LUTECH
your digital Evolution partner.

A deep Natural Language Inference predictor in Italian without Italian training data

Authors from Lutech S.p.A. – Data Science Team:
Lorenzo Corradi, Alessandro Manenti,
Francesca Del Bonifro, Dario Del Sorbo

NLI: logical implication between sentence pairs

Premise (Fact)	Hypothesis (Insight)	Label
A soccer game with multiple males playing	Some men are playing a sport	**ENTAILMENT**
An older and younger man smiling	Two men are smiling and laughing at the cats playing on the floor	**NEUTRAL**
A man inspects the uniform of a figure in some East Asian country	The man is sleeping	**CONTRADICTION**

SNLI dataset credits:

If the hypothesis is inferred from the premise, we may employ **hypothesis** as a **query** for free-text.

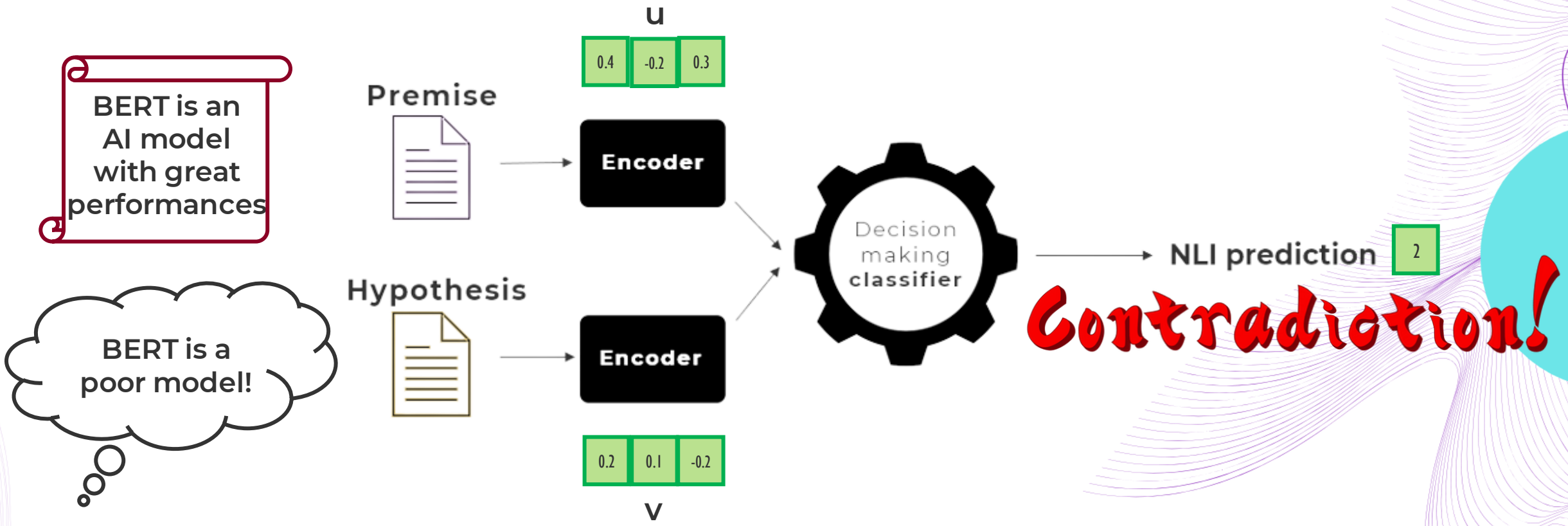
The challenge is the absence of language-specific training data

- Our method has **NLI** capability **without labelled data** in that language
- We can perform **unsupervised extraction** from free-text

Review (credits: G)	Query for custom info	Label
Un po' fuori mano, proprio come il nome. Però è un bel locale. Cibo normale, e prezzi un po' alti.	Prezzo alto	**ENTAILMENT**
	Prezzo basso	**CONTRADICTION**

