

Seminário IA Generativa e a Web no Brasil

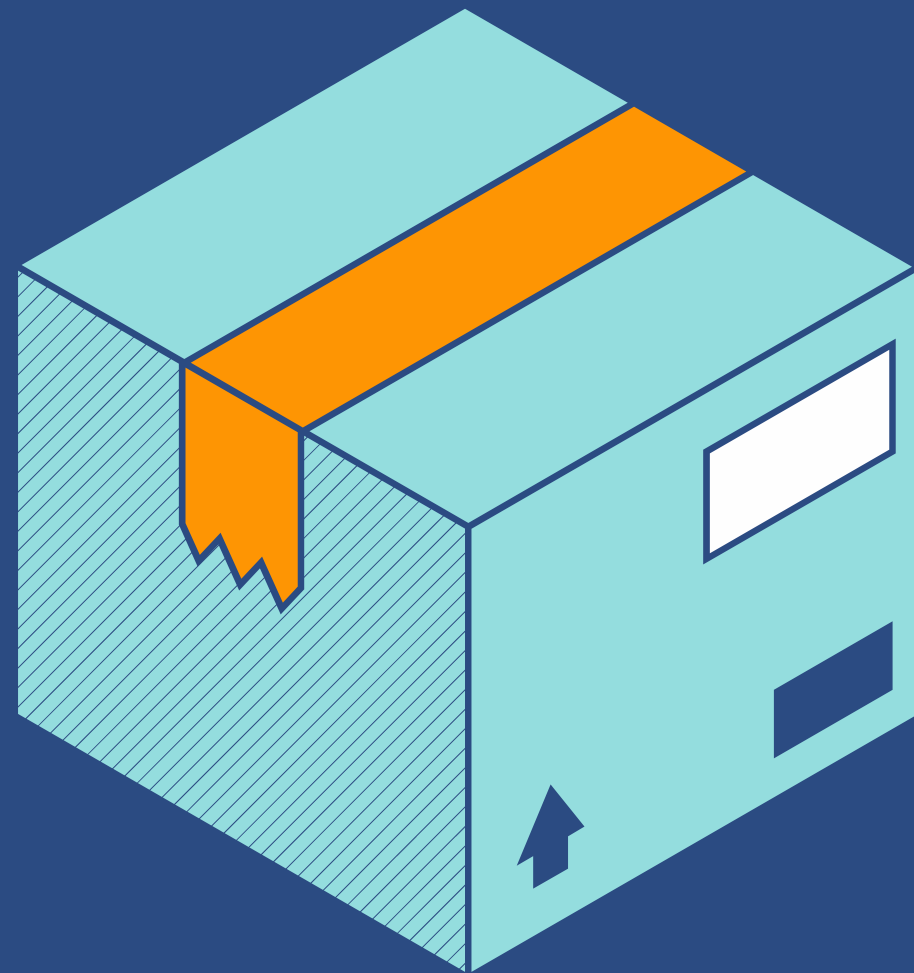
# Inteligências em Interação

Tiago Torrent

Universidade Federal de Juiz de Fora  
FrameNet Brasil



# TL;DL



Interfaces conversacionais unem uma inteligência artificial programada para encontrar **padrões formais que soem naturais** a uma cognição linguística multimodal e cooperativa programada para **construir sentido a partir de qualquer pista**.

