-2.576630e-01 -4.668790e-01  3.466400e-01  5.944420e-01 -1.117638e+00  
 3.680880e-01  4.041270e-01 -1.125877e+00 -2.312270e-01  2.164400e-02  
-1.893950e-01  1.305760e-01 -2.246610e-01  5.550500e-02 -5.592660e-01  
 6.581200e-02 -2.552280e-01 -1.439760e-01  9.989890e-01 -2.378500e-02]
```

Redes Neurais

Multi-layer Perceptron

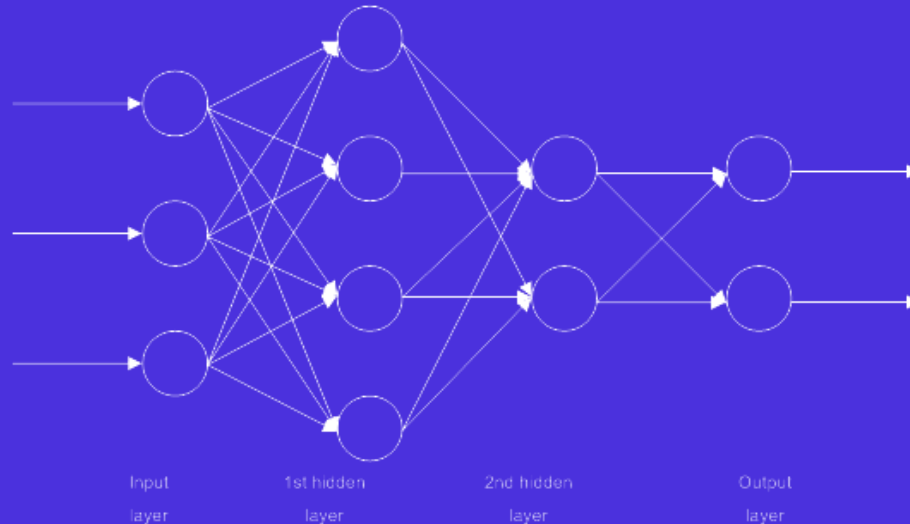
Problemas do Perceptron

- Classificador Linear;
- Muito “fraquinho”;
- Resolve apenas problemas limitados;

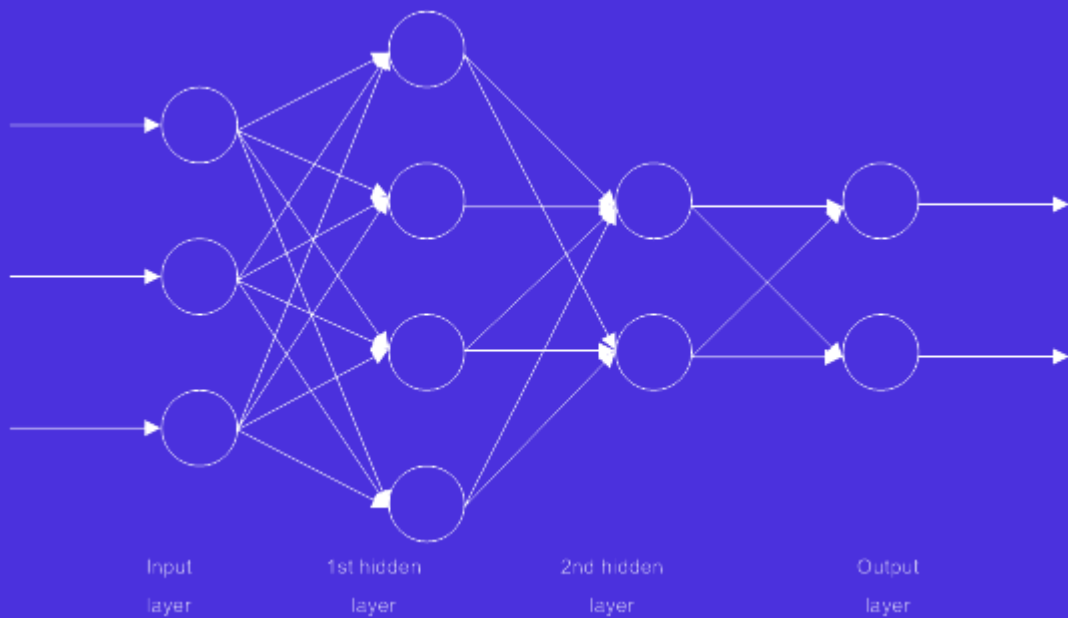


Solução - Multi-layer Perceptron

- Combinar vários Perceptrons!
- “Mais ou menos” como o cérebro funciona;
- Organiza os neurônios em camadas;
- Cada camada processa um “pedaço” do problema;



MLP - Multi-layer Perceptron



MLP - Multi-layer Perceptron

- Como atualizar os pesos?
- Mesma ideia de antes!
- A diferença é que precisa atualizar **todas** as camadas;
- Essa atualização é feita de trás para frente:
 - *Backpropagation*;
- Descida de Gradiente;
 - *Stochastic Gradient Descent*;

Outros Links Úteis

- [Scikit-learn](#) – Documentação da biblioteca;
- [Neural Networks - 3Blue1Brown](#) – Aulas mais a fundo sobre como redes neurais funcionam;
- [Investigate.ai](#) – Casos reais de jornalismo de dados com Inteligência Artificial;
- [Brasileiras em PLN](#) – Comunidade feminina de Processamento de Linguagem Natural.





Valeu!

Imagens adaptadas do [Wikimedia Commons](#)

Henrique Rieger

henrique@nucleo.jor.br

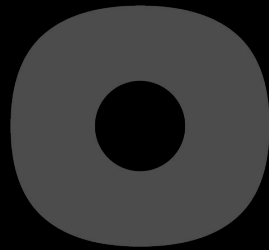
in/henrique-rieger

Michel Gomes

michel@nucleo.jor.br

@jmallone2

in/michel-gomes



NÚCLEO
JORNALISMO • DADOS • TECNOLOGIA