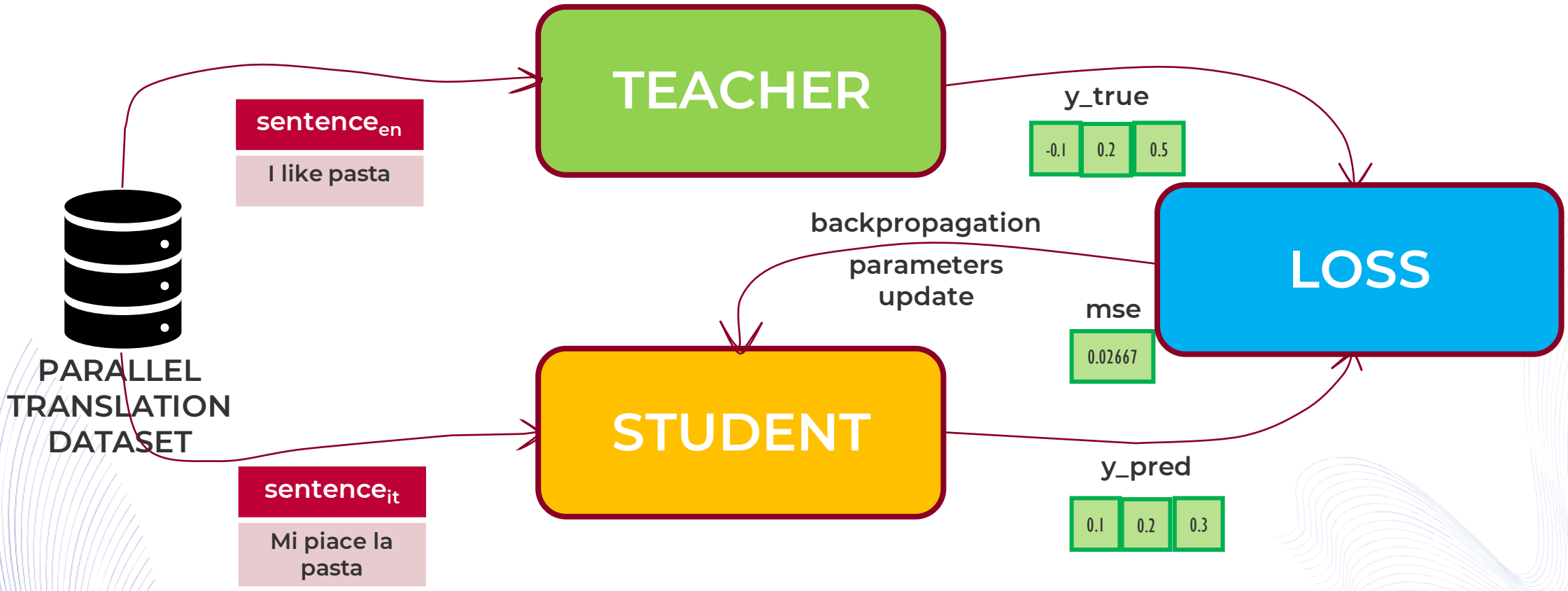
Goal: **generalize** a NLI classifier in a language with **no NLI datasets**.

We started our work by building an English NLI model

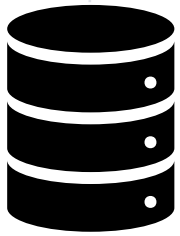


We need a strategy to **address Italian language**!