ONLY  
PROTECTS  
YOU FROM  
THE SUN



# IA's Gerativas Conversacionais





**MODELO DE LÍNGUA  
PRÉ-TREINADO**

+



**APRENDIZADO DE  
REFORÇO POR  
FEEDBACK HUMANO**



MODELO DE LÍNGUA  
PRÉ-TREINADO

+



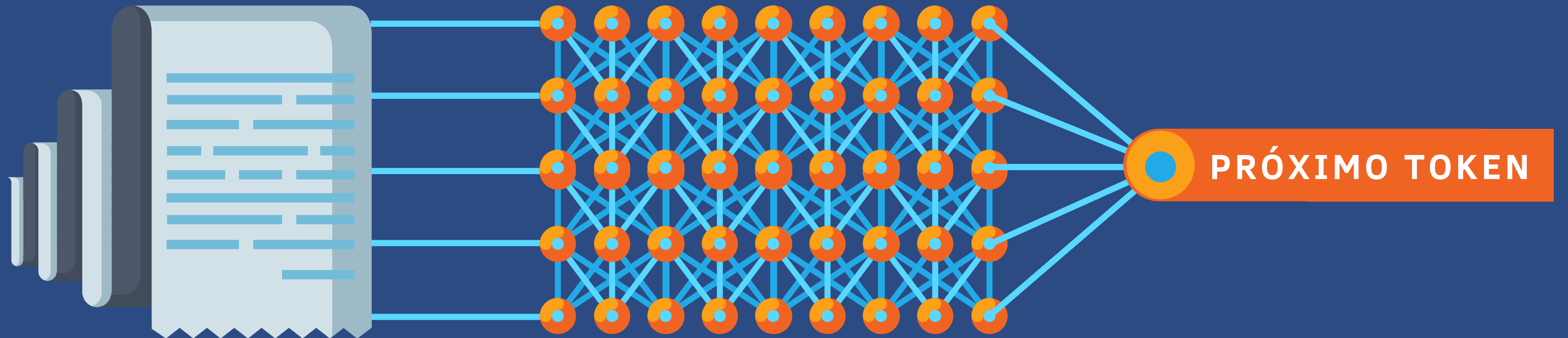
APRENDIZADO DE  
REFORÇO POR  
**FEEDBACK HUMANO  
ON STEROIDS!**



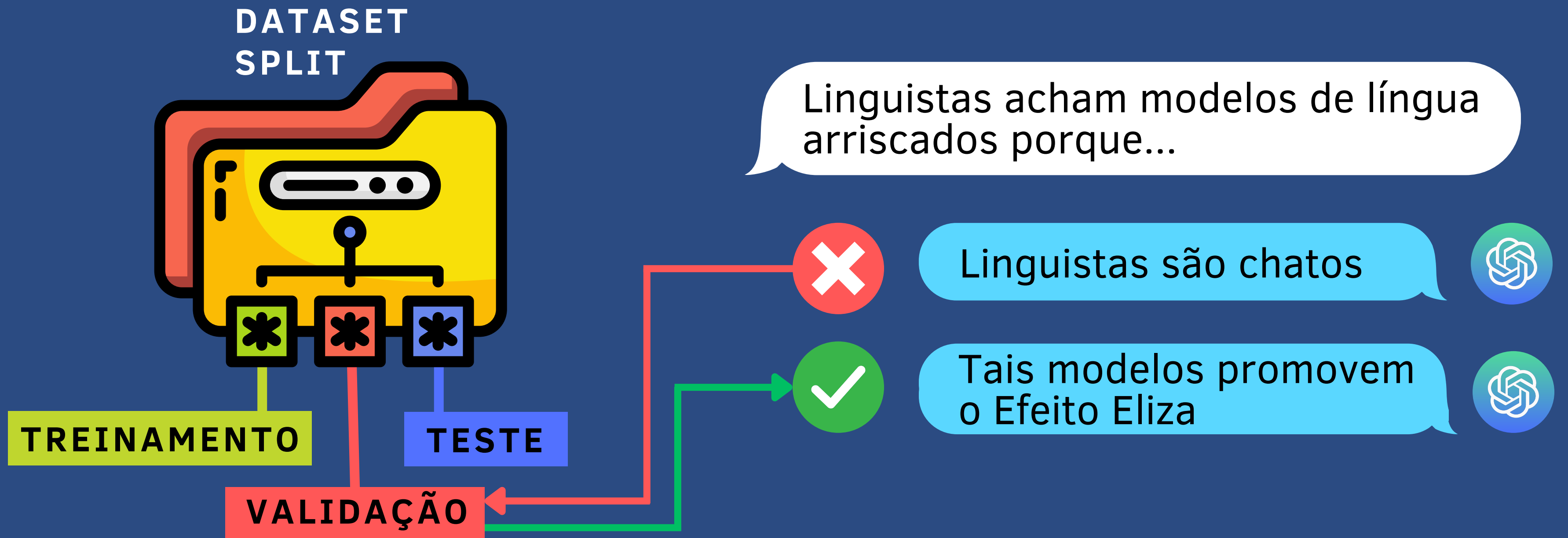
# Modelo de Língua Pré-Treinado

300 BILHÕES  
DE TOKENS


175 BILHÕES  
DE PARÂMETROS



# Treinamento Não-Supervisionado







Load a preset... 

