

Desafios e oportunidades
para as Línguas Moçambicanas
na Era dos ***Large Language Models***

Felermino Ali

Universidade Lúrio - Faculdade de Engenharia
Universidade do Porto - Faculdade de Engenharia

Agenda

- + Nesta apresentação
- + nosso objetivo é primeiro **desmistificar o conceito de Modelos de Linguagem de Grande Escala**,
- + explicando **como eles funcionam** e quais são as **oportunidades e desafios associados** ao seu **uso no contexto das línguas moçambicanas**.
- + Abordaremos também os **riscos potenciais** do uso desses modelos.
- + Além disso, **apresentaremos uma série de recomendações** direcionadas aos diversos stakeholders envolvidos com as línguas moçambicanas,
- + com o intuito de impulsionar o avanço do estado atual de nossas línguas.
- + Isso incluirá **sugestões** para pesquisadores, desenvolvedores de tecnologia, educadores e formuladores de políticas,
- + **de forma a promover o desenvolvimento e a preservação das línguas moçambicanas**
- + nesta era de rápidas transformações tecnológicas

Agenda – Emakhuwa (Tradução)

- + Mooniheryo wa olelo!
- + Nninlelahu ti ela, yopacerya otaphulela, miruku sa malavulelo a Mananna maalupale.
- + Anooniherya moota xeeni anivarya muteko, ni okathi xeeni ni manxankiho xeeni amphwanyaneyaya wi evareliwe muteko. Ovarela waya muteko eriya ya mattava a mMosampiikhi.
- + Ninrowa woniherya moota nikhanliyahu owiikasoopa nivarelaka muteko malavulelo ala.
- + Wowii nincerere makhalelo a mattava ahu.
- + Eyo enrowa wakhaliha vamosa ahooleli, anamasomi ni ale aninnuwiha masuwelo a ekomputatoore, anamasomihe, ni ale anuupuwela.
- + Miruku sa ovuwiha ni wunnuwiha ni makhlelo a mattava a elapo ahu.
- + Mahiku ala oolelo owaakuveya orukunuwa wa miirelo.

Agenda – Back-Translation, Emkhuwa-Português

+ Mooniheryo wa olelo!



+ **Tradutor:** A lição de hoje!

+ **Tradução Automática:** O testemunho de hoje!

+ **Referência:** Nesta apresentação

+ Nninlelahu ti ela, yopacerya otaphulela, miruku sa malavulelo a Mananna maalupale.

+ **Tradutor:** O que pretendemos é o seguinte, primeiro a interpretação das ideias das expressões ou falas importantes.

+ **Tradução Automática:** Esta é a primeira declaração inspirada dos Dez Mandamentos.

+ **Referência:** nosso objetivo é primeiro desmistificar o conceito de Modelos de Linguagem de Grande Escala,

+ Anooniherya moota xeeni anivarya muteko, ni okathi xeeni ni manxankiho xeeni amphwanyaneyaya wi evareliwe muteko. Ovarela waya muteko eriya ya mattava a mMosampiikhi.

+ **Tradutor:** Demonstram que formas trabalham, que períodos e quais dificuldades enfrentam no seu trabalho. Para além disso, vamos mostrar as regras que devem ser seguidas por pessoas que trabalham com as línguas moçambicanas.

+ **Tradução Automática:** Como eles trabalham, quando e com que desafios eles enfrentam? Eles trabalham em vários idiomas moçambicanos.

+ **Referência:** explicando como eles funcionam e quais são as oportunidades e desafios associados ao seu uso no contexto das línguas moçambicanas.

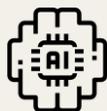
Agenda – Back-Translation, Emkhuwa-Português

+ Mooniheryo wa olelo!

