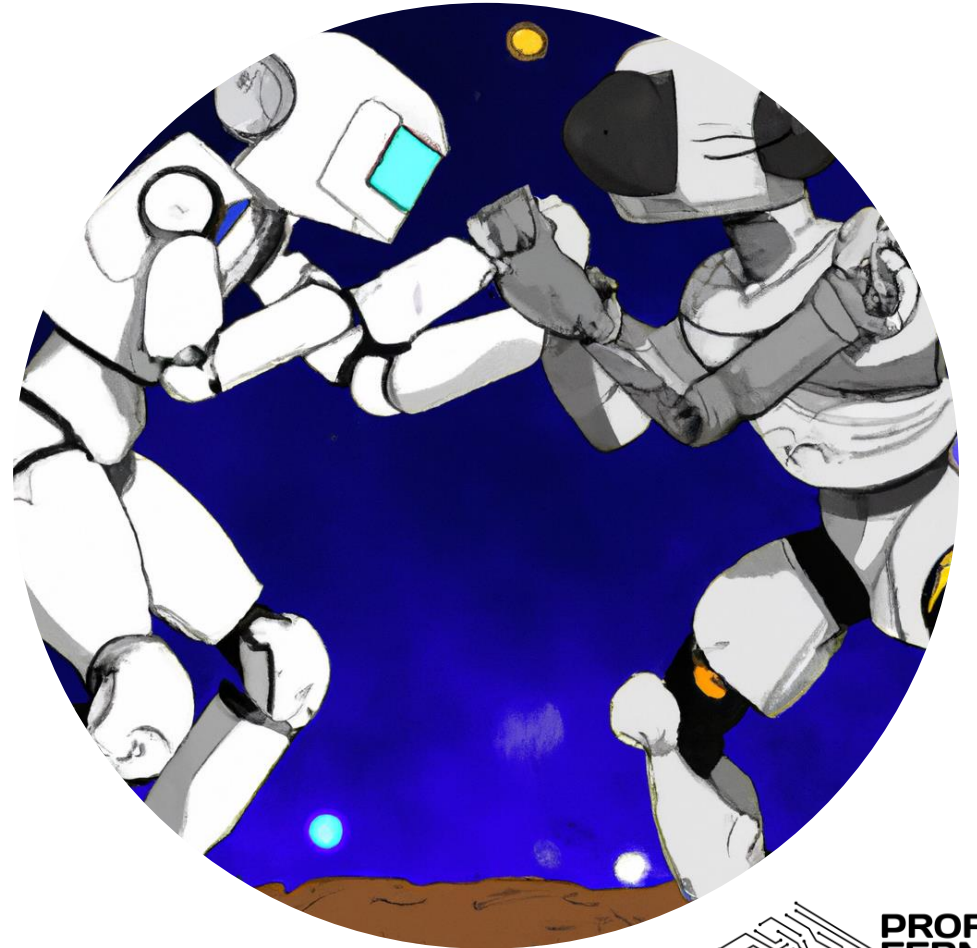


Formação Cientista de Dados

NLP

Word Embedding

- Representações computacionais de texto
 - Independentes de contexto
 - Dependentes de contexto



