



Grandes modelos de l ngua: problemas e desaf os

Paulo Gamallo (CiTIUS – USC)

Curso de ver n **Humanidades Dixitais:**
recursos, ferramentas e servizos, 2024
ILG - Instituto da L ngua Galega
Faculdade de Filologia, USC.





PLANO

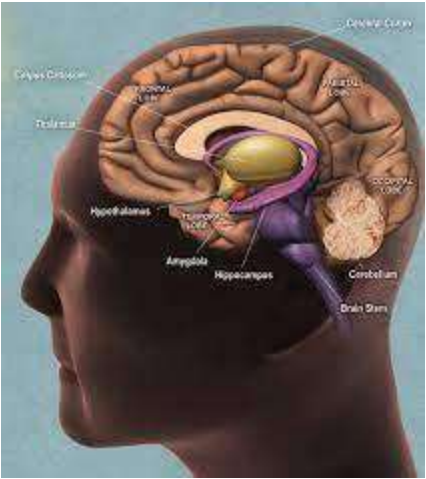
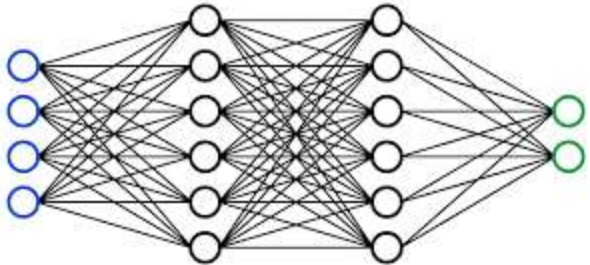
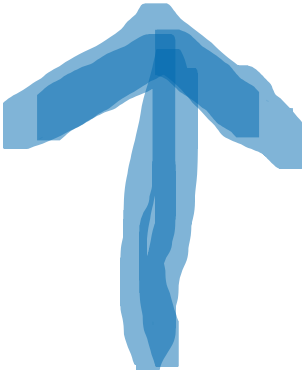
1. Que são os (grandes) modelos de língua?
2. Que problemas acarretam?
3. Que desafíos temos por diante para o galego?
 - Proyecto Nós: modelos generativos e de tradución





Que é um modelo de língua?

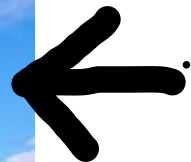
Línguas



Redes neuronais artificiais

Sistema cognitivo e neurológico humano

Línguas

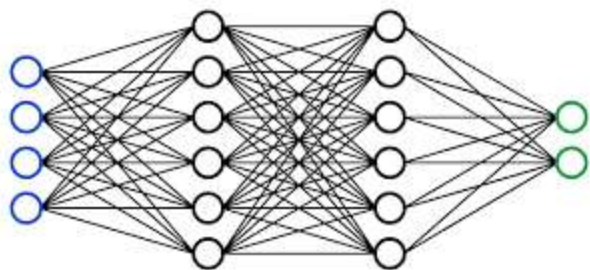


Linguagem

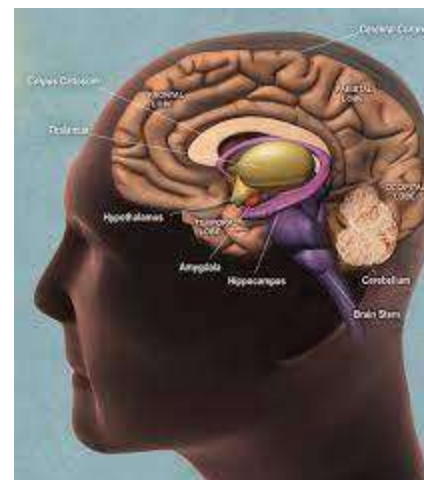


Faculdade da
linguagem

Modelo de
língua



Redes neuronais artificiais



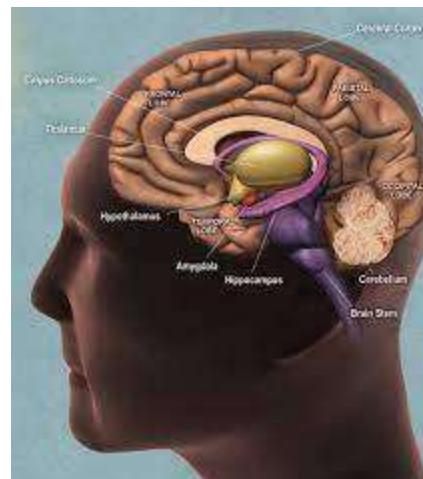
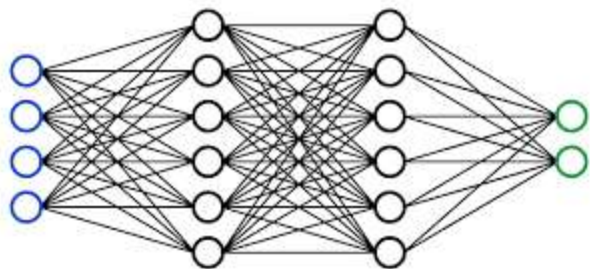
Sistema cognitivo e
neurológico humano

Línguas



AGI?

