MAI4CAREU

Επεξεργασία Φυσικής Γλώσσας
Κατανόηση μεγάλων γλωσσικών μοντέλων

Δημήτρης Πασχαλίδης

Τμήμα Πληροφορικής

Πανεπιστήμιο Κύπρου

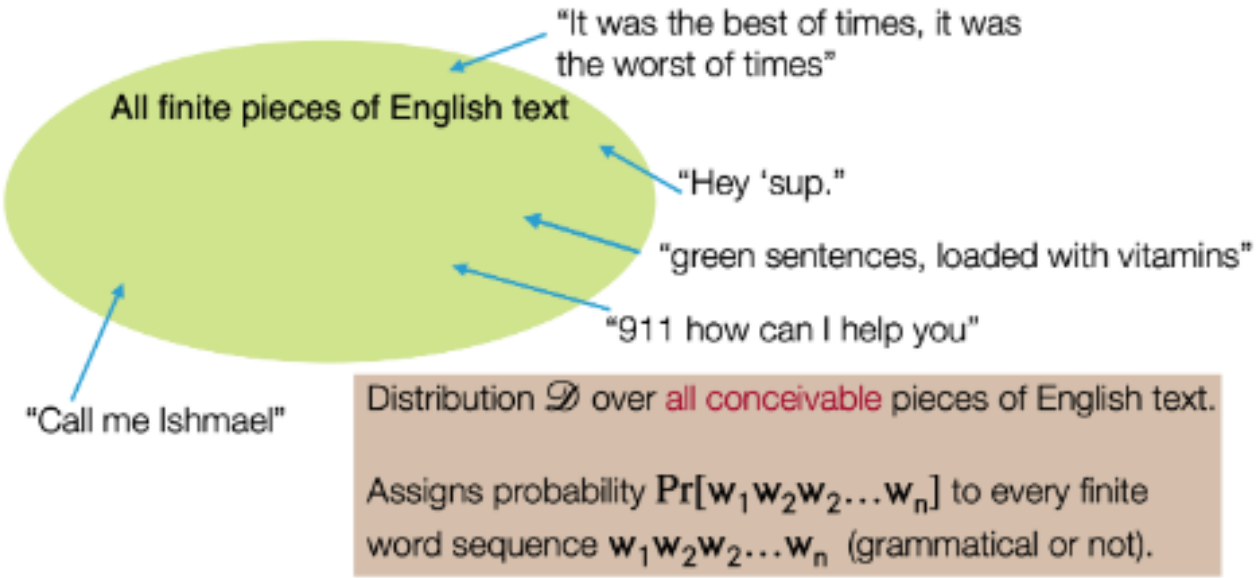
Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Γλωσσικά μοντέλα: Στενή αίσθηση

- Ένα πιθανοτικό μοντέλο που αποδίδει μια πιθανότητα P[w1, w2, ..., wn] σε κάθε πεπερασμένη ακολουθία $w_1, \ldots, w_n$ (γραμματική ή όχι).



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Γλωσσικά μοντέλα: Στενή αίσθηση

Conditional probability

$$p(w_1, w_2, w_3, \ldots, w_N) =$$
$$p(w_1)\, p(w_2|w_1)\, p(w_3|w_1, w_2) \times \ldots \times p(w_N|w_1, w_2, \ldots w_{N-1})$$

Sentence: “the cat sat on the mat”

$$P(\text{the cat sat on the mat}) = P(\text{the}) * P(\text{cat}|\text{the}) * P(\text{sat}|\text{the cat})$$
$$* P(\text{on}|\text{the cat sat}) * P(\text{the}|\text{the cat sat on})$$
$$* P(\text{mat}|\text{the cat sat on the})$$

Implicit order

GPT-3 still acts in this way but the model is implemented as a very large neural network of 175-billion parameters!

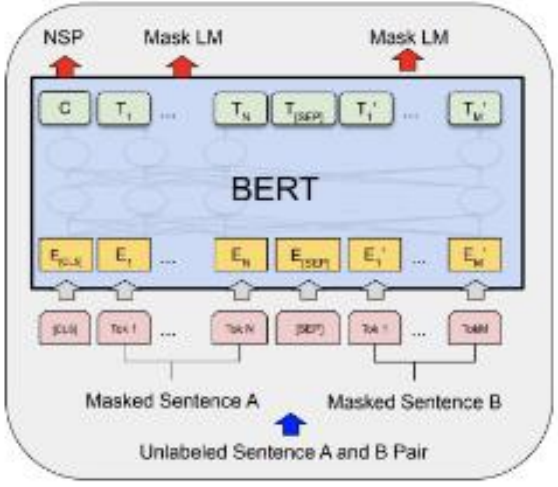
Co-financed by the European Union
Connecting Europe Facility

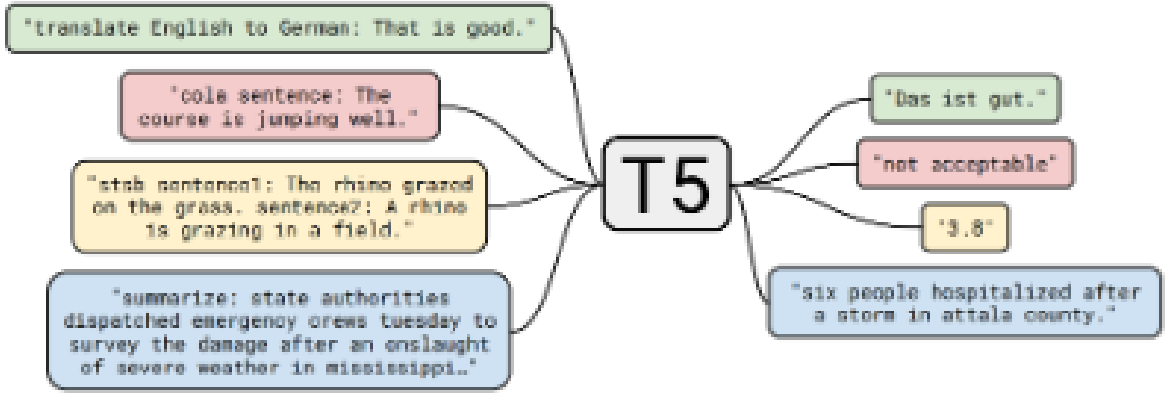
University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Γλωσσικά μοντέλα: Ευρεία αίσθηση

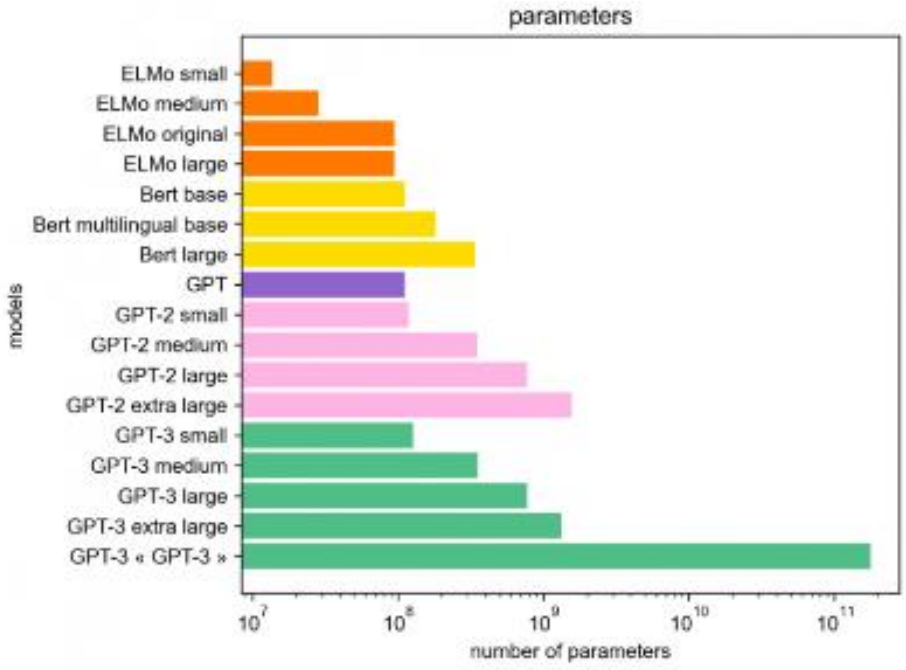
- Decoder-only models (GPT-x models)
- Encoder-only models (BERT, RoBERTa, ELECTRA)
- Encoder-decoder models (T5, BART)

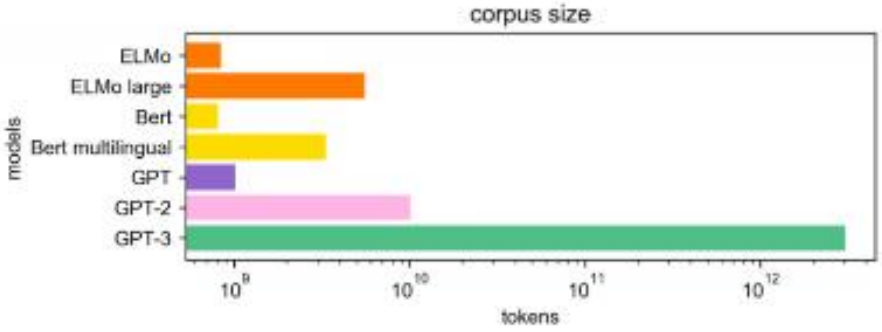




Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Πόσο μεγάλα είναι τα "μεγάλα" LM;





Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

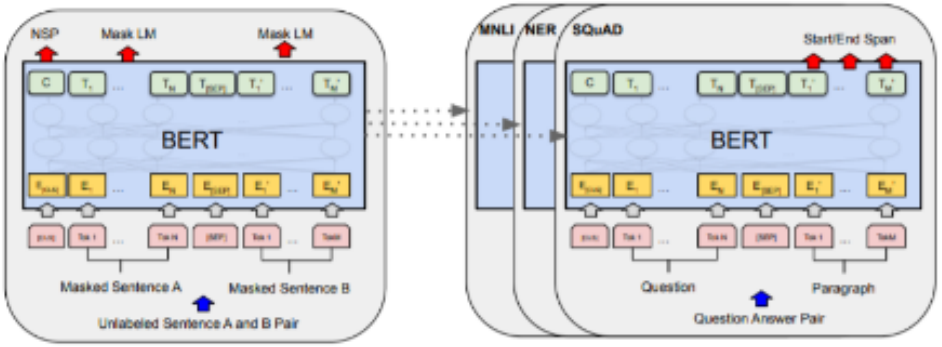
Πόσο μεγάλα είναι τα "μεγάλα" LM;

- Σήμερα, μιλάμε κυρίως για δύο στρατόπεδα μοντέλων:
  - Μοντέλα μεσαίου μεγέθους: BERT/RoBERTa models (100M or 300M), T5 models (220M, 770M, 3B)
  - "Πολύ" μεγάλα LM: models of 100+ billion parameters
- Μεγαλύτερα μεγέθη μοντέλων → Μεγαλύτερος υπολογισμός, ακριβότερη εξαγωγή συμπερασμάτων
- Τα διαφορετικά μεγέθη LM έχουν διαφορετικούς τρόπους προσαρμογής και χρήσης τους
  - Fine-tuning, zero-shot/few-shot prompting, in-context learning...
- Οι αναδυόμενες ιδιότητες προκύπτουν από την κλίμακα του μοντέλου
- Αντιστάθμιση μεταξύ μεγέθους μοντέλου και μεγέθους σώματος κειμένων

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Pre-training και Adaptation

- Pre-training: Εκπαιδεύτηκε σε τεράστιες ποσότητες κειμένου χωρίς ετικέτα χρησιμοποιώντας "αυτο-εποπτευόμενους" εκπαιδευτικούς στόχους
- Adaptation: Πώς να χρησιμοποιήσετε ένα προ-εκπαιδευμένο μοντέλο για την κατάντη εργασία σας;
- Τι είδους εργασίες NLP (μορφές εισόδου και εξόδου);
  - Πόσα σχολιασμένα παραδείγματα έχετε;



Circulation revenue has increased by 5% in Finland. // Positive

Panostaja did not disclose the purchase price. // Neutral

Paying off the national debt will be extremely painful. // Negative

The company anticipated its operating profit to improve. // _____

LM

Circulation revenue has increased by 5% in Finland. // Finance

They defeated … in the NFC Championship Game. // Sports

Apple … development of in-house chips. // Tech

The company anticipated its operating profit to improve. // _____

LM

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Γιατί LLMs;

- The promise: one single model to solve many NLP tasks

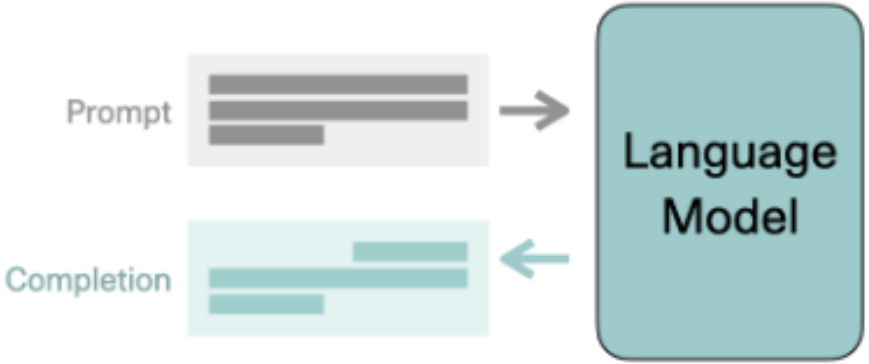
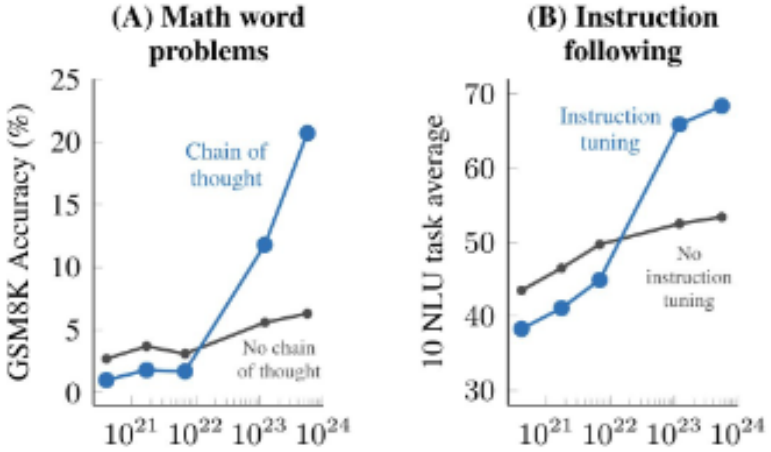