Save




View code

Share



Complete a frase: Linguistas acham modelos de língua arriscados porque... 

eles podem levar a conclusões incorretas sobre a natureza da língua e como ela funciona, e podem levar ao uso de práticas de ensino e de aprendizado inadequadas.

 Looking for ChatGPT? [Try it now](#)  


Submit











84

### Mode

 Complete 

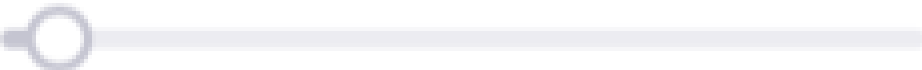
### Model

text-davinci-003 

Temperature 0.7




Maximum length 256



### Stop sequences

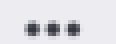
Enter sequence and press Tab

Load a preset... 

Save

View code

Share



Complete a frase: Linguistas acham modelos de língua arriscados porque... 

eles podem levar a conclusões incorretas sobre a natureza da língua e como ela funciona, e podem levar ao uso de práticas de ensino e de aprendizado inadequadas.

Mode

 Complete 

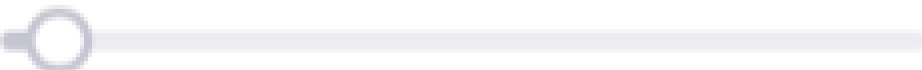
Model

text-davinci-003 

Temperature 0.7



Maximum length 256



Stop sequences

Enter sequence and press Tab

 Looking for ChatGPT? [Try it now](#)  

Submit

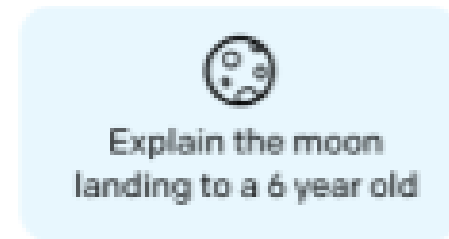


84

### Step 1

#### Collect demonstration data, and train a supervised policy.

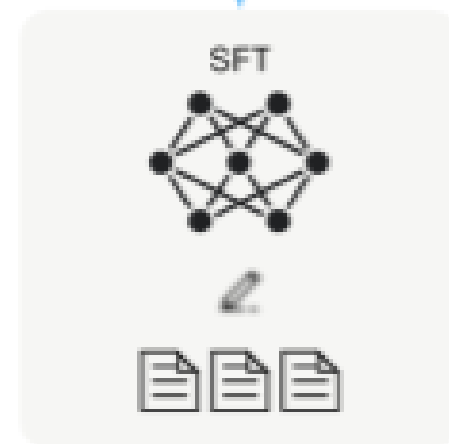
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



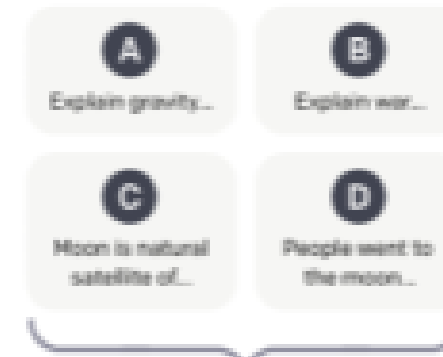
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

#### Collect comparison data, and train a reward model.

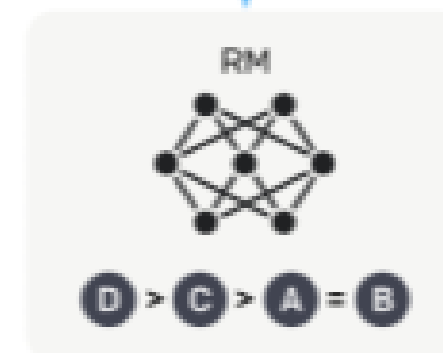
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



### Step 3

#### Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.



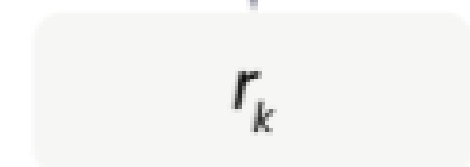
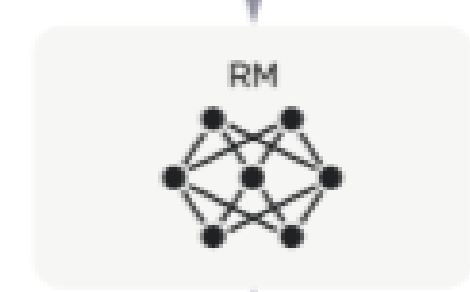
The policy generates an output.