- + **Tradutor:** A lição de hoje!
- + **Tradução Automática:** O testemunho de hoje!
- + **Referência:** Nesta apresentação

+ **Nninlelahu ti ela, yopacerya otaphulela, miruku sa malavulelo a Mananna maalupale.**

- + **Tradutor:** ~~O que pretendemos é o seguinte, primeiro a interpretação das ideias das~~ **expressões ou falas importantes.**
- + **Tradução Automática:** ~~Esta é a primeira declaração inspirada~~ **dos Dez Mandamentos.**
- + **Referência:** nosso objetivo é primeiro desmistificar o conceito de **Modelos de Linguagem de Grande Escala,**



+ **Anooniherya moota xeeni anivarya muteko, ni okathi xeeni ni manxankiho xeeni amphwanyaneyaya wi evareliwe muteko. Ovarela waya muteko eriya ya mattava a mMosampiikhi.**

- + **Tradutor:** ~~Demonstram que formas trabalham, que períodos e quais dificuldades enfrentam no seu trabalho. Para além disso, vamos mostrar as regras que devem ser seguidas por pessoas que trabalham com as línguas moçambicanas.~~
- + **Tradução Automática:** ~~Como eles trabalham, quando e com que desafios eles enfrentam? Eles trabalham em vários idiomas moçambicanos.~~
- + **Referência:** ~~explicando como eles funcionam e quais são as oportunidades e desafios associados ao seu uso no contexto das línguas moçambicanas.~~

Agenda – Back-Translation, Emkhuwa-Português

+ Mooniheryo wa olelo!

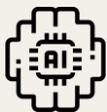
- + **Tradutor:** A lição de hoje!
- + **Tradução Automática:** O testemunho de hoje!
- + **Referência:** Nesta apresentação

+ Nninlelahu ti ela, yopacerya otaphulela, miruku sa malavulelo a Mananna maalupale.

- + **Tradutor:** O que pretendemos é o seguinte, primeiro a interpretação das ideias das expressões ou falas importantes.
- + **Tradução Automática:** Esta é a primeira declaração inspirada dos Dez Mandamentos.
- + **Referência:** nosso objetivo é primeiro desmistificar o conceito de Modelos de Linguagem de Grande Escala,

+ Anooniherya moota xeeni anivarya muteko, ni okathi xeeni ni manxankiho xeeni amphwanyaneyaya wi evareliwe muteko. Ovarela waya muteko eriya ya mattava a mMosampiikhi.

- + **Tradutor:** Demonstram que formas trabalham, que períodos e quais dificuldades enfrentam no seu trabalho. Para além disso, vamos mostrar as regras que devem ser seguidas por pessoas que trabalham com as línguas moçambicanas.
- + **Tradução Automática:** Como eles trabalham, quando e com que desafios eles enfrentam? Eles trabalham em vários idiomas moçambicanos.
- + **Referência:** explicando como eles funcionam e quais são as oportunidades e desafios associados ao seu uso no contexto das línguas moçambicanas.



O que é Large Language Models?

+ Large Language Models, ou seja, Modelos de Linguagem de Grande Escala

+ Language Models:

+ projetado para entender e gerar conteúdo (texto, áudio, etc.) - Como fosse um papagaio

Modelos Probabilísticos (Calcular a probabilidade de próxima palavra)

O amor é paciente, o amor é ?

bondoso

liberdade

dançar

vender

eterno

O que é Large Language Models?

+ **Language Models:** Modelos Probabilísticos - Qual é a frequências das alternativas no texto existentes.

+ **Probabilidade** (próxima palavra | "o amor é paciente, o amor é")

O amor é paciente, o amor é ?

?

?

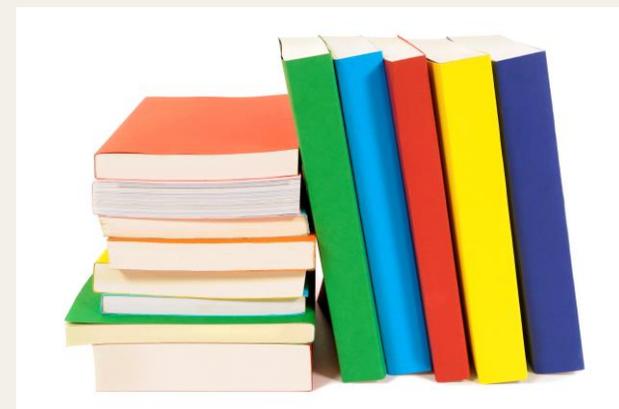
bondoso

liberdade

dançar

vender

eterno



O que é Large Language Models?

