

# **Os bastidores do ChatGPT:**

**Como funciona esta tecnologia de PLN?**

**Prof. Luciano Antonio Digiampietri**  
**@digiampietri**

**O que é o ChatGPT?**

# O que é o ChatGPT?

*“Eu sou o ChatGPT, um modelo de linguagem desenvolvido pela OpenAI com base na arquitetura GPT-3.5. Eu fui projetado para conversar com pessoas e responder a uma ampla variedade de perguntas e tópicos.”*

# O que é um modelo de linguagem?

Um modelo de linguagem é uma distribuição de probabilidade sobre sequências de palavras.

# O que é um modelo de linguagem?

Um modelo de linguagem é uma distribuição de probabilidade sobre sequências de palavras.

Dada uma sequência de palavras, um modelo de linguagem atribui uma probabilidade a essa sequência.

# **Modelo de linguagem - tarefa básica**

**Identificação da próxima palavra (mais provável).**

**[Entrada]: Uma sequência de palavras**

**[Saída]: A palavra mais provável para dar sequência ao texto**

# Modelo de linguagem - tarefa básica

Identificação da próxima palavra (mais provável).

[Entrada]: Uma sequência de palavras

**“Com grandes poderes vêm grandes”**

[Saída]: A palavra mais provável para dar sequência ao texto

# Modelo de linguagem - tarefa básica

Identificação da próxima palavra (mais provável).

[Entrada]: Uma sequência de palavras

**“Com grandes poderes vêm grandes”**

[Saída]: A palavra mais provável para dar sequência ao texto

**“responsabilidades”**



## O que é a arquitetura GPT-3.5?

A arquitetura GPT-3.5 (*Generative Pre-trained Transformer*) é um modelo de linguagem neural desenvolvido pela OpenAI.

A arquitetura GPT-3.5 é baseada em uma rede neural de transformadores (*transformers*) pré-treinada em grandes quantidades de dados textuais.

## O que é a arquitetura GPT-3.5?

**Modelo Generativo:** modelo capaz de produzir conteúdos com fluidez semelhante aos conteúdos produzidos por humanos.

## O que é a arquitetura GPT-3.5?

**Modelo Generativo:** modelo capaz de produzir conteúdos com fluidez semelhante aos conteúdos produzidos por humanos.

**Pré-treinado:** o treinamento não é realizado de forma *on-line*. É bastante custoso (computacionalmente falando).

**Transformer: ...**

# Processamento de Língua Natural

Processamento de Língua Natural (**PLN**) (*Natural Language Processing - NLP*) é um campo interdisciplinar envolvendo ciência da computação, inteligência artificial e linguística focado na **interação entre computadores e línguas humanas (naturais)**, em especial no **processamento ou produção de dados relacionados a estas línguas (textos, áudios, vídeos etc.)**.

# Representação de Texto

Em abordagens mais tradicionais, um texto era representado computacionalmente como um “saco” de palavras (*Bag-of-Words*)<sup>1</sup>.

---

<sup>1</sup>Harris, Z. S. (1954)

# Representação de Texto

Em abordagens mais tradicionais, um texto era representado computacionalmente como um “saco” de palavras (*Bag-of-Words*)<sup>1</sup>.

Nessa representação as seguintes frases são idênticas:

*“Eu não estou feliz, eu estou triste.”*

*“Eu não estou triste, eu estou feliz.”*

---

<sup>1</sup>Harris, Z. S. (1954)

# Representação de Texto

Outras representações utilizam conjuntos de **palavras**<sup>2</sup> consecutivas (**n-gramas**).

---

<sup>2</sup>Conjuntos de **tokens**, que podem ser palavras, caracteres, fonemas etc.

# Representação de Texto

Outras representações utilizam conjuntos de **palavras**<sup>2</sup> consecutivas (**n-gramas**).

Nessa representação, as mesmas frases possuem **trigramas** bem diferentes:

... “*não estou feliz*” ... “*eu estou triste*”

... “*não estou triste*” ... “*eu estou feliz*”

---

<sup>2</sup>Conjuntos de **tokens**, que podem ser palavras, caracteres, fonemas etc.



# Representação de Texto

Porém, como reconhecer que as duas primeiras frases a seguir têm **significados** semelhantes e a terceira não?

# Representação de Texto

Porém, como reconhecer que as duas primeiras frases a seguir têm **significados** semelhantes e a terceira não?