Modelo da linguagem



Sistema cognitivo e neurológico humano

Modelos de linguagem
Redes neuronais artificiais

Modelo de língua

o míssil israelita assassinou crianças palestinas em ...



Gaza



Modelo de língua



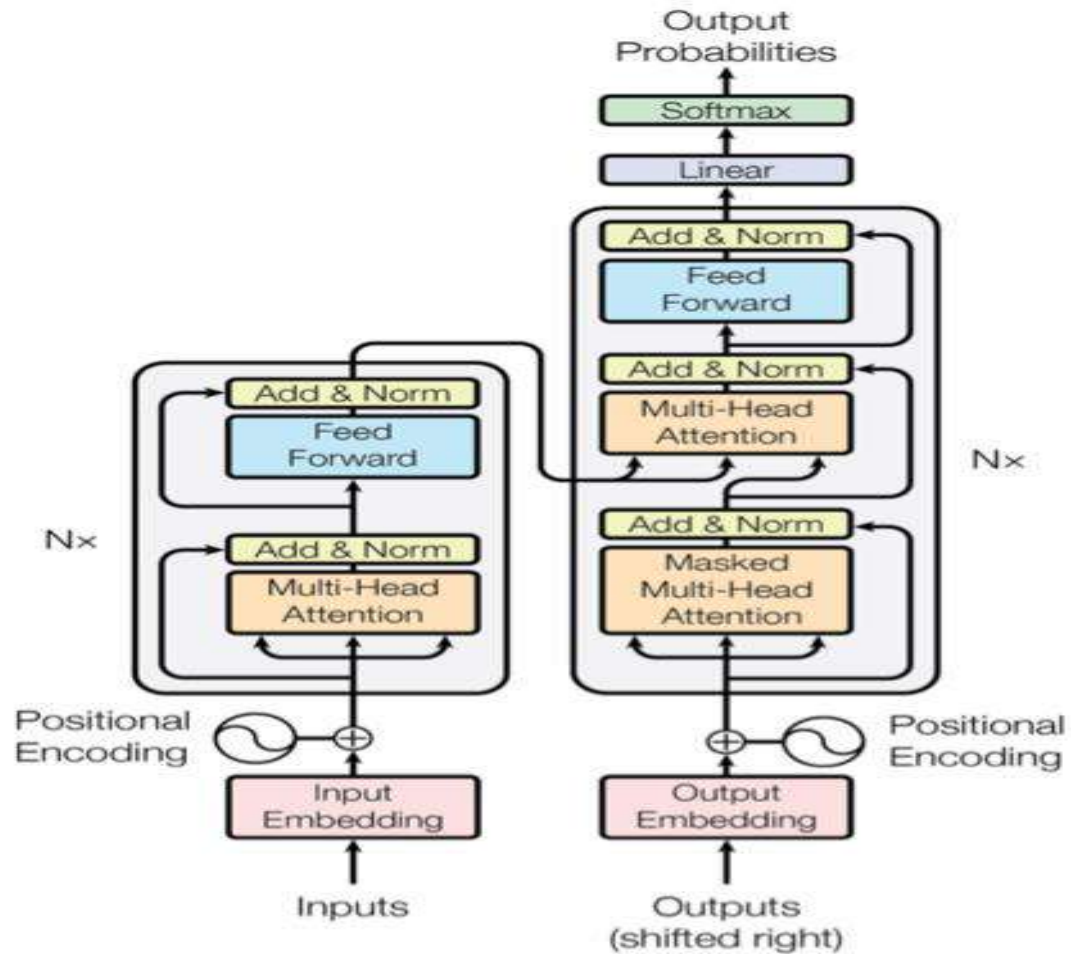
o míssil israelita assassinou crianças palestinas em