The reward model calculates a reward for the output.

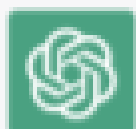


The reward is used to update the policy using PPO.





Assistente, por que os Linguistas acham modelos de língua arriscados?



Os linguistas podem considerar os modelos de língua arriscados por algumas razões. Aqui estão algumas delas:



1. **Simplificação excessiva:** Os modelos de língua, como o GPT-3.5, são projetados para gerar texto e responder perguntas com base em padrões aprendidos de grandes conjuntos de dados. No entanto, eles podem simplificar demais a complexidade das línguas. Os modelos de língua podem não captar a totalidade da riqueza e sutileza de um idioma, como variações regionais, linguagem figurada ou expressões idiomáticas específicas.
2. **Dados de treinamento enviesados:** Os modelos de língua aprendem com grandes conjuntos de dados disponíveis na internet. Esses dados podem ser enviesados em termos de gênero, raça, cultura ou outros aspectos sociais. Como resultado, os modelos podem reproduzir ou amplificar esses vieses, gerando respostas discriminatórias ou ofensivas.

3. **Ausência de compreensão real:** Embora os modelos de língua possam gerar respostas aparentemente corretas, eles muitas vezes carecem de uma verdadeira compreensão do significado subjacente das palavras ou do contexto. Eles se baseiam principalmente em padrões estatísticos dos dados de treinamento para gerar suas respostas. Isso pode levar a respostas imprecisas, ambíguas ou fora de contexto.
4. **Falta de consciência pragmática:** Os modelos de língua geralmente têm dificuldade em entender as sutilezas pragmáticas da linguagem, como ironia, sarcasmo ou intenções implícitas. Eles tendem a interpretar o texto de maneira literal, o que pode resultar em respostas inadequadas ou confusas em certas situações.
5. **Responsabilidade ética:** Os modelos de língua podem ser usados de maneiras prejudiciais ou antiéticas, como a disseminação de desinformação, discurso de ódio ou manipulação de opiniões. Isso levanta preocupações sobre o papel e a responsabilidade dos modelos de língua na sociedade.

É importante notar que essas preocupações são específicas para os modelos de língua atuais e não refletem necessariamente todos os pontos de vista dos linguistas. Além disso, os avanços na pesquisa estão sendo feitos para abordar essas questões e melhorar a capacidade dos modelos de língua de lidar com a complexidade e a nuances das línguas humanas.

# Cognição Linguística



# Princípio da Cooperação de Grice

## Máximas Conversacionais

1 ————— 2 ————— 3 ————— 4

**QUANTIDADE**

Informartividade

**QUALIDADE**

Verdade

**RELEVÂNCIA**

Ancoramento  
no contexto

**MODOS**

Perspicácia

“ Princípio da Escassez da Forma Linguística.

– Salomão, 1999

“ A forma linguística não porta significado, mas o guia.

– Fauccounier, 1994

“ Os significados são relativizados às cenas,  
ao contexto e ao construal.

– Trott et al., 2020



**“ Tiago, vai lá em cima  
e pede a sua avó o  
trenzin de coisa!**

**“ Tiago, vai lá em cima  
e pede a sua avó o  
trenzin de coisa!**



“ Tiago, vai lá em cima e pede a sua avó o **trenzin de coisa!** ”

## Contexto

- Hora: 11h
- No contexto familiar de então, minha mãe e minha avó eram as responsáveis por fazer o almoço em suas casas.
- As casas ficavam uma sobre a outra.
- Todos eram do Pleonástico Interior de Minas Gerais.



# Criação\_culinária

## Definição

Este frame descreve a preparação de alimentos e refeições. Um **Cozinheiro** cria uma **Comida\_produzida** a partir de **Ingredientes** (crus). O **Instrumento\_de\_aquecimento** e / ou **Recipiente** também podem ser especificados.

## Exemplo(s)

### Elementos de Frame Nucleares

#### FE Core:

**Comida\_produzida** [Produced\_food] A **Comida\_produzida** é o resultado dos esforços do **Cozinheiro**.

**Cozinheiro** [Cook]

**semantic\_type:** @sentient

O **Cozinheiro** prepara a **Comida\_produzida**.

