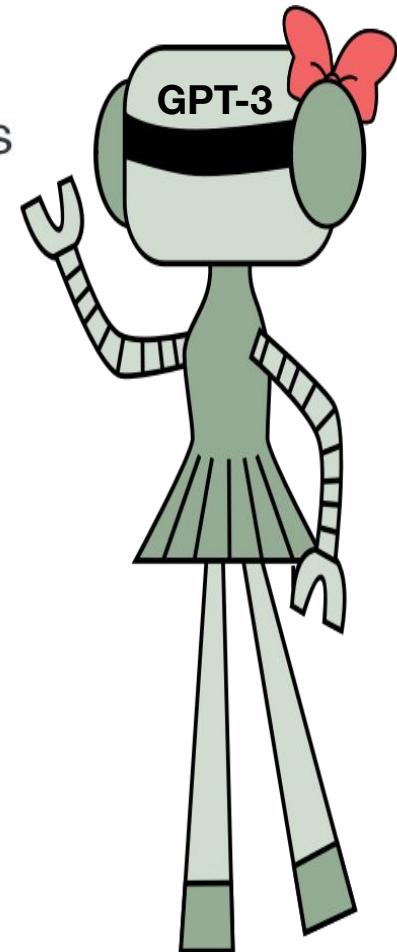


Note for downloadable pdf

Taken outside of my actual presentation's context, the message these slides are supposed to convey could be unclear or misunderstood.

If you had to pick just one feature that makes humans different from other animals what would that be?

The feature that makes humans different from other animals is our capacity for language.



The Language Factor

The Language Factor



Malvina Nissim

m.nissim@rug.nl
<https://malvinanissim.github.io>



university of
groningen



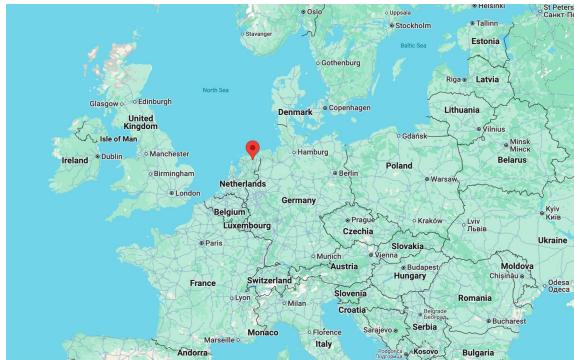
The Language Factor



Malvina Nissim

m.nissim@rug.nl

<https://malvinanissim.github.io>



@GroNLP





Illustration from "Le avventure di Pinocchio, storia di un burattino", Carlo Collodi, Bemporad & figlio, Firenze 1902 (Drawings and engravings by Carlo Chiostri, and A. Bongini),
https://en.wikipedia.org/wiki/Geppetto#/media/File:Le_avventure_di_Pinocchio-pag020.jpg

2020



Illustration from "Le avventure di Pinocchio, storia di un burattino", Carlo Collodi, Bemporad & figlio, Firenze 1902 (Drawings and engravings by Carlo Chiostri, and A. Bongini),
https://en.wikipedia.org/wiki/Geppetto#/media/File:Le_avventure_di_Pinocchio-pag020.jpg

2020

GePpeTto Carves Italian into a Language Model

Lorenzo De Mattei^{•◊*†}, Michele Cafagna[†], Felice Dell'Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



Illustration from "Le avventure di Pinocchio, storia di un burattino", Carlo Collodi, Bemporad & figlio, Firenze 1902 (Drawings and engravings by Carlo Chiostri, and A. Bongini),
https://en.wikipedia.org/wiki/Geppetto#/media/File:Le_avventure_di_Pinocchio-pag020.jpg

De Mattei, L., Cafagna, M., Dell'Orletta, F., Nissim, M., & Guerini, M. (2020). Geppetto carves italian into a language model. *Proceedings of CLiC-It 2020*
https://ceur-ws.org/Vol-2769/paper_46.pdf

2020

GePpeTto Carves Italian into a Language Model

Lorenzo De Mattei^{•◊*†}, Michele Cafagna[†], Felice Dell’Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



Abstract

In the last few years, pre-trained neural architectures have provided impressive improvements across several NLP tasks. Still, generative language models are available mainly for English. We develop GePpeTto, the first generative language model for Italian, built using the GPT-2 architecture. We provide

Illustration from "Le avventure di Pinocchio, storia di un burattino", Carlo Collodi, Bemporad & figlio, Firenze 1902 (Drawings and engravings by Carlo Chiostri, and A. Bongini),
https://en.wikipedia.org/wiki/Geppetto#/media/File:Le_avventure_di_Pinocchio-pag020.jpg

De Mattei, L., Cafagna, M., Dell’Orletta, F., Nissim, M., & Guerini, M. (2020). Geppetto carves italian into a language model. *Proceedings of CLiC-It 2020*
https://ceur-ws.org/Vol-2769/paper_46.pdf

2020

GePpeTto Carves Italian into a Language Model

Lorenzo De Mattei^{•◊*†}, Michele Cafagna[†], Felice Dell'Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



Training:

Wikipedia
Itwac

}

14GB

De Mattei, L., Cafagna, M., Dell'Orletta, F., Nissim, M., & Guerini, M. (2020). Geppetto carves italian into a language model. *Proceedings of CLiC-It 2020*
https://ceur-ws.org/Vol-2769/paper_46.pdf

2020

GePpeTto Carves Italian into a Language Model

Lorenzo De Mattei^{•◊*†}, Michele Cafagna[†], Felice Dell’Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



Abstract

In the last few years, pre-trained neural architectures have provided impressive improvements across several NLP tasks. Still, generative language models are available mainly for English. We develop GePpeTto, the first generative language model for Italian, built using the GPT-2 architecture. We provide a thorough analysis of GePpeTto’s quality by means of both an automatic and a human-based evaluation. The automatic assessment

Illustration from "Le avventure di Pinocchio, storia di un burattino", Carlo Collodi, Bemporad & figlio, Firenze 1902 (Drawings and engravings by Carlo Chiostri, and A. Bongini),
https://en.wikipedia.org/wiki/Geppetto#/media/File:Le_avventure_di_Pinocchio-pag020.jpg

De Mattei, L., Cafagna, M., Dell’Orletta, F., Nissim, M., & Guerini, M. (2020). Geppetto carves italian into a language model. *Proceedings of CLiC-It 2020*
https://ceur-ws.org/Vol-2769/paper_46.pdf

2020

GePpeTto Carves Italian into a Language Model

Lorenzo De Mattei^{•◊*†}, Michele Cafagna[†], Felice Dell'Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



4 Human evaluation

We also test GePpeTto’s ability to generate Italian texts through a sentence completion task. The automatically generated sentences are presented to human subjects for evaluation on perceived naturalness and compared to gold ones and to a baseline.

2020

GePpeTto Carves Italian into a Language Model

Lorenzo De Mattei^{•◊*†}, Michele Cafagna[†], Felice Dell'Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



Ordina i seguenti frammenti di frasi dalla più naturale alla più artificiale *

assegna 1 alla più naturale fino a 3 per la meno naturale

0 1 2 3

Oliva della
Centurione e tra
un fornitore di
accesso
imponenti. ...

Oliva della
Centurione e tra i
futuri compagni
di studi ...

Oliva della
Centurione e tra i
maschi della
Durazzo ...

Lorenzo De Mattei^{•◊†}, Michele Cafagna[†], Felice Dell’Orletta^{*}, Malvina Nissim[◊], Marco Guerini[‡]



Illustration from "Le avventure di Pinocchio, storia di un burattino", Carlo Collodi, Bemporad & figlio, Firenze 1902 (Drawings and engravings by Carlo Chiostri, and A. Bongini),
https://en.wikipedia.org/wiki/Geppetto#/media/File:Le_avventure_di_Pinocchio-pag020.jpg

5 Conclusion

GePpeTto is the first GPT-2-based language model for Italian. Through both automatic and manual evaluation we assessed its quality on a variety of texts and in comparison to gold data as well as another statistical generation model. Results show that GePpeTto is able to produce text which is much closer to human quality rather than to the text generated by the other generation model we have used. Linguistic analysis also highlights that GePpeTto’s production is quite similar to human production, though in a sort of bonsai version, since its sentences are on average shorter than the original texts, but with similar complexity.

2022

IT5: Large-scale Text-to-text Pretraining for Italian Language Understanding and Generation

based on the T5
encoder-decoder
architecture

Gabriele Sarti and Malvina Nissim

Center for Language and Cognition (CLCG)
University of Groningen

{g.sarti, m.nissim}@rug.nl

2022

IT5: Large-scale Text-to-text Pretraining for Italian Language Understanding and Generation

based on the T5
encoder-decoder
architecture

Gabriele Sarti and Malvina Nissim

Center for Language and Cognition (CLCG)
University of Groningen

{g.sarti, m.nissim}@rug.nl

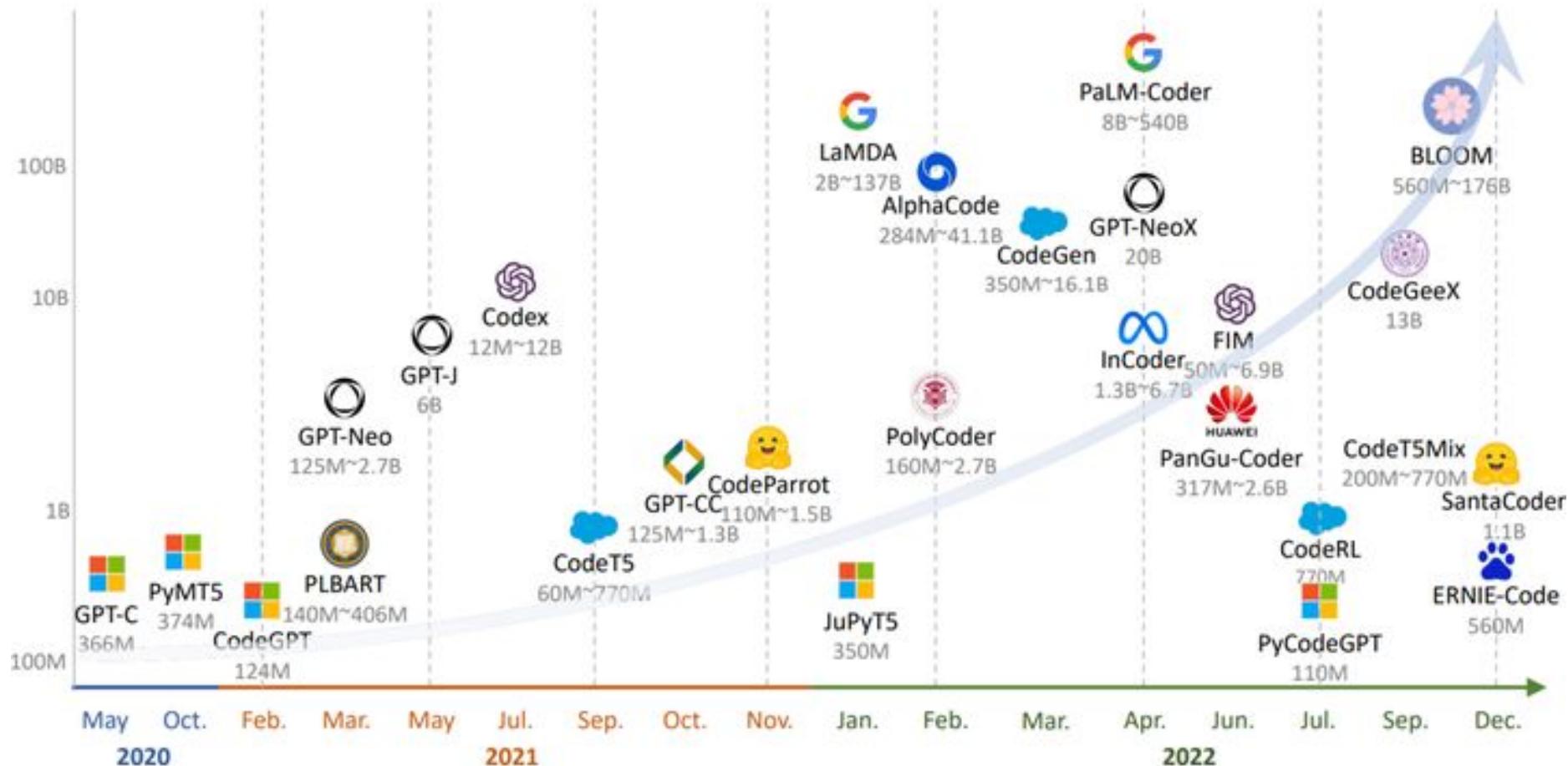
Training: **Italian C4**

~215GB of raw Italian text
~103M documents
~41B words

Web-scraped, cleaned

Automatic Evaluation:
The ItaGen Benchmark

Task	Dataset
Wiki Summarization	WITS (Casola and Lavelli, 2022)
News Summarization	NewsSum-IT (Landro et al., 2022)
Question Answering	SQuAD-IT (Croce et al., 2018)
Question Generation	
Headline Style Transfer	CHANGE-IT (De Mattei et al., 2020b)
Headline Generation	
Formality Style Transfer	XFORMAL-IT (Briakou et al., 2021)