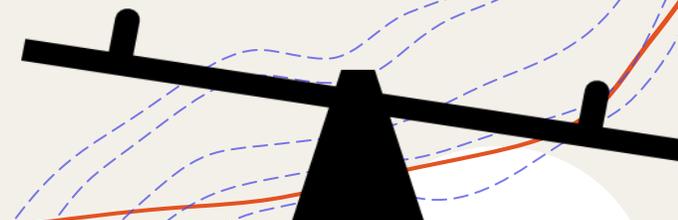
+ **Language Models:** Modelos Probabilísticos - Qual é a frequências das alternativas no texto existentes.

+ **Probabilidade** (próxima palavra | "o amor é paciente, o amor é")



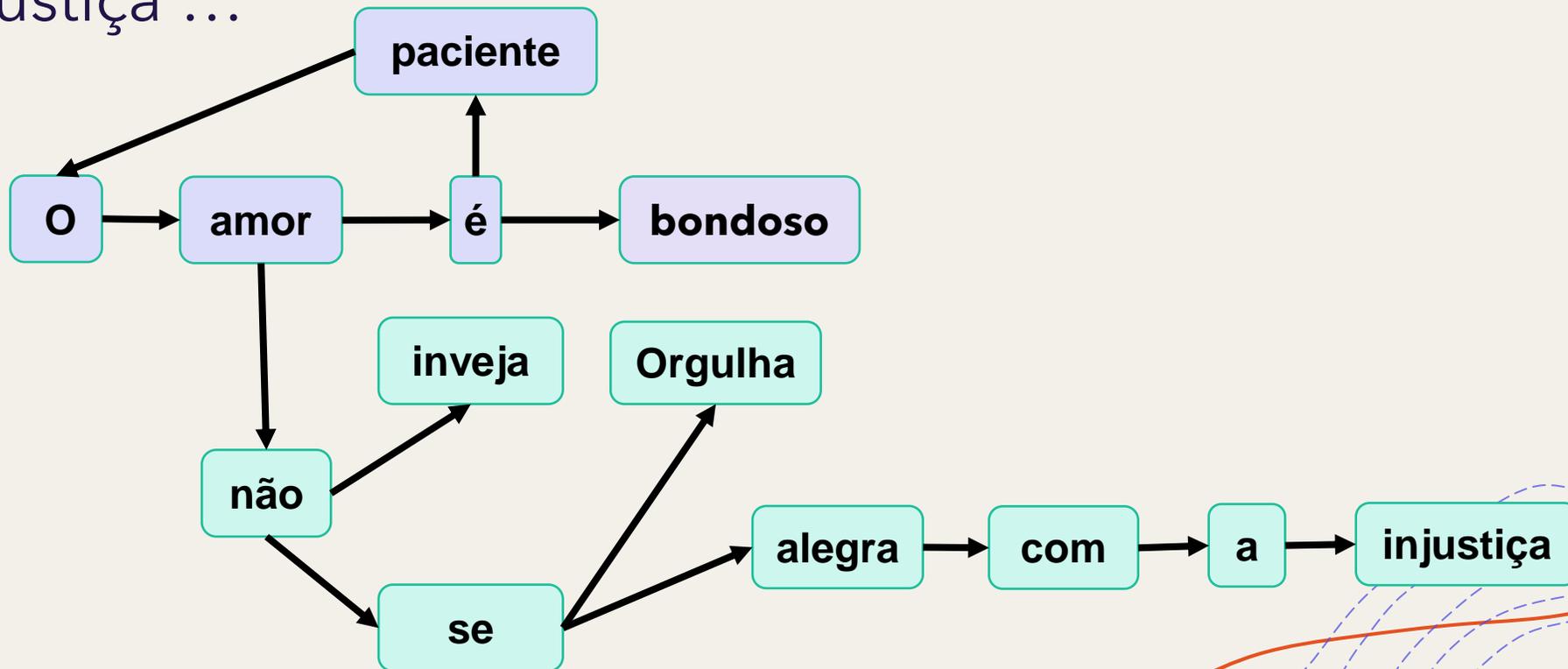
1 Coríntios 13:4

Religião

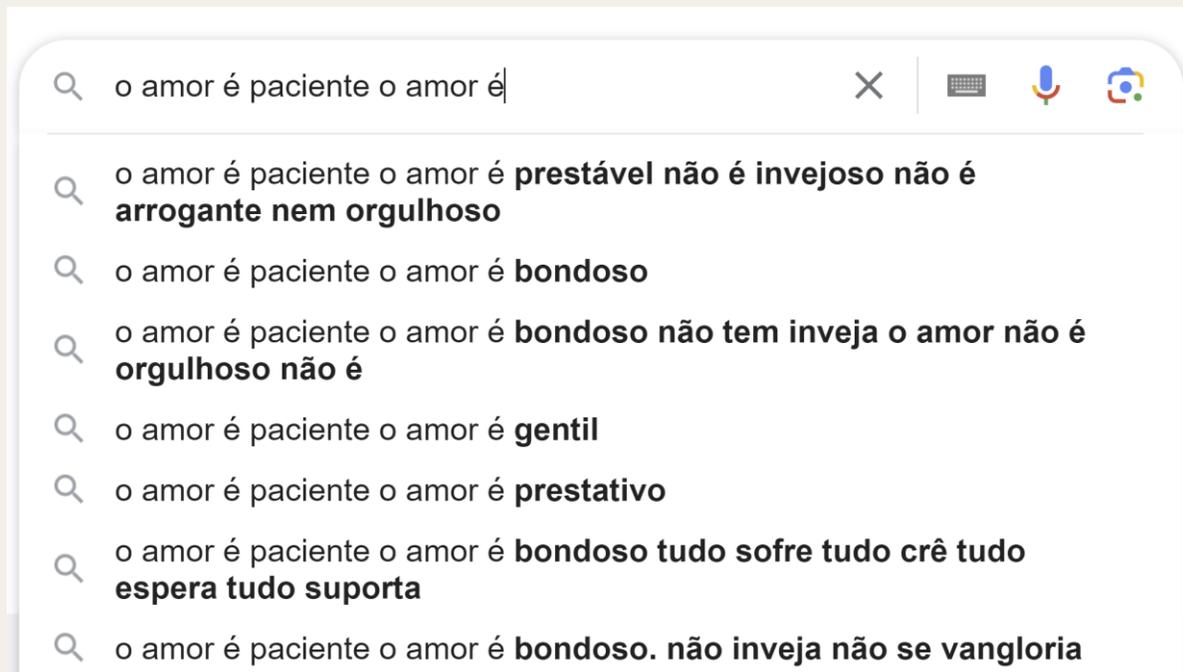
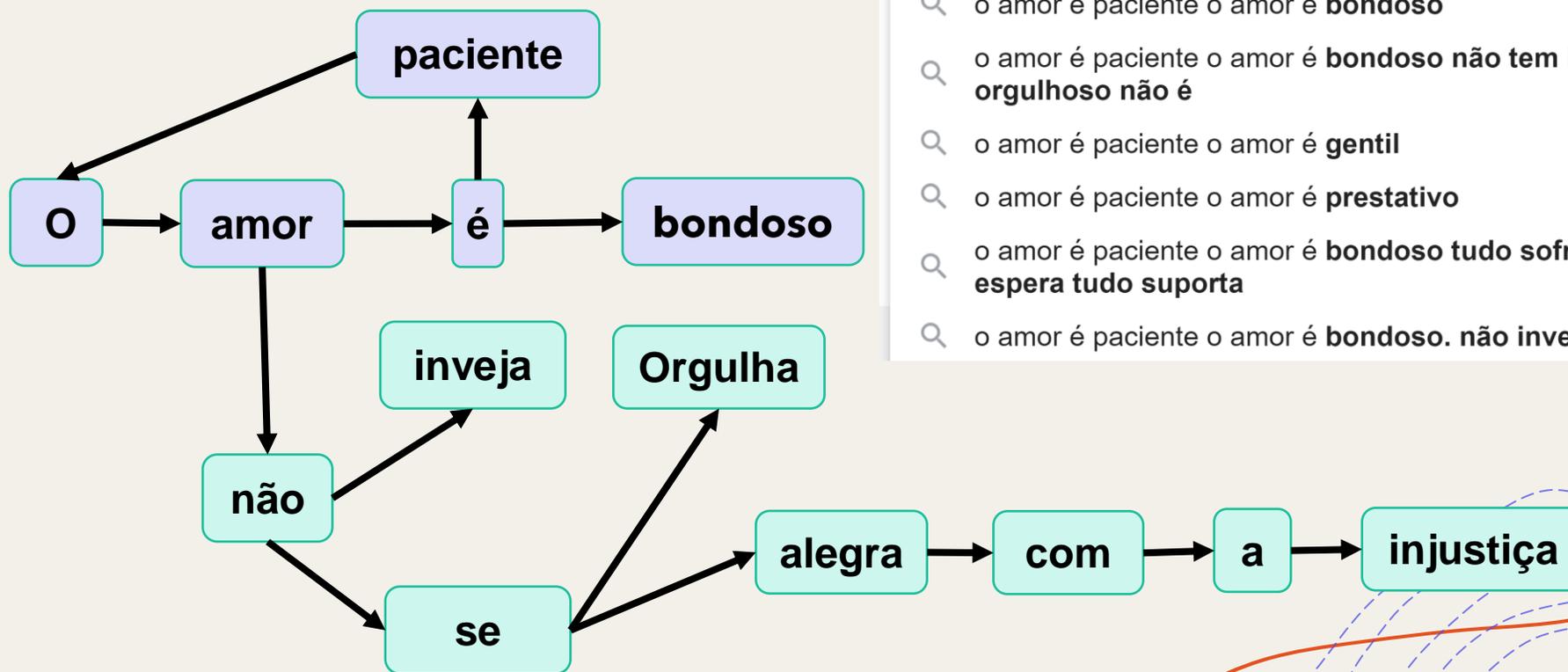