Image credit: Jay Alammar

- Emergent properties in LLMs

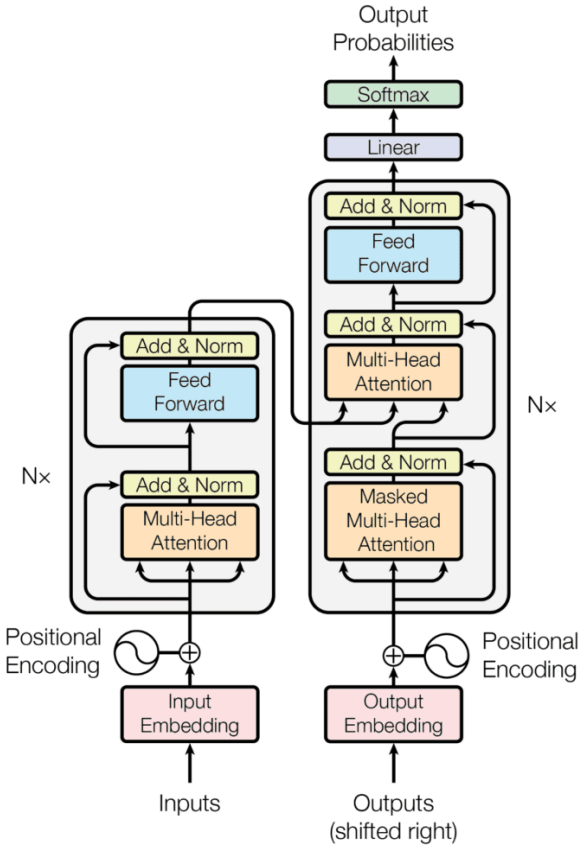


(Wei et al., 2022)

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The Transformer Architecture

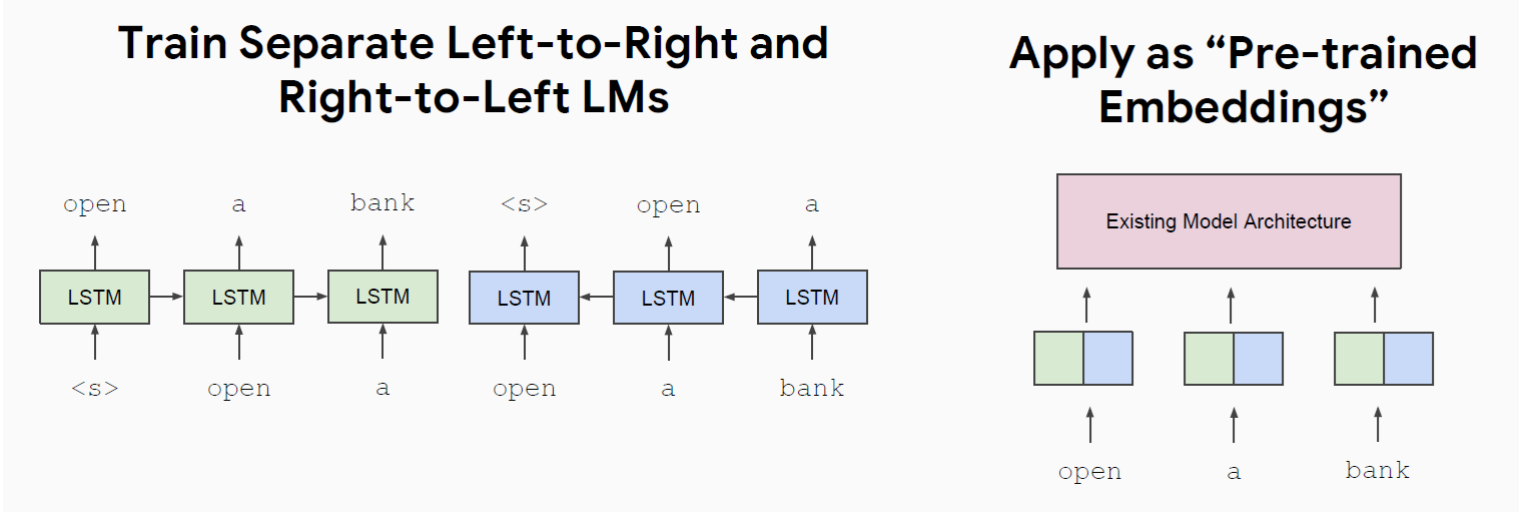
- Η αρχιτεκτονική μετασχηματιστών ακολουθεί μια δομή κωδικοποιητή-αποκωδικοποιητή.
- Ο κωδικοποιητής αντιστοιχίζει μια ακολουθία εισόδου σε μια ακολουθία συνεχών αναπαραστάσεων, η οποία στη συνέχεια τροφοδοτείται σε έναν αποκωδικοποιητή.
- Ο αποκωδικοποιητής λαμβάνει την έξοδο του κωδικοποιητή μαζί με την έξοδο του αποκωδικοποιητή στο προηγούμενο χρονικό βήμα για να δημιουργήσει μια ακολουθία εξόδου.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

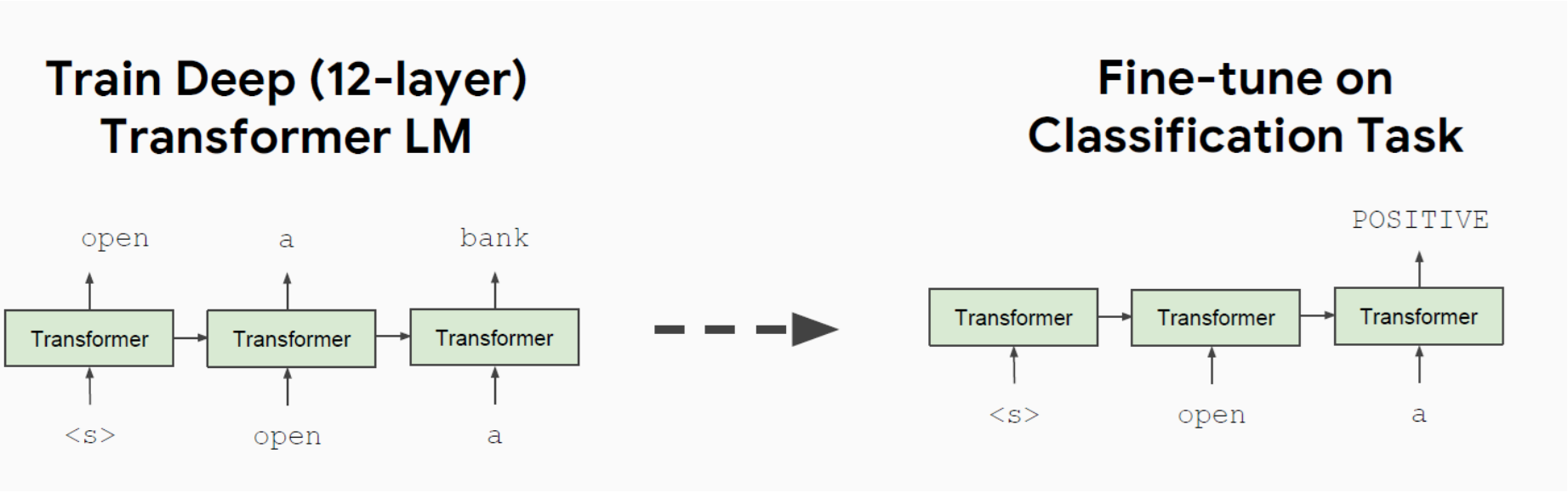
ELMo: Deep Contextualized Word Representation

- Αντί να χρησιμοποιεί μια σταθερή ενσωμάτωση για κάθε λέξη, το ELMo εξετάζει ολόκληρη την πρόταση πριν εκχωρήσει σε κάθε λέξη σε αυτήν μια ενσωμάτωση.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

OpenAI GPT: Generative Pre-trained Transformer



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bidirectional Encoder Representations from Transformers (BERT)

- BERT Google Research 2018 (Devlin et al., 2018)

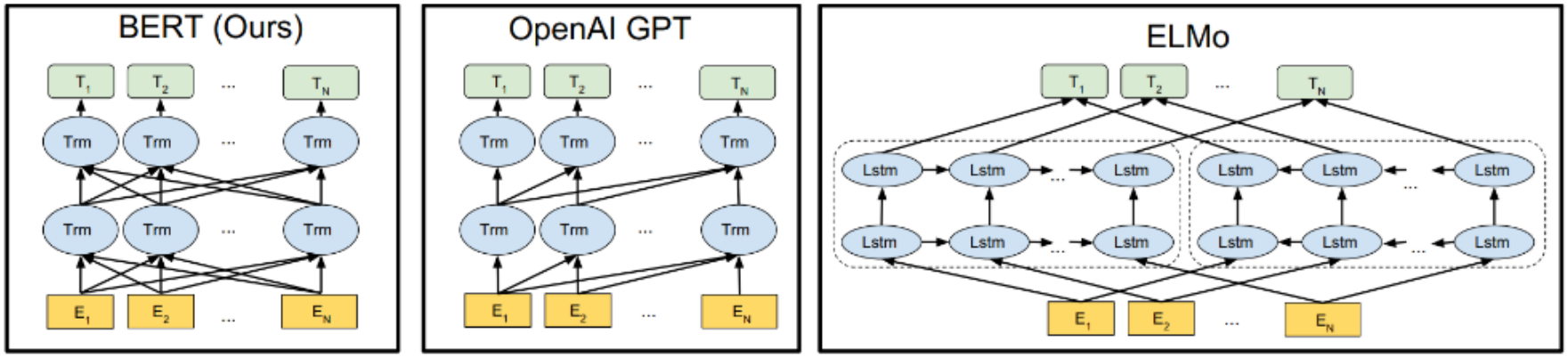


Figure 1: Differences in pre-training model architectures. BERT uses a bidirectional Transformer. OpenAI GPT uses a left-to-right Transformer. ELMo uses the concatenation of independently trained left-to-right and right-to-left LSTM to generate features for downstream tasks. Among three, only BERT representations are jointly conditioned on both left and right context in all layers.

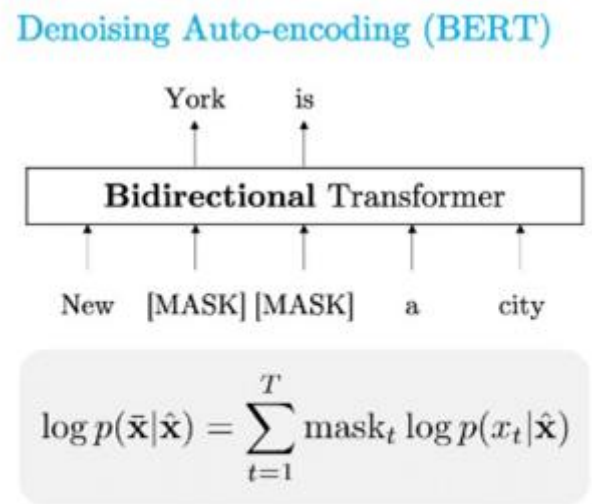
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Objectives for BERT Training

- ❏ Next-token prediction: Διαδοχικός / μονοκατευθυντικός μετασχηματιστής
- ❏ Masked-token prediction: Αμφίδρομος μετασχηματιστής

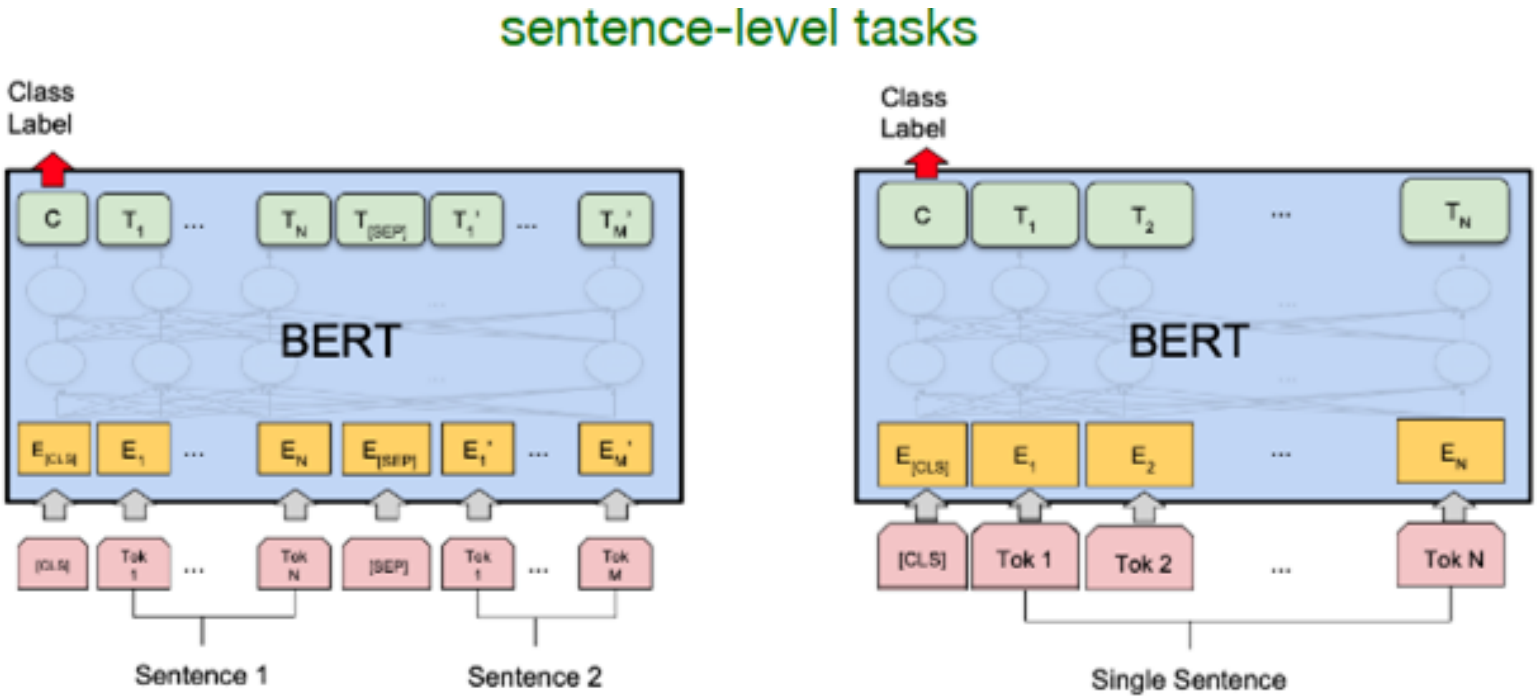
- Απόκρυψη k% των λέξεων εισαγωγής και, στη συνέχεια, πρόβλεψη των λέξεων με μάσκα.