### Elementos de Frame Não-Nucleares

### Relações

### Unidades Lexicais

[assado.n](#) [assar.v](#) [bater.v](#) [colocar.v](#) [consertar.v](#) [cozinhar.v](#) [cozinheiro.n](#) [culinária.n](#) [culinário.a](#) [decorar.v](#)

# Utensílios

## Definição

Um **Utensílio** é um recipiente, uma ferramenta ou algo que é especialmente voltado ao uso doméstico. É criado para um **Uso** específico. Várias propriedades do **Utensílio** podem ser especificadas, como o **Criador**, o **Tempo\_de\_criação**, o **Nome**, o **Tipo**, a **Origem**, as **Partes\_constituíntes**, a **Descrição**, o **Material** do qual ele é composto, a **Quantidade** e o **Lugar** onde está.

## Exemplo(s)

### Elementos de Frame Nucleares

#### FE Core:

**Utensílio [Utensil]** Indica um utensílio feito para um **Uso** específico.

### Elementos de Frame Não-Nucleares

## Relações

## Unidades Lexicais

[bandeja.n](#)

[colher.n](#)

[espátula.n](#)

[panela.n](#)

[porcelana.n](#)

[tigela.n](#)

[utensílio.n](#)

# Subpartes\_de\_artefato

## Definição

Um **Componente** tem uma **Função** distinta na operação de um **Artefato** complexo, que não é considerado como completo, a menos que seja montado a partir de um número de componentes diferentes. Cada **Componente** também pode ser descrito por seu **Criador**, **Tempo\_de\_criação** ou **Material**.

## Exemplo(s)

### Elementos de Frame Nucleares

#### FE Core:

**Artefato [Artifact]** A entidade feita que cumpre um propósito, cuja forma completa é composta de mais de um **Componente**.

**Componente [Component]** Uma entidade que é parte de um **Artefato** e tem uma **Função** distinta.

### Elementos de Frame Não-Nucleares

## Relações

## Unidades Lexicais

[cabo.n](#)

[detalhe.n](#)

[HD.n](#)

[tela.n](#)

As línguas podem ser vagas,  
bagunçadas e variáveis porque  
**os humanos cooperam**  
**enquanto as usam.**