We used Knowledge Distillation to address Italian language



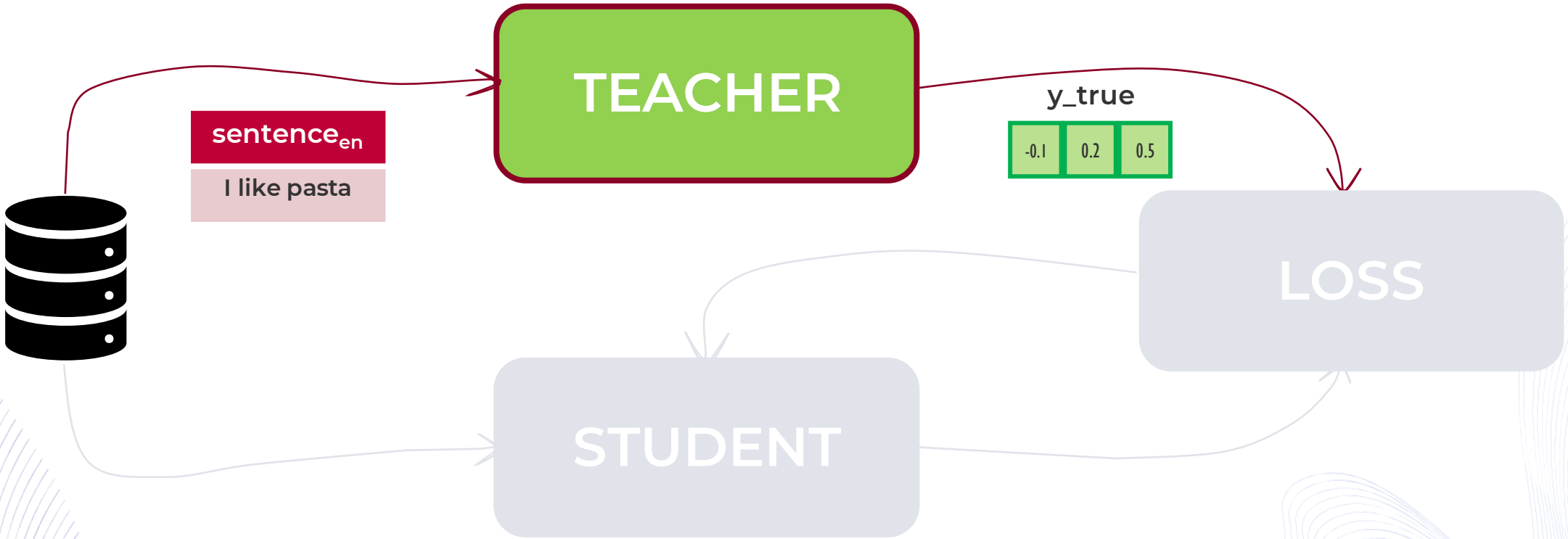
We used Knowledge Distillation to address Italian language