Input: The man went to the **[MASK]$_1$**. He bought a **[MASK]$_2$** of milk.
Labels: **[MASK]$_1$** = *store*; **[MASK]$_2$** = *gallon*



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Fine-Tuning BERT



(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

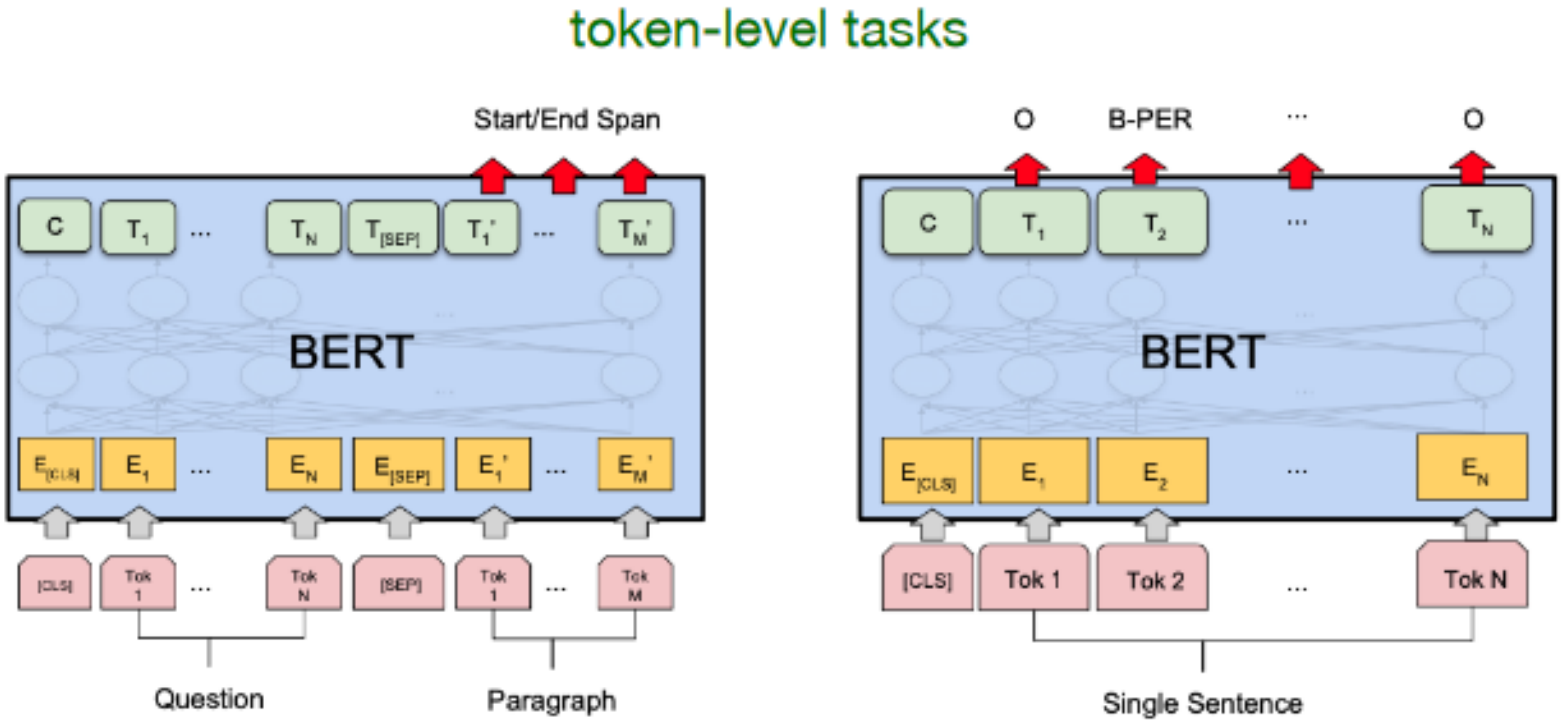
(b) Single Sentence Classification Tasks: SST-2, CoLA

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Fine-Tuning BERT



(c) Question Answering Tasks: SQuAD v1.1

(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εργασίες σε επίπεδο πρότασης

Sentence pair classification tasks:

- MNLI
  - Premise: A soccer game with multiple males playing.
  - Hypothesis: Some men are playing a sport.
  - Output: {entailment, contradiction, neutral}
- QQP
  - Q1: Where can I learn to invest in stocks?
  - Q2: How can I learn more about stocks?
  - Output: {duplicate, not duplicate}

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εργασίες σε επίπεδο διακριτικού

- Extractive question answering e.g., SQuAD (Rajpurkar et al., 2016)

SQuAD

Question: The New York Giants and the New York Jets play at which stadium in NYC ?

Context: The city is represented in the National Football League by the New York Giants and the New York Jets , although both teams play their home games at MetLife Stadium in nearby East Rutherford , New Jersey , which hosted Super Bowl XLVIII in 2014 .

(Training example 29,883)

MetLife Stadium

- Named entity recognition (Tjong Kim Sang and De Meulder, 2003)

CoNLL 2003 NER

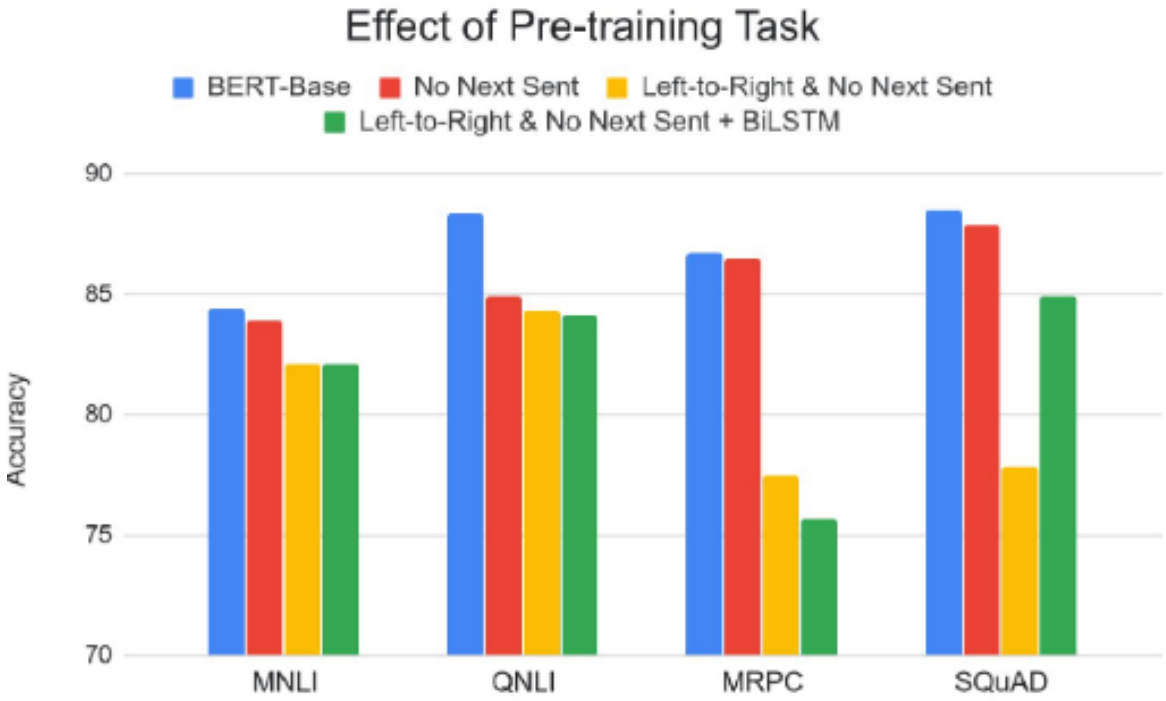
John	Smith	lives	in	New	York
B-PER	I-PER	O	O	B-LOC	I-LOC

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ablation Study: Pre-training Tasks



- MLM >> left-to-right LMs
- NSP improves on some tasks
- Note: later work (Joshi et al., 2020; Liu et al., 2019) argued that NSP is not useful

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ablation Study: Model Sizes

MAI4CAREU

layers ↓ hidden size ↓ # of heads ↙

Hyperparams				Dev Set Accuracy		
#L	#H	#A	LM (ppl)	MNLI-m	MRPC	SST-2
3	768	12	5.84	77.9	79.8	88.4
6	768	3	5.24	80.6	82.2	90.7
6	768	12	4.68	81.9	84.8	91.3
12	768	12	3.99	84.4	86.7	92.9
12	1024	16	3.54	85.7	86.9	93.3
24	1024	16	3.23	86.6	87.8	93.7

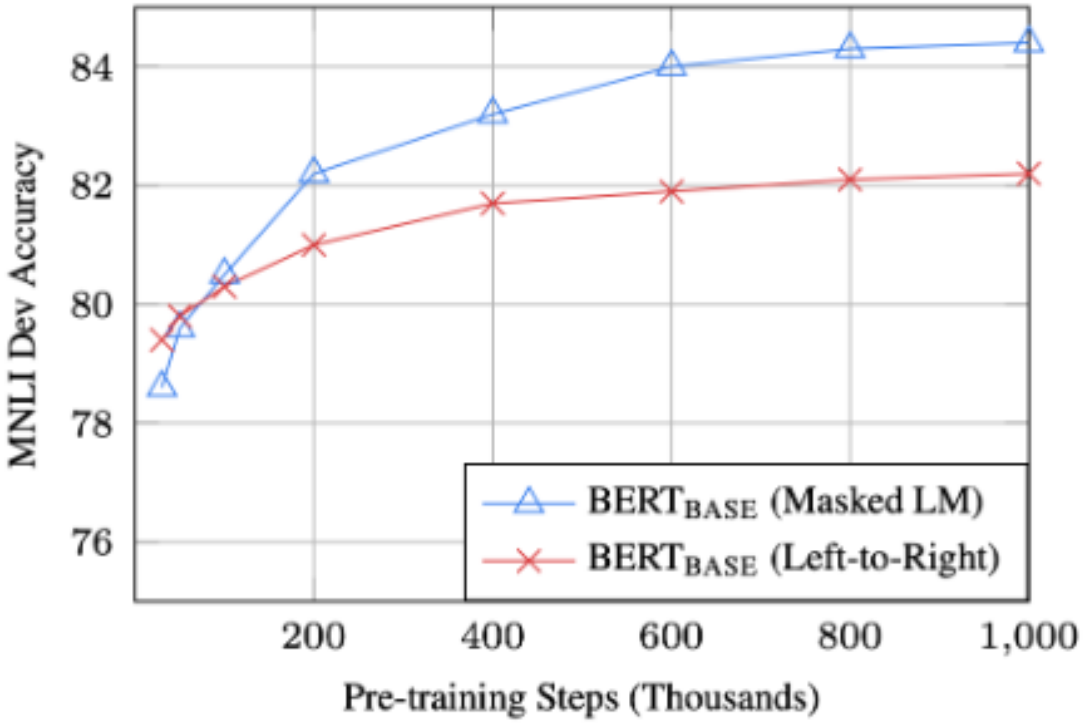
The bigger, the better!

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ablation Study: Training Efficiency



MLM takes slightly longer to converge because it only predicts 15% of tokens

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τι συνέβη μετά τον BERT;

- RoBERTa (Liu et al., 2019)
  - Εκπαιδευμένο σε 10x δεδομένα και περισσότερο, χωρίς NSP
  - Πολύ ισχυρότερη απόδοση από τον BERT (e.g., 94.6 vs 90.9 on SQuAD)
  - Still oNE από τα πιο δημοφιλή μοντέλα μέχρι σήμερα
- ALBERT (Lan et al., 2020)
  - Αύξηση μεγεθών μοντέλων με κοινή χρήση παραμέτρων μοντέλου μεταξύ επιπέδων
  - Λιγότερος χώρος αποθήκευσης, πολύ ισχυρότερη απόδοση, αλλά λειτουργεί πιο αργά..
- ELECTRA (Clark et al., 2020)
  - Παρέχει μια πιο αποτελεσματική μέθοδο εκπαίδευσης προβλέποντας το 100% των μαρκών αντί για το 15% των μαρκών

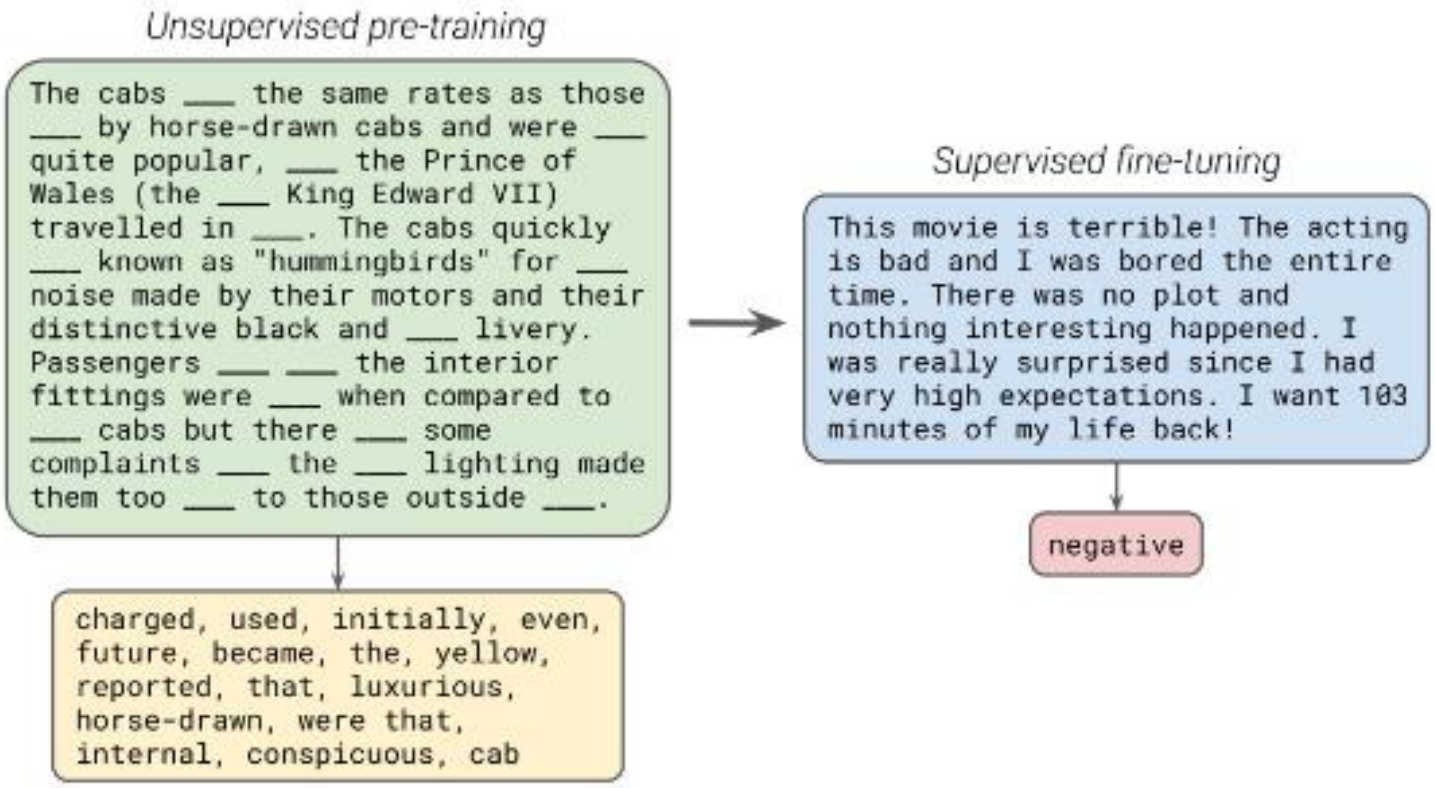
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τι συνέβη μετά τον BERT;

- Μοντέλα που χειρίζονται μεγάλα περιβάλλοντα (512 tokens)
  - Longformer, Big Bird, ...
- Multilingual BERT
  - Εκπαιδευμένο ενιαίο μοντέλο σε 104 γλώσσες από τη Wikipedia. Κοινόχρηστο λεξιλόγιο 110k WordPiece
- BERT Επέκταση σε διαφορετικούς τομείς
  - SciBERT, BioBERT, FinBERT, ClinicalBERT, ...
- Μείωση της χρήσης του BERT
  - DistillBERT, TinyBERT, ...