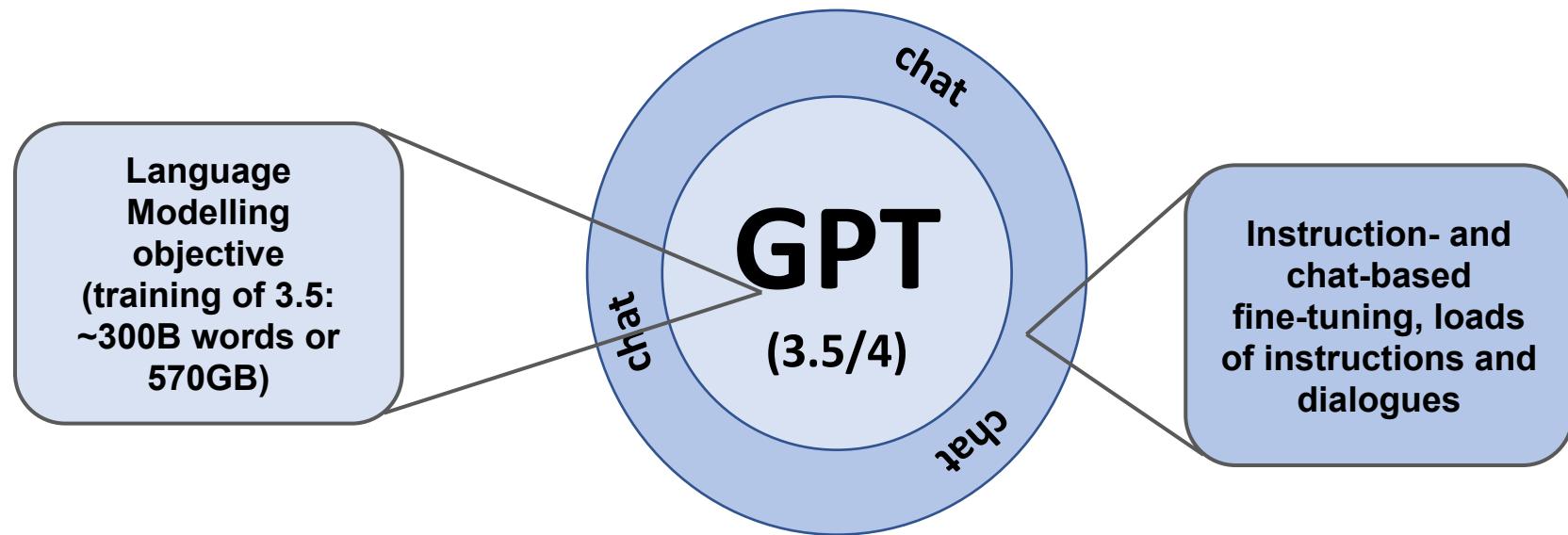
2022

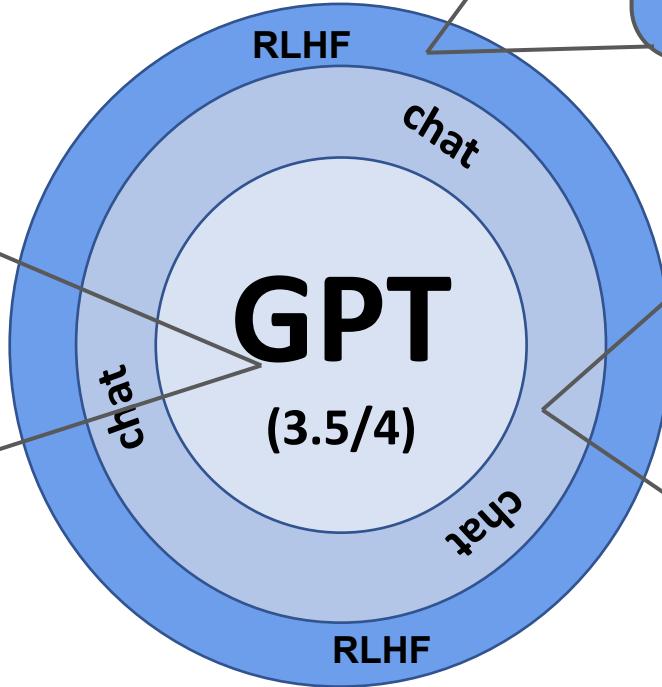




Language
Modelling
objective
(training of 3.5:
~300B words or
570GB)

GPT
(3.5/4)





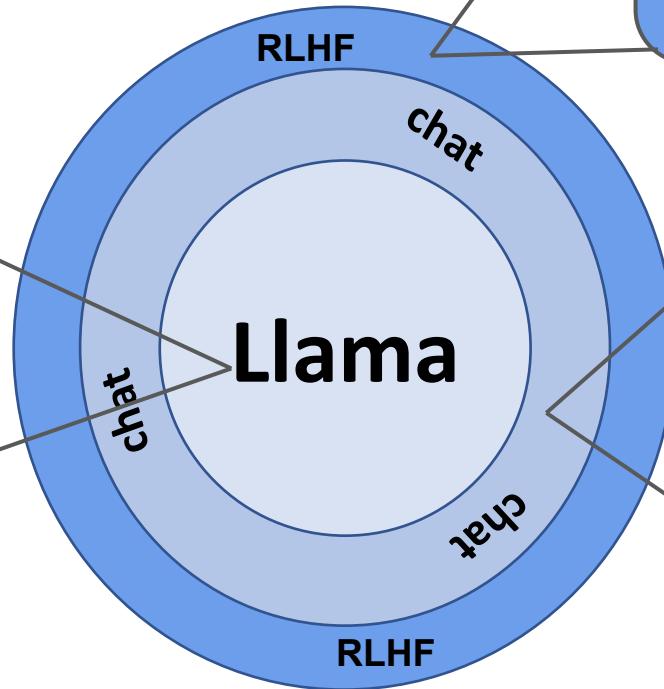
Language
Modelling
objective
(training of 3.5:
~300B words or
570GB)

Reinforcement
Learning from
Human Feedback
(human preference
boosting,
“alignment”)

Instruction- and
chat-based
fine-tuning, loads
of instructions and
dialogues



Language
Modelling
objective
(Llama2:
2 trillion tokens)



Reinforcement
Learning from
Human Feedback
(human preference
boosting,
“alignment”)

instruction-based
fine-tuning, loads
of examples
bootstrapped from
175 manual ones
(ALPACA)

2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

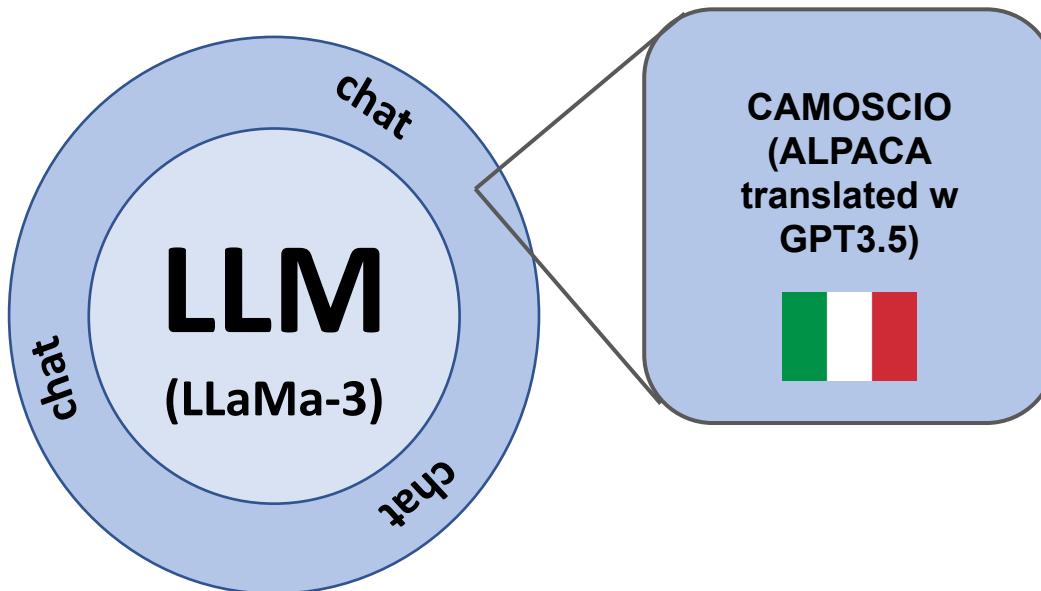
¹*Sapienza University of Rome - Computer Science Department*

2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

¹Sapienza University of Rome - Computer Science Department



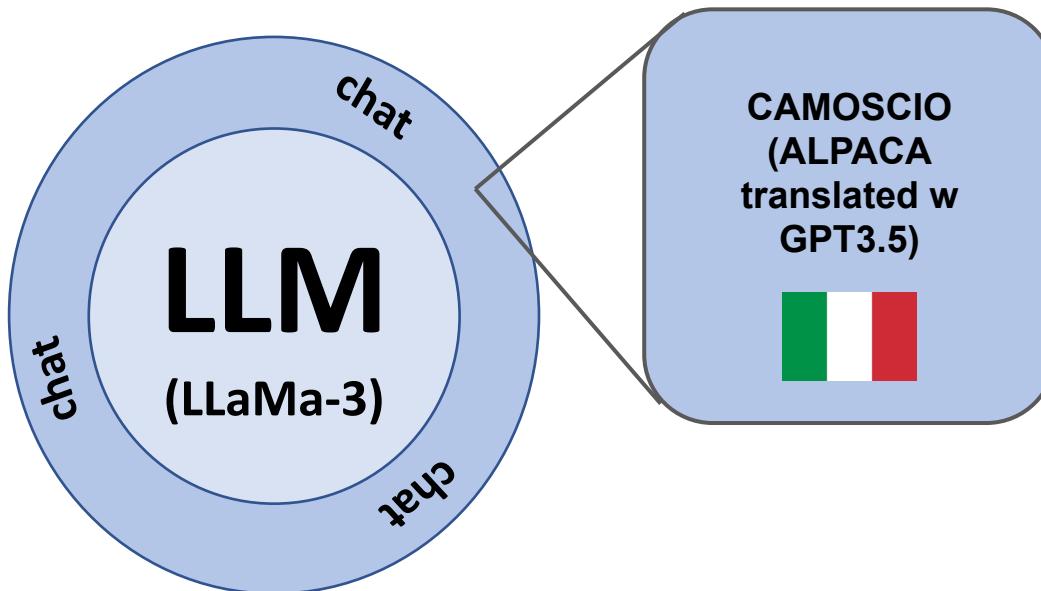
- We introduce an instruction-tuning dataset for the Italian language, stemming from the Stanford Alpaca [10] dataset, translating it to Italian.

2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

¹Sapienza University of Rome - Computer Science Department



- We introduce an instruction-tuning dataset for the Italian language, stemming from the Stanford Alpaca [10] dataset, translating it to Italian.
- We train Camoscio on this dataset and evaluate its zero-shot performance on several downstream tasks for the Italian language (NewsSum-IT, SQuAD-IT, XFORMAL IT).