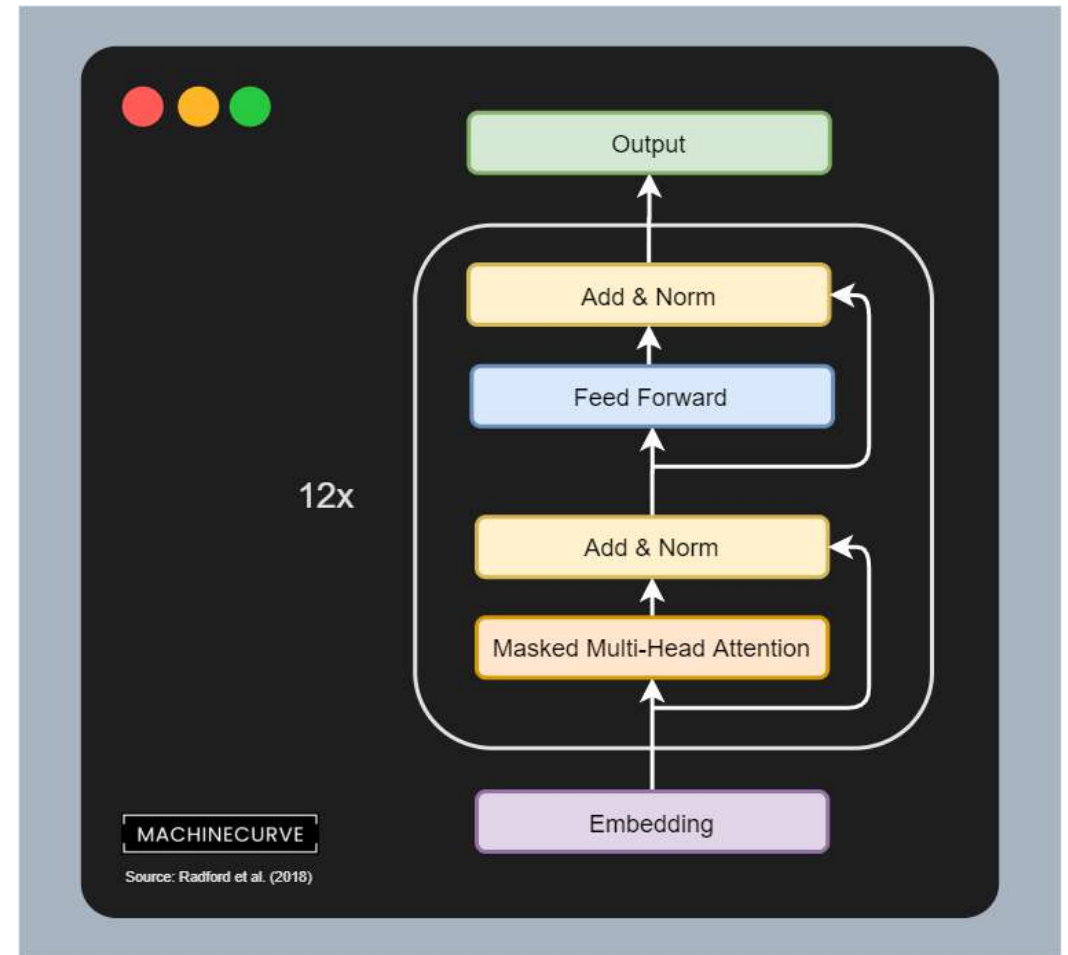
**E que é um
GRANDE
modelo de
lingua?**

Arquitectura neuronal artificial

Transformers



Vaswani et al., (2017)



Radford et al., (2018)