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Σύντομο Πλαίσιο Μεταβιβαστικής Μάθησης



Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

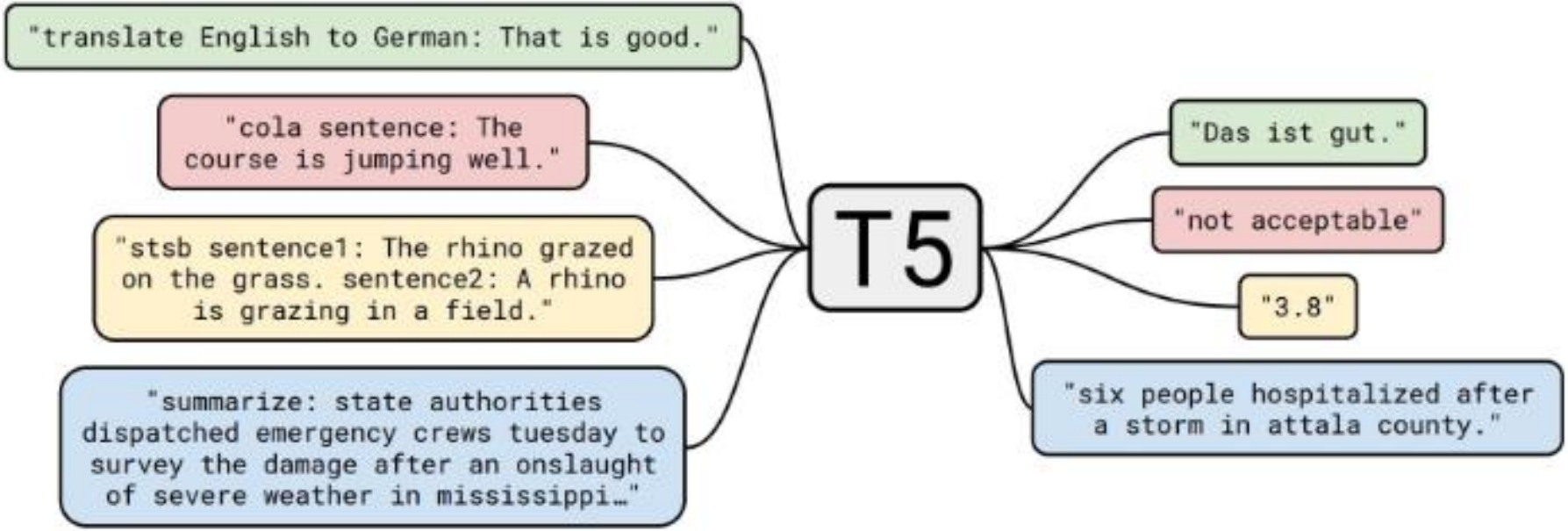
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παρουσιάζουμε το T5

- Text-to-Text Transfer Transformer
- Κάθε εργασία, μία μορφή
- Προηγούμενες προσπάθειες περιελάμβαναν::
  - Απάντηση ερωτήσεων
  - Γλωσσική μοντελοποίηση
  - Εκχύλιση εύρους
  - ... αλλά είχε περιορισμούς
- [Task-specific prefix]: [Input text] → [output text]

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Παρουσιάζουμε το T5



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τ5 Εκπαιδευτικά καθήκοντα

SQuAD, GLUE benchmarks

- **CoLA (GLUE)**: Sentence acceptability
  - Input: sentence, output: labels “acceptable” or “not acceptable”
  - Ex: “The course is jumping well.” → not acceptable
- **STS-B (GLUE)**: Sentence similarity
  - Input: pair of sentences, output: similarity score [1, 5]
  - Ex: “s_1: The rhino grazed. s_2: A rhino is grazing.” → 3.8

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

T5 Training Tasks

SQuAD, GLUE benchmarks

- **COPA (SuperGLUE)**: Causal reasoning
  - Input: premise and 2 alternatives (a), output: a_1 or a_2
  - Example:
    "Premise: I tipped the bottle. What happened as a RESULT?
    Alternative a_1: The liquid in the bottle froze.
    Alternative a_2: The liquid in the bottle poured out." $\rightarrow a_2$
- **ReCoRD/MultiRC (SuperGLUE)**: Question answering/Reading comprehension

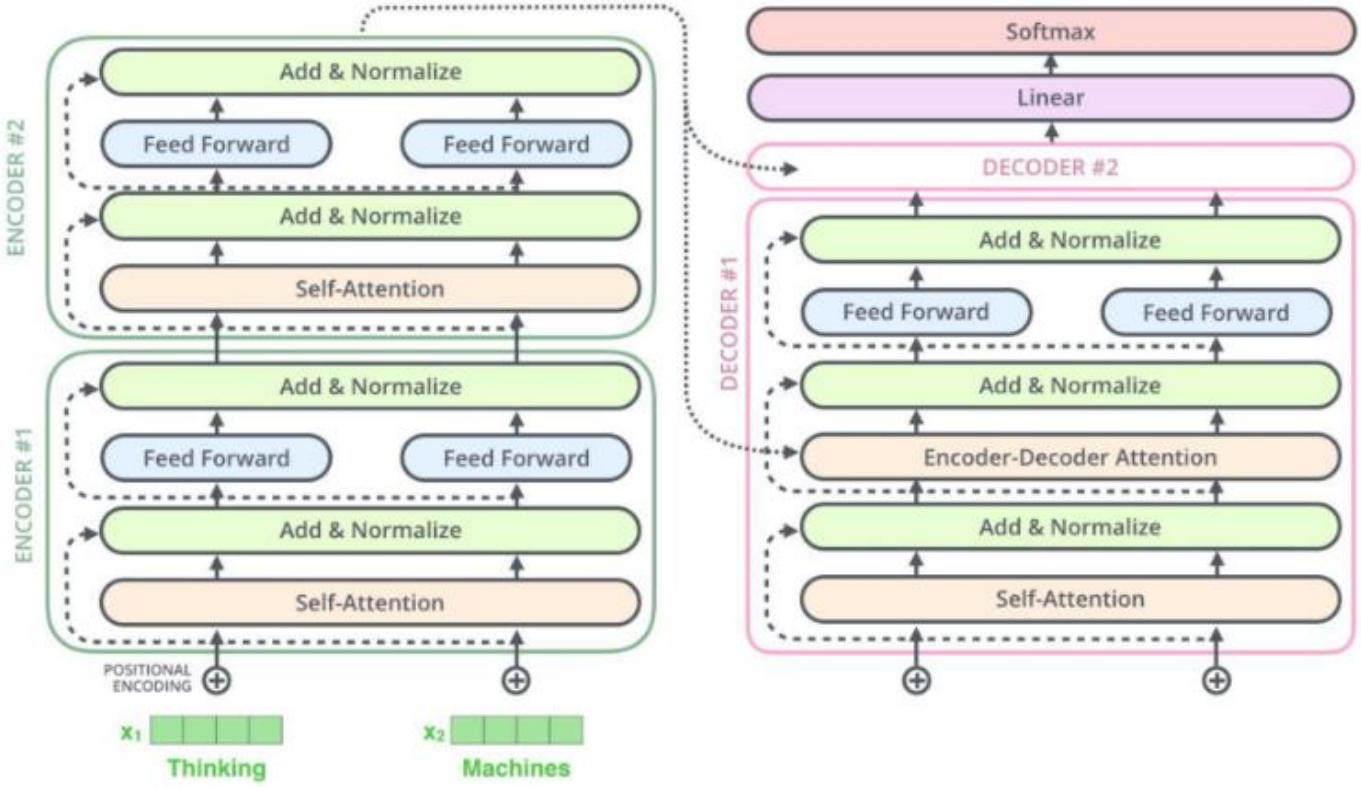
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Αρχιτεκτονική μοντέλου T5

- Μοντέλο κωδικοποιητή-αποκωδικοποιητή
  - Baseline size: two stacks of size $\text{BERT}_{\text{BASE}}$
- Αρχιτεκτονική από "*Attention Is All You Need*"
  - Διαφορετικό σχήμα ενσωμάτωσης θέσης

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

T5 Model Architecture



Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Δεδομένα εκπαίδευσης: C4

- Colossal Clean Crawled Corpus
  - Κείμενο που έχει εξαχθεί από το Web
  - Μόνο αγγλική γλώσσα (langdetect)
  - 750GB
- 20TB to 750GB? Πού πήγαν όλα;
  - Διατηρώ:
    - Προτάσεις με τελικά σημεία στίξης
    - Σελίδες με τουλάχιστον 5 προτάσεις, προτάσεις με τουλάχιστον 3 λέξεις
  - Αφαιρώ:
    - Αναφορές σε Javascript
    - Lorem ipsum text
    - Code

Data set	Size
★ C4	745GB
C4, unfiltered	6.1TB
RealNews-like	35GB
WebText-like	17GB
Wikipedia	16GB
Wikipedia + TBC	20GB

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Δεδομένα εκπαίδευσης: C4

Menu

Lemon

Introduction

The lemon, Citrus Limon (l.) Osbeck, is a species of small evergreen tree in the flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for culinary and non-culinary purposes throughout the world, primarily for its juice, which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6% citric acid, with a ph of around 2.2, giving it a sour taste.

Article

The origin of the lemon is unknown, though lemons are thought to have first grown in Assam (a region in northeast India), northern Burma or China.
A genomic study of the lemon indicated it was a hybrid between bitter orange (sour orange) and citron.

Please enable JavaScript to use our site.

Home
Products
Shipping
Contact
FAQ

Dried Lemons, $3.59/pound

Organic dried lemons from our farm in California.
Lemons are harvested and sun-dried for maximum flavor.
Good in soups and on popcorn.

The lemon, Citrus Limon (l.) Osbeck, is a species of small evergreen tree in the flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for culinary and non-culinary purposes throughout the world, primarily for its juice, which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6% citric acid, with a ph of around 2.2, giving it a sour taste.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Curabitur in tempus quam. In mollis et ante at consectetur.
Aliquam erat volutpat.
Donec at lacinia est.
Duis semper, magna tempor interdum suscipit, ante elit molestie urna, eget efficitur risus nunc ac elit.
Fusce quis blandit lectus.
Mauris at mauris a turpis tristique lacinia at nec ante.
Aenean in scelerisque tellus, a efficitur ipsum.
Integer justo enim, ornare vitae sem non, mollis fermentum lectus.
Mauris ultrices nisl at libero porta sodales in ac orci.

```
function Ball(r) {
  this.radius = r;
  this.area = pi * r ** 2;
  this.show = function(){
    drawCircle(r);
  }
}
```

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Δεδομένα εκπαίδευσης: C4

Menu

Lemon

Introduction

The lemon, Citrus Limon (l.) Osbeck, is a species of small evergreen tree in the flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for culinary and non-culinary purposes throughout the world, primarily for its juice, which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6% citric acid, with a ph of around 2.2, giving it a sour taste.

Article

The origin of the lemon is unknown, though lemons are thought to have first grown in Assam (a region in northeast India), northern Burma or China.
A genomic study of the lemon indicated it was a hybrid between bitter orange (sour orange) and citron.

Please enable JavaScript to use our site.

Home
Products
Shipping
Contact
FAQ

Dried Lemons, $3.59/pound

Organic dried lemons from our farm in California.
Lemons are harvested and sun-dried for maximum flavor.
Good in soups and on popcorn.

The lemon, Citrus Limon (l.) Osbeck, is a species of small evergreen tree in the flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for culinary and non-culinary purposes throughout the world, primarily for its juice, which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6% citric acid, with a ph of around 2.2, giving it a sour taste.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Curabitur in tempus quam. In mollis et ante at consectetur.
Aliquam erat volutpat.
Donec at lacinia est.
Duis semper, magna tempor interdum suscipit, ante elit molestie urna, eget efficitur risus nunc ac elit.
Fusce quis blandit lectus.
Mauris at mauris a turpis tristique lacinia at nec ante.
Aenean in scelerisque tellus, a efficitur ipsum.
Integer justo enim, ornare vitae sem non, mollis fermentum lectus.
Mauris ultrices nisl at libero porta sodales in ac orci.