Word Embedding

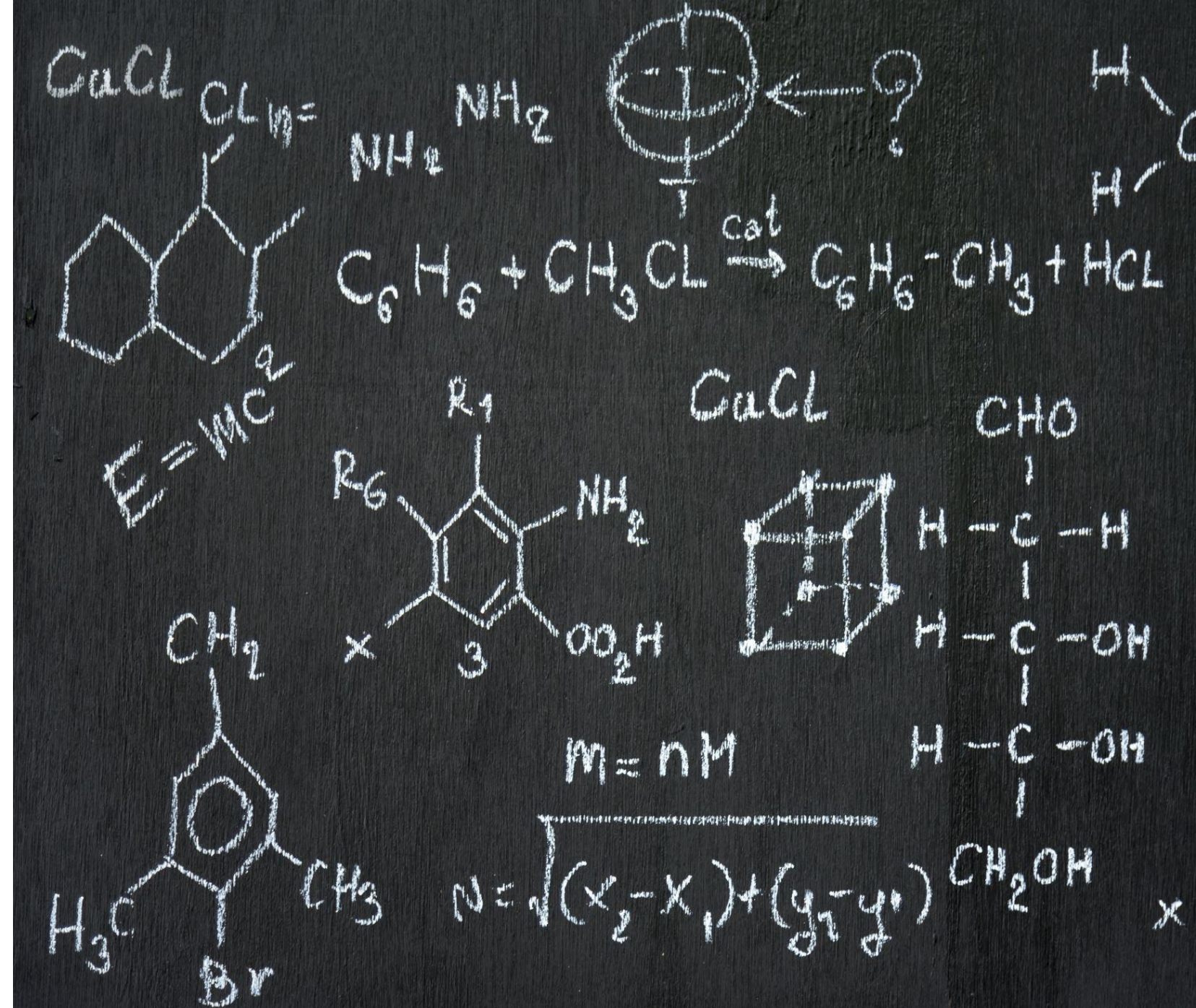
- **Propriedade Semânticas:** Vetor de Gato é semelhante ao de Cachorro
- **Propriedades sintáticas:** Função gramatical da palavra: Substantivo e Verbo
- **Propriedades Distributivas:** Palavras que aparecem em contexto semelhante

Word Embedding					
Gato	0.3	0.4	0.1	-1.3	0.0
Cachorro	0.4	0.2	0.7	-1.1	0.2
Muro	0.2	0.2	0	-0.8	0.1

Semelhança de cosseno			
	Gato	Cachorro	Muro
Gato	1	0.9	0.5
Cachorro	0.9	1	0.7
Muro	0.5	0.7	1

Word2Vec

- Através de um processo de treinamento, produz um vetor demonstrando matematicamente a relação entre palavras



BANANA.VECTOR

```
array([2.02280000e-01, -7.66180009e-02,  3.70319992e-01,
       3.28450017e-02, -4.19569999e-01,  7.20689967e-02,
      -3.74760002e-01,  5.74599989e-02, -1.24009997e-02,
       5.29489994e-01, -5.23800015e-01, -1.97710007e-01,
      -3.41470003e-01,  5.33169985e-01, -2.53309999e-02,
       1.73800007e-01,  1.67720005e-01,  8.39839995e-01,
       5.51070012e-02,  1.05470002e-01,  3.78719985e-01,
       2.42750004e-01,  1.47449998e-02,  5.59509993e-01,
       1.25210002e-01, -6.75960004e-01,  3.58420014e-01,
      # ... and so on ...
       3.66849989e-01,  2.52470002e-03, -6.40089989e-01,
      -2.97650009e-01,  7.89430022e-01,  3.31680000e-01,
      -1.19659996e+00, -4.71559986e-02,  5.31750023e-01], dtype=float32)
```


Transformers

- Arquitetura de RNA utilizada para NLP
- Mecanismos “self-attention”
- Ótima em Tarefas de Contexto
- Pode utilizar word embeddings como entrada para a RNA