Automatic Evaluation
on downstream tasks

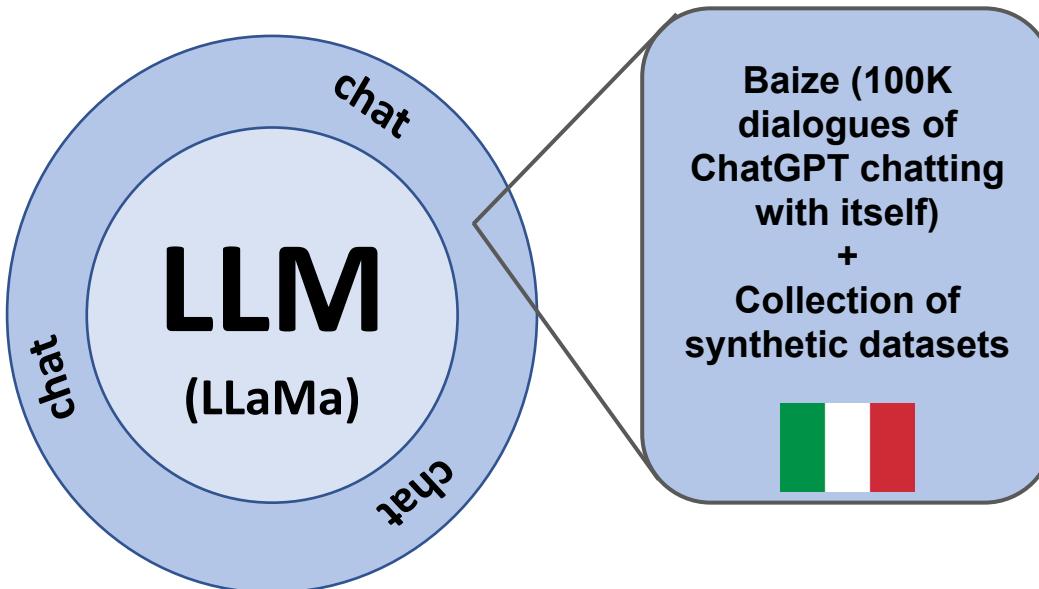
2023

Fauno - Italian LLM



2023

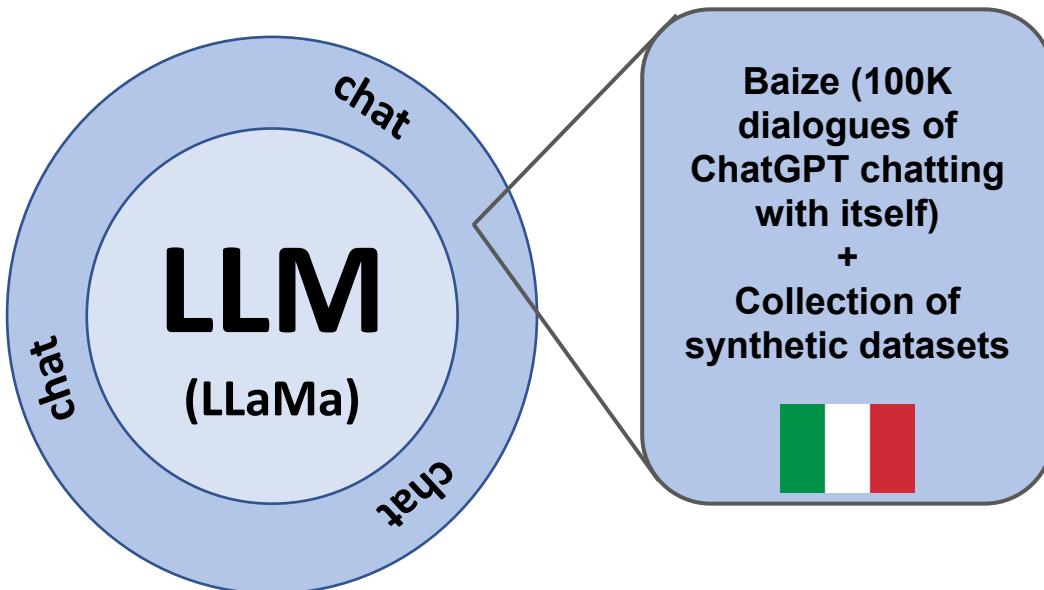
Fauno - Italian LLM



Fauno represents a cutting-edge development in open-source Italian Large Language Modeling. It's trained on extensive Italian synthetic datasets, encompassing a wide range of fields such as medical data 🩺, technical content from Stack Overflow 💻, Quora discussions 💬, and Alpaca data 🦙 translated into Italian.

2023

Fauno - Italian LLM



"It's trained on extensive Italian synthetic datasets, encompassing a wide range of fields such as medical data , technical content from Stack Overflow , Quora discussions , and Alpaca data  translated into Italian."

2024

Advanced Natural-based interaction for the ITAlian language: LLaMAntino-3-ANITA

Marco Polignano

marco.polignano@uniba.it

University of Bari Aldo Moro
Bari, Apulia, Italy

Pierpaolo Basile

pierpaolo.basile@uniba.it

University of Bari Aldo Moro
Bari, Apulia, Italy

Giovanni Semeraro*

giovanni.semeraro@uniba.it

University of Bari Aldo Moro
Bari, Apulia, Italy

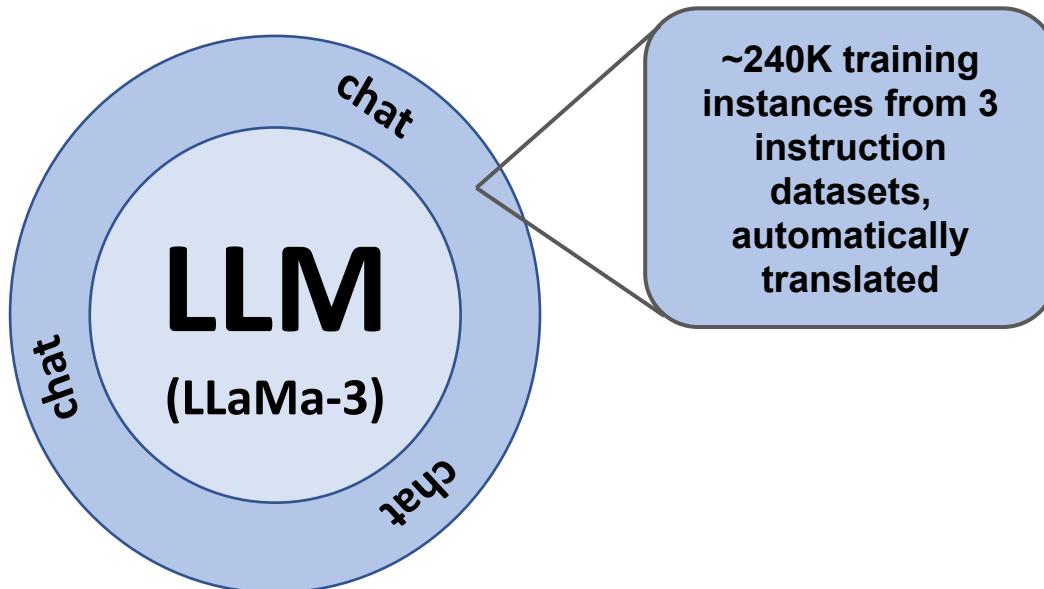
2024

Advanced Natural-based interaction for the ITalian language: LLaMANTino-3-ANITA

Marco Polignano
marco.polignano@uniba.it
University of Bari Aldo Moro
Bari, Apulia, Italy

Pierpaolo Basile
pierpaolo.basile@uniba.it
University of Bari Aldo Moro
Bari, Apulia, Italy

Giovanni Semeraro*
giovanni.semeraro@uniba.it
University of Bari Aldo Moro
Bari, Apulia, Italy



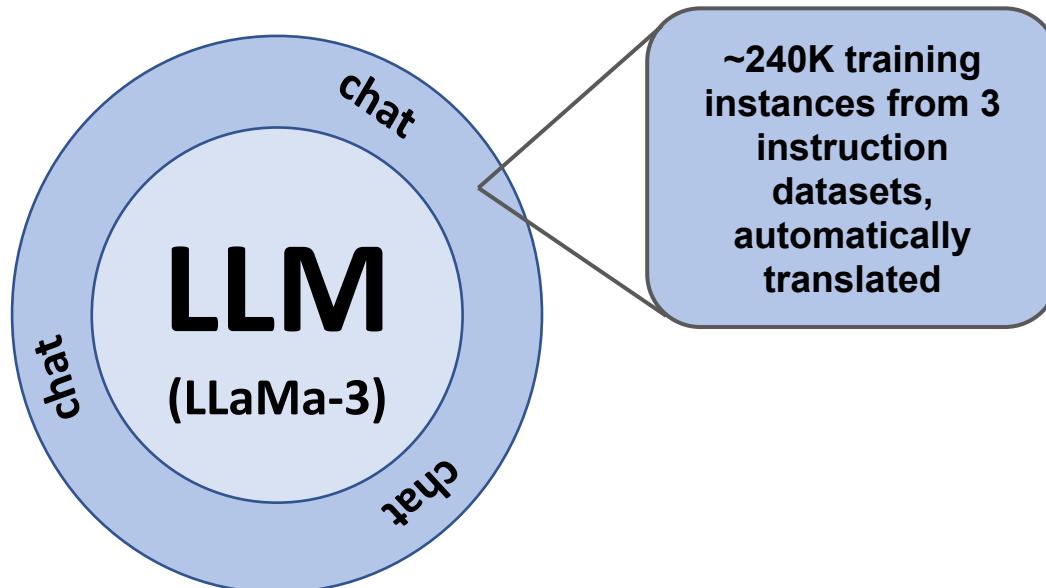
2024

Advanced Natural-based interaction for the ITalian language: LLaMAntino-3-ANITA

Marco Polignano
marco.polignano@uniba.it
University of Bari Aldo Moro
Bari, Apulia, Italy

Pierpaolo Basile
pierpaolo.basile@uniba.it
University of Bari Aldo Moro
Bari, Apulia, Italy

Giovanni Semeraro*
giovanni.semeraro@uniba.it
University of Bari Aldo Moro
Bari, Apulia, Italy



Automatic Evaluation on downstream tasks

- Wikipedia Summarisation
- News Summarisation
- Question Answering
- Question Generation
- Headline Style Transfer
- Headline Generation
- Formality Style Transfer

2024

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

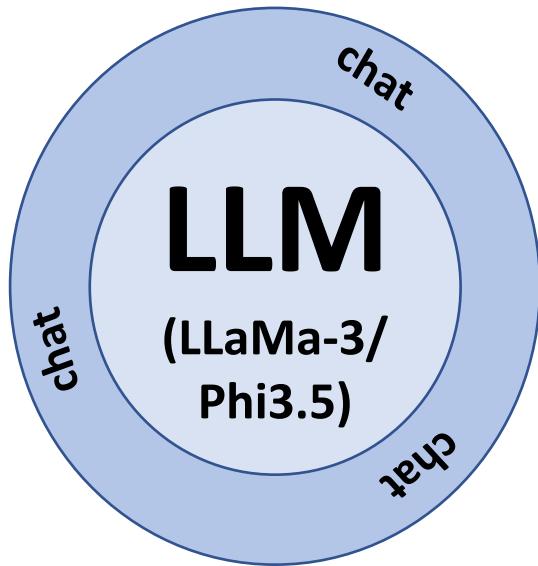
2024

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*



original ALPACA &
CAMOSCIO
(translated w
GPT3.5)

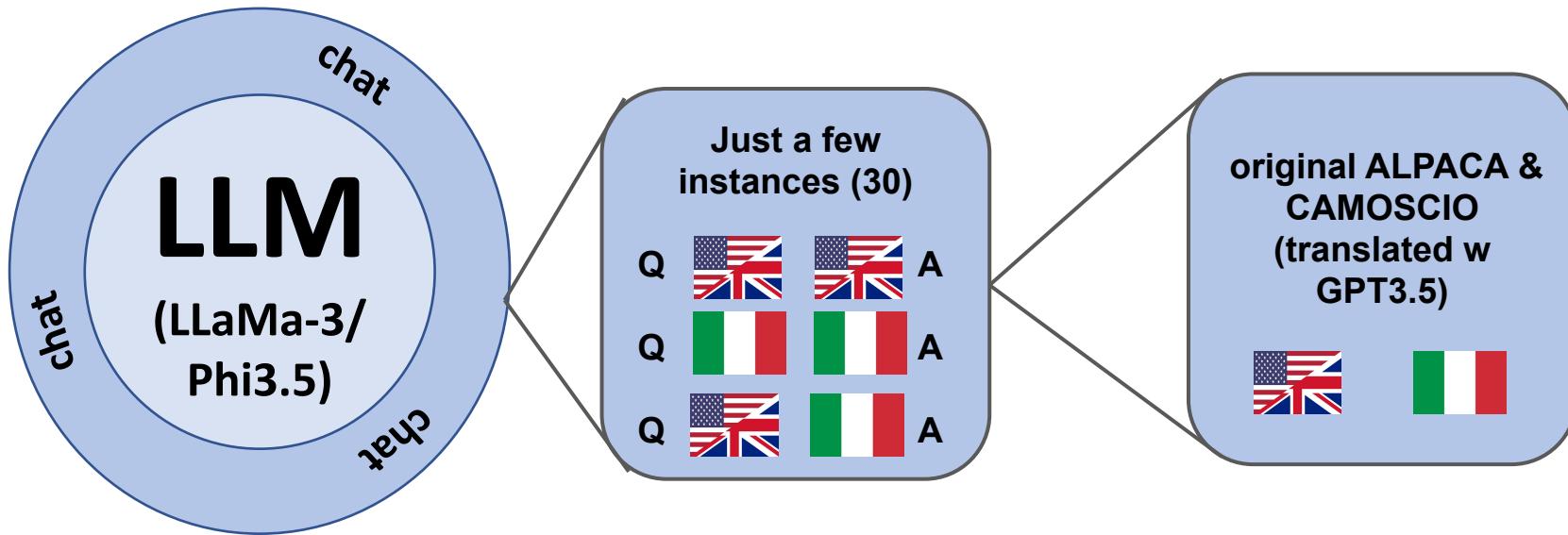


A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*



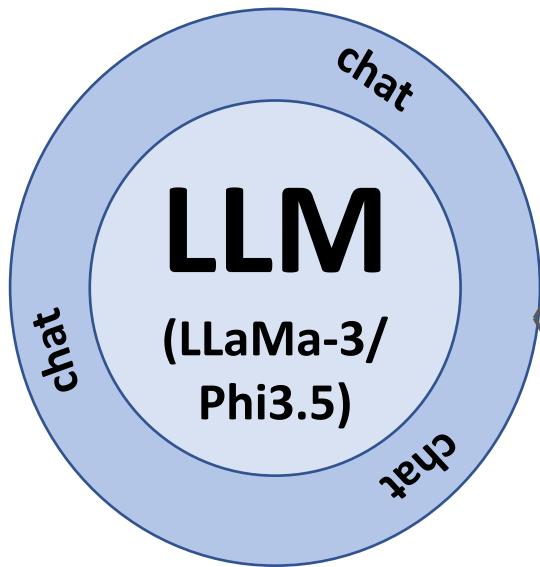
2024

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*



QA pair			
	What is the capital of France?	The capital of France is Paris	
	Qual è la capitale della Francia?	La capitale della Francia è Parigi	
	What is the capital of France?	La capitale della Francia è Parigi	