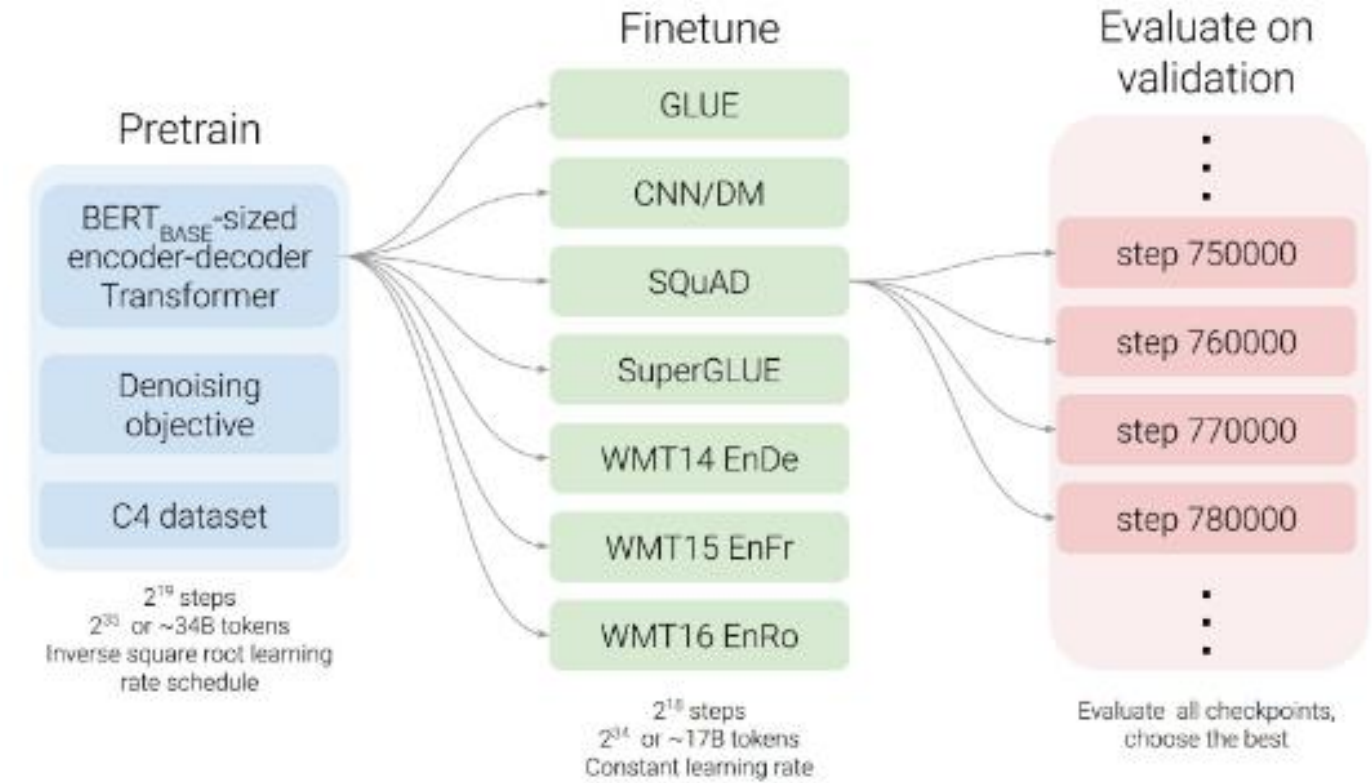
```
function Ball(r) {
  this.radius = r;
  this.area = pi * r ** 2;
  this.show = function(){
    drawCircle(r);
  }
}
```

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ροή εργασίας



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

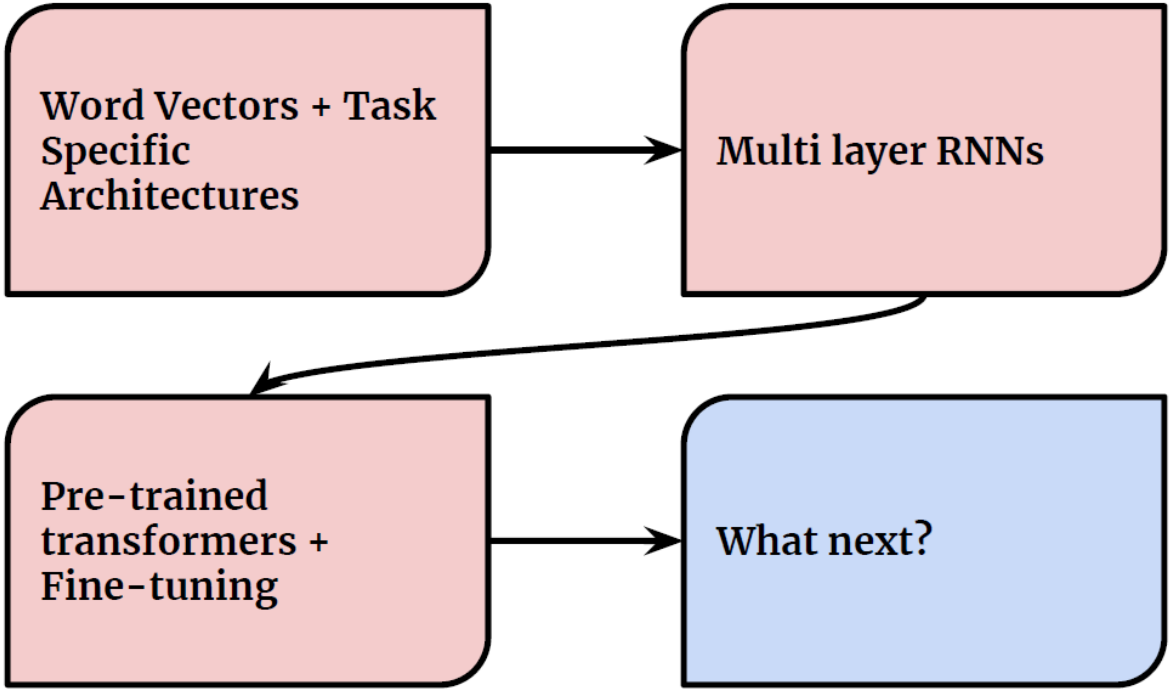
Εκτέλεση

	GLUE	CNNDM	SQuAD	SGLUE	EnDe	EnFr	EnRo
★ Baseline average	**83.28**	**19.24**	**80.88**	**71.36**	**26.98**	**39.82**	**27.65**
Baseline standard deviation	0.235	0.065	0.343	0.416	0.112	0.090	0.108
No pre-training	66.22	17.60	50.31	53.04	25.86	**39.77**	24.04

- GLUE/SuperGLUE είναι σύνολα εργασιών, όπως: CoLA, STS-B, etc.
- CNNDM είναι μια εργασία σύνοψης
- EnDe/EnFr/EnRo είναι εργασίες μετάφρασης

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Μεταβαλλόμενα παραδείγματα στον NLP



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Pre-training → Fine-tuning Limitations

Πρακτικά θέματα

- ❑ Χρειάζεστε μεγάλο σύνολο δεδομένων για συγκεκριμένες εργασίες για τελειοποίηση
- ❑ Συλλογή δεδομένων για την εργασία A → Fine-tune για την επίλυση της εργασίας A → Επανάληψη για την εργασία B → Επαναλάβετε για την εργασία C → …
- ❑ Καταλήξτε με πολλά "αντίγραφα" του ίδιου μοντέλου

Overfitting

- ❑ Μεγάλα μοντέλα με ακρίβεια σε πολύ στενές κατανομές εργασιών
- ❑ Τα μοντέλα είναι υπερβολικά κατάλληλα για διανομές προπόνησης και δεν γενικεύουν καλά έξω από αυτό
- ❑ Τα μοντέλα είναι καλά στο σύνολο δεδομένων, όχι τόσο καλά στην υποκείμενη εργασία.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Pre-training → Fine-tuning Limitations

Οι άνθρωποι δεν χρειάζονται μεγάλα εποπτευόμενα σύνολα δεδομένων

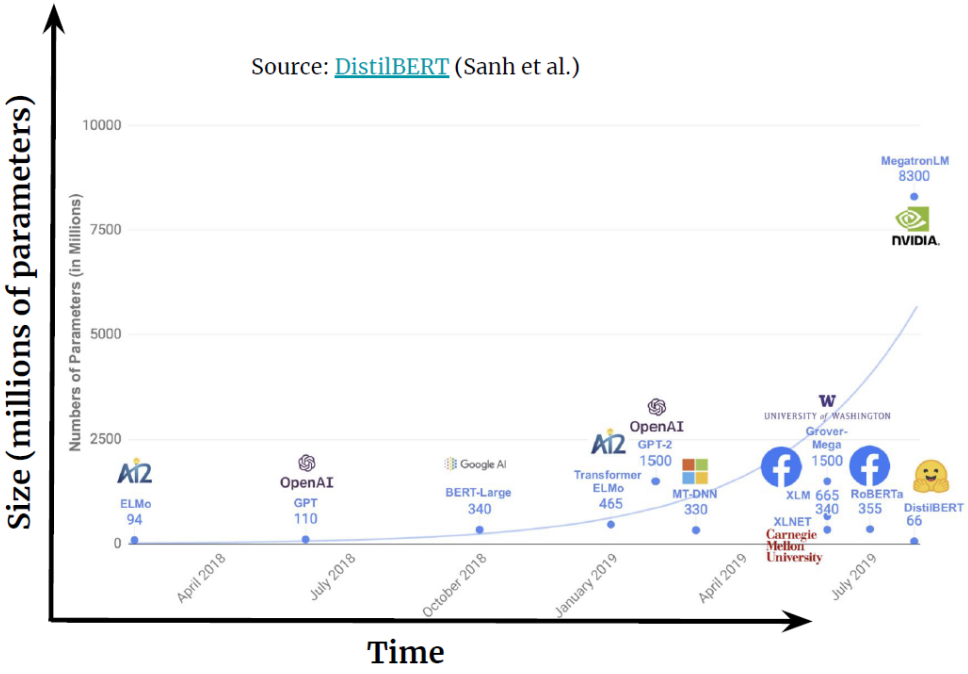
- Οι άνθρωποι μπορούν να μάθουν από απλές οδηγίες
- Επιτρέπει στους ανθρώπους να συνδυάζουν και να ταιριάζουν δεξιότητες + εναλλαγή μεταξύ εργασιών εύκολα
- Hope for NLP systems to function with the same fluidity
- **Αντιμετώπιση αυτών των περιορισμών**
  a. Scaling-up
  b. In-Context-Learning

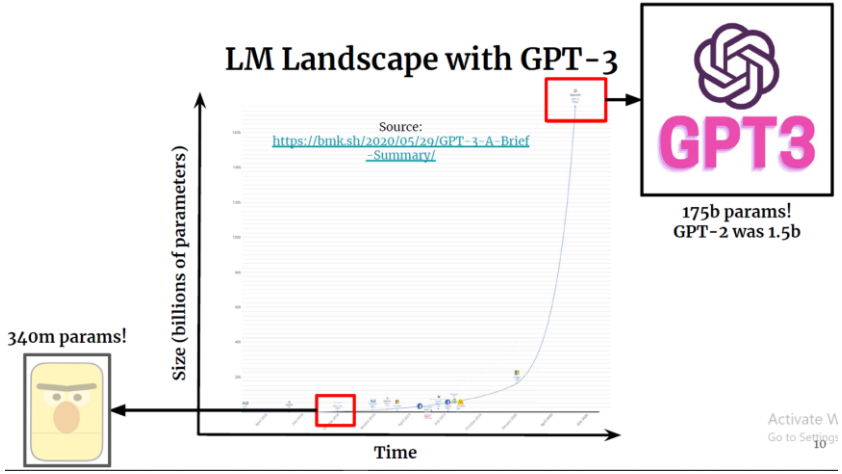
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

LM Scale-up

LM Landscape pre GPT-3

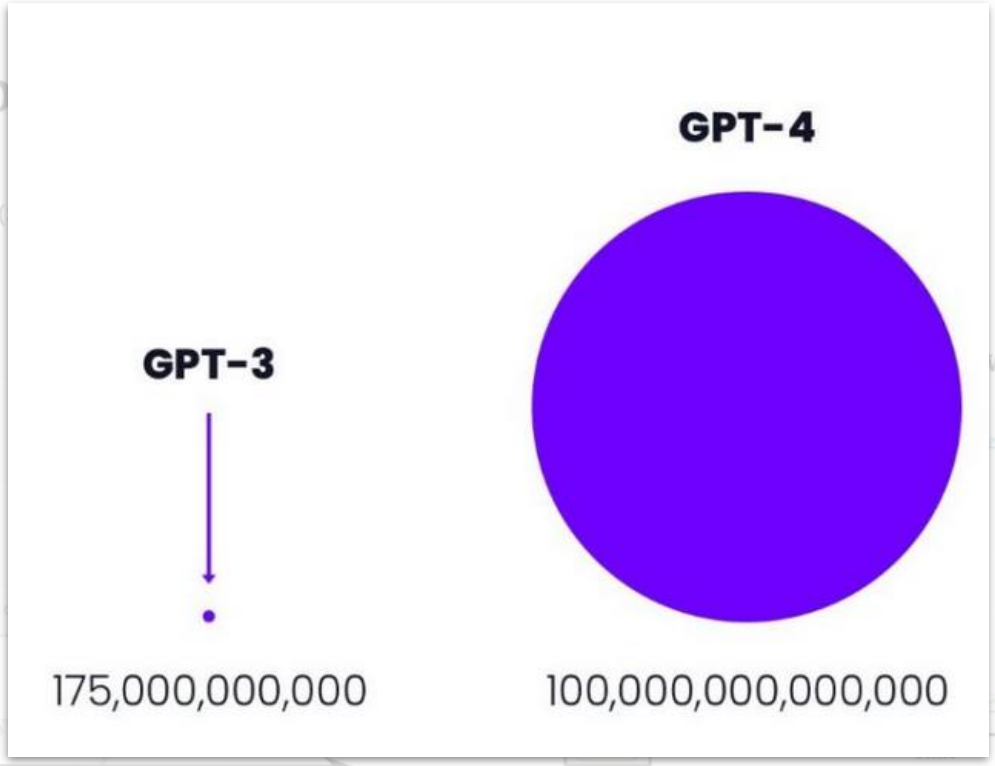
Source: DistilBERT (Sanh et al.)





Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

LM Scale-up

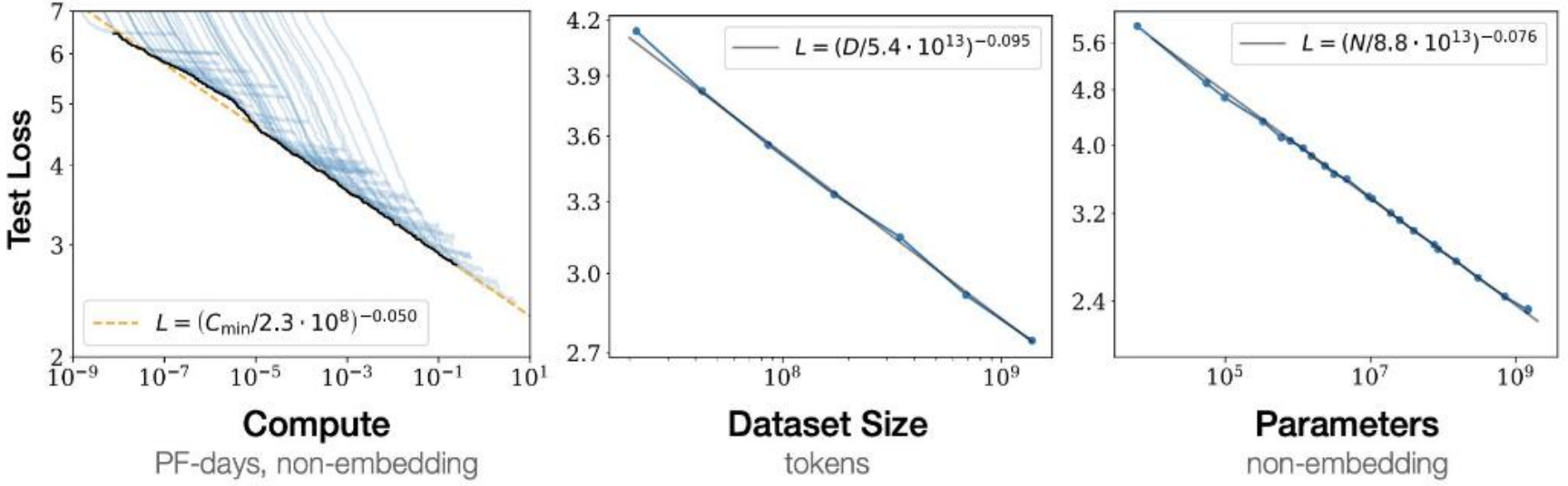


Γιατί να κλιμακώσετε τα LM;

- Μελέτη που διεξήχθη από την OpenAI → Νόμοι Ομοιοθεσίας για Μοντέλα Νευρωνικής Γλώσσας (Kaplan et al. 2020)
- Μερικά βασικά ευρήματα:
  - Η απόδοση εξαρτάται σε μεγάλο βαθμό από την κλίμακα, ασθενώς από το σχήμα του μοντέλου
  - Ομαλοί νόμοι ισχύος (y=axk) μεταξύ εμπειρικής απόδοσης και N-παραμέτρων, μεγέθους συνόλου δεδομένων D, υπολογισμού C.
  - Η μεταφορά βελτιώνεται με την απόδοση της δοκιμής
  - Τα μεγαλύτερα μοντέλα είναι πιο αποδοτικά ως προς το δείγμα

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Γιατί να κλιμακώσετε τα LM;



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Μάθηση εντός πλαισίου

	No Prompt	Prompt
Zero-shot (0s)	skicts = sticks	Please unscramble the letters into a word, and write that word: skicts = sticks
1-shot (1s)	chiar = chair skicts = sticks	Please unscramble the letters into a word, and write that word: chiar = chair skicts = sticks
Few-shot (FS)	chiar = chair [...] pciinc = picnic skicts = sticks	Please unscramble the letters into a word, and write that word: chiar = chair [...] pciinc = picnic skicts = sticks

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μάθηση εντός πλαισίου

- Η μάθηση εντός πλαισίου είναι η διαδικασία εκμάθησης ποικίλων δεξιοτήτων και δευτερευουσών εργασιών κατά τη διάρκεια της διαδικασίας προ-κατάρτισης που μπορεί στη συνέχεια να αξιοποιηθεί με την προτροπή του μοντέλου κατά το χρόνο συμπερασμάτων χρησιμοποιώντας οδηγίες φυσικής γλώσσας ή / και επιδείξεις ("λήψεις").
- Σε αντίθεση με την τελειοποίηση, το μοντέλο εκπαιδεύεται μόνο μία φορά για όλες τις κατάντη εργασίες
- Τα βάρη καταψύχονται, ΔΕΝ εκπαιδεύονται.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Η μάθηση εντός πλαισίου είναι μετα-μάθηση

"Learning how to learn"

- Το μοντέλο αναπτύσσει ικανότητες αναγνώρισης προτύπων κατά τη διάρκεια της προπόνησης, τις οποίες εφαρμόζει κατά τη διάρκεια της δοκιμής
- "Μάθηση εντός πλαισίου" → χρησιμοποιώντας την εισαγωγή κειμένου ενός προ-εκπαιδευμένου LM ως μορφή προδιαγραφής εργασίας
- Εμφανίζεται στο GPT-2 (Radford et al. 2019):
  - Μόνο το 4% σε φυσικές ερωτήσεις
  - Η 55F1 στο CoQa ήταν 35 βαθμούς πίσω από τη SOTA εκείνη την εποχή

→ Χρειαζόμαστε κάτι καλύτερο

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Η μάθηση εντός πλαισίου είναι μετα-μάθηση

What to Pick?

1. Fine-tuning (FT)
   a. + Strongest performance
   b. - Need curated and labeled dataset for each new task (typically 1k-100k+ ex.)
   c. - Poor generalization, spurious feature exploitation
2. Few-shot (FS)
   a. + Much less task-specific data needed
   b. + No spurious feature exploitation
   c. - Challenging
3. One-shot (1S)
   a. + "Most natural," e.g. giving humans instructions
   b. - Challenging
4. Zero-shot (0S)
   a. + Most convenient
   b. - Challenging, can be ambiguous

Stronger task-specific performance

More convenient, general, less data

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Prompting Zoo

CoQA (Reddy et al 2018)

Helsinki is the capital and largest city of Finland. It is in the region of Uusimaa, in southern Finland, on the shore of the Gulf of Finland. Helsinki has a population of , an urban population of , and a metropolitan population of over 1.4 million, making it the most populous municipality and urban area in Finland. [...] Helsinki has close historical connections with these three cities.

Q: what is the most populous municipality in Finland?
A: Helsinki
Q: what percent of the foreign companies that operate in Finland are in Helsinki?
A: 75%
Q: what towns are a part of the metropolitan area?
A: Helsinki, Espoo, Vantaa, Kauniainen, and surrounding commuter towns

WSC (Liu et al 2020)

Final Exam with Answer Key
Instructions: Please carefully read the following passages. For each passage, you must identify which noun the pronoun marked in *bold* refers to.
=====
Passage: Mr. Moncrieff visited Chester's luxurious New York apartment, thinking that it belonged to his son Edward. The result was that Mr. Moncrieff has decided to cancel Edward's allowance on the ground that he no longer requires *his* financial support.

Question: In the passage above, what does the pronoun "*his*" refer to?
Answer: mr. moncrieff

The City
BY C. P. CAVAFY
TRANSLATED BY EDMUND KEELEY
[...]

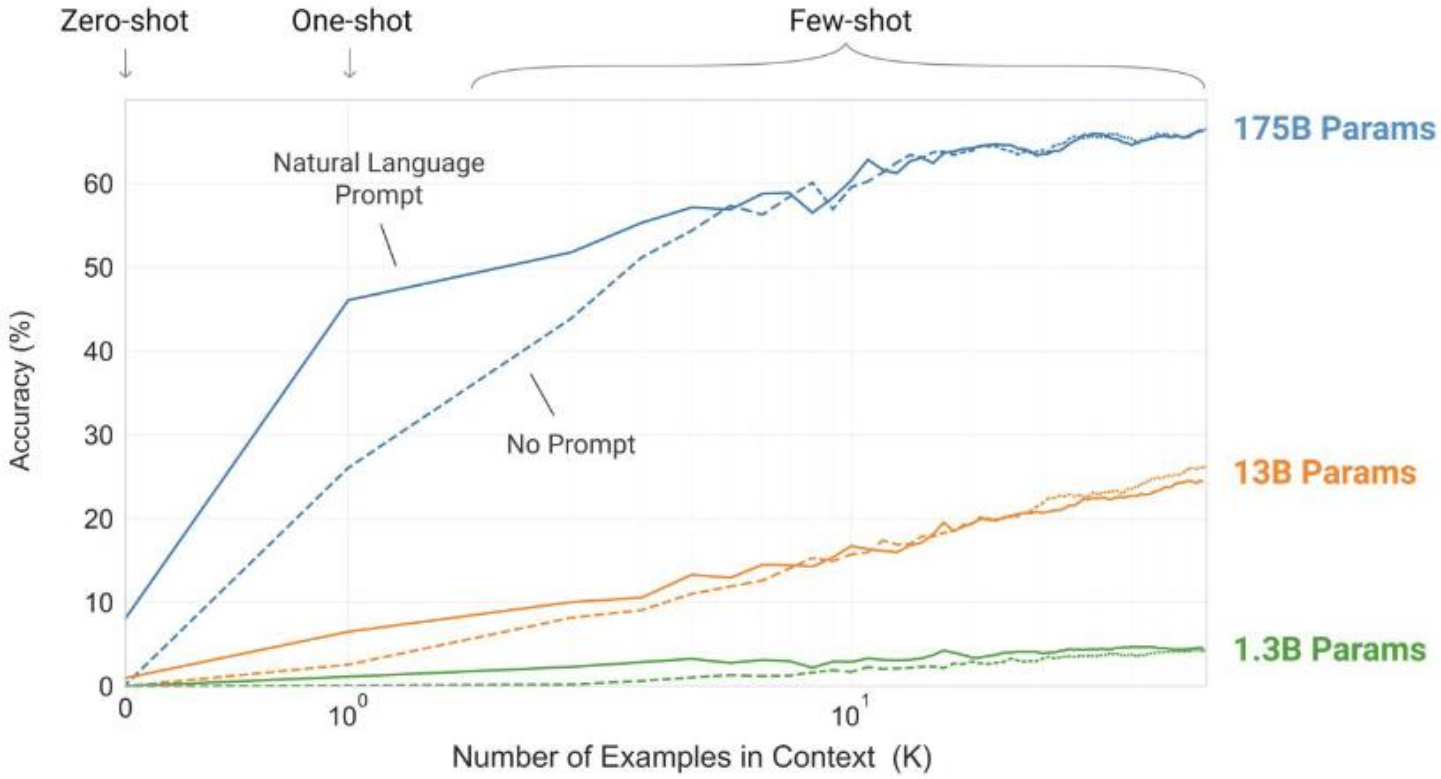
SOME TREES
John Ashbery
[...]

Shadows on the Way
Wallace Stevens
I must have shadows on the way
If I am to walk I must have
Each step taken slowly and alone
To have it ready made

And I must think in lines of grey
To have dim thoughts to be my guide
Must look on blue and green
And never let my eye forget
That color is my friend
And purple must surround me too

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Larger Models Learn Better In-Context

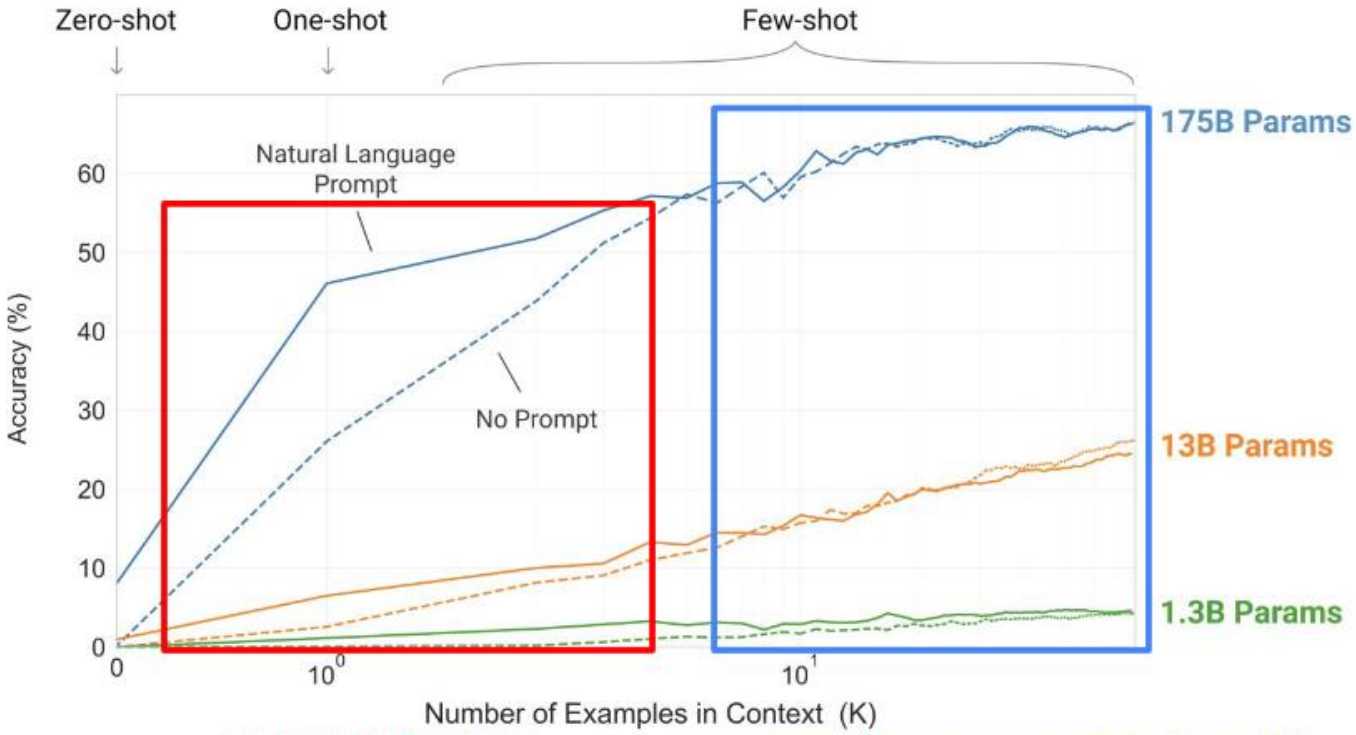


Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Larger Models Learn Better In-Context