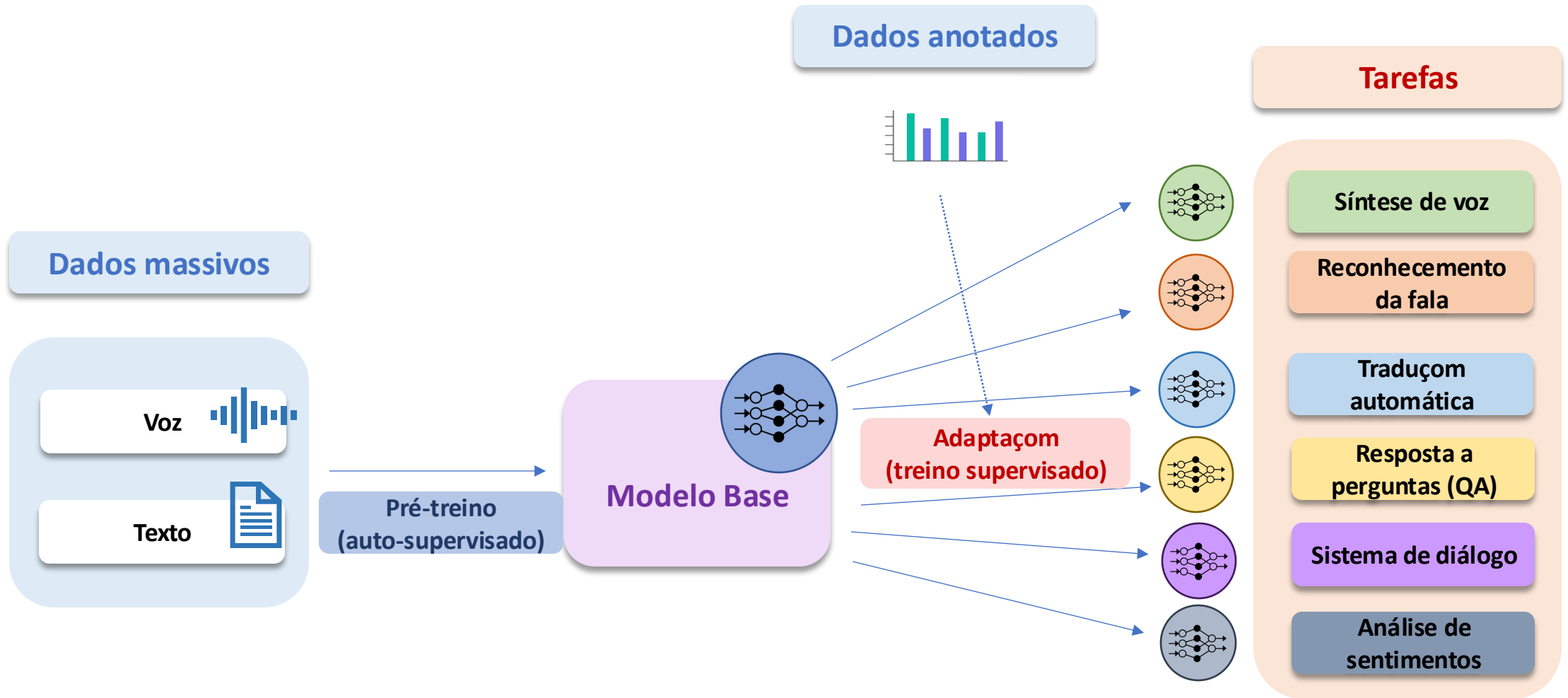
Transformers

Modelo base

Modelo
adaptado a
uma tarefa

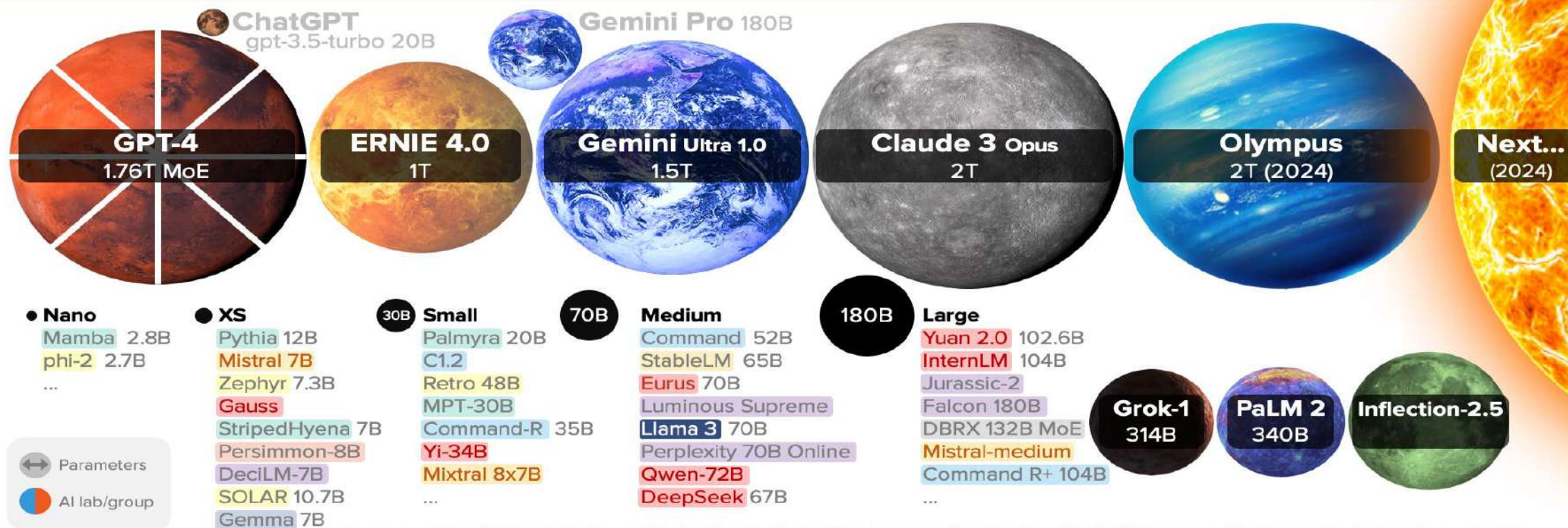


Adaptação de grandes modelos de língua



Modelos de língua por tamanho

LARGE LANGUAGE MODEL HIGHLIGHTS (APR/2024)



Sizes linear to scale. Selected highlights only. All models are available. All models are Chinchilla-aligned (20:1 tokens:parameters) <https://liferchitect.ai/chinchilla/> All 300+ models: <https://liferchitect.ai/models-table/> Alan D. Thompson, 2023-2024.