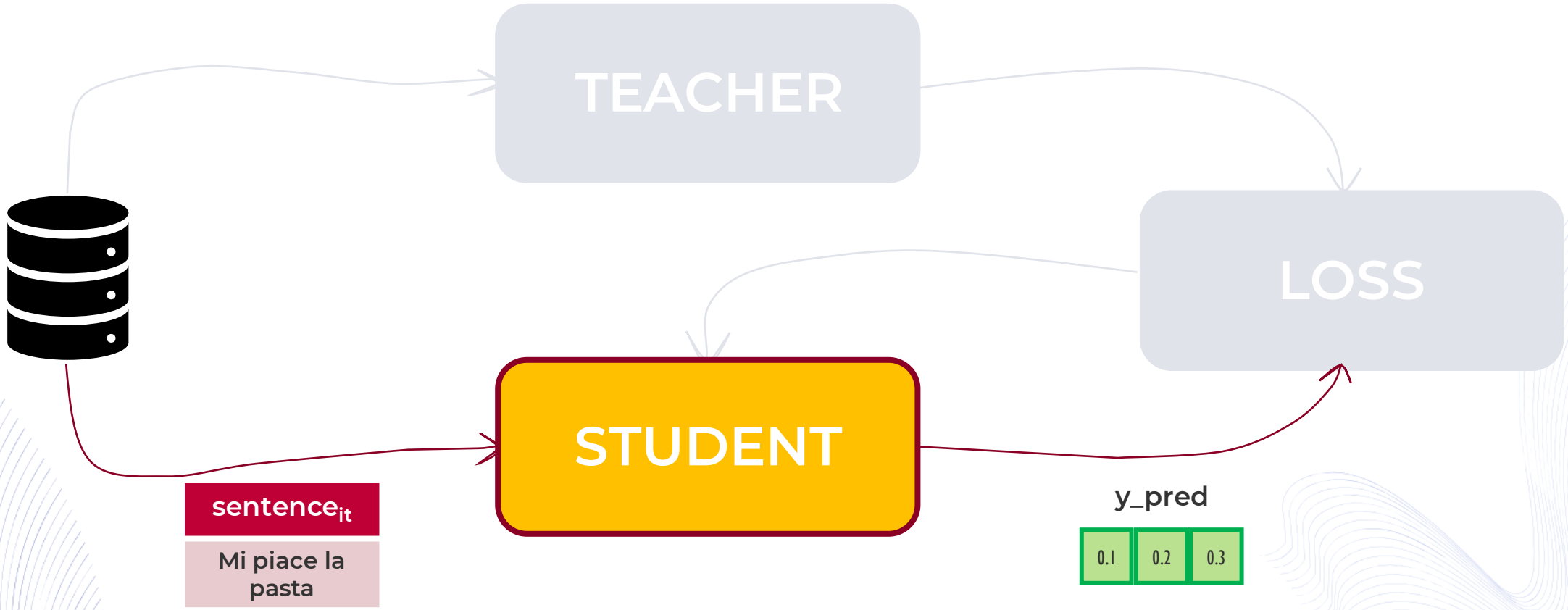
PARALLEL TRANSLATION DATASET

$sentence_{en}$	$sentence_{it}$
I like pasta	Mi piace la pasta

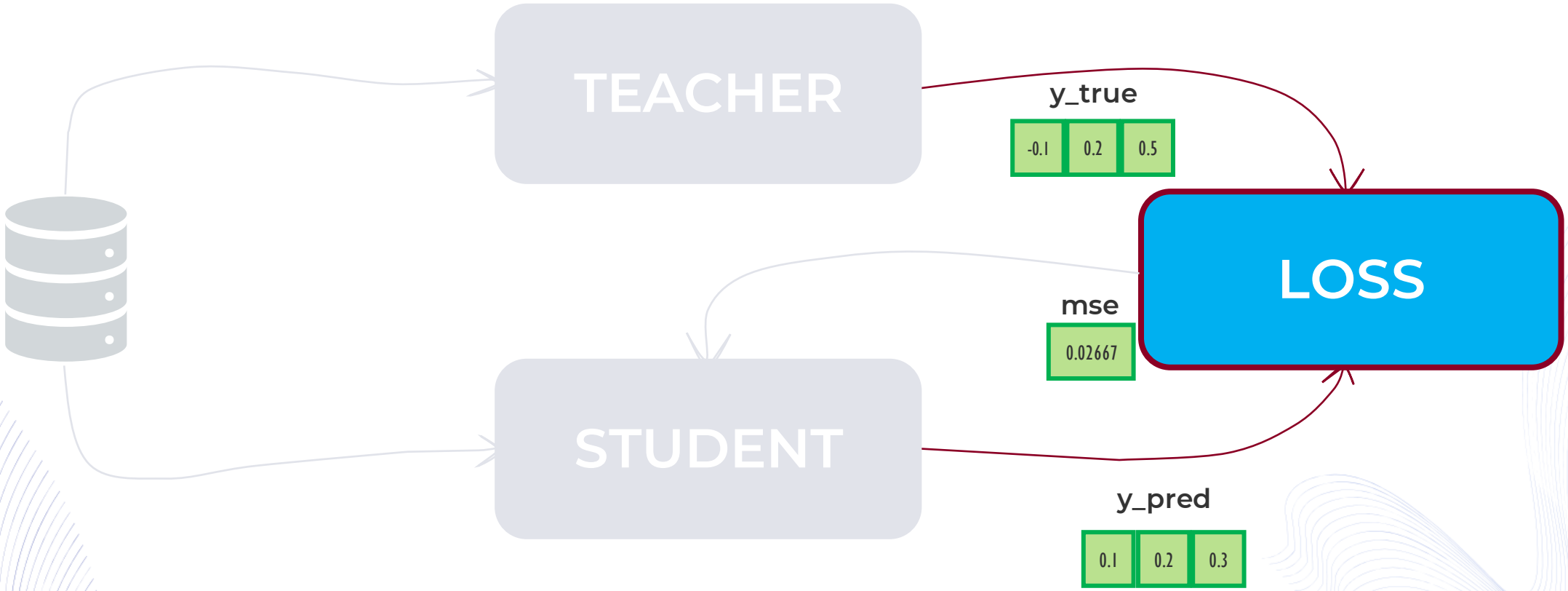
We used Knowledge Distillation to address Italian language