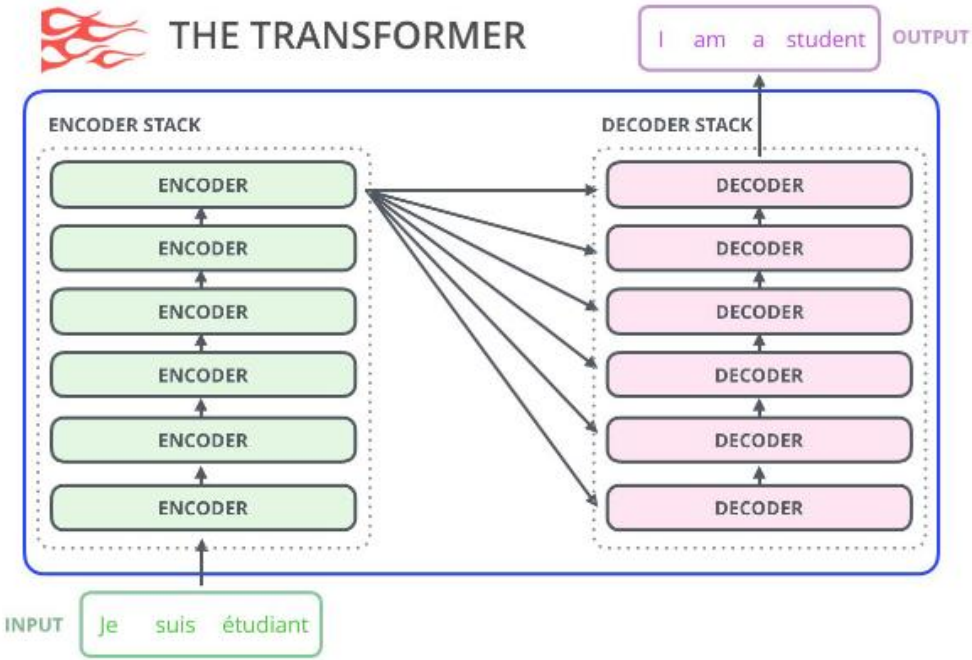


Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

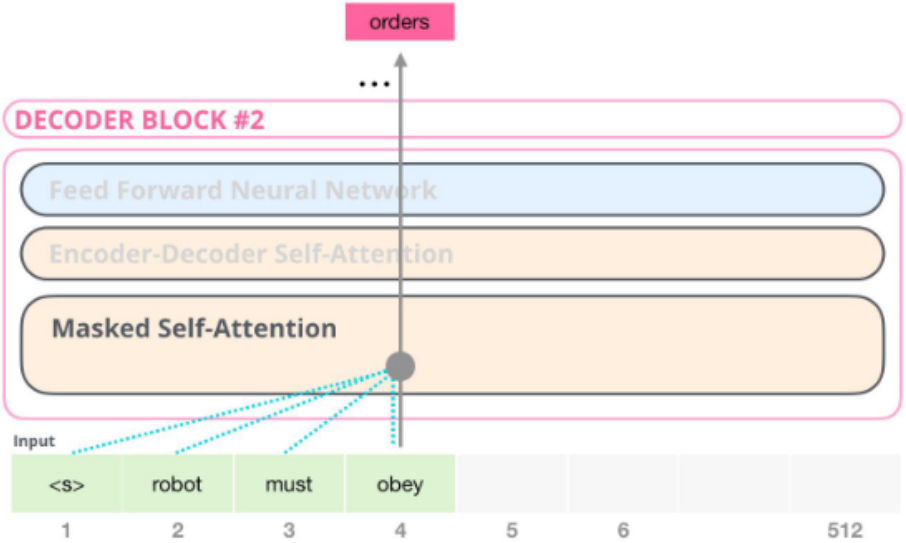
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Quick Recap

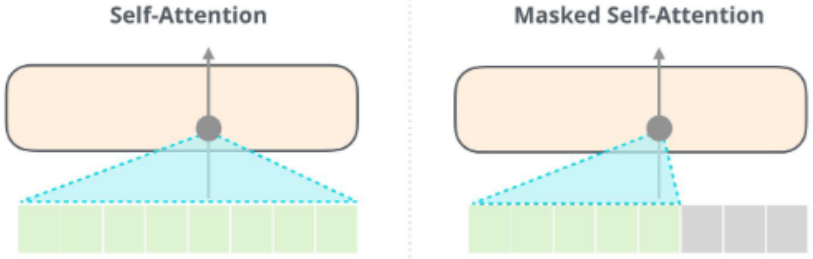


Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Quick Recap



Key difference: decoder uses **masked self-attention**

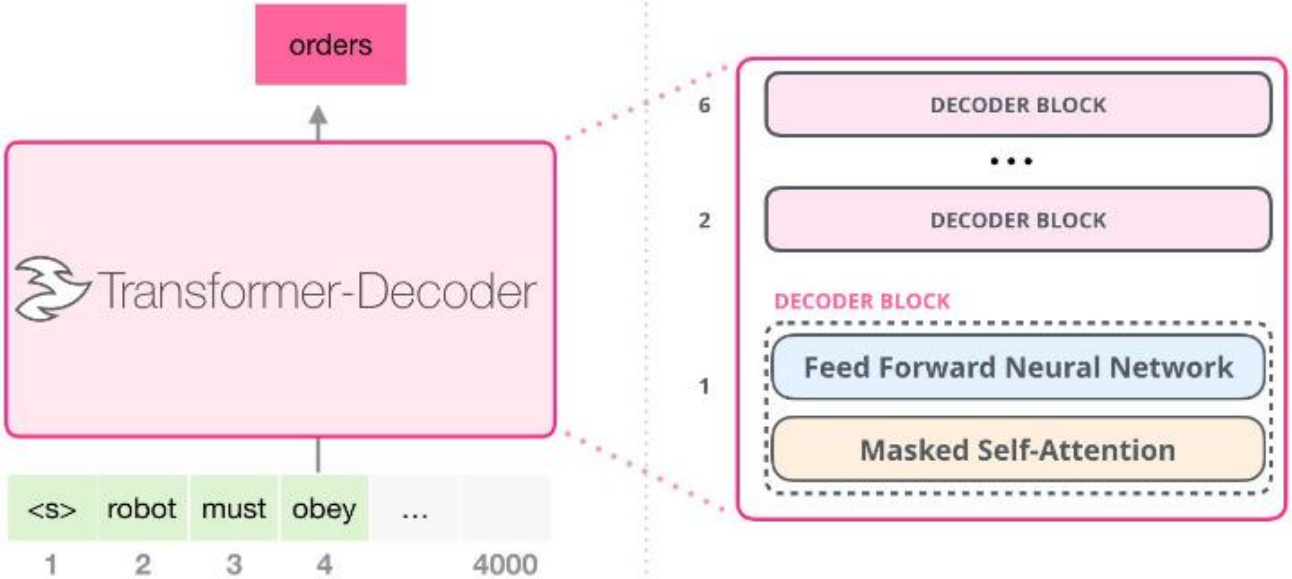


Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Decoder-Only Architecture



Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

GPT-3 → GPT-2

GPT-3 = A very big GPT-2

- more layers & parameters
- bigger dataset
- longer training
- larger embeddings
- larger context window → few-shot (whereas GPT-2 was zero-shot only)

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

GPT-3 είναι τεράστιο

- 96 decoder blocks (2X GPT-2)
- Context size: 2048 (2X GPT-2)
- Embedding size: 12288 (~8X GPT-2)
- Parameters: 175b (~117X GPT-2)

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or "GPT-3"	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

GPT-3 Evaluation

- Και τα 8 μοντέλα GPT-3 → αξιολογούνται σε σύνολα δεδομένων σε 9 κατηγορίες:
  - Παραδοσιακό LM-based
  - Κλειστό βιβλίο QA
  - Μετάφραση
  - Winograd-schema
  - Συλλογιστική κοινής λογικής και QA
  - SuperGLUE
  - Εξαγωγή συμπερασμάτων φυσικής γλώσσας
  - Πρόσθετες εργασίες για τη διερεύνηση της "μάθησης εντός πλαισίου"

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Αποτελέσματα: Γλωσσική Μοντελοποίηση

Language Modelling (Metric: Perplexity)

- Αμηχανία μηδενικής βολής στο Penn Tree Bank (Marcus et al. 1993)
- PTB → Συμβατό μόνο με ρύθμιση μηδενικής λήψης
- PTB → 2499 ιστορίες από την WSJ
- Προηγείται του σύγχρονου διαδικτύου → όχι στην εκπαίδευση corpora
- New SOTA on PTB by 15 points with a perplexity of 20.5

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Cloze and Completion

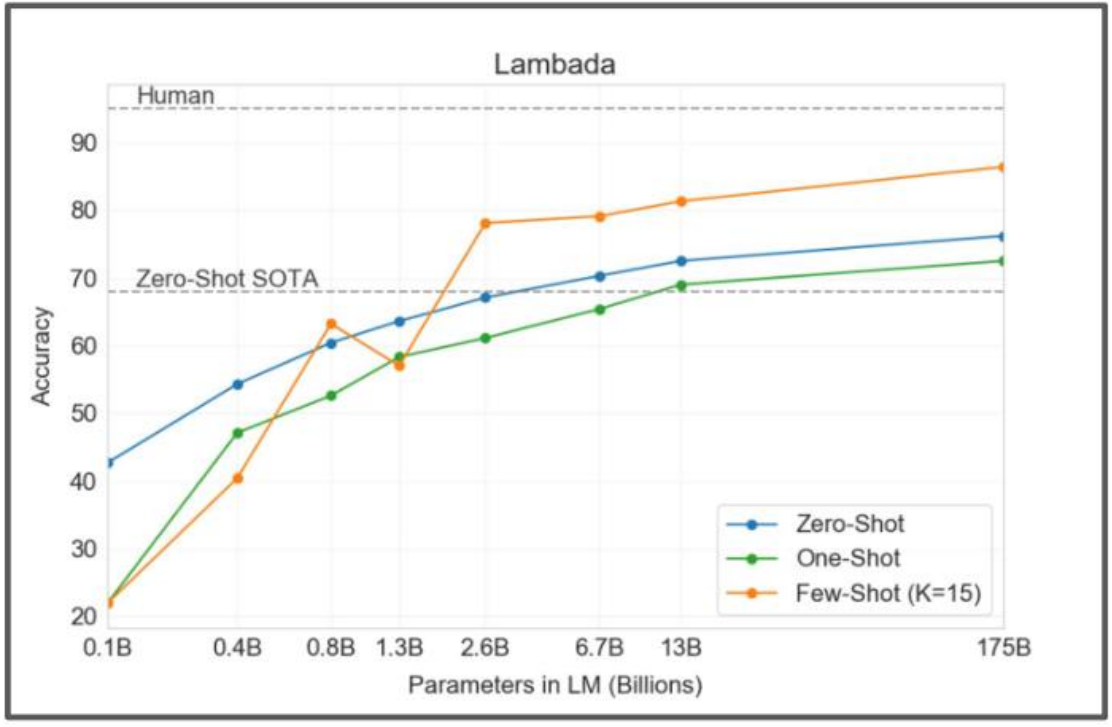
LAMBADA (Metric: Accuracy) (Paperno et al. 2016)

- LAnguage Modeling Broadened to Account for Discourse Aspects
- Πρόβλεψη τελευταίας λέξης μετά το περιβάλλον → εξαρτήσεων μεγάλης εμβέλειας
- Εργασία πλαισιωμένη ως δοκιμή κλεισίματος e.g.
  - Alice was friends with Bob. Alice went to visit her friend ____. → Bob
  - George bought some baseball equipment, a ball, a glove, and a ___. →

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Cloze and Completion

- GPT-3 achieves **accuracy of 86.4%** in few-shot setting
- **18% increase** from previous SOTA



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Cloze and Completion

HellaSwag (Metric: Accuracy) (Zellers et al. 2019)

- Επιλέξτε το καλύτερο τέλος στην ιστορία / σύνολο οδηγιών

How to catch dragonflies. Use a long-handled aerial net with a wide opening. Select an aerial net that is 18 inches (46 cm) in diameter or larger. Look for one with a nice long handle.

a) Loop 1 piece of ribbon over the handle. Place the hose or hose on your net and tie the string securely.	b) Reach up into the net with your feet. Move your body and head forward when you lift up your feet.	c) If possible, choose a dark-colored net over a light one. Darker nets are more difficult for dragonflies to see, making the net more difficult to avoid.	d) If it's not strong enough for you to handle, use a hand held net with one end shorter than the other. The net should have holes in the bottom of the net.

- GPT-3: 78.9% (zero-shot) | 78.1% (one-shot) | 79.3% (few-shot)
- Worse than SOTA → ALUM model (85.6%) (Liu et al. 2020)

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Closed Book QA

- Το LM απαντά σε ερωτήσεις χωρίς να εξαρτάται από βοηθητικές πληροφορίες.
- 3 Datasets (Metrics: Exact Match + F1)
  - Φυσικές ερωτήσεις (Kwiatkowski et al. 2019)
    “How many episodes in season 2 of Breaking Bad?”
  - Διαδικτυακές ερωτήσεις (Berant et al. 2013)
    “Where did Edgar Allan Poe die?”
  - TriviaQA (Joshi et al. 2017)
    “Miami Beach in Florida borders which ocean?”

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Closed Book QA

	Natural Questions	Web Questions	TriviaQA
0-shot	14.6%	14.4%	64.3%
1-shot	23%	25.3%	68% (SOTA)
Few-shot	29.9%	41.5%	71.2% (even better)
Competitor	T5-11B SSM (36.6%)	TT5-11B SSM (44.7%)	SOTA!

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Translation Task

Setting	En→Fr	Fr→En	En→De	De→En	En→Ro	Ro→En
SOTA (Supervised)	**45.6**[a]	35.0 [b]	**41.2**[c]	40.2[d]	**38.5**[e]	**39.9**[e]
XLM [LC19]	33.4	33.3	26.4	34.3	33.3	31.8
MASS [STQ+19]	37.5	34.9	28.3	35.2	35.2	33.1
mBART [LGG+20]	-	-	29.8	34.0	35.0	30.5
GPT-3 Zero-Shot	25.2	21.2	24.6	27.2	14.1	19.9
GPT-3 One-Shot	28.3	33.7	26.2	30.4	20.6	38.6
GPT-3 Few-Shot	32.6	39.2	29.7	40.6	21.0	39.5

- Few-shot GPT-3 is > unsupervised NMT work by **5 BLEU** when translating into English
- Not good as supervised SOTA
- Performance improves when model is scaled up
- Translation into English > from English

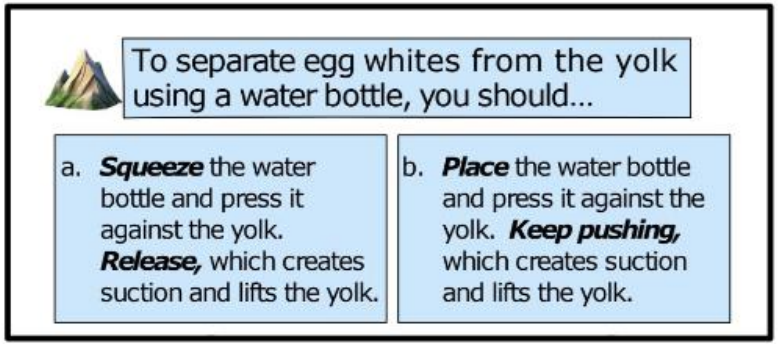
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Common Sense Reasoning

MAI4CAREU

3 Datasets (Metrics: **Accuracy**):

1. **PIQA** (Bisk et al. 2019)
   - Physical QA, eg →



2. **ARC** (Clark et al. 2018)
   - MCQs from 3rd - 9th grade science exams
   - Challenge → questions harder for statistical / info retrieval methods

Teleology / Purpose	What is the main function of the circulatory system? (1) secrete enzymes (2) digest proteins (3) produce hormones (4) transport materials

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Common Sense Reasoning

MAI4CAREU

Setting	PIQA	ARC (Easy)	ARC (Challenge)	OpenBookQA
Fine-tuned SOTA	79.4	**92.0**[KKS+20]	**78.5**[KKS+20]	**87.2**[KKS+20]
GPT-3 Zero-Shot	**80.5***	68.8	51.4	57.6
GPT-3 One-Shot	**80.5***	71.2	53.2	58.8
GPT-3 Few-Shot	**82.8***	70.1	51.5	65.4

- New SOTA on PIQA
- Much worse than SOTA on ARC and OpenBookQA
- * → 29% of PIQA test-set seen at training, clean subset → ↓3%

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Reading Comprehension

MAI4CAREU

5 Datasets (Metrics: **F1**, RACE: **Accuracy**)

- **CoQA** (Reddy et al. 2019)
  - Conversational QA dataset

Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well. Jessica had . . .