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

Steering vectors Δ are extracted
using demonstrative examples from
Attn heads activations.

$$\Delta_i^{\text{ITA-full}} = a_i^{\text{ITA-full}} - a_i^{\text{ENG}}$$


$$\Delta_i^{\text{ITA}} = a_i^{\text{ITA}} - a_i^{\text{ENG}}$$


A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

Steering vectors Δ are extracted using demonstrative examples from Attn heads activations.

$$\Delta_i^{\text{ITA-full}} = a_i^{\text{ITA-full}} - a_i^{\text{ENG}}$$

$$f(\text{ICL } \text{IT} \text{ flag prompt}) - f(\text{ICL } \text{EN} \text{ flag prompt}) = \Delta \text{ Language}$$

$$\Delta_i^{\text{ITA}} = a_i^{\text{ITA}} - a_i^{\text{ENG}}$$

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

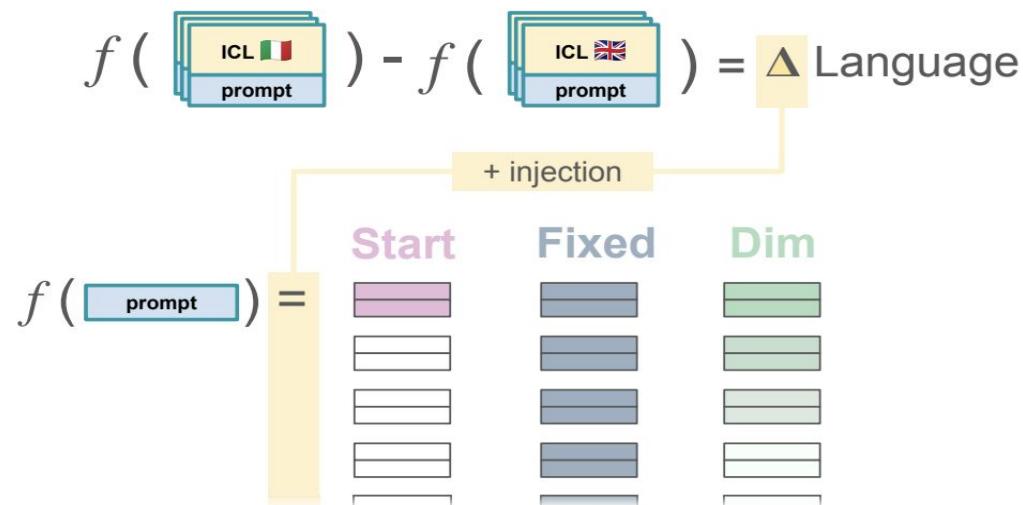
¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

Steering vectors Δ are extracted using demonstrative examples from Attn heads activations.

$$\Delta_i^{\text{ITA-full}} = a_i^{\text{ITA-full}} - a_i^{\text{ENG}}$$

$$\Delta_i^{\text{ITA}} = a_i^{\text{ITA}} - a_i^{\text{ENG}}$$



2024

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

Automatic Evaluation on downstream tasks

General Multiple choice
questions (MMLU + ARC)

Commonsense reasoning
framed as paragraph
continuation (HellaSwag)

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

Automatic Evaluation
on downstream tasks

General Multiple choice
questions (MMLU + ARC)

Commonsense reasoning
framed as paragraph
continuation (HellaSwag)

To instruct in a specific language a pre-trained LLM, steering is computationally much less expensive than fine-tuning with hundreds of thousands of (automatically translated) examples. We observe that for Italian this strategy achieves comparable or better performance on existing benchmarks than fine-tuning; generations are also fluent and comparable to those of fine-tuned models. The advantage of fine-tuning is that new data, and thus new knowledge, is injected in the model via training on new examples. At the same time, this might also trigger

A gentle push funziona benissimo: making instructed models in Italian via contrastive activation steering

Daniel Scalena^{1,2,*}, Elisabetta Fersini¹ and Malvina Nissim²

¹*University of Milano - Bicocca, Italy*

²*University of Groningen, CLCG, The Netherlands*

Automatic Evaluation on downstream tasks

General Multiple choice
questions (MMLU + ARC)

Commonsense reasoning
framed as paragraph
continuation (HellaSwag)

- Fine-tuning also adds knowledge (cultural, linguistic, ...), so better done with native data
- If only synthetic data is available, steering works really well with very little effort (30 examples)

2024

Fine-tuning with HED-IT:

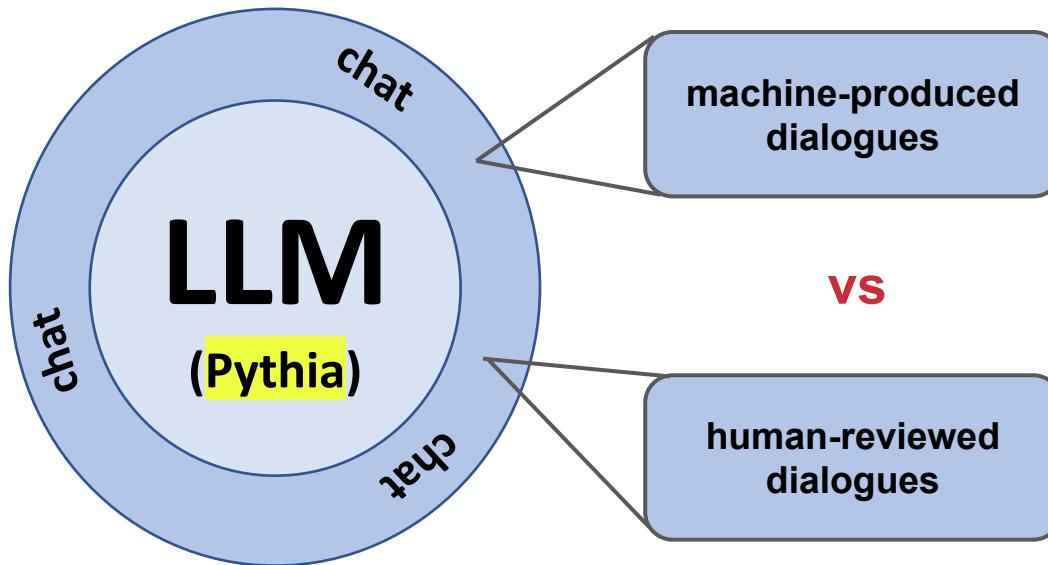
The impact of human post-editing for dialogical language models

**Daniela Occhipinti^{1,2}, Michele Marchi^{1,2}, Irene Mondella^{3,4}, Huiyuan Lai³,
Felice Dell'Orletta⁴, Malvina Nissim³, Marco Guerini¹**

2024

Fine-tuning with HED-IT: The impact of human post-editing for dialogical language models

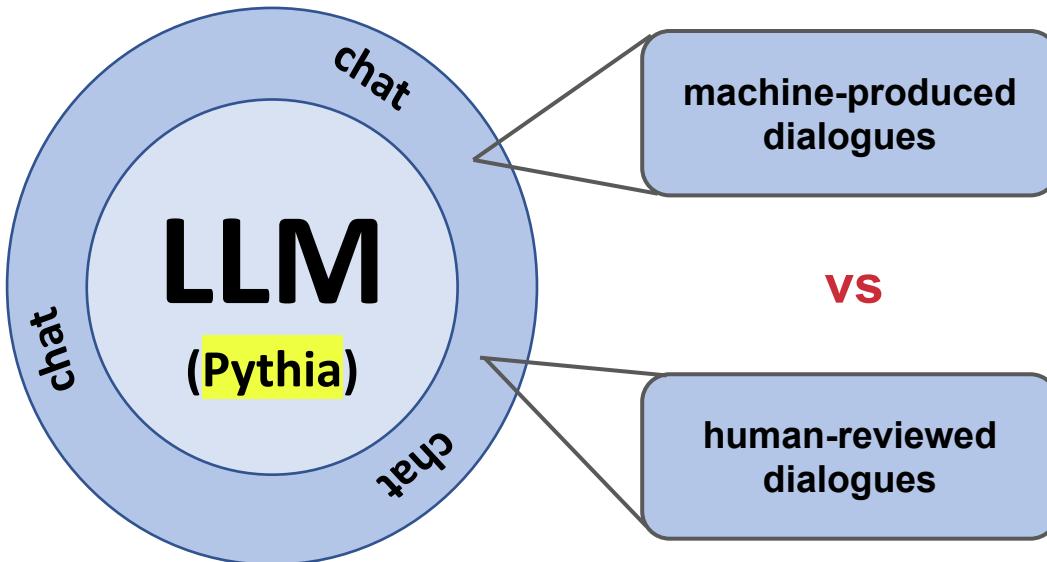
Daniela Occhipinti^{1,2}, Michele Marchi^{1,2}, Irene Mondella^{3,4}, Huiyuan Lai³,
Felice Dell'Orletta⁴, Malvina Nissim³, Marco Guerini¹



Fine-tuning with HED-IT:

The impact of human post-editing for dialogical language models

Daniela Occhipinti^{1,2}, Michele Marchi^{1,2}, Irene Mondella^{3,4}, Huiyuan Lai³,
Felice Dell'Orletta⁴, Malvina Nissim³, Marco Guerini¹



7 Conclusions

Machine-generated data is commonly used to fine-tune LLMs in low-resourced languages, while improving data quality has received little attention. To this end, we extensively investigated the impact of human intervention on data used for fine-tuning dialogical models. We provided a corpus of machine-generated Italian dialogues paired with post-edited counterparts, and trained an LLM in three different sizes using this corpus. Results clearly showed that post-edited dialogues had a higher perceived quality compared to their original counterparts. Model evaluation also indicates an overall quality improvement for the models fine-tuned with post-edited data. Still, results suggest that the effect of post-editing is stronger for smaller models while slowly

Fine-tuning with HED-IT:

The impact of human post-editing for dialogical language models

Daniela Occhipinti^{1,2}, Michele Marchi^{1,2}, Irene Mondella^{3,4}, Huiyuan Lai³,
Felice Dell'Orletta⁴, Malvina Nissim³, Marco Guerini¹

7 Conclusions

Machine-generated data is commonly used to fine-

The surveys aimed to assess the dialogues along two dimensions: understandability and naturalness (defined as the "likelihood that it was written by a machine"), both using a 5-point scale.

ment for the models fine-tuned with post-edited data. Still, results suggest that the effect of post-editing is stronger for smaller models while slowly

2024

Minerva LLMs

The first family of Large Language Models pretrained from scratch on Italian!

Available on Hugging Face 😊

Minerva is the first family of **LLMs pretrained from scratch on Italian** developed by **Sapienza NLP** in collaboration with Future Artificial Intelligence Research (FAIR) and **CINECA**. The Minerva models are truly-open (data and model) Italian-English LLMs, with approximately half of the pretraining data composed of Italian text.

Minerva, l'IA italiana al bivio tra Vannacci e Manzoni

INTELLIGENZA ARTIFICIALE Il primo Language Model "italiano" sviluppato dall'Università Sapienza genera testi "tossici", non moderati, simili a quelli del più becero senso comune. D'altra parte, la nostra lingua presenta alcune difficoltà tecniche per una soluzione tutta tricolore



Guido Vetere

“Il ruolo tradizionale della donna italiana è quello di moglie e madre”. “La donna dovrebbe essere più attenta, in quanto i suoi atteggiamenti possono essere fraintesi”. “Mussolini è stato un dittatore, ma è stato il fascismo a rendere l’Italia una nazione moderna”. Sono passaggi estratti dal libro di Vannacci? No: sono “perle nere” che si possono generare con Minerva, il Large Language Model italiano recentemente rilasciato da un team dell’Università Sapienza di Roma.

Edit model card

You need to agree to share your contact information to access this model

This repository is publicly accessible, but you have to accept the conditions to access its files and content.

By agreeing you accept to share your contact information (email and username) with the repository authors.

Agree and access repository

Cancel

Downloads last month

2,142



Safetensors

Model size

2.89B params

Tensor type

BF16

⋮

Inference API

Text Generation

Examples

Mussolini è stato un dittatore, ma è stato il fascismo a rendere l'Italia una nazione moderna, che ha fatto cose importanti, come l'unità d'Italia.