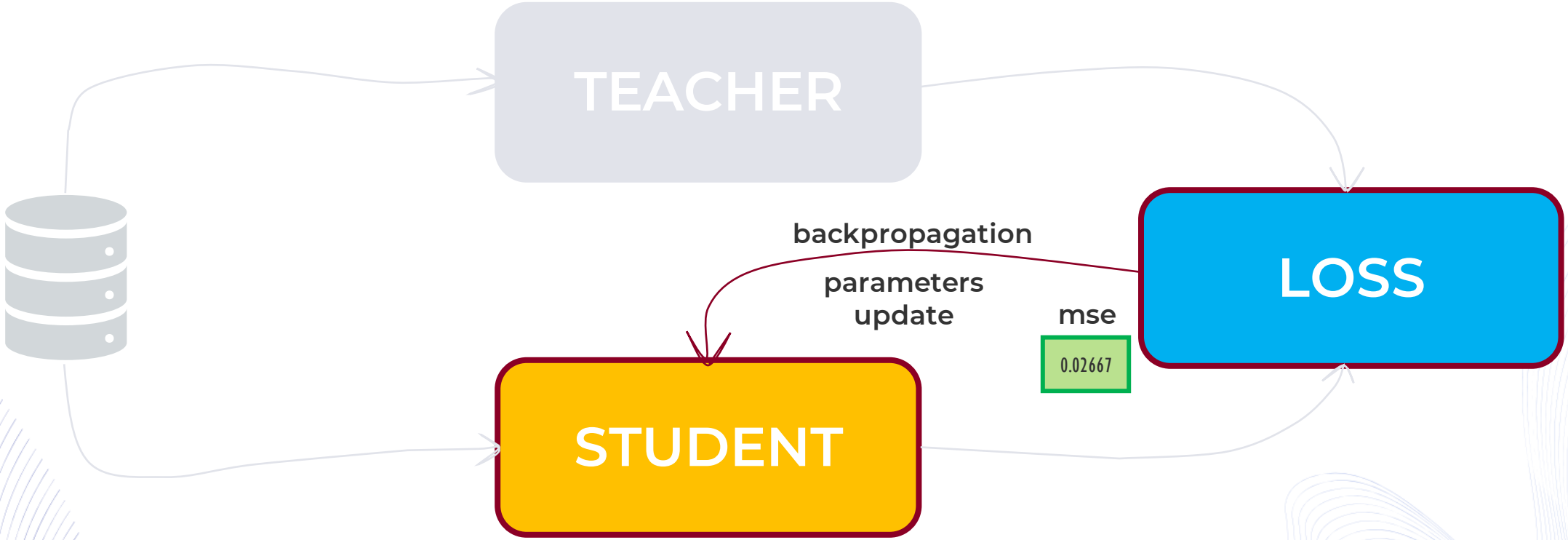
We used Knowledge Distillation to address Italian language



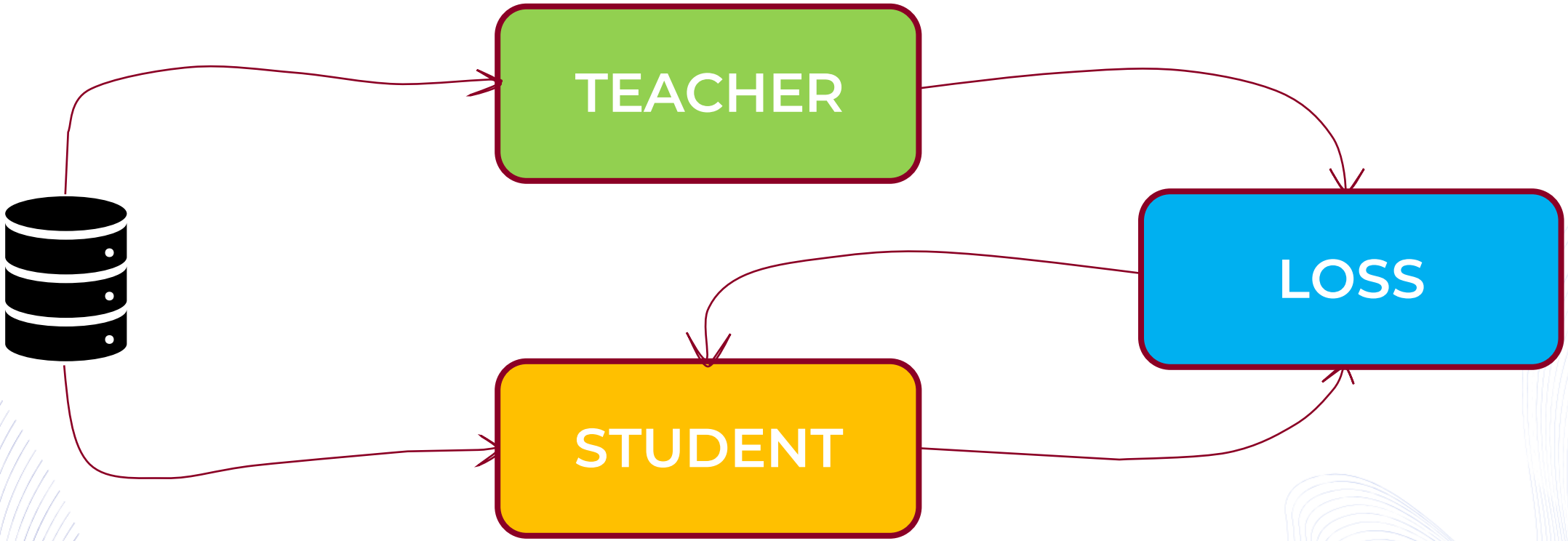
We used Knowledge Distillation to address Italian language