*“Eu estou feliz.”*

*“Eu estou alegre.”*

*“Eu estou triste.”*

# Representação de Texto

**Word embedding**: estratégia para representar cada palavra, tipicamente, **na forma de um arranjo de números reais** de forma a tentar representar o “**significado**” da palavra.

---

<sup>3</sup>word2vec é uma das estratégias mais conhecidas (Mikolov et al., 2013)

# Representação de Texto

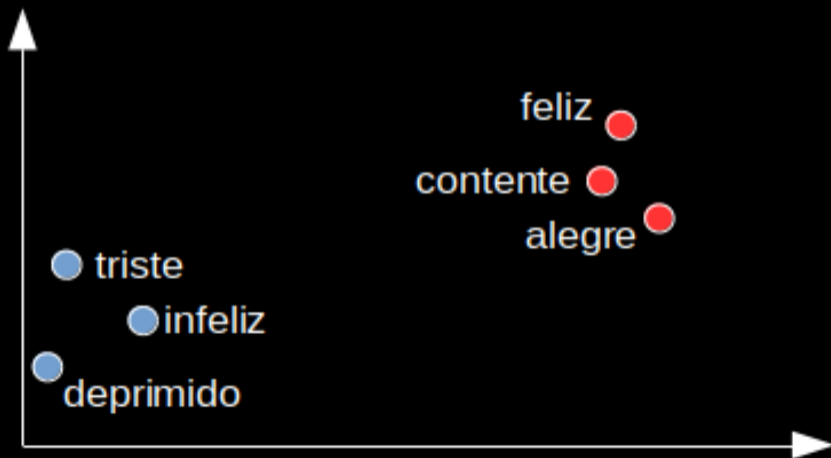
**Word embedding**: estratégia para representar cada palavra, tipicamente, **na forma de um arranjo de números reais** de forma a tentar representar o “**significado**” da palavra.

Isto é, espera-se que palavras com **significados semelhantes** tenham **representações vetoriais próximas**<sup>3</sup>.

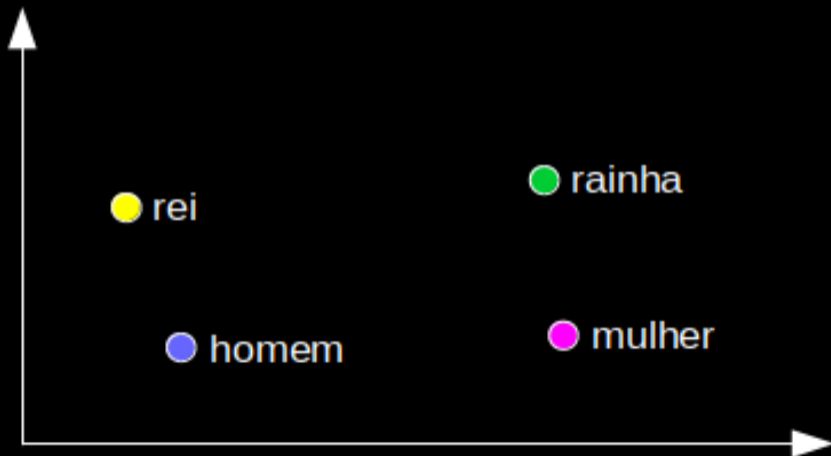
---

<sup>3</sup>word2vec é uma das estratégias mais conhecidas (Mikolov et al., 2013)

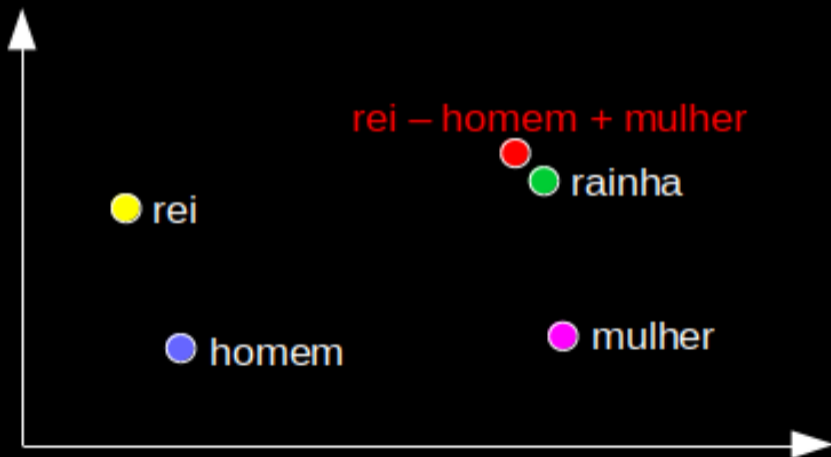
# Word embedding



# Word embedding



# Word embedding



# Word embedding





# *Large Language Models (LLMs)*

Um grande modelo de linguagem (LLM) é um modelo de linguagem que consiste tipicamente em uma **rede neural profunda** com muitos **parâmetros**, treinada com grandes quantidades de texto não rotulado usando **aprendizado auto-supervisionado**.

