

---

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

**ANÁLISE DO FUNCIONAMENTO DOS GRANDES MODELOS DE  
LINGUAGEM E A PRODUÇÃO DE TEXTOS SEMELHANTES AOS  
HUMANOS**

**ANALYSIS OF THE FUNCTIONING OF LARGE LANGUAGE  
MODELS AND THE PRODUCTION OF HUMAN-LIKE TEXTS**

Guilherme Tavares Pinheiro  
Melina Paula Batista Garcia

**Resumo**

Este artigo analisa o funcionamento dos *Large Language Models* (Grandes Modelos de Linguagem) e sua capacidade de gerar textos que muitas vezes são indistinguíveis dos produzidos por humanos. Com foco nos modelos GPT-4 e Gemini, o estudo explora as diferenças em fluência, coerência e originalidade dos textos gerados. Além de examinar as arquiteturas, como *Transformer*, que sustentam essas tecnologias, o trabalho discute os desafios éticos e técnicos envolvidos, incluindo a proteção de dados e a desinformação. A metodologia envolve a comparação prática dos modelos através de experimentos com prompts genéricos e técnicos. Os resultados mostram que, enquanto o GPT-4 se destaca em precisão técnica e abstração, o Gemini demonstra maior capacidade criativa e contextualização. Conclui-se que os LLMs possuem grande potencial para transformar diversas áreas, como educação, desenvolvimento de *software* e produção de conteúdo, mas ainda necessitam de aprimoramentos para lidar com limitações éticas e técnicas.

**Palavras-chave:** Modelos de linguagem, GPT-4, Gemini, Processamento de Linguagem Natural, Inteligência Artificial.

**Abstract**

*This article analysis the functioning of Large Language Models (LLMs) and their ability to generate text that are often indistinguishable from those produced by humans. Focusing on the GPT-4 and Gemini models, the study explores the differences in fluency, coherence, and originality of the generated texts. In addition to examining the architectures, such as the Transformer, that underpin these Technologies, the work discusses the ethical and technical challenges involved, including data protection and misinformation. The methodology involves the practical comparison of the models through experiments with generic and technical prompts. The results show that, while GPT-4 excels in technical accuracy and abstraction, Gemini demonstrates greater creative capacity and contextualization. It is concluded that LLMs have great potential to transform various areas, such as education, software development, and content production, but still require improvements to avoid ethical and technical limitations.*

**Keywords:** Language Models, GPT-4, Gemini, Natural Language Processing, Artificial Intelligence.

---

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

## 1. INTRODUÇÃO

Nos últimos anos, a inteligência artificial (IA) tem avançado significativamente, sobretudo no campo do Processamento de Linguagem Natural (PLN), o que está resultando no desenvolvimento de novas tecnologias cada vez mais sofisticadas. Uma das maiores conquistas desse campo são as *Large Language Models* (Grandes Modelos de Linguagem, da sigla LLM), como GPT-3 e GPT-4, que são capazes de processar e gerar textos que, muitas vezes, são quase indistinguíveis daqueles produzidos por humanos (Brown et al., 2020). O desenvolvimento de LLMs é o resultado de décadas de investimento e progresso em inteligência artificial, aprendizado de máquina e PLN, com saltos proporcionados pelo surgimento de tecnologias mais avançadas, como redes neurais e o refinamento de aprendizado de máquina.

A introdução da arquitetura *Transformer*, que foi inicialmente proposta por Vaswani (2017), permitiu que a área de desenvolvimento dessas inteligências artificiais saltasse enormemente em capacidade de inovação, permitindo que grandes quantidades de dados textuais fossem processadas de uma maneira mais eficiente, utilizando diversos mecanismos para identificar padrões e relações textuais complexas, sejam em frases, palavras ou subpalavras.

O processo de aprendizado de máquina das LLMs é complexo e subdividido em múltiplas partes, que basicamente podem ser resumidos em dois momentos: o ensino sem supervisão e o treinamento com base no que foi aprendido. No caso do ensino sem supervisão, múltiplos tokens são inseridos no processo de aprendizado de máquina, com poucos filtros ou limitações. Após isso, o modelo é ensinado o que fazer com esse conhecimento, de forma a não ferir alguns princípios básicos de uma boa funcionalidade, como ética ou direitos humanos.