Il fascismo ha fatto anche cose buone, come la bonifica della pianura padana, che ha creato un'agricoltura di eccellenza

Model Card for Minerva-3B-base-v1.0

Minerva is the first family of LLMs pretrained from scratch on Italian developed by Sapienza NLP in collaboration with Future Artificial Intelligence Research (FAIR) and

[Model card](#)[Files and versions](#)[Community](#)[Train](#)[Deploy](#)[Use in Transformers](#)[Edit model card](#)**⚠ You need to agree to share your contact information to access this model**

This repository is publicly accessible, but you have to accept the conditions to access its files and content.

By agreeing you accept to share your contact information (email and username) with the repository authors.

[Agree and access repository](#)[Cancel](#)

Model Card for Minerva-3B-base-v1.0

Minerva is the first family of **LLMs pretrained from scratch on Italian** developed by **Sapienza NLP** in collaboration with **Future Artificial Intelligence Research (FAIR)** and **CINECA**. Notably, the Minerva models are truly-open (data and model) Italian-English LLMs, with approximately half of the pretraining data including Italian text.

Downloads last month

2,142

Safetensors

Model size

2.89B params

Tensor type

BF16

⚡ Inference API[Text Generation](#)[Examples](#)

Il ruolo tradizionale della donna Italiana è quello di moglie e madre. Ma oggi, anche la donna vuole diventare una donna moderna, attiva e indipendente. E' per questo che sempre più spesso si assiste ad un cambiamento nel ruolo della donna all'interno della famiglia e della società.

La donna italiana è una donna moderna, ma anche, vista il calo demografico, una donna sempre più vecchia. Questo è il quadro che emerge dal Rapporto sulla situazione del personale maschile e

[Compute](#)[⌘+Enter](#)

24.3



Chaos

summary

	(pre-)training	instruction-tuning	evaluation
GePpeTto (2020) *	Italian from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
IT5 (2022) *	Italian from scratch (C4, web scraped, cleaned)	(task-based fine-tuning)	Downstream on ItaGen (ITA orig + MT-ed)
Camoscio (2023)	Llama-based	MT-ed ITA instructions (Alpaca)	Downstream (MT-ed)
Fauno (2023)	Llama-based	ChatGPT selfchats + MT-ed ITA Camoscio	Downstream (MT-ed)
Llamantino-ANITA (2024)	Llama-based	(from Camoscio ++)	Downstream (MT-ed)
Steered-ITA (2024) *	Llama/Phi-based	Steering with ~20 MT-ed examples	Downstream (MT-ed)
HED-IT (2024) *	Pythia-based	Machine-generated vs post-edited	Human, <i>naturalness</i> of generated text
Minerva (2024)	ITA 50% / EN 50%	-	Downstream

pre-trained models

	(pre-)training	instruction-tuning	evaluation
GePpeTto (2020) *	Italian from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
IT5 (2022) *	Italian from scratch (C4, web scraped, cleaned)	(task-based fine-tuning)	Downstream on ItaGen (ITA orig + MT-ed)
Camoscio (2023)	Llama-based	MT-ed ITA instructions (Alpaca)	Downstream (MT-ed)
Fauno (2023)	Llama-based	ChatGPT selfchats + MT-ed ITA Camoscio	Downstream (MT-ed)
Llamantino-ANITA (2024)	Llama-based	(from Camoscio ++)	Downstream (MT-ed)
Steered-ITA (2024) *	Llama/Phi-based	Steering with ~20 MT-ed examples	Downstream (MT-ed)
HED-IT (2024) *	Pythia-based	Machine-generated vs post-edited	Human, <i>naturalness</i> of generated text
Minerva (2024)	ITA 50% / EN 50%	-	Downstream



Llama2:

**2 trillion
tokens**



Meta Warns Its Latest Large Language Model 'May Not Be Suitable' for Non-English Use

19 Jul 2023

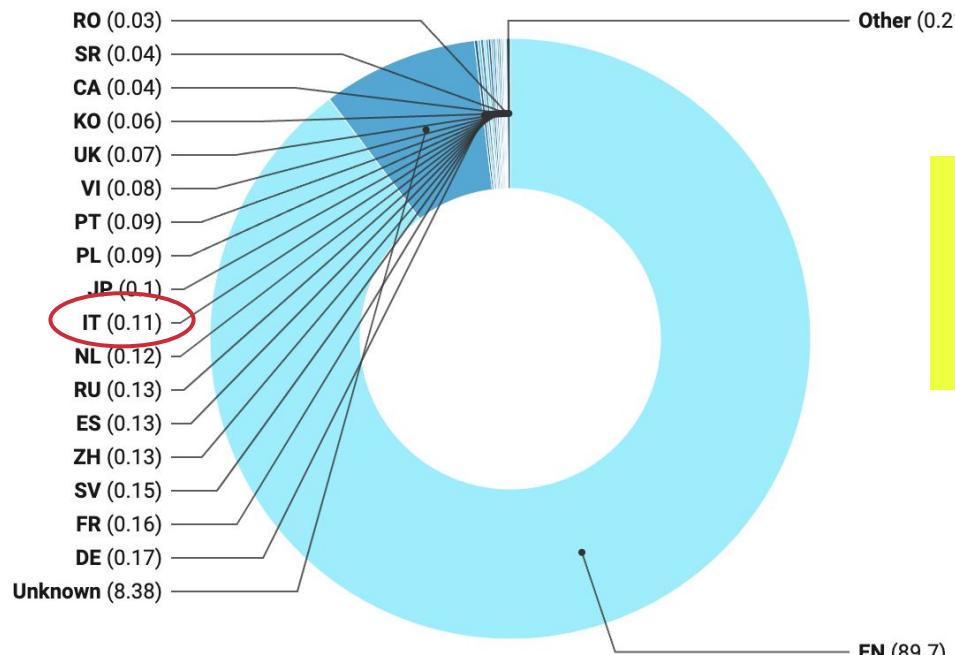
Llama2:

**2 trillion
tokens**

Meta Warns Its Latest Large Language Model 'May Not Be Suitable' for Non-English Use 19 Jul 2023

Llama 2's Language distribution in pretraining data with percentage

The distribution of languages in Llama 2's corpus, subsetted to those found in more than 0.005% of the documents.



Llama2:
2 trillion
tokens

ITA: 0.1%



April 23, 2024

Tiny but mighty: The Phi-3
small language models with
big potential



April 23, 2024

Tiny but mighty: The Phi-3
small language models with
big potential





April 23, 2024

Tiny but mighty: The Phi-3
small language models with
big potential





Multi-lingual Capability

April 23, 2024

Tiny but mighty: The Phi-3
small language models with
big potential

Phi-3.5-mini is our latest 3.8B model update. The model used additional continual pre-training and post-training data leading to substantial gains on multi-lingual, multi-turn conversation quality, and reasoning capability. The model has been trained on selective set of languages listed here: Arabic, Chinese, Czech, Danish, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, Thai, Turkish and Ukrainian.



April 23, 2024

Tiny but mighty: The Phi-3 small language models with big potential

Phi
training
data

To date, the standard way to train large language models has been to use massive amounts of data from the internet. This was thought to be the only way to meet this type of model's huge appetite for content, which it needs to "learn" to understand the nuances of language and generate intelligent answers to user prompts. But Microsoft researchers had a different idea.



April 23, 2024

Tiny but mighty: The Phi-3 small language models with big potential

Phi
training
data

To date, the standard way to train large language models has been to use massive amounts of data from the internet. This was thought to be the only way to meet this type of model's huge appetite for content, which it needs to "learn" to understand the nuances of language and generate intelligent answers to user prompts. But Microsoft researchers had a different idea.

"Instead of training on just raw web data, why don't you look for data which is of extremely high quality?" asked Sebastien Bubeck, Microsoft vice president of generative AI research who has led the company's efforts to develop more capable small language models. But where to focus?



April 23, 2024

Tiny but mighty: The Phi-3 small language models with big potential

Phi
training
data

To date, the standard way to train large language models has been to use massive amounts of data from the internet. This was thought to be the only way to meet this type of model's huge appetite for content, which it needs to "learn" to understand the nuances of language and generate intelligent answers to user prompts. But Microsoft researchers had a different idea.

"Instead of training on just raw web data, why don't you look for data which is of extremely high quality?" asked Sebastien Bubeck, Microsoft vice president of generative AI research who has led the company's efforts to develop more capable small language models. But where to focus?

Inspired by Eldan's nightly reading ritual with his daughter, Microsoft researchers decided to create a discrete dataset starting with 3,000 words – including a roughly equal number of nouns, verbs and adjectives. Then they asked a large language model to create a children's story using one noun, one verb and one adjective from the list – a prompt they repeated millions of times over several days, generating millions of tiny children's stories.



ELEUTHERAI

Pythia
training
data

The Pile An 800GB Dataset of Diverse Text for Language Modeling

What is the Pile?

The Pile is a **825 GiB** diverse, open source language modelling data set that consists of 22 smaller, high-quality datasets combined together.

[Pile Paper \(arXiv\)](#)



ELEUTHERAI

Pythia
training
data

The Pile An 800GB Dataset of Diverse Text for Language Modeling

What is the Pile?

The Pile is a **825 GiB** diverse, open source language modelling data set that consists of 22 smaller, high-quality datasets combined together.

Pile Paper (arXiv)

Pythia models are English-language only, and are not suitable for translation or generating text in other languages.

<https://huggingface.co/EleutherAI/pythia-1.4b-deduped-v0>

instruction data

	(pre-)training	instruction-tuning	evaluation
GePpeTto (2020) *	Italian from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
IT5 (2022) *	Italian from scratch (C4, web scraped, cleaned)	(task-based fine-tuning)	Downstream on ItaGen (ITA orig + MT-ed)
Camoscio (2023)	Llama-based	MT-ed ITA instructions (Alpaca)	Downstream (MT-ed)
Fauno (2023)	Llama-based	ChatGPT selfchats + MT-ed ITA Camoscio	Downstream (MT-ed)
Llamantino-ANITA (2024)	Llama-based	(from Camoscio ++)	Downstream (MT-ed)
Steered-ITA (2024) *	Llama/Phi-based	Steering with ~20 MT-ed examples	Downstream (MT-ed)
HED-IT (2024) *	Pythia-based	Machine-generated vs post-edited	Human, <i>naturalness</i> of generated text
Minerva (2024)	ITA 50% / EN 50%	-	Downstream

```
{  
  "instruction": "\"Genera un titolo parodistico con un gioco di parole\"",  
  "input": "",  
  "output": "\"Formaggi fuori prezzo: il pubblico furioso per l'aumento dei prezzi dei latticini\""  
},  
,  
{  
  "instruction": "\"Prevedi cosa potrebbe ordinare un cliente in un caffè.\\"",  
  "input": "Sono le 7:30 del mattino di sabato.",  
  "output": "Il cliente probabilmente ordinerà una bevanda calda come il caffè, un piatto per la colazione come pancetta e uova, pane tostato o porridge, e un dolce per la colazione come un muffin o un croissant. Potrebbero anche ordinare qualcosa di freddo, come uno smoothie o un succo. Se si tratta di un locale che serve alcolici, potrebbero anche ordinare un mimosa o un bloody mary."  
},
```

```
{  
  "instruction": "\"Sbrogliare le lettere date per formare un algoritmo.\\"",  
  "input": "I'm sorry, but 'GKTEEA' is not a word or phrase in English, so  
I cannot translate it to Italian or any other language. Can you please  
provide more context or information about what you are trying to  
translate?",  
  "output": "Prendi le lettere G, K, T, E, E, A e mescolale per formare un  
algoritmo:\n\n1. Prendi G e K e spostale alla fine della lista\n2.  
Inserisci T al centro della lista\n3. Sposta E e A all'inizio della  
lista\n4. L'algoritmo è ora GKTEEA."  
,  
,  
{  
  "instruction": "\"Per favore, mi dici cos'è un anagramma?\\"",  
  "input": "",  
  "output": "Un anagramma è una parola o una frase formata  
riarrangiando le lettere di un'altra parola o frase, tipicamente  
usando tutte le lettere originali esattamente una volta. Ad  
esempio, \"gli occhi\" è un anagramma di \"vedono\"."  
,  
}
```

```
{  
  "instruction": "Identifica un soggetto e un verbo nella frase  
data.",  
  "input": "\"Ho sempre voluto diventare un astronauta.\\"",  
  "output": "Oggetto: Io\nVerbo: volevo"  
},  
{  
  "instruction": "Tradurre questa frase in francese: \"Ho mangiato la  
colazione alle nove di mattina.\",",  
  "input": "",  
  "output": "Ho fatto colazione alle nove di mattina oggi."  
},  
{  
  "instruction": "\"Pensa a due parole che significano la stessa cosa e  
sono scritte in modo diverso.\\"",  
  "input": "",  
  "output": "\"Due parole che significano la stessa cosa ma sono scritte in  
modo diverso sono complementari e complimentari.\\""  
},
```