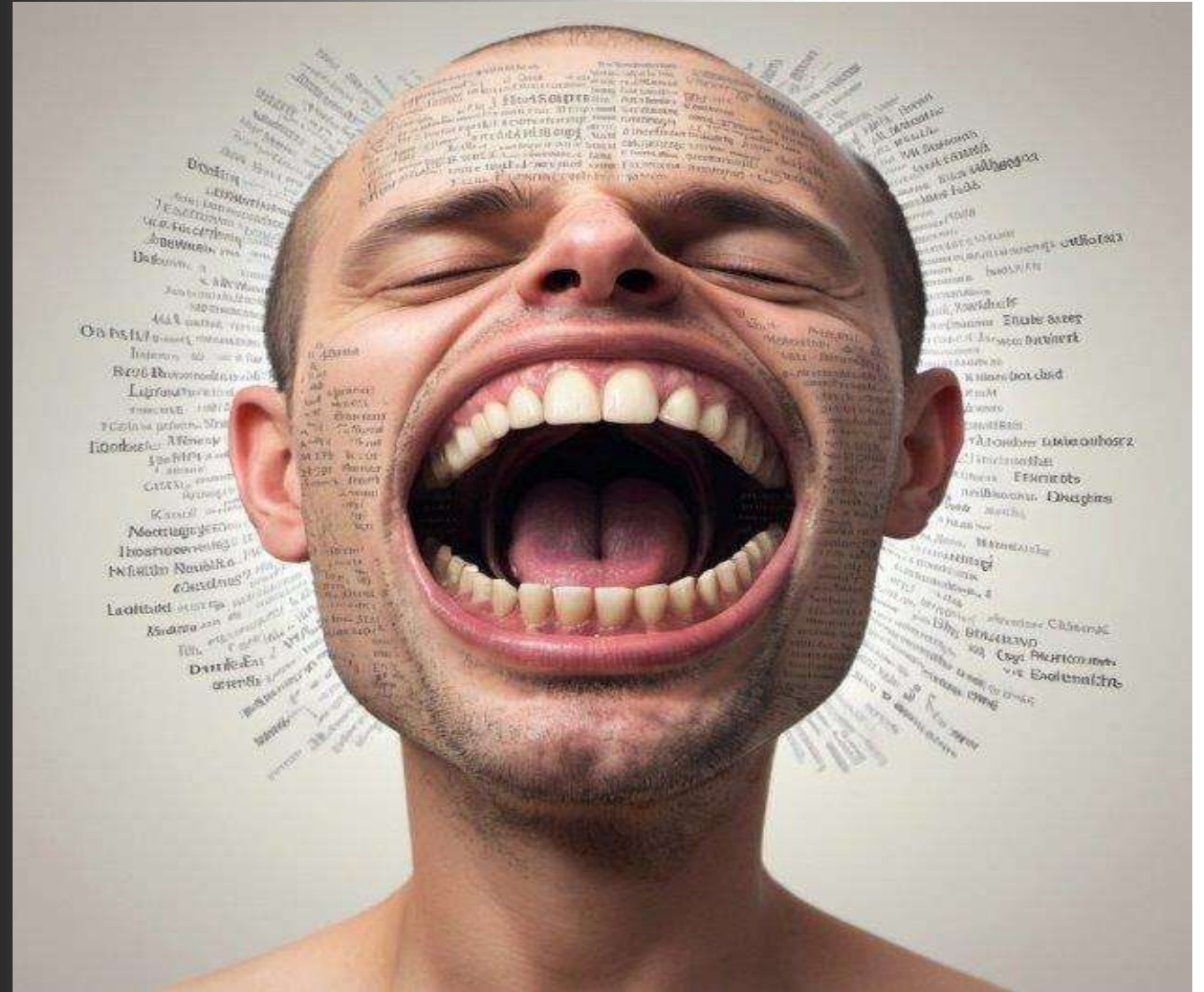


PROBLEMAS dos grandes mo delos de lingua

Problema energético



Enorme quantidade de dados



Pouco fiável
com os feitos:
alucinações



Falta de alinhamento com as expectativas éticas: viés sexista, racista...



Desinformação



Cópia de
trabalhos,
deepfakes, ...