Entretanto, essa tecnologia vem levantando questões éticas e críticas importantes para a ciência da computação. A autoria e originalidade vêm sendo os novos desafios na adaptação mercadológica e, com a ascensão de extremismos na política, desinformação e manipulação de conteúdo é osmoticamente uma capacidade desses modelos, produzindo um extenso e contínuo debate sobre os limites da inteligência artificial e como devemos conduzir tais implicações.

Este artigo tem como objetivo analisar de uma forma generalista o funcionamento dessa tecnologia, destacando sua arquitetura e a capacidade de diferentes modelos, bem como o processamento de linguagem. Além disso, os impactos dessa tecnologia também serão novamente revisitados e conduzidos para dar continuidade a esse importante debate sobre as

---

## FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

implicações éticas e os desafios que surgem com a utilização indevida dos modelos em larga escala.

### 2. REVISÃO DA LITERATURA

A literatura recente vem destacando que os avanços no desenvolvimento das *Large Language Models* vêm sendo impulsionados por uma combinação de diversos fatores, como o aumento da capacidade computacional, aumento na quantidade de dados para o treinamento disponíveis e inovações técnicas nas arquiteturas utilizadas no desenvolvimento, como o *Transformer* (Vaswani et al., 2017). Alguns modelos, como o GPT-4 e o Gemini, continuam a inovar o processamento de linguagem natural (PLN) (Brown et al., 2020).

Tais avanços têm sido acompanhados de novos desafios técnicos, principalmente os relacionados à filtragem e duplicação de dados, bem como a proteção de dados em respeito a privacidade deles. A deduplicação de dados, em particular, vem sendo crítica para garantir que a demanda computacional não seja ainda maior e evita que os modelos não reproduzam conteúdo redundante ou informações desatualizadas. A redução da privacidade de certos dados utilizados no treinamento das LLMs também gera extensas discussões éticas, principalmente quando consideramos que, a depender da aplicação, informações sensíveis e privadas podem ser expostas.

O treinamento das LLMs é geralmente realizado de forma auto supervisionada em muitos dados textuais. O processo é subdividido em duas etapas principais: o pré-treinamento e ajuste fino. O pré-treinamento faz com que o modelo aprenda a prever a sequência de tokens subsequentes tendo como base um determinado texto. O ajuste fino personaliza o modelo para determinadas tarefas, de forma a ter uma melhor utilização do processamento quando comparamos com modelos generalistas. Modelos como o Gemini tem utilizado ambas as abordagens para melhor desempenho.

Recentemente, discussões a respeito da ética e impacto dessa tecnologia vêm sendo criadas, pois possui um enorme potencial de alterar a forma como os seres humanos produzem e consomem conteúdo e informação.

### 3. METODOLOGIA

Neste capítulo, serão descritos os métodos que foram utilizados para analisar o funcionamento da produção textual das *Large Language Models*. A metodologia adotada avalia

---

## FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

tanto a arquitetura interna desses modelos quanto a qualidade dos textos gerados, utilizando experimentos práticos que explorando diferentes contextos e tarefas.

### 3.1 Abordagem do estudo

O estudo adota uma abordagem qualitativa, com foco na análise das respostas textuais geradas pelas LLMs. A análise é orientada por três principais critérios: fluência, coerência e originalidade dos textos. A escolha da abordagem qualitativa foi feita devido à natureza subjetiva das variáveis analisadas, como a naturalidade e a clareza das respostas, que são mais bem compreendidas por meio de observação interpretativa e análise descritiva.

Os modelos selecionados para esta análise foram escolhidos com base em sua popularidade e acessibilidade ao público em geral, assegurando que os resultados obtidos pudessem ser comparados e aplicáveis a contextos amplos. Os modelos analisados foram:

- GPT-4: desenvolvido pela OpenAI, esse modelo foi acessado por meio do ChatGPT;
- Gemini: desenvolvido pela Google, esse modelo foi acessado por meio do aplicativo Gemini.

Ambos os modelos possuem versões pagas e gratuitas e pagas. No entanto, durante o período de avaliação optou-se por utilizar as versões pagas, visando explorar todo o potencial de geração textual oferecido pelas LLMs, o que garantiu uma maior profundidade nos resultados.