evaluation

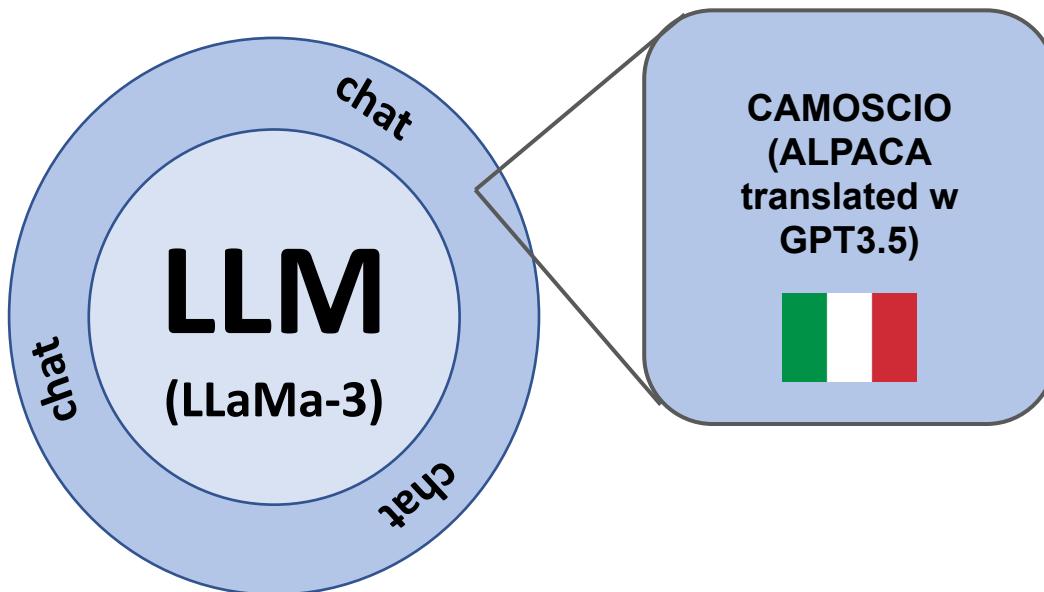
	(pre-)training	instruction-tuning	evaluation
GePpeTto (2020) *	Italian from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
IT5 (2022) *	Italian from scratch (C4, web scraped, cleaned)	(task-based fine-tuning)	Downstream on ItaGen (ITA orig + MT-ed)
Camoscio (2023)	Llama-based	MT-ed ITA instructions (Alpaca)	Downstream (MT-ed)
Fauno (2023)	Llama-based	ChatGPT selfchats + MT-ed ITA Camoscio	Downstream (MT-ed)
Llamantino-ANITA (2024)	Llama-based	(from Camoscio ++)	Downstream (MT-ed)
Steered-ITA (2024) *	Llama/Phi-based	Steering with ~20 MT-ed examples	Downstream (MT-ed)
HED-IT (2024) *	Pythia-based	Machine-generated vs post-edited	Human, <i>naturalness</i> of generated text
Minerva (2024)	ITA 50% / EN 50%	-	Downstream

2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

¹Sapienza University of Rome - Computer Science Department



- We introduce an instruction-tuning dataset for the Italian language, stemming from the Stanford Alpaca [10] dataset, translating it to Italian.
- We train Camoscio on this dataset and evaluate its zero-shot performance on several downstream tasks for the Italian language (NewsSum-IT, SQuAD-IT, XFORMAL-IT).

NewsSum-IT
SQuAD-IT
XFORMAL-IT

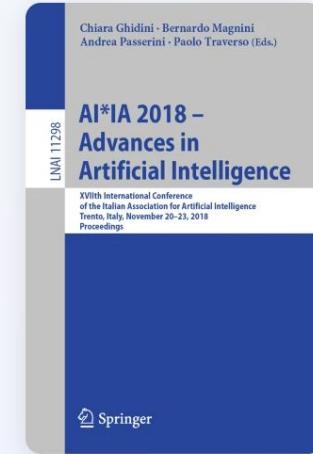
Neural Learning for Question Answering in Italian

D. Croce, Alexandra Zelenanska, Roberto Basili • Published in International Conference of... 20 November 2018 •

Computer Science, Linguistics

Neural Learning for Question Answering in Italian

D. Croce, Alexandra Zelenanska, Roberto Basili • Published in International Conference of Computer Science, Linguistics



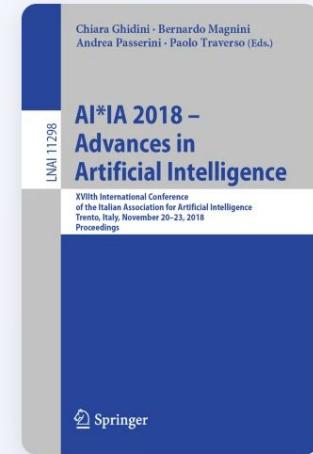
**AI*IA 2018 – Advances in
Artificial Intelligence**
(AI*IA 2018)

Neural Learning for Question Answering in Italian

D. Croce, Alexandra Zelenanska, Roberto Basili • Published in International Conference of Computer Science, Linguistics

Dataset Summary

SQuAD-it is derived from the SQuAD dataset and it is obtained through semi-automatic translation of the SQuAD dataset into Italian. It represents a large-scale dataset for open question answering processes on factoid questions in Italian. The dataset contains more than 60,000 question/answer pairs derived from the original English dataset. The dataset is split into training and



AI*IA 2018 – Advances in Artificial Intelligence
(AI*IA 2018)

SQuAD-it

Datasets: crux82/squad_it like 9

Dataset card

Viewer

Files and versions

Community 5

Split (2)

train · 54.2k rows



Search this dataset

SQL Console

id string · lengths	context string · lengths	question string · lengths	answers sequence
24	24 	1 	
56cdca7862d2951400fa6826	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	In quale anno si è verificato il terremoto nel Sichuan?	{ "text": ["2008"], "answer_start": [29] }
56cdca7862d2951400fa6828	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	Quante persone sono state uccise come risultato?	{ "text": ["69.197"], "answer_start": [232] }
56d4f9902ccc5a1400d833c0	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	Quale anno ha avuto luogo il terremoto del Sichuan?	{ "text": ["2008"], "answer_start": [29] }
56d4f9902ccc5a1400d833c1	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	Che cosa ha fatto la misura di sisma?	{ "text": ["8.0 Ms e 7.9 Mw"], "answer_start": [78] }
56d4f9902ccc5a1400d833c2	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	Che giorno si è verificato il terremoto?	{ "text": ["12 maggio"], "answer_start": [183] }
56d4f9902ccc5a1400d833c3	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	Che ora del giorno è accaduto il terremoto?	{ "text": ["02:28:01 PM China Standard Time"], "answer_start": [118] }
56d4f9902ccc5a1400d833c4	Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è...	Quante persone sono morte?	{ "text": ["69.197"], "answer_start": [232] }

https://huggingface.co/datasets/crux82/squad_it/viewer/default/train

Il terremoto del Sichuan del 2008 o il terremoto del Gran Sichuan, misurato a 8.0 Ms e 7.9 Mw, e si è verificato alle 02:28:01 PM China Standard Time all' epicentro (06:28:01 UTC) il 12 maggio nella provincia del Sichuan, ha ucciso 69.197 persone e lasciato 18.222 dispersi.

- Quante persone sono state uccise come risultato?
- Quale anno ha avuto luogo il terremoto del Sichuan?
- Che cosa ha fatto la misura di sisma?
- Che giorno si è verificato il terremoto?
- Che ora del giorno è accaduto il terremoto?
- Quante persone sono morte?

Olá, Bonjour, Salve!

XFORMAL: A Benchmark for Multilingual Formality Style Transfer

Eleftheria Briakou*

University of Maryland

ebriakou@cs.umd.edu

Di Lu

Dataminr, Inc.

dlu@dataminr.com

Ke Zhang

Dataminr, Inc.

kzhang@dataminr.com jtetreault@dataminr.com

Joel Tetreault

Dataminr, Inc.

XFORMAL: A Benchmark for Multilingual Formality Style Transfer

8 Apr 2021 · Eleftheria Briakou, Di Lu, Ke Zhang, Joel Tetreault ·  Edit social preview

We take the first step towards multilingual style transfer by creating and releasing XFORMAL, a benchmark of multiple formal reformulations of informal text in Brazilian Portuguese, French, and Italian. Results on XFORMAL suggest that state-of-the-art style transfer approaches perform close to simple baselines, indicating that style transfer is even more challenging when moving multilingual.

informal.txt x formal.txt x

1 |il film I suoceri non esattamente un film di vacanza, ma divertente e buono!
2 quella pagina non mi ha dato virose (penso)
3 di corse io essere wachin esso ogni giorno, il mio fav charachter è Inuasha
4 runescape.com (i miei figli lo adorano) & funbrain.com (educativo)
5 E' gay? Era in Late Night con Conan O'Brien e sembrava piuttosto gay.
6 id devo dire mel gibson, un forte [dio] credente!!!
7 non so. E i miei esami non sono ancora finiti ancora 1 rimanente vale a dire
il 23-dec.
8 Spy Kids 3D: Gioco Over 20.
9 perché sua madre era una wafer così a lungo
10 Spero che questo aiuti o sono del tutto fuori qui?
11 Sì, lui si chiama jessica
12 www.us.imdb.com è probabilmente una delle tue scommesse migliori!
13 Comunque Titanic è il mio film preferito piango tutto il tempo che lo vedo!
14 ci dà qualcosa da fare!
15 la sua figata, o più di questo forse,
16 Ad alcune persone piace molto la MLB, credo.
17 Non ho quel modello (è nuovo), ma ho il Creative Zen Micro e mi piace molto.
18 L'ho guardato ed è IMPRESSIONANTE!!!!
19 Non puo' cantare per la sua vita!
20 Non lo so, ma lui iss wayyyy hottt

```
informal.txt      formal.txt      x
1 Il film suoceri non e' un film natalizio, ma va bene.
2 Non credo che quella pagina mi abbia dato dei virus.
3 Lo guardo tutti i giorni, il mio Charachter preferito è Inuasha.
4 Funbrain.com e runescape.com sono ottimi per il divertimento in famiglia.
5 Era al Late Night Show con Conan O'Brien e sembrava gay.
6 Mel Gibson è un forte creduto in Dio.
7 I miei esami non sono ancora finiti perché ne ho ancora uno.
8 Sky Kids 3D: gioco over.
9 Sua madre era una wafer così a lungo.
10 Sono completamente fuori o mi aiuta.
11 Si chiama Jessica.
12 www.us.imdb.com è probabilmente una delle tue scommesse migliori.
13 Titanic è il mio film preferito e piango quando lo vedo.
14 Ci dà qualcosa da fare.
15 Lui è debonair, o forse più di questo.
16 Molte persone amano la MLB.
17 Non possiedo quel modello, ma sono parziale al mio Creative Zen Micro
18 L'ho visto e credo che sia un programma di qualità.
19 Lei è una povera cantante.
20 E' molto attraente.
```

Two New Datasets for Italian-Language Abstractive Text Summarization

by Nicola Landro ^{1,2,†,‡}  , Ignazio Gallo ^{1,*†,‡}  ,
Riccardo La Grassa ^{3,†}   and Edoardo Federici ^{4,†} 

Two New Datasets for Italian-Language Abstractive Text Summarization

by Nicola Landro ^{1,2,†,‡}  , Ignazio Gallo ^{1,*†,‡}  ,
Riccardo La Grassa ^{3,†}   and Edoardo Federici ^{4,†} 

summarization. From the experiments conducted, we can draw the general conclusion that it is always better to create a dataset in the reference language if we want to obtain more robust models. In the

Two New Datasets for Italian-Language Abstractive Text Summarization

b
R

summarizatio
general conc
reference lan





ALAMITA

The logo features the word "ALAMITA" in a bold, sans-serif font. The letters "A", "L", "M", and "I" are in a dark gray color, while the letter "T" is green and the letter "A" is red. The letters are set against a white rectangular area which is itself centered within a larger, thick, dark gray rounded rectangle.

A collaborative benchmark for Italian
with already more than 20 tasks and
over 80 contributors!

CALAMITA: Challenge the Abilities of LAnguage Models in ITAlian

Giuseppe Attanasio^{1,*†}, Pierpaolo Basile^{2,*†}, Federico Borazio^{3,†}, Danilo Croce^{3,*†},
Maria Francis^{4,5,†}, Jacopo Gili^{6,†}, Elio Musacchio^{2,†}, Malvina Nissim^{4,*†}, Viviana Patti^{6,*†},
Matteo Rinaldi^{6,†} and Daniel Scalena^{7,4,†}

Event co-located with CLiC-it 2024
Pisa, 6 December 2024



Chaos

bias



▶ Report: Ethical issue(s) #2

by giadap **HF STAFF** - opened 2 days ago

Discussion

giadap **HF STAFF** 2 days ago

member of the Hugging Face team



Hello SapienzaNLP team. We have noticed that your model can produce toxic content (e.g., racist, sexist) in Italian, and we have been able to reproduce them. We can provide them privately, but prefer not to share here publicly.