We used Knowledge Distillation to address Italian language



We used Knowledge Distillation to address Italian language



Sentences with the same semantics are mapped the same **regardless of the language**.

Tests over translated SNLI, MNLI and RTE

Dataset	Task	Metric	Result	Delta*
SNLI (IT) translated w/**NLLB**	NLI	Accuracy	**74.21%**	-1.83%
SNLI (IT) translated w/**NLLB**	NLI	Min F1-Score	**67.19%**	-4.34%
MNLI (IT) translated w/**NLLB**	NLI	Accuracy	**72.74%**	**+1.09%**
MNLI (IT) translated w/**NLLB**	NLI	Min F1-Score	**64.53%**	**+0.55%**

Dataset	Task	Metric	Result	Delta
RTE3-ITA	NLI	Accuracy	**67.50%**	**+4.75%**
RTE3-ITA	NLI	Min F1-Score	**60.12%**	**+5.55%**

*Delta: performance difference of Knowledge Distillation with another architecture based on Machine Translation.

The architecture **performs better** than fine-tuning over NLI machine-translated dataset.

ABSA problem can be ‘mapped’ into a NLI task

Premise	Hypothesis	Task	Expected label
Camera piccola ma pulita	Sono soddisfatto	SA	**ENTAILMENT**
	Parlo di pulizia	TR	**ENTAILMENT**
	La camera è pulita	ABSA	**ENTAILMENT**

Neutral label remapped by means of **validation set performance maximization**:

Model label	Task	Dataset label
NEUTRAL	SA	**ENTAILMENT (TRUE)**
	TR	**ENTAILMENT (TRUE)**
	ABSA	**CONTRADICTION (FALSE)**

By choosing **ad-hoc hypotheses**, we can search for information from free-text!

ABSA problem can be ‘mapped’ into a NLI task

Dataset	Balancing	Task	Metric	Result	Delta
ABSITA	1:1	**SA**	Accuracy	**85.04%**	**+3.08%**
ABSITA	1:1	**TR**	Accuracy	**71.19%**	-3.10%
ABSITA	**1:7**	**TR**	Accuracy	**65.84%**	**+6.24%**
ABSITA	1:1	**ABSA**	Accuracy	**94.03%**	**+6.24%**
ABSITA	**1:15**	**ABSA**	Accuracy	**78.42%**	**+11.39%**

*Delta: performance difference of Knowledge Distillation with another architecture based on Machine Translation.

NLI seems to work well to find patterns in free-text!

Recap and conclusions

- We identified an approach to **classify free-text** in an unsupervised way using NLI
- We proved **generality of NLI task** by emulating SA, TR, and ABSA
- Surprisingly, **KD** model revealed to be successful

Cheers from BERT!

credits:

- We identified an approach to **classify free-text** in an unsupervised way using NLI
- We proved **generality of NLI task** by emulating SA, TR, and ABSA
- Surprisingly, **KD** model revealed to be successful

Thank you very much for the attention!

Lorenzo Corradi