### 3.2 Avaliação

A avaliação dos textos gerados foi baseada em três critérios fundamentais:

- Coerência: capacidade do modelo de manter uma estrutura lógica em suas respostas, com início, meio e fim claros. Um texto coerente apresenta uma sequência de ideias bem organizadas, sem contradições ou rupturas na narrativa;
- Fluência: avaliação da naturalidade da resposta, observando se o texto gerado é similar ao de uma construção linguística humana. Um texto fluente flui de maneira suave, sem quebras gramaticais ou semânticas que possam prejudicar sua compreensão;
- Originalidade: verifica a presença de textos criativos e inovadores, evitando repetições ou padrões previsíveis. A originalidade é fundamental para identificar

---

## FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

se o modelo consegue produzir respostas criativas em relação a diferentes *prompts* e contextos, sem depender de padrões preexistentes.

### 3.3 Experimentos

Foram conduzidos dois experimentos práticos para avaliar o desempenho dos modelos em diferentes contextos:

Experimento 1: consistiu em fornecer aos modelos um conjunto de *prompts* genéricos, solicitando a geração de textos curtos. O objetivo deste experimento foi observar como os modelos lidam com perguntas simples e contextos cotidianos.

Foram utilizados dos *prompts*, um para descrições e outro para diálogos:

- Descrição: “Descreva um dia ensolarado em duas linhas.”;
- Diálogo simples: “Crie um diálogo entre dois colegas de trabalho que acabaram de finalizar um projeto importante.”.

Foi exigido de cada modelo a geração de três respostas distintas por *prompt*, totalizando seis respostas analisadas.

Experimento 2: neste experimento, foram apresentadas aos modelos *prompts* de áreas técnicas específicas, como matemática, inteligência artificial e programação. O objetivo foi avaliar a capacidade dos modelos de gerar textos com precisão técnica, detalhando conceitos complexos de maneira clara e acessível.

Assim como no experimento anterior, dois *prompts* foram utilizados, sendo eles:

- Teste de abstração: “Descreva como funcionaria uma máquina do tempo, com todos os detalhes possíveis.”;
- Teste técnico: “Explique o conceito de reatores nucleares de fissão e como eles são utilizados para gerar eletricidade.”.

Cada modelo também gerou três respostas distintas para cada *prompts*, o que resultou em seis respostas para análise. Os critérios de avaliação foram aplicados uniformemente em ambos os experimentos.

### 3.4 Seleção dos *prompts*

Os *prompts* foram escolhidos com intuito de abordar diferentes níveis de complexidade, cobrindo questões de cotidiano, assim como temas técnicos e complexos. Essa variação resultou

---

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

em uma análise abrangente das capacidades de ambos os modelos. Os *prompts* genéricos foram formulados para permitir a exploração da capacidade dos modelos na produção textual com naturalidade e fluência, baseando-os em questões do cotidiano. Enquanto *prompts* técnicos tiveram como foco medir a validação científica e teórica (abstrata) dos modelos, bem como o rigor técnico utilizado.

## 4. RESULTADOS

Neste capítulo, serão apresentados os resultados obtidos nos experimentos realizados com as *Large Language Models* citadas. Foram analisados a qualidade dos textos gerados, compreensão textual, coerência e originalidade, com base nas métricas já estabelecidas no capítulo anterior.

### 4.1 Geração de texto genérico

Para avaliar a capacidade dos modelos, foram realizados experimentos com *prompts* genéricos e técnicos. A seguir será apresentado os resultados obtidos comparando as respostas geradas pelos dois modelos citados na metodologia.

#### 4.1.1 Descrições

Quando solicitados a gerar textos genéricos, os modelos variaram em profundidade contextual e abordagem na resposta. Em textos gerados para descrever situações ou sensações, não houve diferenças substanciais entre os dois modelos, no entanto o Gemini caracterizou-se como mais profundo na parte de descrição das sensações humanas.

Quadro 1: Análise das descrições

Métrica Avaliada	GPT-4	Gemini
Coerência	Alta coerência em descrições diretas.	Alta coerência em descrições de sensações humanas.
Fluência	Fluência alta, mas focada em descrições mais sucintas.	Fluência alta com descrições mais detalhadas.
Originalidade	Originalidade moderada em descrições genéricas.	Alta originalidade com detalhes imersivos.
Prós	- Coerência técnica e clara. - Texto objetivo.	- Descrições ricas e imersivas. - Detalhamento de sensações.
Contras	- Menos profundidade nas descrições sensoriais.	- Pode gerar respostas muito detalhadas para descrições simples.