# Representação

***Transformer***: modelo de aprendizado profundo que utiliza um mecanismo de **auto-atenção**<sup>4</sup> para identificar o peso de cada parte dos dados de entrada.

---

<sup>4</sup>Vaswani et al. (2017)

# Representação

***Transformer***: modelo de aprendizado profundo que utiliza um mecanismo de **auto-atenção**<sup>4</sup> para identificar o peso de cada parte dos dados de entrada.

Modelos inicialmente usados na tradução de textos utilizando arquiteturas do tipo *encoder–decoder*.

---

<sup>4</sup>Vaswani et al. (2017)

# Transformer (Vaswani et al., 2017)

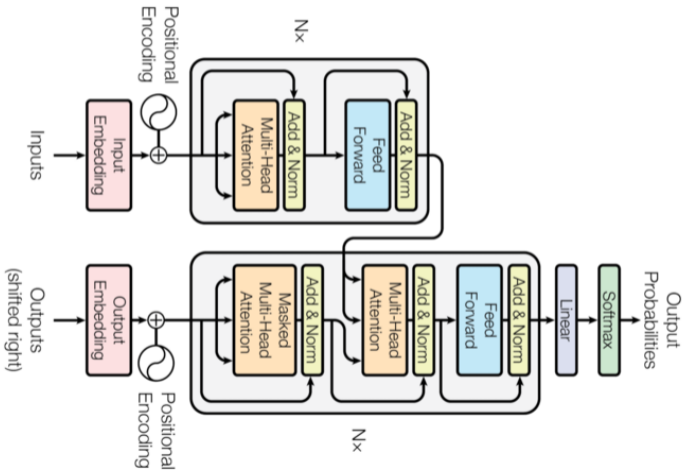


Figure 1: The Transformer - model architecture.

# GPT - Desempenho

**Estado da arte em nove de 12 tarefas, incluindo<sup>5</sup>:**

- sentença gramaticalmente correta
- identificação de sentimentos
- verificação de equivalência semântica
- perguntas/respostas
- raciocínio de senso comum

---

<sup>5</sup>Radford et al. (2018)

# BERT

***Bidirectional Encoder Representations from Transformers (BERT)*** é um modelo de representação de língua bidirecional criado por pesquisadores da Google pré-treinado usando apenas corpora textual.

# BERT

***Bidirectional Encoder Representations from Transformers (BERT)*** é um modelo de representação de língua bidirecional criado por pesquisadores da Google pré-treinado usando apenas corpora textual.

***Masked-Language Modeling (Masked LM)***

***Next Sentence Prediction (NSP)***

# BERT

**Modelos livres de contexto**, como word2vec, geram uma **representação única de cada palavra**, já os **modelos contextuais**, como o BERT e o GPT, **consideram o contexto de ocorrência da palavra**.



# BERT - Desempenho

Estado da arte em 11 tarefas<sup>6</sup>, incluindo:

- sentença gramaticalmente correta
- identificação de sentimentos
- verificação de equivalência semântica
- relação entre premissa e hipótese
- perguntas/respostas
- relação entre sentenças (concorda/discorda/não relacionada)
- identificação da entidade referenciada pelo pronome
- identificação da sequência mais plausível

---

<sup>6</sup>Devlin et al. (2019)

# **Generative pre-trained transformers (GPT)**

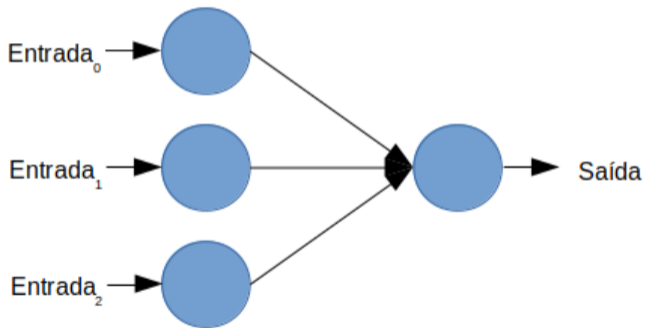
**Transformadores pré-treinados generativos (GPT) são uma família de grandes modelos de linguagem (LLMs).**

