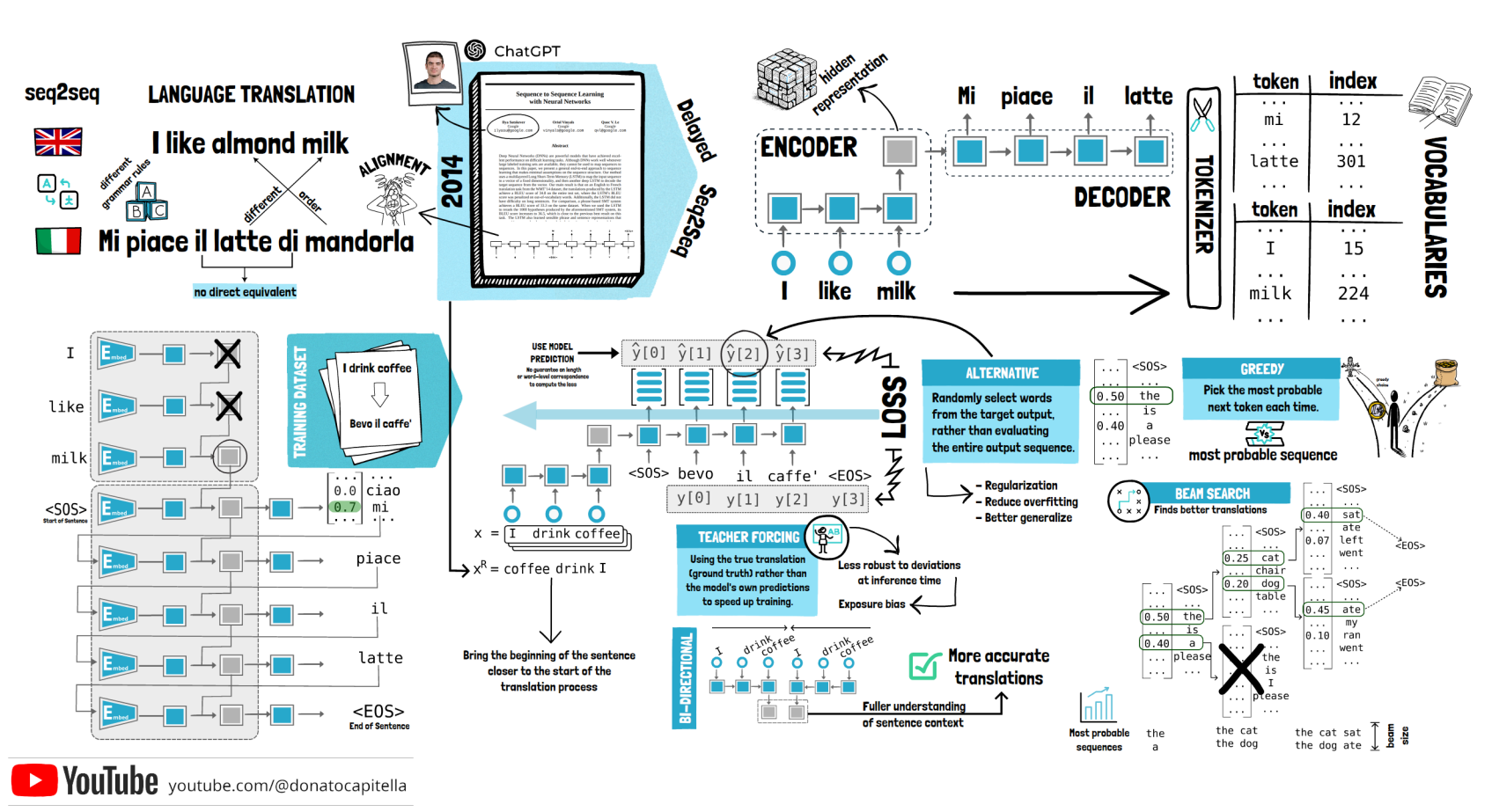
seq2seq
LANGUAGE TRANSLATION
I like almond milk
different grammar rules
different order
ALIGNMENT
Mi piace il latte di mandorla
no direct equivalent
ChatGPT
2014
Sequence to Sequence Learning with Neural Networks
Abstract
Delayed Seq2Seq
hidden representation
ENCODER
DECODER
Mi piace il latte
I like milk
TOKENIZER
token index
mi 12
latte 301
I 15
milk 224
VOCABULARIES
I
like
milk
<SOS>
Start of Sentence
ciao
mi
piace
il
latte
<EOS>
End of Sentence
TRAINING DATASET
I drink coffee
Bevo il caffe'
USE MODEL PREDICTION
Be guarantee on length or word-level correspondence to compute the loss
ŷ[0] ŷ[1] ŷ[2] ŷ[3]
LOSS
<SOS> bevo il caffe' <EOS>
y[0] y[1] y[2] y[3]
x = I drink coffee
x^R = coffee drink I
Bring the beginning of the sentence closer to the start of the translation process
ALTERNATIVE
Randomly select words from the target output, rather than evaluating the entire output sequence.
- Regularization
- Reduce overfitting
- Better generalize
TEACHER FORCING
Using the true translation (ground truth) rather than the model's own predictions to speed up training.
Less robust to deviations at inference time
Exposure bias
BI-DIRECTIONAL
I drink coffee I drink coffee
Fuller understanding of sentence context
More accurate translations
GREEDY
Pick the most probable next token each time.
most probable sequence
BEAM SEARCH
Finds better translations
Most probable sequences
the
a
the cat
the dog
the cat sat
the dog ate
beam size
YouTube
youtube.com/@donatocapitella