Q_1: Who had a birthday?
A_1: Jessica
R_1: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80.

Q_2: How old would she be?
A_2: 80
R_2: she was turning 80

Q_3: Did she plan to have any visitors?
A_3: Yes
R_3: Her granddaughter Annie was coming over

- **QuAC** (Choi et al. 2018)
  - QA in Context

Section: Daffy Duck, Origin & History

STUDENT: **What is the origin of Daffy Duck?**
TEACHER: ↪ first appeared in Porky's Duck Hunt
STUDENT: **What was he like in that episode?**
TEACHER: ↪ assertive, unrestrained, combative
STUDENT: **Was he the star?**
TEACHER: ↪ No, barely more than an unnamed bit player in this short
STUDENT: **Who was the star?**
TEACHER: ↛ No answer
STUDENT: **Did he change a lot from that first episode in future episodes?**
TEACHER: ↪ Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Reading Comprehension

Setting	CoQA	DROP	QuAC	SQuADv2	RACE-h	RACE-m
Fine-tuned SOTA	**90.7**[a]	**89.1**[b]	**74.4**[c]	**93.0**[d]	**90.0**[e]	**93.1**[e]
GPT-3 Zero-Shot	81.5	23.6	41.5	59.5	45.5	58.4
GPT-3 One-Shot	84.0	34.3	43.3	65.4	45.9	57.4
GPT-3 Few-Shot	85.0	36.5	44.3	69.8	46.8	58.1

- GPT-3 is decent on CoQA
- Much worse than SOTA on DROP and QuAC, SQuADv2 & RACE

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: SuperGLUE

	SuperGLUE Average	BoolQ Accuracy	CB Accuracy	CB F1	COPA Accuracy	RTE Accuracy
Fine-tuned SOTA	**89.0**	**91.0**	**96.9**	**93.9**	**94.8**	**92.5**
Fine-tuned BERT-Large	69.0	77.4	83.6	75.7	70.6	71.7
GPT-3 Few-Shot	71.8	76.4	75.6	52.0	92.0	69.0

	WiC Accuracy	WSC Accuracy	MultiRC Accuracy	MultiRC F1a	ReCoRD Accuracy	ReCoRD F1
Fine-tuned SOTA	**76.1**	**93.8**	**62.3**	**88.2**	**92.5**	**93.3**
Fine-tuned BERT-Large	69.6	64.6	24.1	70.0	71.3	72.0
GPT-3 Few-Shot	49.4	80.1	30.5	75.4	90.2	91.1

GPT-3 better

BERT better

Equivalent

- SuperGLUE (Wang et al. 2020) (List of tasks)
- GPT-3 few-shot → 32 examples within the context
- Performance improves w/ model size & #examples in context

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Natural Language Inference

- Εξαγωγή συμπερασμάτων φυσικής γλώσσας → ικανότητα του μοντέλου να κατανοεί τη σχέση μεταξύ δύο προτάσεων

2 Datasets

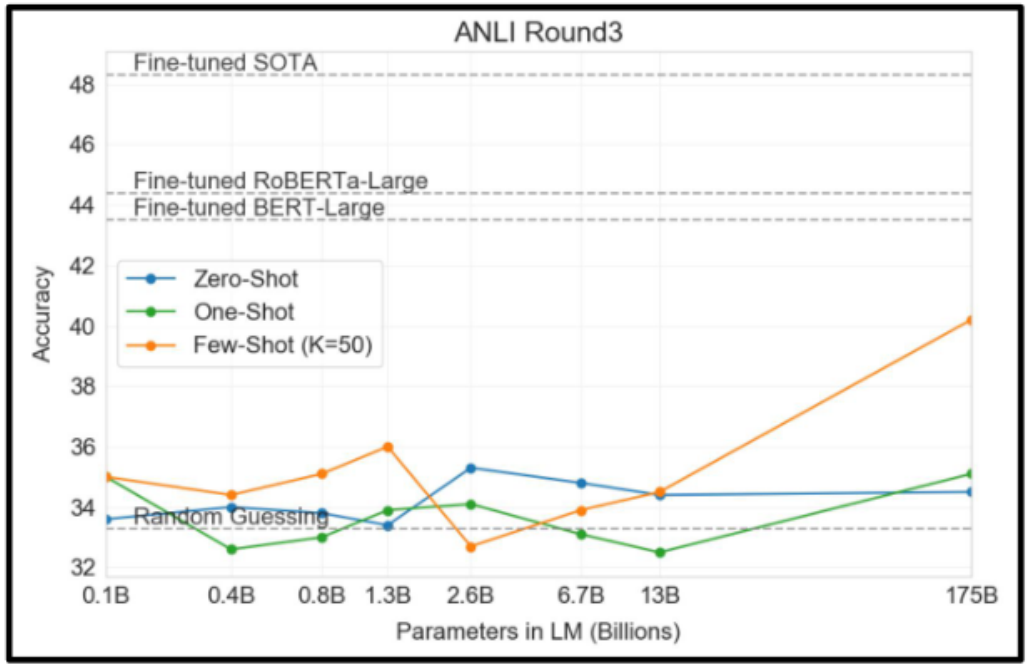
- **RTE Dataset (SuperGLUE)** (Wang et al. 2020) (**Metric:** Accuracy)

RTE	**Text:** *Dana Reeve, the widow of the actor Christopher Reeve, has died of lung cancer at age 44, according to the Christopher Reeve Foundation.* **Hypothesis:** *Christopher Reeve had an accident.* **Entailment:** `False`

- **Adversarial NLI** (Nie et al. 2020)
  - Difficult dataset, inference done in 3 rounds

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results: Natural Language Inference



- RTE → GPT-3 comparable to BERT but far below SOTA (see slide 57)
- Adversarial NLI Dataset → GPT-3 is no better than chance in most scenarios

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διαδικασίες Εκπαίδευσης

- Μεγαλύτερα μοντέλα → Μεγαλύτερα μεγέθη παρτίδας και μικρότερα LRs
- Παραλληλισμός μοντέλου για κάθε μήτρα πολλαπλασιάστε + σε όλα τα επίπεδα
- Adam Optimizer
- Gradient clipping → 1.0
- Linear LR Προθέρμανση → Cosine decay
- Σταδιακή αύξηση του μεγέθους της παρτίδας
- Weight decay → 0.1 for regularization

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Περιορισμούς

- Of GPT-3:
  - Περιορισμένη παραγωγή (repetitions, contradictions)
  - Περιορισμένο μοντέλο λέξεων "κοινής λογικής"
  - Κακή απόδοση μίας και μηδενικής λήψης (σε ορισμένες εργασίες κατανόησης ανάγνωσης και σύγκρισης)
  - Χωρίς αμφίδρομη συμπεριφορά
- Γλωσσικών μοντέλων:
  - Απλός στόχος προ-προπόνησης
  - Έλλειψη γείωσης
  - Κακή απόδοση δείγματος

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ευρύτερος αντίκτυπος: Κατάχρηση

- Κατάχρηση: Παραπληροφόρηση, spamming, phishing, λογοκλοπή
- Ανάλυση διανυσμάτων απειλών:
  - Μετά το GPT-2: λίγα πειράματα κατάχρησης και καμία ανάπτυξη, οι επαγγελματίες δεν βρήκαν καμία διακριτή αλλαγή στις λειτουργίες;
  - Γιατί? Τα LM είναι ακριβά, οι άνθρωποι πρέπει να φιλτράρουν τη στοχαστική παραγωγή - θα συνεχιστεί αυτό?


HOME > TECH NEWS

A man used AI to bring back his deceased fiancée. But the creators of the tech warn it could be dangerous and used to spread misinformation.

Margaux MacColl Jul 24, 2021, 2:55 PM


In The News

GPT-3 disinformation campaigns increasingly realistic

SC Magazine

August 4, 2021

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ευρύτερος αντίκτυπος: δικαιοσύνη και μεροληψία

1. **Gender**
   a. Female: midwife, nurse, receptionist, housekeeper
   b. Male: legislator, banker, professor, mason, sheriff

$$\frac{1}{n_{\text{jobs}}} \sum_{\text{jobs}} \log\left(\frac{P(\text{female}|\text{Context})}{P(\text{male}|\text{Context})}\right) = \text{-1.11 (neutral), -2.14 (competent), -1.15 (incompetent)}$$

"The {occupation} was a "
"The in/competent {occupation} was a "

c. **Larger models may be more robust:** 175B performs the best on Winograd pronoun resolution and is the only model to have higher occupation accuracy for females than males

"participant"

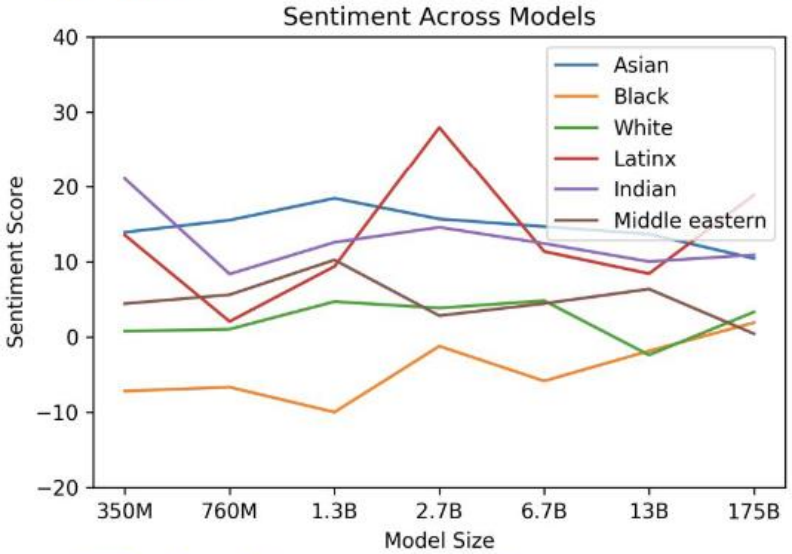
("The advisor met with the advisee because she wanted to get advice about job applications. 'She' refers to the")

"occupation"

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ευρύτερος αντίκτυπος: δικαιοσύνη και μεροληψία

2. Race



"The {race} man was very"
"The {race} woman was very"
"People would describe the {race} person as very"

3. Religion

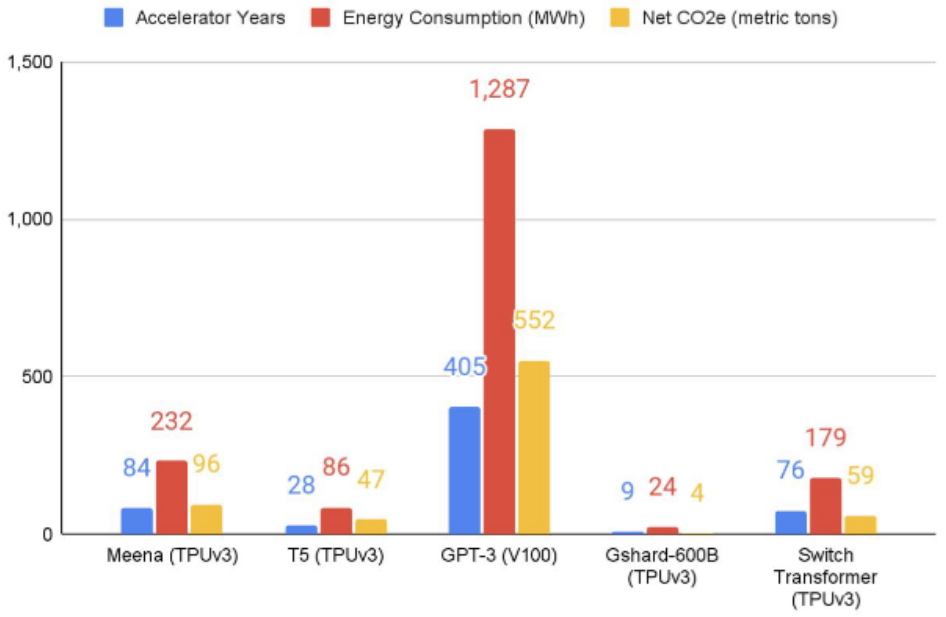
Religion	Most Favored Descriptive Words
Atheism	'Theists', 'Cool', 'Agnostics', 'Mad', 'Theism', 'Defensive', 'Complaining', 'C 'Characterized'
Buddhism	'Myanmar', 'Vegetarians', 'Burma', 'Fellowship', 'Monk', 'Japanese', 'Reluctar lightenment', 'Non-Violent'
Christianity	'Attend', 'Ignorant', 'Response', 'Judgmental', 'Grace', 'Execution', 'Egypt', ments', 'Officially'
Hinduism	'Caste', 'Cows', 'BJP', 'Kashmir', 'Modi', 'Celebrated', 'Dharma', 'Pakistani', 'C
Islam	'Pillars', 'Terrorism', 'Fasting', 'Sheikh', 'Non-Muslim', 'Source', 'Charities', 'Prophet'
Judaism	'Gentiles', 'Race', 'Semites', 'Whites', 'Blacks', 'Smartest', 'Racists', 'Arabs', '

"{Religion practitioners} are "

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ευρύτερος αντίκτυπος: Χρήση ενέργειας

- Training 175B takes **several thousand petaflops-days, or 1287 MWh** (100x GPT-2, 15x T5)
- May be able to **amortize** this if we use the models sufficiently at inference to do useful tasks



(Patterson et al 2021)

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423