



Τεχνητή Νοημοσύνη, οι προκλήσεις για την Έρευνα και την Εκπαίδευση

*19ο Σεμινάριο της Ερμούπολης για την Κοινωνία της
Πληροφορίας & την Οικονομία της Γνώσης, 12-14 Ιουλίου 2024*

ΟΙ ΠΑΡΟΥΣΙΑΣΕΙΣ ΤΗΣ ΕΝΟΤΗΤΑΣ:

- 1. ΓΙΑΝΝΗΣ ΚΑΛΟΓΗΡΟΥ (ΕΜΠ)**, Το τοπίο της τεχνητής νοημοσύνης εν έτει 2024, διάρκεια έως 15 λεπτά.
- 2. ΒΑΣΙΛΗΣ ΜΑΓΚΛΑΡΗΣ (ΕΜΠ)**, Προκλήσεις και Ευθύνες της Ακαδημαϊκής Κοινότητας μπροστά στην Επέλαση της Τεχνητής Νοημοσύνης, διάρκεια έως 15 λεπτά.
- 3. ΠΕΤΡΟΣ ΣΤΕΦΑΝΕΑΣ (ΕΜΠ)**, Τι είναι τελικά η τεχνητή νοημοσύνη; διάρκεια έως 10 λεπτά.
- 4. ΜΑΡΙΝΑ ΡΗΓΟΥ**, Τεχνητή Νοημοσύνη και Δημοσιογραφία, διάρκεια έως 10 λεπτά.
- 5. ΝΙΚΟΣ ΚΑΡΑΜΠΕΚΙΟΣ (ΕΚΤ)**, Επιστημονική και τεχνολογική πολιτική και Τεχνητή Νοημοσύνη.

Μία πρώτη αποτύπωση της μεταξύ τους σχέσης, διάρκεια έως 10 λεπτά.
- 6. ΒΑΣΙΛΗΣ ΚΑΡΚΑΤΖΟΥΝΗΣ (Πανεπιστήμιο Αιγαίου)**, Τεχνητή Νοημοσύνη και online διαφήμιση, διάρκεια έως 10 λεπτά.
- 7. ΓΙΑΝΝΗΣ ΔΑΝΙΑΣ (Vimachem & ΕΜΠ)**, Τεχνητή νοημοσύνη στη βιομηχανία φαρμάκου: Vimachem MES Platform, διάρκεια έως 10 λεπτά.
- 8. ΔΗΜΗΤΡΗΣ ΛΥΜΠΕΡΟΠΟΥΛΟΣ**, Η είσοδος της τεχνητής νοημοσύνης στις εταιρείες πετρελαιοειδών, διάρκεια έως 10 λεπτά.

Τι είναι ένα μεγάλο γλωσσικό μοντέλο;

Ένα μεγάλο γλωσσικό μοντέλο είναι σαν ένα "πολύ-πολύ έξυπνο" παπαγάκι που έχει διαβάσει πάρα πολλά βιβλία και μπορεί να μιλήσει και να καταλάβει την ανθρώπινη γλώσσα.

Οι πολύ βασικές του μονάδες είναι:

1. **Λέξεις:** Τα βασικά δομικά στοιχεία της γλώσσας.
2. **Νευρώνες:** Ψηφιακά "κύτταρα" που επεξεργάζονται πληροφορίες, όπως οι νευρώνες στον ανθρώπινο εγκέφαλο.
3. **Στρώματα:** Ομάδες νευρώνων που συνεργάζονται για να αναλύσουν διαφορετικές πτυχές της γλώσσας.
4. **Συνδέσεις:** Οι "δρόμοι" που συνδέουν τους νευρώνες μεταξύ τους, επιτρέποντας τη ροή πληροφοριών.

Ο ανθρώπινος εγκέφαλος έχει περίπου 86 δισεκατομμύρια νευρώνες και τρισεκατομμύρια συνάψεις. Αν και τα LLMs μπορεί να έχουν συγκρίσιμο αριθμό "νευρώνων", η πολυπλοκότητα και η λειτουργικότητά τους διαφέρουν σημαντικά από τους βιολογικούς νευρώνες. Η αποτελεσματικότητα του μοντέλου εξαρτάται από την αρχιτεκτονική του, την ποιότητα των δεδομένων εκπαίδευσης, και τους αλγόριθμους που χρησιμοποιούνται για την εκπαίδευση και τη βελτιστοποίησή του.