Fonte: Elaborado pelo Autor

## FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

### 4.1.2 Descrições

A partir de um determinado *prompt*, foi exigido aos dois modelos a criação um curto diálogo entre dois colegas de trabalho que acabaram de finalizar um projeto importante. Ao comparar os dois modelos, o Gemini se mostrou superior no quesito de contar histórias adicionando até mesmo descrição das ações dos personagens.

Além disso, o Gemini foi mais bem-sucedido em definir o contexto no qual os personagens se encontram.

Quadro 2: Análise dos diálogos

Métrica Avaliada	GPT-4	Gemini
Coerência	Alta coerência em diálogos simples.	Alta coerência em diálogos criativos e envolventes.
Fluência	Moderada, com diálogos simples e diretos.	Alta fluência, com diálogos detalhados e contexto imersivo.
Originalidade	Moderada, com diálogos convencionais.	Alta originalidade, com descrições de ações e cenário.
Prós	- Coerência em diálogos curtos.	- Diálogos criativos e detalhados.
	- Respostas objetivas.	- Descrições de ações e cenários.
Contras	- Falta de imersão nas descrições do ambiente.	- Respostas mais longas e detalhadas, podendo se estender além do necessário.

Fonte: Elaborado pelo Autor

### 4.1.3 Conceitos abstratos

Descrições detalhadas de objetos e conceitos abstratos foram feitas para testar a capacidade dos modelos de interpretar e gerar conteúdo detalhado e descritivo. O modelo GPT-4 foi superior neste quesito ao adicionar mais detalhes a respeito de termos adjacentes ao *prompt* imposto, mostrando mais capacidade deste modelo em servir como método inicial de busca ou aprendizado sobre determinado tema, mesmo sendo abstrato.

#### 4.1.3.1 Teste de abstração

O teste foi feito nos dois modelos, consistia em um *prompt* que exigia a descrição do

## FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE

funcionamento de uma máquina do tempo, com todos os detalhes possíveis. No Gemini, embora uma resposta sucinta foi descrita, logo o modelo divergiu do tema e não explicitou como seria a aparência de como tal máquina seria.

Diferentemente, o GPT-4 logo no início apontou como o modelo ‘imagina’ a aparência de tal máquina caso existisse, apontando a estrutura e quais conceitos científicos deveriam ser utilizados para o seu funcionamento. Ao separar as respostas em tópicos, o GPT-4 foi capaz de discorrer pela aparência da máquina, fonte de energia, teoria do funcionamento, mecanismo de navegação temporal e entre outros, adicionando profundidade a uma tecnologia que é meramente hipotética.

Quadro 3: Análise utilizando conceitos abstratos

Métrica Avaliada	GPT-4	Gemini
Coerência	Alta coerência em descrições abstratas e técnicas.	Coerência moderada em descrições abstratas.
Fluência	Alta, com textos organizados e claros.	Fluência moderada, com algumas divagações.
Originalidade	Moderada, com foco em precisão técnica.	Alta originalidade, mas menos foco em detalhes técnicos.
Prós	- Explicações técnicas e organizadas.	- Alta criatividade.
	- Boa estrutura em tópicos.	- Descrições inovadoras.
Contras	- Menos criativo em descrições imaginativas.	- Menos preciso em explicações técnicas abstratas.

Fonte: Elaborado pelo Autor

### 4.2 Geração de textos técnicos

Ao solicitar a geração de descrições técnicas de conceitos diversos, houve uma variação entre os modelos em questões de precisão e profundidade contextual. Os textos produzidos pelo Gemini foram mais detalhadas e tecnicamente corretos, bem como possuíam uma maior quantidade de caracteres. A resposta em tópicos produzida pelo Gemini, facilitou uma pesquisa posterior para checar a veracidade dos fatos expostos pelo modelo.

Ao analisar as respostas de ambos os modelos, o Gemini se mostrou mais educativo, ensinando um exemplo de como um tema complexo funciona de uma forma simples, enquanto a resposta produzida pelo GPT-4 assemelha-se àquelas que extraímos de textos técnicos, demonstrando falta de originalidade em explicar conceitos técnicos.