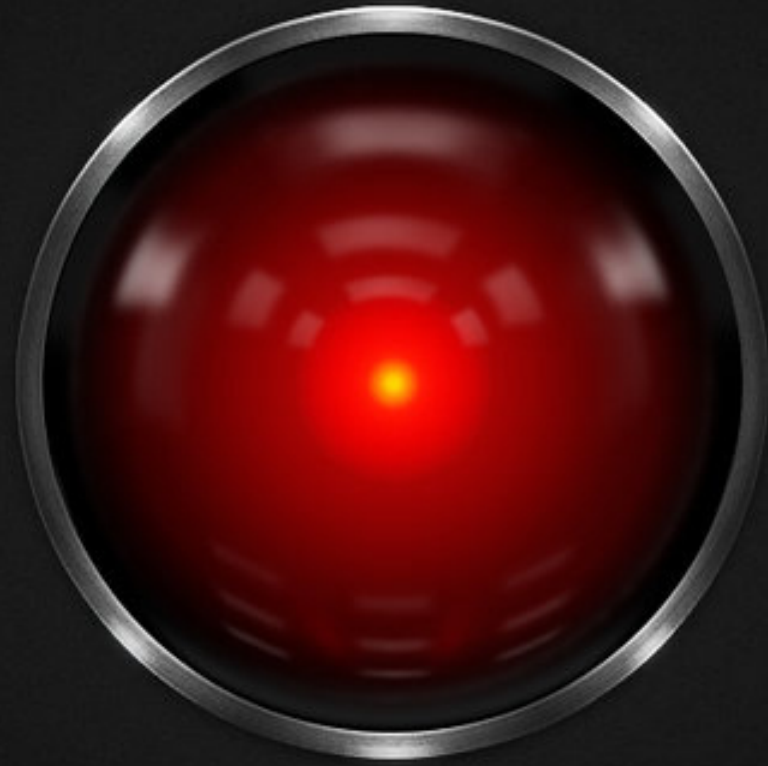


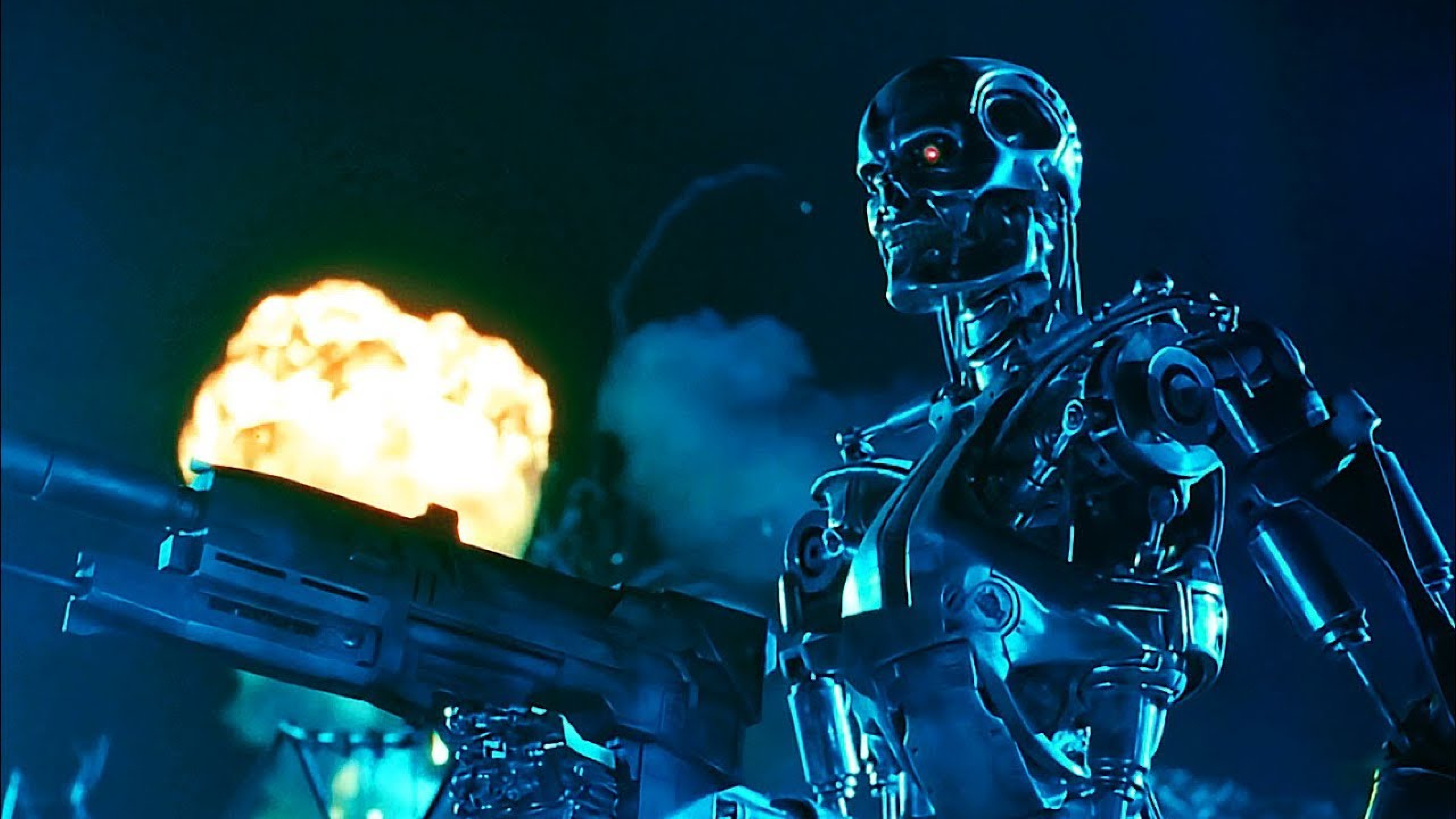
# Riscos das AIs Gerativas Conversacionais





HAL 9000







O risco das IAs gerativas  
conversacionais é a sua  
**antropomorfização comunicativa.**

**A projeção dos princípios de  
cooperação e de construção do sentido  
para IAs gerativas conversacionais é a  
raiz do Efeito Eliza.**

É necessário promover um  
reenquadramento das IAs gerativas  
para aquilo que elas de fato são:

**FERRAMENTAS**

# Obrigado!



<http://tiagotorrent.com>

 @TorrentTiago

FINANCIAMENTO