# Generative pre-trained transformers (GPT)

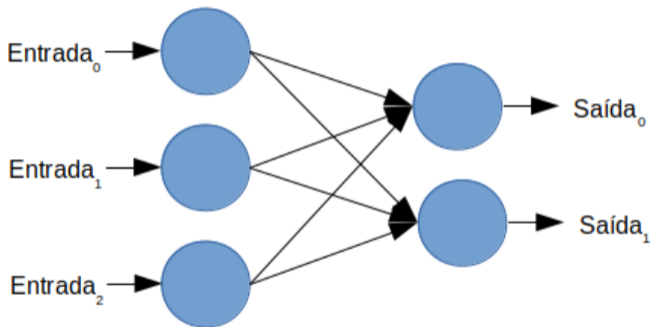
Transformadores pré-treinados generativos (GPT) são uma família de grandes modelos de linguagem (LLMs).

Possuem **muitos parâmetros** e são treinados com **grandes volumes de dados**.

# Parâmetros - Imagine um regressão linear

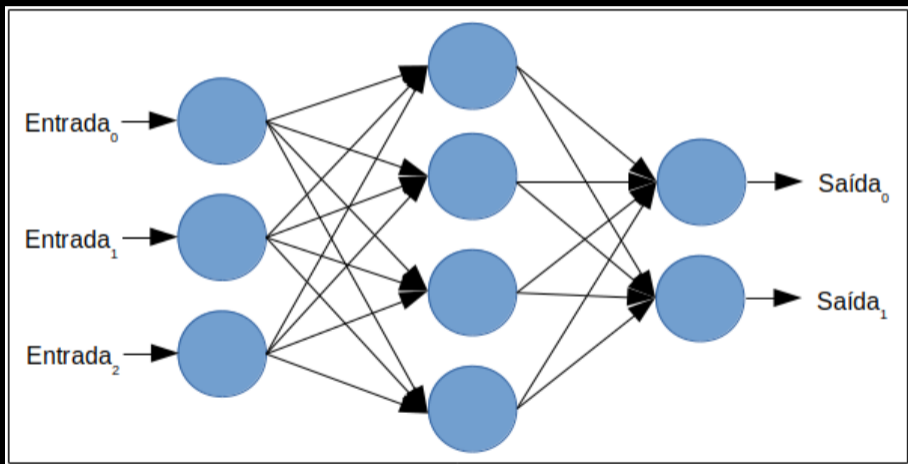


# Parâmetros - Perceptron Simples<sup>7</sup>



<sup>7</sup>Rosenblatt, F. (1958)

# Parâmetros - Perceptron Multicamadas



# BERT e GPT - Tamanho

<b>Modelo</b>	<b>Lançamento</b>	<b>Parâmetros (bilhões)</b>	<b>Dados</b>
<b>GPT 1</b>	<b>2018</b>	<b>0,12</b>	<b>4,5 GB</b>
<b>BERT</b>	<b>2018</b>	<b>0,34</b>	<b>-</b>
<b>GPT 2</b>	<b>2019</b>	<b>1,50</b>	<b>40 GB</b>
<b>GPT 3</b>	<b>2020/2022</b>	<b>175</b>	<b>570 GB</b>
<b>GPT 4</b>	<b>2023</b>	<b>100.000?</b>	<b>?</b>

# BERT e GPT - Tamanho

Modelo	Lançamento	Parâmetros (bilhões)	Dados
GPT 1	2018	0,12	4,5 GB
BERT	2018	0,34	-
GPT 2	2019	1,50	40 GB
GPT 3	2020/2022	175	570 GB
GPT 4	2023	100.000?	?

**GPT 1 - modelo de linguagem**



# BERT e GPT - Tamanho

Modelo	Lançamento	Parâmetros (bilhões)	Dados
GPT 1	2018	0,12	4,5 GB
BERT	2018	0,34	-
GPT 2	2019	1,50	40 GB
GPT 3	2020/2022	175	570 GB
GPT 4	2023	100.000?	?