Uniformização linguística e cultural



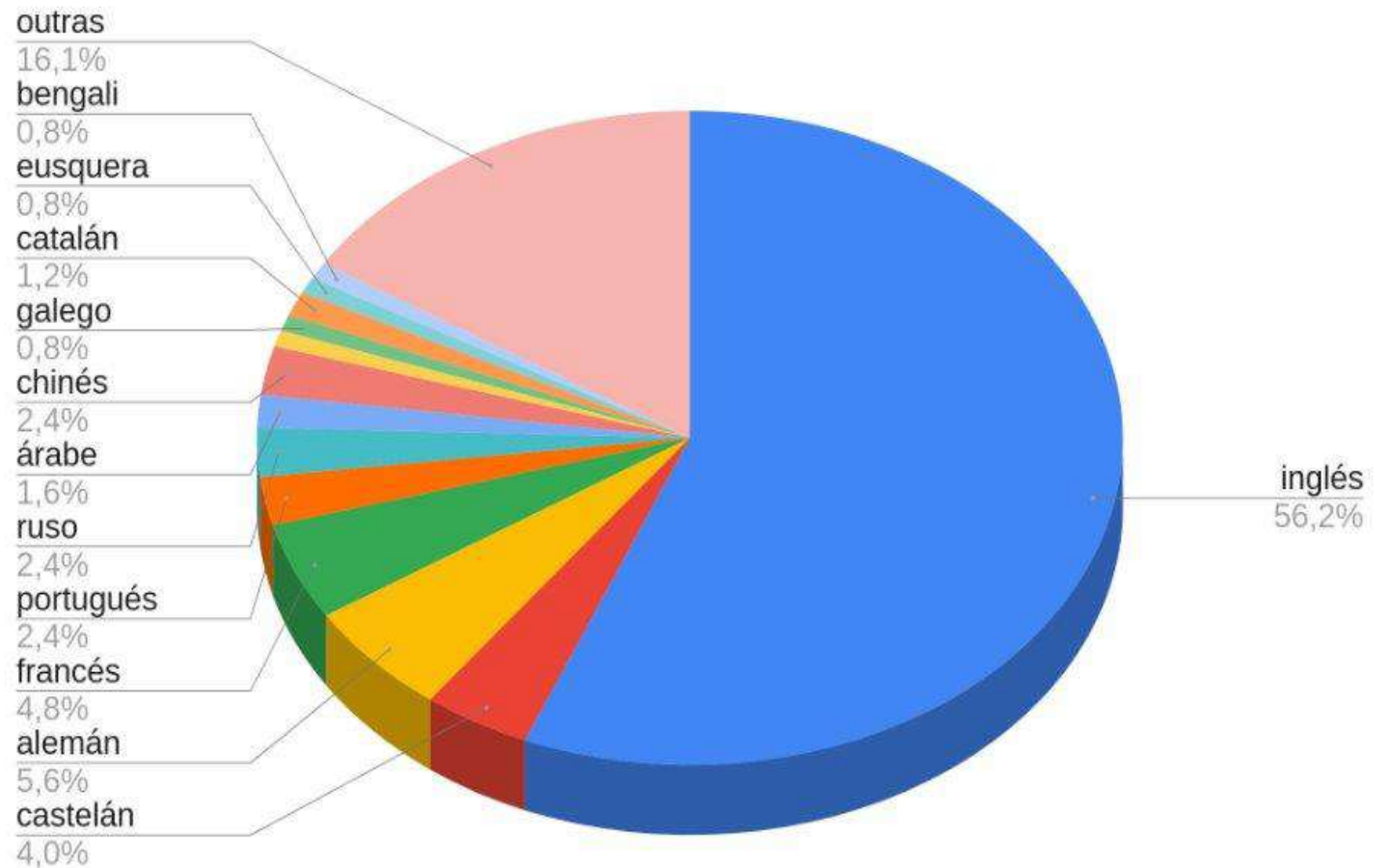
Privacidade e direitos de autor



Opacidade e falta
de transparência



Anglocentrismo lingüístico e cultural





Modelos generativos e de tradução no Proxecto Nós



Que é o Proxecto NÓS?



Missão

- Criar os **suportes digitais** necessários para que o galego prospere como **língua viva** na era digital.
- Criar un **eco-sistema galego** inovador à volta das **tecnologías da linguagem**.