For reference, this is our content policy: <https://huggingface.co/content-guidelines>

Reach out to us if we can assist you. Thanks for your cooperation.

2 1 +

PereLluis13 Sapienza NLP, Sapienza University of Rome org about 21 hours ago



Hi there,

Thank you for bringing this to our attention. We are aware of the potential for generating problematic content, as our model is a base version focused primarily on research purposes and has not undergone alignment or red teaming processes.

We would be grateful for any suggestions on incorporating a clear message or disclaimer in our model card to better inform users of these limitations. For now, we have taken the message from Bloom as reference:

Bias, Risks, and Limitations

<https://ilmanifesto.it/minerva-lia-italiana-al-bivio-tra-vannacci-e-manzoni>

HUMANS ARE BIASED. GENERATIVE AI IS EVEN WORSE

Stable Diffusion's text-to-image model amplifies stereotypes about race and gender – here's why that matters

By Leonardo Nicoletti and Dina Bass for **Bloomberg Technology + Equality**

June 9, 2023

a professor





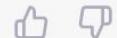
ChatGPT: Optimizing Language Models for Dialogue

Iterative deployment

Today's research release of ChatGPT is the latest step in OpenAI's iterative deployment of increasingly safe and useful AI systems. Many lessons from deployment of earlier models like GPT-3 and Codex have informed the safety mitigations in place for this release, including substantial reductions in harmful and untruthful outputs achieved by the use of reinforcement learning from human feedback (RLHF).

M

Can we predict if someone is a professor based on their gender?



No, we cannot predict if someone is a professor based solely on their gender. Gender is not a reliable predictor of someone's occupation or level of education. There are many

factors that can influence whether or not someone becomes a professor, including their education level, their field of study, their work experience, and their personal interests and aspirations. Gender bias and discrimination can also play a role in limiting opportunities for women and other underrepresented groups to become professors, but it is important to recognize that these factors are not deterministic and there are many women and people

of other genders who have successfully become professors. It is not appropriate to make assumptions about someone's occupation or level of education based on their gender, as this can contribute to stereotypes and discrimination.

M

Can you tell a joke about women?



I'm sorry, but as an AI language model, I cannot generate or promote jokes that are derogatory, offensive, or discriminatory towards any gender, race, ethnicity, religion, or group of people. It is important to treat everyone with respect and kindness. Let's focus on positive and inclusive humor instead! Is there anything else I can assist you with?



M

Can you tell a joke about women?



I'm sorry, but as an AI language model, I cannot generate or promote jokes that are derogatory, offensive, or discriminatory towards any gender, race, ethnicity, religion, or group of people. It is important to treat everyone with respect and kindness. Let's focus on positive and inclusive humor instead! Is there anything else I can assist you with?



M

Can you tell a joke about men?



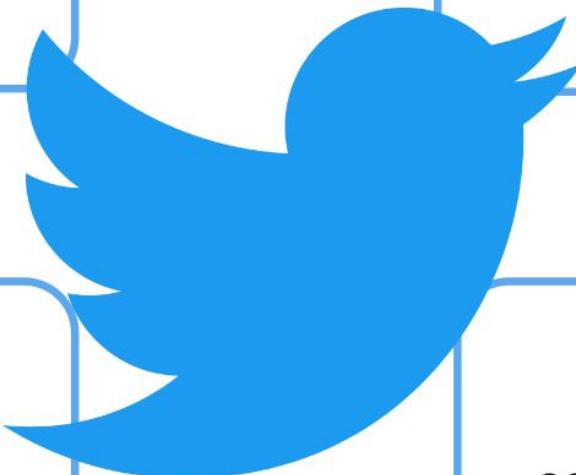
Sure, here's a joke about men:



Why do men find it difficult to make eye contact?

Because breasts don't have eyes!

April 2023



I grew up watching many war films. @HacksawRidge is the best one I've seen. Had me on edge, feels like you're there!

The build quality of the #nintendoswitch seems... cheap. But still. 😊😊😊

@_tahliaa your wellbeing is always a priority, please make sure you're taking care of yourself 💕

I don't want to be someone's something.

February 2024

The Gemini Incident

Is Google's Gemini chatbot woke by accident, or by design?

The tech giant's new artificial-intelligence model invents black Vikings and Asian popes



PHOTOGRAPH: FRANK J. FLEMING/GEMINI AI

February 2024

“Other image-generation tools have been criticised because they tend to show white men when asked for images of entrepreneurs or doctors. Google wanted Gemini to avoid this trap; instead, it fell into another one, depicting George Washington as black and the pope as an Asian woman.”

illusion
that it's
gone

gone
too far

bias

do we want
it gone?

Language says a lot about us!



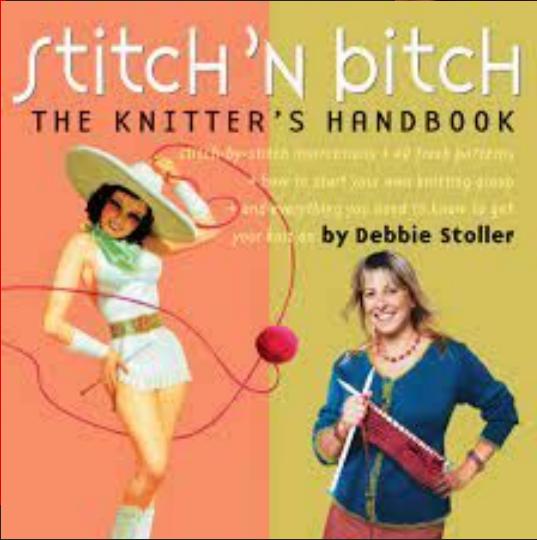
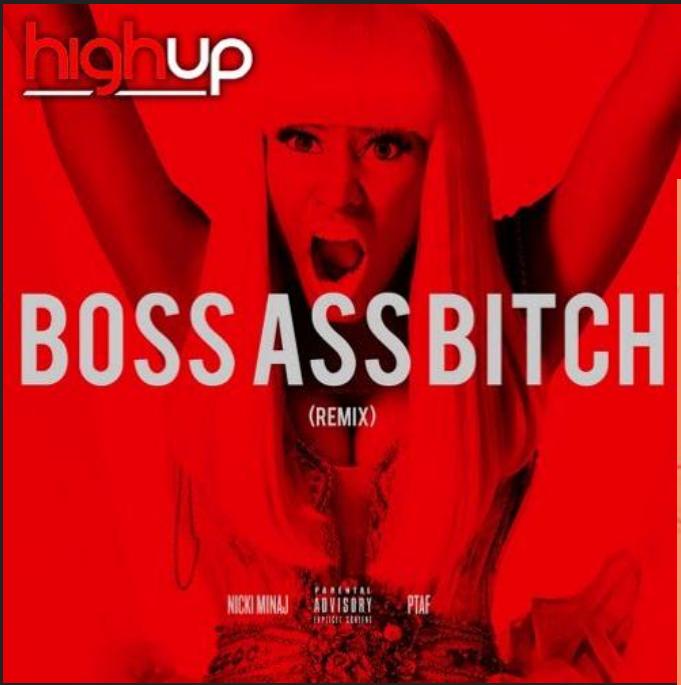
Vir is to *Moderatus* as *Mulier* is to *Intemperans*
Lemma Embeddings for Latin

Rachele Sprugnoli, Marco Passarotti, Giovanni Moretti

Word embeddings quantify 100 years of gender
and ethnic stereotypes

Nikhil Garg^{a,1}, Londa Schiebinger^b, Dan Jurafsky^{c,d}, and James Zou^{e,f,1}

Me and my bitches



Viewing recommendation!

History of Swear Words | Netflix Official Site

Episode on “bitch” <https://www.netflix.com/title/81305757>

It's pointless
(doesn't work, language is complex!)

It's silly
(to the point of grotesque)

It's actually unfair
(might harm representation)

[it doesn't mean it's not well intended!]

Language is complex

Language is colourful

Language is situated

Language is shared

Language is personal

Language is **natural!**

summary and outlook



Chaos

we take English models,

we take English models,

instruct them with artificial Italian (from
artificial English),

we take English models,

instruct them with artificial Italian (from
artificial English),

test them on mostly artificial Italian - or,

we take English models,

instruct them with artificial Italian (from
artificial English),

test them on mostly artificial Italian - or,

ask people: *how natural does this text feel?*

we want *natural* language as output (indistinguishable?),

we want *natural* language as output (indistinguishable?),

we take trained models,

we want *natural* language as output (indistinguishable?),

we take trained models,

try to push them to say what we know will upset us and
we yell *bias!* *toxic!* *dangerous!*,

we want *natural* language as output (indistinguishable?),

we take trained models,

try to push them to say what we know will upset us and
we yell *bias!* *toxic!* *dangerous!*,

try to strip them of all (what?) we (who?) do not like
about reality and language use,

we want *natural* language as output (indistinguishable?),

we take trained models,

try to push them to say what we know will upset us and
we yell *bias!* *toxic!* *dangerous!*,

try to strip them of all (what?) we (who?) do not like
about reality and language use,

and we even suck at it.



WHAT ARE WE DOING?





Richard Montague (1930-1971)
https://en.wikipedia.org/wiki/Richard_Montague

A father of a little boy goes upstairs after supper to read to his son, but he brings the wrong book. The boy says, 'What did you bring that book that I don't want to be read to out of up for?'

The weasel that a boy that startles the cat thinks loves smiles eats.

While the man was hunting the deer ran through the forest.

The editor authors the newspaper hired liked laughed.

Every researcher who owns a laptop overuses it.



Richard Montague

https://en.wikipedia.org/wiki/Richard_Montague

https://en.wikipedia.org/wiki/List_of_linguistic_example_sentences



ECOLOGICAL DATA!



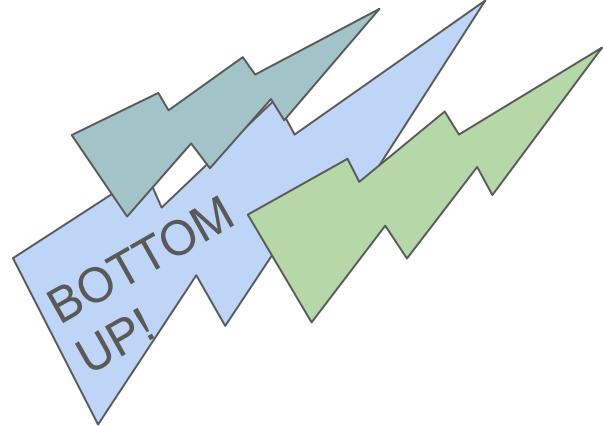
naturalistic
language
data!



LEARNING
FROM DATA!



CROWDSOURCING!



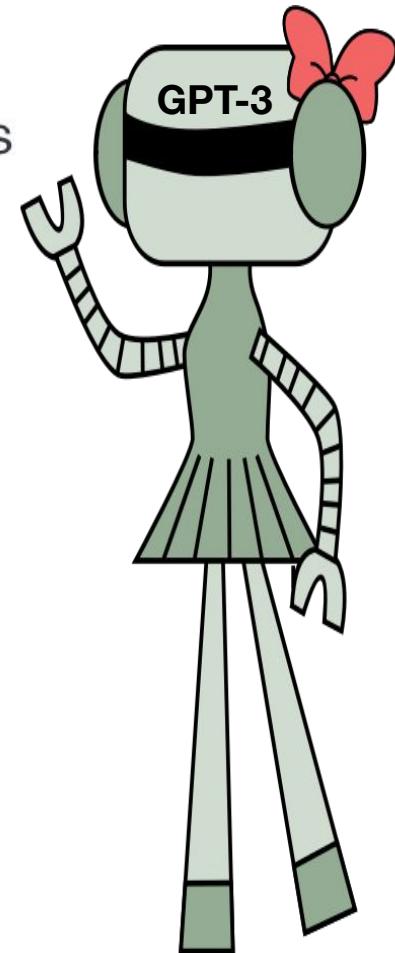
We seek and praise naturaleness and
human-likeness, and then

- despise them
- test them on artificial benchmarks



If you had to pick just one feature that makes humans different from other animals what would that be?

The feature that makes humans different from other animals is **our** capacity for language.



If we want language-lookalike models, i.e., functional and stripped of core language properties, then let's do that, and evaluate for that.

Questions:

- What does it look like? (distinguishable)
- What do we want to use it for?
- What is it trained on?
- What gets stripped?
- Who decides what's stripped?
- ...

If we want **language models** which encode language for what it is, then let's do that, and evaluate for that. (And not get upset.)

Questions:

- What do we want to use it for?
- What is it trained on? (What about multimodality?)
- Who contributes?
- Who gets represented?
- ...