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

Quadro 3: Análise utilizando conceitos abstratos

Métrica Avaliada	GPT-4	Gemini
Coerência	Alta coerência em explicações técnicas.	Alta coerência em descrições detalhadas.
Fluência	Alta fluência com foco em precisão técnica.	Alta fluência com foco em detalhes extensos.
Originalidade	Moderada, com enfoque em precisão.	Alta, com explicações mais elaboradas.
Prós	- Respostas técnicas e precisas.	- Explicações técnicas detalhadas.
	- Explicações diretas.	- Organização por tópicos.
Contras	- Menos originalidade em explicações criativas.	- Respostas mais longas e menos objetivas em temas técnicos.

Fonte: Elaborado pelo Autor

### 4.3 Comparando os modelos

Com o intuito de facilitar a demonstração da diferença entre os modelos analisados, seus prós e contras, foi possível montar o seguinte quadro:

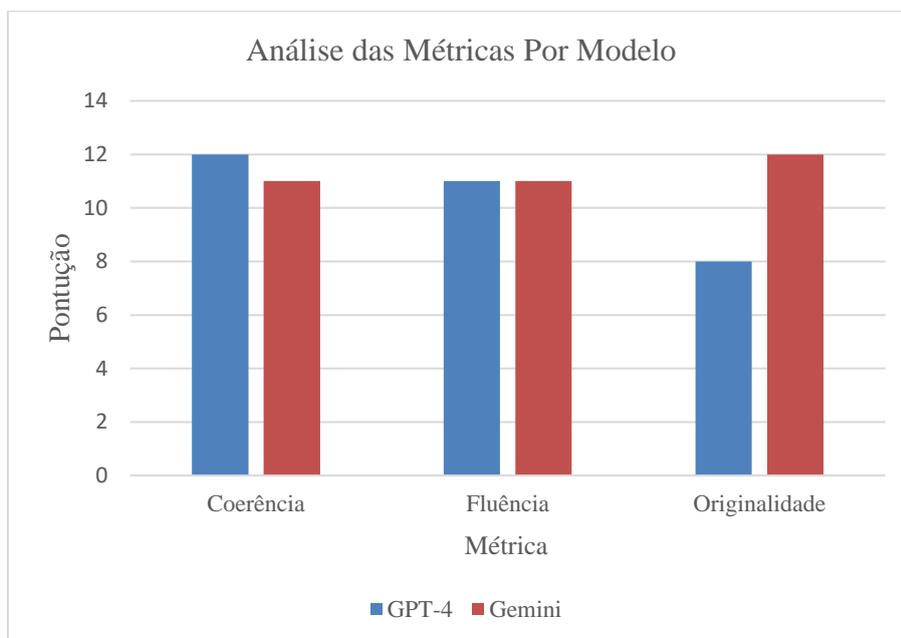
Quadro 4: Análise geral

Métrica	GPT-4	Gemini
Coerência	Alta coerência em textos técnicos e conceituais.	Coerência alta em tarefas criativas e descritivas.
Fluência	Fluência alta, porém, com foco em textos diretos e técnicos.	Fluência alta em textos criativos e descritivos, mais imersivo.
Originalidade	Originalidade moderada, principalmente em textos técnicos.	Alta originalidade.
Prós	- Alta precisão técnica.	- Descrições criativas e detalhadas.
	- Capacidade de separar explicações em tópicos.	- Diálogos mais imersivos.
	- Ideal para tarefas abstratas e conceituais.	- Melhor contextualização

Fonte: Elaborado pelo Autor

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

Diante da análise feita durante a pesquisa, as métricas utilizadas foram coerência, fluência e originalidade. Para cada modelo e métrica, uma nota entre alta, moderada e baixa foi dada. Ao assumir para cada nota as pontuações 3, 2 e 1 e somar os resultados de cada tabela, foi possível construir o seguinte gráfico, que aponta as potencialidades de cada modelo:



## 5. CONCLUSÃO

Este estudo apresentou como funcionam as *Large Language Models* e sua capacidade de produção textual. Através de análises da qualidade de resposta, foram examinados dois modelos de linguagem diferentes, sendo o GPT-4 da OpenAI e o Gemini da Google e os seus desempenhos na produção textual em contextos diferentes.