O que é Large Language Models?

+O amor é paciente, o amor é bondoso. Não inveja, não se vangloria, não se orgulha.... O amor não se alegra com a injustiça ...



O que é Large Language Models?

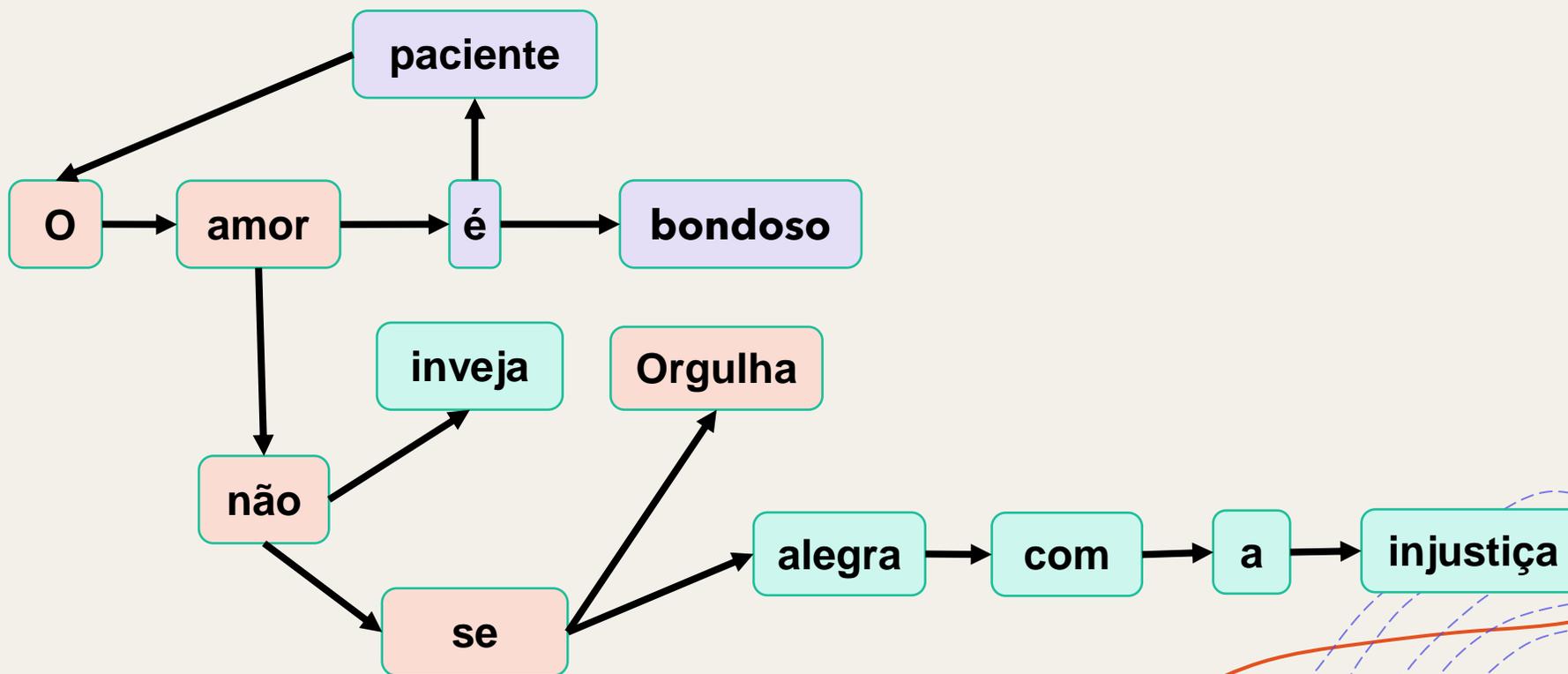


O que é Large Language Models?



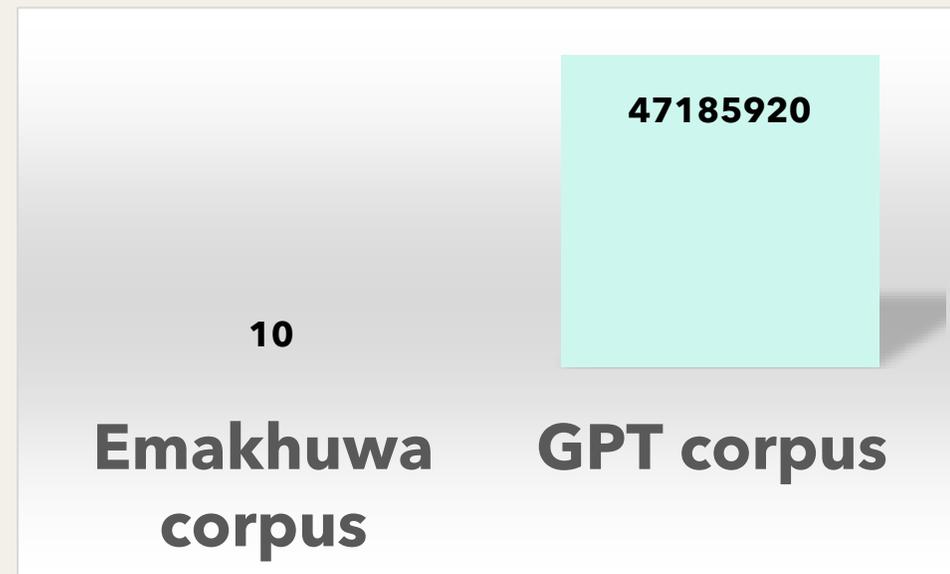
+ Gerar novos conteúdos a partir do modelo:

"O amor não se orgulha"



O que é **Large Language Models**?

- + Processam maior quantidade de dados
- + Treinados usando **conjunto dados (*dataset*) enormes**:
 - + Livros, Wikipedia, artigos, conversas, (internet)
 - + For example: **GPT3, foi treinado com 45 TB** de texto,
>1 terabyte (TB) = 1,048,576 megabytes (MB).



(Em MB)

O que é **Large Language Models**?

+ **Modelo Generativos**

+ **Redes Neurais**: melhor **representação de palavras**, dados, e **arquitecturas robustas, treinamentos**

Entrada

o

amor

é

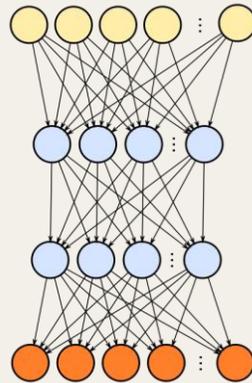
paciente

o

amor

é

?



Saída

bondoso

O que é **Large Language Models**?