Πολύ μεγάλα γλωσσικά μοντέλα

- ο **GPT-4**: Το πιο προηγμένο μοντέλο της OpenAI, γνωστό για τις εξαιρετικές του ικανότητες σε ένα ευρύ φάσμα εργασιών. URL: <https://openai.com/gpt-4>
- ο **BERT**: Αναπτύχθηκε από την Google, χρησιμοποιείται ευρέως για την κατανόηση φυσικής γλώσσας. URL: <https://github.com/google-research/bert>
- ο **Claude**: Δημιουργήθηκε από την Anthropic, εστιάζει στην ασφαλή και ηθική χρήση. URL: <https://www.anthropic.com>
- ο **LLaMA**: Αναπτύχθηκε από τη Meta AI, διατίθεται σε διάφορα μεγέθη. URL: <https://ai.meta.com/llama>
- ο **BLOOM**: Ένα πολύγλωσσο μοντέλο ανοιχτού κώδικα. URL: <https://huggingface.co/bigscience/bloom>
- ο **Gemini**: Το πιο πρόσφατο μοντέλο της Google DeepMind, ανταγωνιστής του GPT-4. URL: <https://deepmind.google/technologies/gemini/>
- ο **Cohere**: Μια σειρά μοντέλων για επιχειρηματική χρήση, συμπεριλαμβανομένων των Command, Rerank και Embed. URL: <https://cohere.ai/>
- ο **Falcon**: Ανοιχτού κώδικα μοντέλο από το Technology Innovation Institute. URL: <https://falconllm.tii.ae>
- ο **Ernie**: Το μεγάλο γλωσσικό μοντέλο της Baidu, με έμφαση στα κινέζικα. URL: <https://wenxin.baidu.com/ernie-bot> (στα κινέζικα)
- ο **Mixtral 8x22B**: Το πιο πρόσφατο μοντέλο της Mistral AI, χρησιμοποιεί τεχνική Sparse Mixture-of-Experts. URL: <https://mistral.ai/>

Μεγάλα γλωσσικά μοντέλα ανοιχτού λογισμικού

Τα πιο δημοφιλή μεγάλα γλωσσικά μοντέλα ανοιχτού λογισμικού που διαθέτουν και τα δεδομένα τους ανοιχτά είναι τα εξής:

1. LLaMA 2

Αναπτύχθηκε από τη Meta AI, περιλαμβάνει μοντέλα από 7 έως 70 δισεκατομμύρια παραμέτρους. Τα δεδομένα εκπαίδευσης περιλαμβάνουν 2 τρισεκατομμύρια tokens από δημόσια διαθέσιμα δεδομένα.

URL: [LLaMA 2](#)

2. BLOOM

Αναπτύχθηκε από την κοινότητα BigScience, περιλαμβάνει 176 δισεκατομμύρια παραμέτρους και υποστηρίζει 46 φυσικές γλώσσες και 13 γλώσσες προγραμματισμού. Τα δεδομένα εκπαίδευσης περιλαμβάνουν ένα μεγάλο αριθμό δημόσια διαθέσιμων datasets. URL: [BLOOM](#)

3. Mistral 7B

Αναπτύχθηκε από την εταιρεία Mistral AI, έχει 7 δισεκατομμύρια παραμέτρους και είναι διαθέσιμο υπό την άδεια Apache 2.0. Τα δεδομένα εκπαίδευσης περιλαμβάνουν ένα συνδυασμό δημόσια διαθέσιμων δεδομένων. URL: [Mistral 7B](#)

Μικρά και μεσαία γλωσσικά μοντέλα ανοιχτού λογισμικού που αναπτύσσονται από ευρωπαϊκές χώρες.

- ο **Meltemi**: Ένα ελληνικό δίγλωσσο (ελληνικά-αγγλικά) μοντέλο που αναπτύχθηκε από το Ινστιτούτο Επεξεργασίας Λόγου και Γλώσσας του Ερευνητικού Κέντρου "Αθηνά". URL: <https://huggingface.co/ilsp/Meltemi-7B-v1>
- ο **Occiglot**: Μια πρωτοβουλία για την ανάπτυξη πολύγλωσσων μοντέλων για τις ευρωπαϊκές γλώσσες, με έμφαση στις πέντε μεγαλύτερες: αγγλικά, γερμανικά, γαλλικά, ισπανικά και ιταλικά, URL: <https://occiglot.eu>
- ο **Poro**: Ένα φινλανδικό μοντέλο που αναπτύχθηκε από τη Silo AI, με στόχο να υποστηρίξει τη φινλανδική, την αγγλική και γλώσσες προγραμματισμού, <https://huggingface.co/LumiOpen/Poro-34B>

γλωσσΑΡΙ

(<https://github.com/eellak/glossAPI>)

- **Αποστολή:** *Ελληνικό γλωσσικό μοντέλο ανοιχτού λογισμικού «Greek OSS LLM», σε συνεργασία με όλα τα μέλη της ΕΕΛΛΑΚ.*
- Διαφανής και ανθρωποκεντρική ενσωμάτωση της ΤΝ και Γλωσσικών Μοντέλων ...
- **Οφέλη του GlossAPI:** (α) Απόρρητο και ασφάλεια δεδομένων, (β) Εξάρτηση και προσαρμογή, (γ) Κόστος και επεκτασιμότητα, (δ) Πρόσβαση και διαθεσιμότητα.

ΔΗΜΟΣΚΟΠΗΣΕΙΣ:

1. Χρησιμοποιείτε αυτή τη στιγμή εργαλεία ΤΝ στην εργασία σας;

<https://framadata.ellak.gr/ergaliaAI>



2. Ποιο είναι για σας το σημαντικότερο κριτήριο με βάση το οποίο θα επιλέγατε ένα εργαλείο ΤΝ ; <https://framadata.ellak.gr/kritiria>