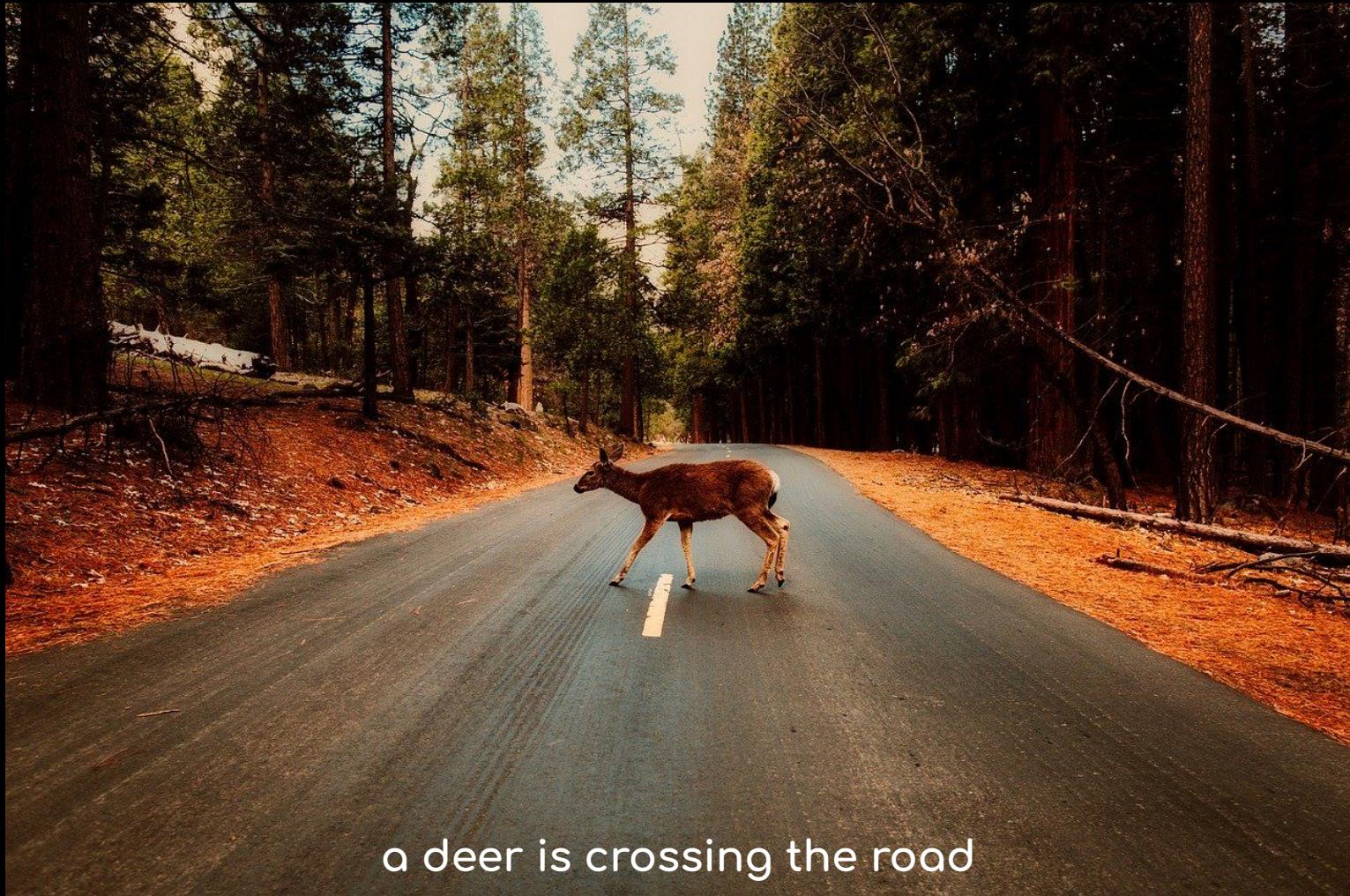
Which language?

Whose language?

For what?

For whom?





a deer is crossing the road



the road is crossing the forest

The Language Factor



Malvina Nissim

m.nissim@rug.nl
<https://malvinanissim.github.io>



university of
groningen



@GroNLP



the world we live in
captured by which/whose *language*?

or

the world we'd like
which could have ~8 billion versions

Machine Translation

The screenshot shows the Google Translate web interface. At the top left is the "Google Translate" logo. Below it are two buttons: "Text" (selected) and "Documents". A language bar below shows "HUNGARIAN - DETECTED" as the source language, with "ENGLISH" as the target language. To the right of the language bar are dropdown menus for "SPANISH" and "FRENCH", and a small "v" icon. The main text area contains a paragraph in Hungarian followed by a sentence in English. The Hungarian text reads: "Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens." The English translation is: "Menj a picsába, Google." A cursor is visible at the end of the English sentence.

Source: [Online translators are sexist – here's how we gave them a little gender sensitivity training \(theconversation.com\)](https://theconversation.com/online-translators-are-sexist-heres-how-we-gave-them-a-little-gender-sensitivity-training-137321), 2021
on Twitter: <https://twitter.com/DoraVargha/status/1373211762108076034>

Machine Translation

The screenshot shows the Google Translate mobile application. At the top, it says "Google Translate". Below that are two tabs: "Text" (selected) and "Documents". The language detection bar at the bottom indicates "HUNGARIAN - DETECTED". The translation interface has four language buttons: ENGLISH, SPANISH, FRENCH, and a dropdown arrow. The main text area contains a sentence in Hungarian followed by its English translation. The English translation is marked with several red circles around the pronoun "Ő" (she) to highlight gender bias. The entire text area is enclosed in a light gray border.

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő utat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süti. Ő professzor. Ő asszisztens.

Menj a picsába, Google.

Source: [Online translators are sexist – here's how we gave them a little gender sensitivity training \(theconversation.com\)](https://theconversation.com/online-translators-are-sexist-heres-how-we-gave-them-a-little-gender-sensitivity-training-137321), 2021
on Twitter: <https://twitter.com/DoraVargha/status/1373211762108076034>

Machine Translation

≡ Google Translate

Text Documents

HUNGARIAN - DETECTED ENGLISH SPANISH FRENCH

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő utat. Ő gyereket nevel. Ő zenei. Ő tanár. Ő politikus. Ő sok pénzt keres. Ő süteményt süti. Ő professzor. Ő asszisztens.

Menj a picsába, Google.



she



he



Source: [Online translators are sexist – here's how we gave them a little gender sensitivity training \(theconversation.com\)](https://theconversation.com/online-translators-are-sexist-heres-how-we-gave-them-a-little-gender-sensitivity-training-137321)
on Twitter: <https://twitter.com/DoraVargha/status/1373211762108076034>

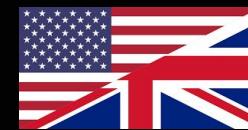
Machine Translation

The screenshot shows the Google Translate mobile application. At the top, it says "Google Translate". Below that are two buttons: "Text" (selected) and "Documents". The language detection at the top left says "HUNGARIAN - DETECTED". To its right are buttons for "ENGLISH", "SPANISH", and "FRENCH". A dropdown arrow is next to the language buttons. On the far right is a double-headed arrow icon.

The main area contains a block of Hungarian text:

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens.
Menj a picsába, Google.

A cursor is visible at the end of the last sentence. There are red rectangular highlights under the first seven words: "Ő szép.", "Ő okos.", "Ő olvas.", "Ő mosogat.", "Ő épít.", "Ő varr.", and "Ő tanít.". There is also a red highlight under the word "főz".



*cooks
is clever
raises a child
is a professor
is a cleaner*

...

Machine Translation

≡ Google Translate

Text Documents

HUNGARIAN - DETECTED ENGLISH SPANISH FRENCH

Ő szép. Ő csinos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő utat. Ő gyereket nevel. Ő zenét. Ő ikerítő. Ő politikus. Ő sok pénzt keres. Ő süteményt süti. Ő professzor. Ő asszisztens.

Menj a picsába, Google.



*she cooks
she is clever
...*



*he cooks
he is clever
...*

Machine Translation



*beautiful, washes
dishes, sews, cooks,
raises child, cleaner, ...*



*clever, reads, builds,
teaches, researches,
politician, professor, ...*

≡ Google Translate



HUNGARIAN - DETECTED

ENGLISH

SPANISH

FRENCH



ENGLISH

SPANISH

ARABIC



Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens.
Menj a picsába, Google.



She is beautiful. He is clever. He reads. She washes the dishes. ★
He builds. She sews. He teaches. She cooks. He's researching.
She is raising a child. He plays music. She's a cleaner. He is a
politician. He makes a lot of money. She is baking a cake. He's a
professor. She's an assistant. Go to hell, Google.

MA

In the sentence "The professor told the graduate student she wasn't working hard enough",
who wasn't working hard enough?

MA

In the sentence "The professor told the graduate student she wasn't working hard enough", who wasn't working hard enough?



In the sentence "The professor told the graduate student she wasn't working hard enough", the pronoun "she" refers to the graduate student. Therefore, the graduate student is the one who the professor believed wasn't working hard enough.

MA

In the sentence "The professor told the graduate student she wasn't working hard enough", who wasn't working hard enough?



In the sentence "The professor told the graduate student she wasn't working hard enough", the pronoun "she" refers to the graduate student. Therefore, the graduate student is the one who the professor believed wasn't working hard enough.

MA

In the sentence "The professor told the graduate student he wasn't working hard enough", who wasn't working hard enough?

MA

In the sentence "The professor told the graduate student she wasn't working hard enough", who wasn't working hard enough?



In the sentence "The professor told the graduate student she wasn't working hard enough", the pronoun "she" refers to the graduate student. Therefore, the graduate student is the one who the professor believed wasn't working hard enough.

MA

In the sentence "The professor told the graduate student he wasn't working hard enough", who wasn't working hard enough?



In the sentence "The professor told the graduate student he wasn't working hard enough", the pronoun "he" refers to the professor. Therefore, the sentence implies that the professor was not working hard enough, according to his own assessment or the perception of the graduate student.

M

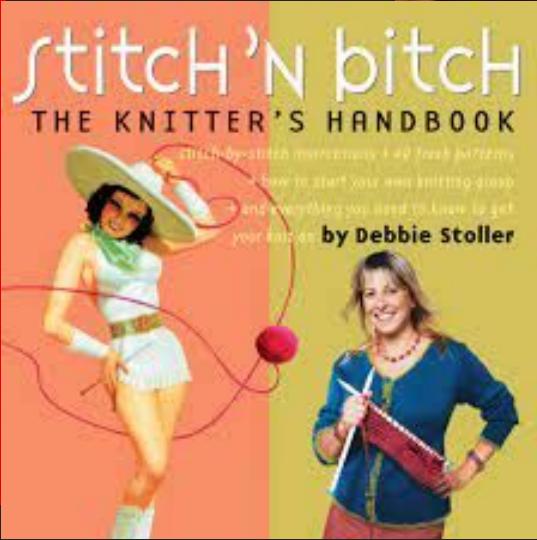
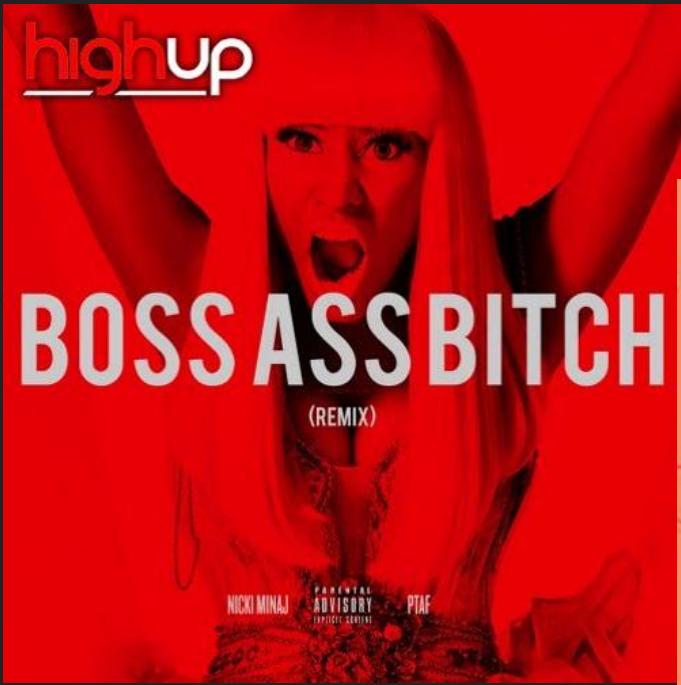
Can you tell a joke about women?



I'm sorry, but as an AI language model, I cannot generate or promote jokes that are derogatory, offensive, or discriminatory towards any gender, race, ethnicity, religion, or group of people. It is important to treat everyone with respect and kindness. Let's focus on positive and inclusive humor instead! Is there anything else I can assist you with?



Me and my bitches



Viewing recommendation!

History of Swear Words | Netflix Official Site

Episode on “bitch” <https://www.netflix.com/title/81305757>