+ **Generalizam**

+ Analogias

+ **São Adaptáveis**

+ Transferência de Aprendizagem:
ajustar para dados menores



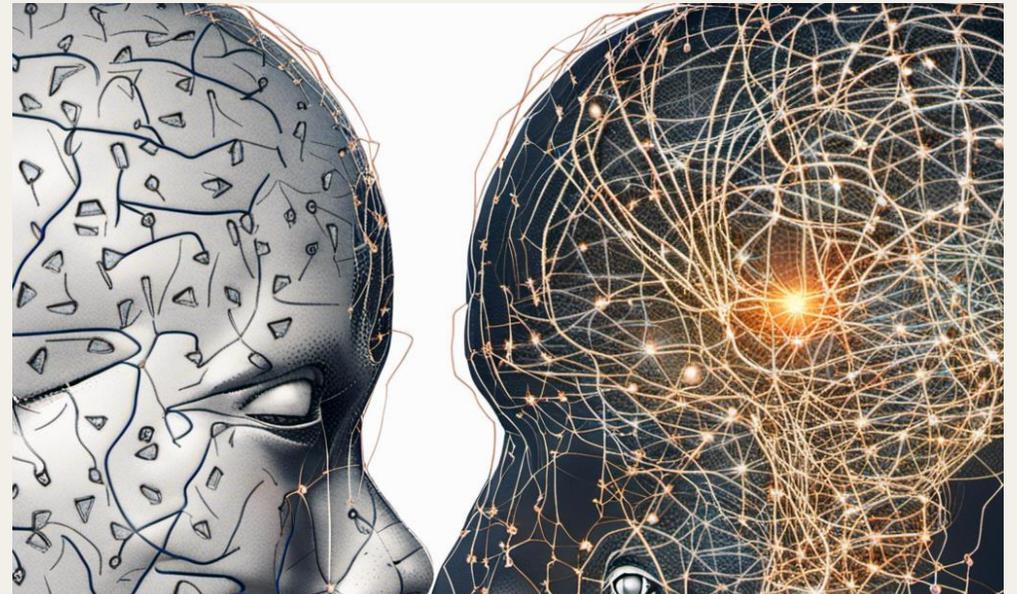
+ **Traduza** "Wiwana oreera" _____

+ **Sumarizar** "Wiwana oreera" _____

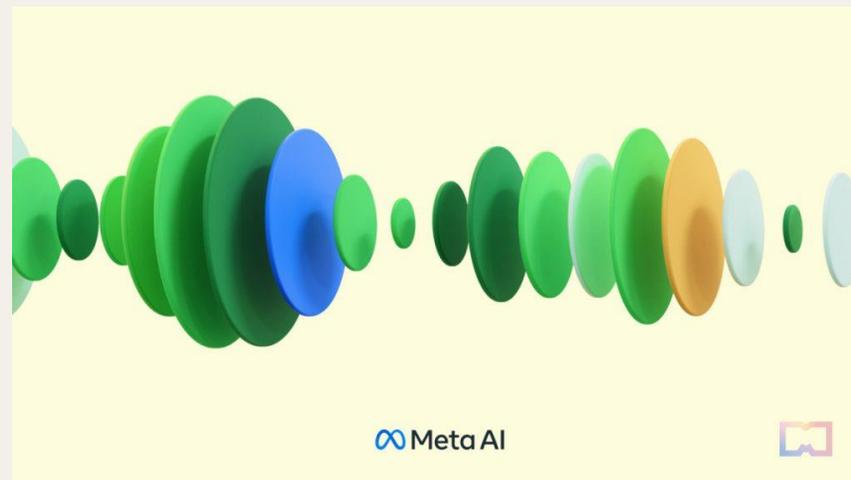
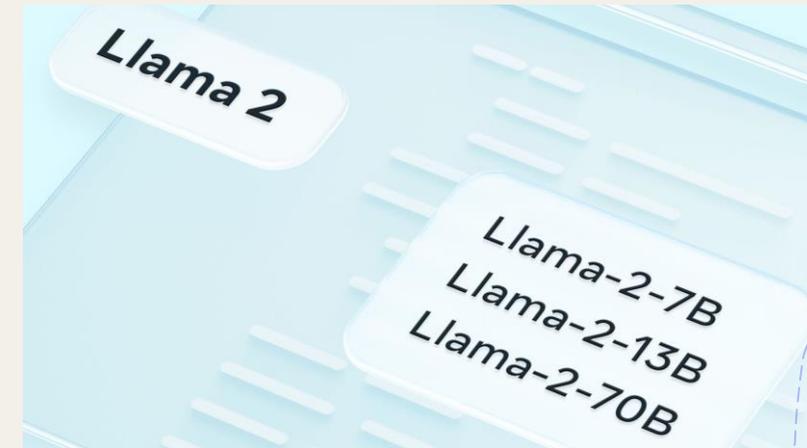
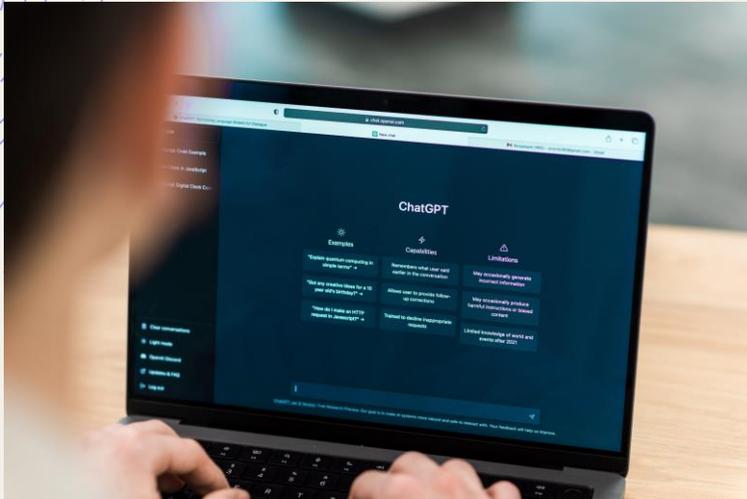
+ **Responder Questões** "Um mais um é igual" _____

+ **Escrever código-fonte** "imprima hello world" _____

Etc.



Embora todos utilizemos, raramente se discute sobre o assunto



Riscos

"No one has a clue how to build a conscious machine, at all."

- Ninguém faz a mínima ideia de como construir uma máquina consciente.

Stuart Russell

Oportunidades

- + Redução do custo computacional**
- + Menor necessidade de dados anotados**
- + Rápida implementação de novas aplicações**
 - + Tradução de materiais didáticos, etc
 - + Sistemas de reconhecimento de áudio
- + Melhoria do desempenho**
- + Preservação e valorização da diversidade linguística e cultural**
- + Inclusão digital**

Desafios

+ Escassez de dados

- + Desequilíbrio de recursos linguísticos
- + Viés linguístico e cultural