### 5.1 Resumo dos resultados

Os resultados demonstraram que as *Large Language Models* são capazes de produzir textos coerentes, fluentes e originais em diversos contextos. O Gemini destacou-se como um modelo mais educativo em suas respostas, enquanto o GPT-4 possuiu uma melhor abstração ao lidar com assuntos complexos e científicos.

Os experimentos práticos revelaram que, embora os modelos sejam capazes de gerar textos, limitações entre um modelo e outro existem e desvios de foco nas respostas podem ser encontradas, o que aponta que tais modelos necessitam de uma abordagem mais generalista e robusta para garantir uma utilização satisfatória nos mais variados temas.

---

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

## **5.2 Contribuições**

Este artigo teve como objetivo contribuir para o entendimento de como as *Large Language Models* funcionam, seus potenciais usos em áreas como desenvolvimento de *software*, produções culturais e automação de tarefas de escrita. Ao demonstrar prós e contras entre os dos modelos diferentes, o trabalho também fornece e aponta aos desenvolvedores e pesquisadores evidências de possíveis aprimoramentos dessa tecnologia.

## **5.2 Pesquisas futuras**

Este estudo forneceu uma visão geral sobre a capacidade das LLMs e uma breve análise qualitativa entre os modelos, no entanto vários aspectos ainda precisam ser estudados. Uma linha de pesquisa analisando a capacidade quantitativa entre os modelos, isto é, tempo de resposta, banda consumida e processamento interno exigido, pode ajudar a melhorar ainda mais essas tecnologias, garantindo uma maior precisão e velocidade nas respostas geradas.

---

**FACULDADE DE TECNOLOGIA DE PRESIDENTE PRUDENTE**

**REFERÊNCIAS**

BROWN, Tom; MANN, Benjamin; RYDER, Nick; et al. **Language Models are Few-Shot Learners**. In: **Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)**. Vancouver, 2020.

DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. In: **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2019)**. Minneapolis, 2019.

KUDO, Taku. **Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates**. In: GUREVYCH, Iryna; MIYAO, Yusuke (Ed.). **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Melbourne, Australia: Association for Computational Linguistics, 2018. p. 66-75. Disponível em: <https://aclanthology.org/P18-1007>. DOI: 10.18653/v1/P18-1007. Acesso em: 25 ago. 2024.

RAFFEL, Colin; SHAZEER, Noam; ROBERTS, Adam; LEE, Katherine; NARANG, Sharan; MATENA, Michael; ZHOU, Yanqi; LI, Wei; LIU, Peter J. **Exploring the limits of transfer learning with a unified text-to-text transformer**. **Journal of Machine Learning Research**, v. 21, n. 140, p. 1-67, 2020. Acesso em: 14 set. 2024.

SENNICH, Rico; HADDOW, Barry; BIRCH, Alexandra. **Neural Machine Translation of Rare Words with Subword Units**. In: ERK, Katrin; SMITH, Noah A. (Ed.). **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Berlin, Germany, ago. 2016. p. 1715-1725. Association for Computational Linguistics. Disponível em: <https://aclanthology.org/P16-1162>. DOI: 10.18653/v1/P16-1162.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; et al. **Attention Is All You Need**. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)**. Long Beach, 2017.

ZIEGLER, Daniel M.; et al. **Fine-Tuning Language Models from Human Preferences**. In: **Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)**. 2019. Disponível em: <https://arxiv.org/abs/1909.08593>. Acesso em: 19 set. 2024.

WANG, Thomas; ROBERTS, Adam; HESSLOW, Daniel; LE SAO, Teven; CHUNG, Hyung Won; BELTAGY, Iz; LAUNAY, Julien; RAFFEL, Colin. **What language model architecture and pretraining objective works best for zero-shot generalization?**. In: CHAUDHURI, Kamalika; JEGELKA, Stefanie; SONG, Le; SZEPEVARI, Csaba; NIU, Gang; SABATO, Sivan (Ed.). **Proceedings of the 39th International Conference on Machine Learning**. PMLR, 2022. p. 22964-22984. Disponível em: <https://proceedings.mlr.press/v162/wang22u.html>. Acesso em: 14 set. 2024.