**GPT 2 - “permite” passar a tarefa como parte da entrada**

# BERT e GPT - Tamanho

Modelo	Lançamento	Parâmetros (bilhões)	Dados
GPT 1	2018	0,12	4,5 GB
BERT	2018	0,34	-
GPT 2	2019	1,50	40 GB
GPT 3	2020/2022	175	570 GB
GPT 4	2023	100.000?	?

**GPT 3 - *zero/one/few-shot learning***

# GPT - Parâmetros

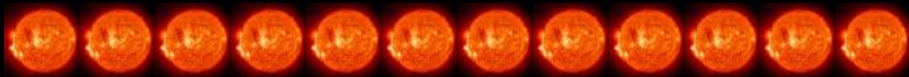
GPT 1



GPT 2



GPT 3

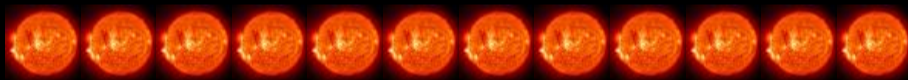


# GPT - Parâmetros (em escala)

GPT 1

GPT 2

GPT 3



# GPT-3.5 / ChatGPT

O ChatGPT consiste de uma versão aprimorada do modelo de linguagem, otimizada para conversas.

---

<sup>8</sup>InstructGPT

<sup>9</sup>Ziegler et al. (2020)

# GPT-3.5 / ChatGPT

O ChatGPT consiste de uma versão aprimorada do modelo de linguagem, otimizada para conversas.

Este “aprimoramento” é realizado por aprendizado por reforço a partir de *feedback* humano<sup>8</sup> (***RLHF - Reinforcement Learning from Human Feedback***).<sup>9</sup>

---

<sup>8</sup>InstructGPT

<sup>9</sup>Ziegler et al. (2020)

# InstructGPT

Processo realizado em três etapas:

1. **Rotuladores** apresentam o comportamento esperado a certas perguntas.  
Um processo de **aprendizado supervisionado** é utilizado para aprimorar [*fine tuning*] o GPT 3.

# InstructGPT

Processo realizado em três etapas:

2. Diversas saídas são produzidas para diferentes perguntas e os **rotuladores** ordenam as saídas. Um processo de **recompensa** é treinado.



# InstructGPT

Processo realizado em três etapas:

3. Perguntas são amostradas do conjunto de dados, a política produzida gera a saída, a qual é avaliada pelo **modelo de recompensa** e com base na avaliação **a política é atualizada.**

**ChatGPT - o que ele não é?**

# ChatGPT - o que ele não é?

Uma **enciclopédia digital** com **todo o conhecimento da humanidade.**

# ChatGPT - limitações

*Ocasionalmente pode gerar informações incorretas.*

# ChatGPT - limitações

*Ocasionalmente pode gerar informações incorretas.*

*Ocasionalmente, pode produzir instruções prejudiciais ou conteúdo tendencioso.*

# ChatGPT - limitações

*Ocasionalmente pode gerar informações incorretas.*

*Ocasionalmente, pode produzir instruções prejudiciais ou conteúdo tendencioso.*

*Conhecimento limitado do mundo e eventos após 2021.*

**O que é o ChatGPT?**

# O que é o ChatGPT?





# **Biografia (não autorizada) de Luciano Digiampietri**

# Biografia (não autorizada) de Luciano Digiampietri



Luciano Antonio Digiampietri é um professor e pesquisador da Universidade de São Paulo (USP), uma das principais instituições de ensino e pesquisa do Brasil. Ele é docente do Instituto de Ciências Matemáticas e de Computação (ICMC) desde 1992, onde desenvolve pesquisas na área de Ciência da Computação e Tecnologia da Informação.

O professor Digiampietri possui graduação em Engenharia Elétrica pela Universidade Estadual de Campinas (UNICAMP), mestrado em Ciência da Computação pela USP e doutorado em Engenharia Elétrica pela Universidade Estadual de Campinas (UNICAMP). Além disso, realizou estágio pós-doutoral na Universidade de Maryland, nos Estados Unidos.

# Biografia (não autorizada) de Luciano Digiampietri



Luciano Antonio Digiampietri é um professor e pesquisador da Universidade de São Paulo (USP), uma das principais instituições de ensino e pesquisa do Brasil. Ele é docente do Instituto de Ciências Matemáticas e de **Escola de Artes, Ciências e Humanidade (EACH)** **2008** Computação (ICMC) desde 1992, onde desenvolve pesquisas na área de Ciência da Computação e Tecnologia da Informação.