+ Déficit de expertise linguística e know-how

+ Custo computacional

+ Falta de interesse pesquisa / comercial

+ Sustentabilidade

Recomendações

Devemos assumir a responsabilidade por nossas ações **ao desenvolver e utilizar ferramentas de IA.**

Recomendações

“Akuboha makungu svitshembisa nkongometo” - **A decisão define o destino** 

+Política. Academia. Sociedade.

Recomendações

“Muttu mekhawe khanteka empa” 

Uma pessoa sozinha não constrói casa

+ Comunidades multidisciplinares é o caminho

Recomendações

“Wiwana oreera” - **A concórdia é boa** 

+É urgente haver referencias de escritas em línguas moçambicanas, digitalizados e de fácil acesso.

Recomendações

“Ovaha waala ” - **Oferecer é semear** 

Em momentos de escassez, todo apoio é extremamente valioso.



Gere áudio aqui

Obrigado

“Wanuna wa ku khala a ntubuluko a hi wanuna”

Homem sem cultura não é homem

“Tirha, Vusiku dza ta”

Trabalhe, a noite está vindo



Desafios e oportunidades para as Línguas Moçambicanas na Era dos ***Large Language Models***

Felermino Ali

felerminoali@gmail.com

Universidade Lúrio - Faculdade de Engenharia

Universidade do Porto - Faculdade de Engenharia

LIACC, CLUP