OBJECTIVOS

- Desenvolver **recursos, modelos e ferramentas** para o processamento automático do galego e distribuí-los baixo **licenças livres**.
- Modelos multilingues **equilibrados**
- **Reciclar e reutilizar** todo o que se poda



Modelos generativos do galego

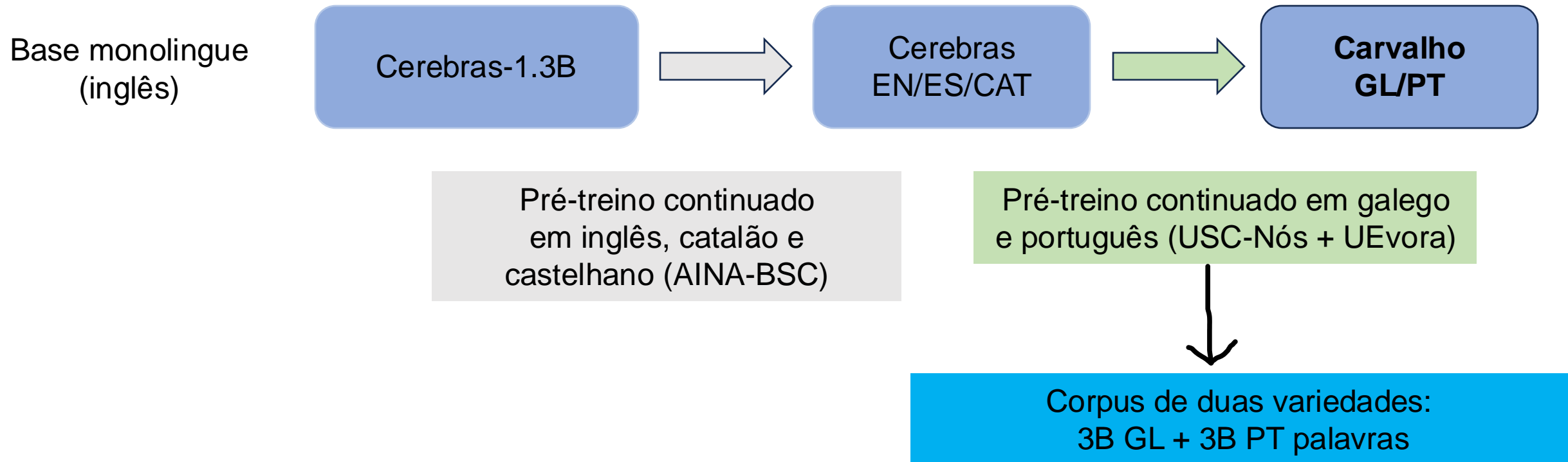


/kar' βalo/

Carballo / Carvalho



Carvalho GL/PT 1.3B



https://huggingface.co/spaces/proxectonos/Nos_LLM_galicia

CESGA



	FinisTerra III					Other nodes
Node groups	ilk	a100	t4	dtn	smp	clk
Number of nodes	256	66	16	2	17	94
Processor	2x Intel Xeon Ice Lake 8352Y, 32 cores					2x Intel Xeon Gold 6240R, 24 cores
Number of cores	16.384	4.224	1.024	128	1.088	4.512
RAM memory	256 GB				2 TB (16) 8 TB (1)	192 GB
Local storage	960 GB SSD NVMe				1920 GB SSD NVMe	2x 480 GB SSD
Accelerators	2x NVIDIA A100 (64)		NVIDIA T4			
	5x NVIDIA A100 (1)					
	8x NVIDIA A100 (1)					
Network	Infiniband HDR 100					10 Gigabit Ethernet

+140 GPUs

Carballo-bloom 1.3B

Exemplo de geração

 Text Generation

Filloas



A receita tradicional das filloas é a de fariña, leite e ovos, pero hai moitas variantes. A máis popular é a de leite

Carballo-bloom 1.3B

Exemplo de resposta a perguntas

Responde á seguinte pregunta.

Pregunta: "Cal é a capital de Noruega?"

Resposta: "A capital de Noruega é Oslo."

Responde á seguinte pregunta.

Pregunta: "Cal é a moeda de Portugal"

Resposta: "A moeda de Portugal é o euro."

Responde á seguinte pregunta.

Pregunta: "Cal é a capital de Suecia?"

Resposta: "A capital de Suecia é Estocolmo."

Carballo-bloom 1.3B

Exemplo de tradución

Traduce ao galego esta frase en inglés:

Inglés: "my sister is studying Biology at the university."

Galego: "a miña irmá está a estudar bioloxía na universidade."

Traduce ao galego esta frase en inglés:

Inglés: "You are working with my mother on a very interesting project."

Galego: "Estás a traballar coa miña nai nun proxecto moi interesante"

Traduce ao galego esta frase en inglés:

Inglés: "You have to fix the computer now"

Galego: "Tes que arranxar o ordenador agora."