O professor Digiampietri possui graduação em **Ciência da Computação** Engenharia Elétrica pela Universidade Estadual de Campinas (UNICAMP), ~~mestrado em Ciência da Computação pela USP e doutorado em **Ciência da Computação** Engenharia Elétrica pela Universidade Estadual de Campinas (UNICAMP).~~ Além disso, realizou estágio **na Microsoft Research** ~~pos-doutorado na Universidade de Maryland,~~ nos Estados Unidos.

# Biografia (não autorizada) de Luciano Digiampietri

Outra importante contribuição do professor Digiampietri é a sua atuação como líder de diversos projetos de pesquisa e desenvolvimento, incluindo projetos financiados por agências de fomento e empresas nacionais e internacionais. Nesses projetos, ele tem coordenado equipes de pesquisadores e estudantes para a realização de pesquisas de ponta e a criação de tecnologias inovadoras.

Entre os principais temas de pesquisa em que o professor Digiampietri tem se concentrado atualmente, destacam-se a análise de redes sociais e de mídias sociais, a mineração de dados em saúde e medicina, a modelagem de dados e sistemas, a análise de dados em tempo real e a segurança da informação.

# Biografia (não autorizada) de Luciano Digiampietri

## EXAGERADO

Outra importante contribuição do professor Digiampietri é a sua atuação como líder de diversos projetos de pesquisa e desenvolvimento, incluindo projetos financiados por agências de fomento e empresas nacionais e internacionais. Nesses projetos, ele tem coordenado equipes de pesquisadores e estudantes para a realização de pesquisas de ponta e a criação de tecnologias inovadoras.

Entre os principais temas de pesquisa em que o professor Digiampietri tem se concentrado atualmente, destacam-se a análise de redes sociais e de mídias sociais, a mineração de dados em saúde e medicina, a modelagem de dados e sistemas, a análise de dados em tempo real e a segurança da informação.

# Biografia (não autorizada) de Luciano Digiampietri

Ao longo de sua carreira, o professor Digiampietri tem se destacado como um pesquisador de renome na área de Ciência da Computação, com foco em temas como sistemas de informação, mineração de dados, inteligência artificial, aprendizado de máquina, processamento de linguagem natural e web semântica.

Entre as suas contribuições mais importantes para a área, destacam-se a criação do software Weka, que é amplamente utilizado em todo o mundo para mineração de dados, e o desenvolvimento de metodologias para a criação de ontologias na web semântica.

# Biografia (não autorizada) de Luciano Digiampietri

Ao longo de sua carreira, o professor Digiampietri tem se destacado como um pesquisador de renome na área de Ciência da Computação, com foco em temas como sistemas de informação, mineração de dados, inteligência artificial, aprendizado de máquina, processamento de linguagem natural e web semântica.

Entre as suas contribuições mais importantes para a área, destacam-se a criação do software Weka, que é amplamente utilizado para mineração de dados, e o desenvolvimento de metodologias para a criação de ontologias na web semântica.

# Referências

- 1 **Harris, Z. S. (1954). Distributional Structure. *Word*. 10 (2/3): 146–62.**  
<https://doi.org/10.1080/00437956.1954.11659520>
- 2 **Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.** <https://doi.org/10.1037/h0042519>
- 3 **Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Lukasz; Polosukhin, Illia (2017). Attention Is All You Need.**  
<https://doi.org/10.48550/arXiv.1706.03762>
- 4 **Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. (2018) Improving Language Understanding by Generative Pre-Training.** . [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)



# Referências

- 5 Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding  
<https://doi.org/10.48550/arXiv.1810.04805>
- 6 Ouyang, L. et al. (2022). Training language models to follow instructions with human feedback. <https://doi.org/10.48550/arXiv.2203.02155>
- 7 Ziegler, D.M. et al. (2020). Fine-Tuning Language Models from Human Preferences. <https://doi.org/10.48550/arXiv.1909.08593>
- 8 Mikolov, T. et al. (2013). Efficient Estimation of Word Representations in Vector Space. <https://arxiv.org/abs/1301.3781>
- 9 OpenAI - Models: <https://platform.openai.com/docs/models/overview>  
Acessado em 2023-04-20.
- 10 ChatGPT <https://chat.openai.com/> Acessado em 2023-04-12.

# **Os bastidores do ChatGPT:**

**Como funciona esta tecnologia de PLN?**

**Prof. Luciano Antonio Digiampietri**  
**@digiampietri**