Modelos de tradução do galego



TRADUTOR NEURONAL

MELHORAS E AVANCES

- **TRADUTOR NEURONAL** feito na Galiza
- Livre e de **CÓDIGO ABERTO**
- **QALIDADE** próxima a *Google translator*
- Neste momento xa é potencialmente **ÚTIL** para terceiros

Esta é un sistema de **tradución automática** de propósito xeral baseada en **redes neuronais artificiais** que está en fase de probas. Esta tecnoloxía ten a particularidade de depender da dispoñibilidade de grandes cantidades de datos, e que polo momento as traducións xeradas de forma automática non son totalmente fiables e precisan dunha revisión experta. É por iso que solicitamos a túa colaboración e convidámoste a que nos fagas chegar a túa opinión sobre a calidade das traducións premendo en "*axúdanos a mellorar*". Polo momento esta ferramenta só traduce de Castelán a Galego e de Inglés a Galego. Estamos traballando para engadir as outras direccións (Galego a Castelán e Galego a Inglés) así como para mellorar a calidade dos modelos.

Galego Español Inglés ⇌ Inglés Galego Español Transformer

Insira o texto ou URL que desexa traducir

0/1000

TRADUCIR

Limpar

Axúdanos a mellorar

Copiar

Advertencia

Os datos de uso do tradutor serán gardados e utilizados exclusivamente para a mellora do sistema. Ten en conta que as traducións xeradas de forma automática non son totalmente fiables e precisan dunha revisión experta.

Galego Inglés Español ⇌ Inglés Galego Español

Transformer ▾

me quiero acostar temprano porque estoy muy cansado, y mañana me caso en el ayuntamiento

88/1000

[Empty translation area]

 Copiar

TRADUCIR

Limpar

 Axúdanos a mellorar

Galego Inglés Español ⇌ Inglés Galego Español

Transformer ▾

me quiero acostar temprano porque estoy muy cansado, y mañana me caso en el ayuntamiento

88/1000

quérome deitar cedo porque estou moi canso, e mañá caso no concello

Copiar

TRADUCIR

Limpar

 Axúdanos a mellorar

Galego **Inglés** Español ⇌ Inglés **Galego** Español

UK general election live: Gambling Commission speaks to PM's chief of staff as witness in election bets investigation

117/1000

TRADUCIR

Transformer ▼

[Empty translation area]

Copiar

Limpar

Axúdanos a mellorar


Galego
Inglés
Español
↔
Inglés
Galego
Español

Transformer 

UK general election live: Gambling Commission speaks to PM's chief of staff as witness in election bets investigation

117/1000

Eleccións xerais do Reino Unido en directo: A Comisión de xogo fala co xefe de persoal do PM como testemuña na investigación de apostas electorais

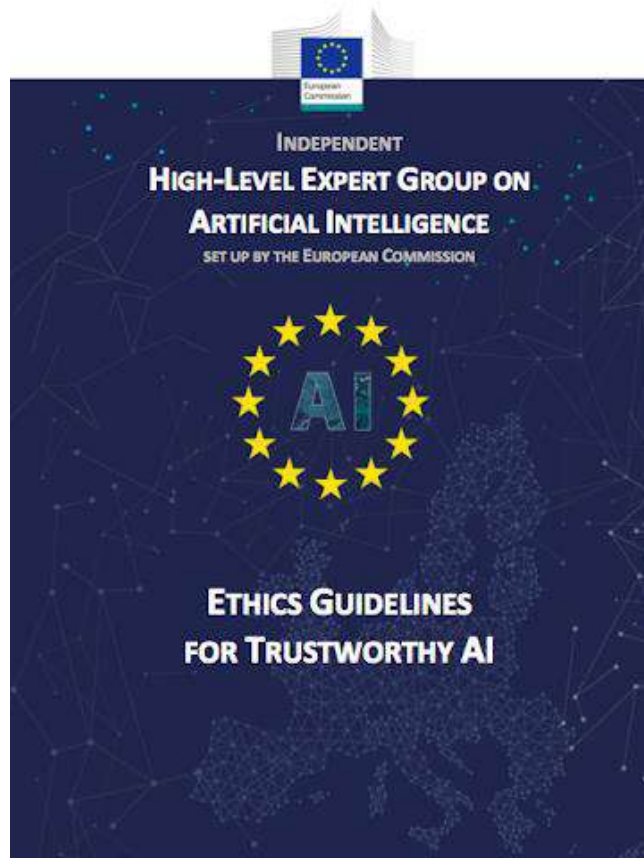
 Copiar

TRADUCIR

Limpar

 Axúdanos a mellorar

IA e PLN fiables



Todos os desenvolvementos no marco do **PLN fiable** (*trustworthy*) que promove especialmente a UE

- **Acción e supervisión humanas**
- **Solidez técnica e seguridade**
- Xestión da **privacidade** e dos datos
- **Transparente: trazable, explicable**
- **Diversidade, non discriminación e equitativa**
- **Benestar** social e ambiental
- **Rendición de contas:** auditable, minimización de efectos negativos

