

# Εισαγωγή στην Τεχνητή Νοημοσύνη

Μανόλης Κουμπάρκης

# Τι είναι Τεχνητή Νοημοσύνη?

- **Τεχνητή Νοημοσύνη - TN (Artificial Intelligence - AI)** είναι η επιστήμη που έχει στόχο την ανάπτυξη ευφυσών συστημάτων.
- Τι είναι τα **ευφυή συστήματα**?

# Τι είναι Τεχνητή Νοημοσύνη?

- Στη διαδρομή της στο χρόνο, η ΤΝ είχε (και έχει!) στόχο την ανάπτυξη συστημάτων που
  - ... ενεργούν σαν τον άνθρωπο.
  - ... σκέφτονται σαν τον άνθρωπο.
  - ... σκέφτονται ορθολογικά (rationally).
  - ... ενεργούν ορθολογικά.
  - ... αφού εκπαιδευτούν σε ένα μεγάλο σύνολο δεδομένων, μαθαίνουν από αυτά και έτσι γίνονται ευφυή. Αυτή είναι η προσέγγιση της Μηχανικής Μάθησης.

# Ανθρώπινη Δράση: Η Δοκιμασία Turing



- Η **δοκιμασία Turing (Turing test)** προτάθηκε από τον Alan Turing το 1950.
- Για να περάσει αυτή τη δοκιμασία ένας υπολογιστής, πρέπει να έχει τις εξής ικανότητες:
  - επεξεργασία φυσικής γλώσσας (natural language processing)
  - αναπαράσταση γνώσης (knowledge representation)
  - αυτοματοποιημένη συλλογιστική (automated reasoning)
  - μηχανική μάθηση (machine learning)
  - μηχανική όραση (computer vision)
  - ρομποτική (robotics)

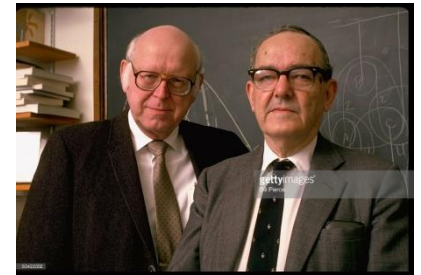
# Πέρα από τη Δοκιμασία Turing

- Οι ερευνητές της TN δεν έχουν αφιερώσει μεγάλη προσπάθεια στη δημιουργία μηχανών που μπορούν να περάσουν τη δοκιμασία Turing. Γιατί;
  - Ένα πρόγραμμα μπορεί απλά να μας κοροϊδέψει χωρίς να είναι ευφυές. Δείτε, για παράδειγμα, το παρακάτω άρθρο για το πρόγραμμα Eugene Goostman το οποίο κατάφερε το 2015 να πείσει 10 στους 30 κριτές ότι είναι άνθρωπος:  
<https://www.theguardian.com/technology/2014/jun/09/scientists-disagree-over-whether-turing-test-has-been-passed>
  - Υπάρχουν αρκετές **πιο συγκεκριμένες δοκιμασίες** για να μετρήσουμε την πρόοδο των σημερινών υπολογιστών. Δείτε την σχετική έκδοση του AI magazine (Vol 37, No. 1, Spring 2016).

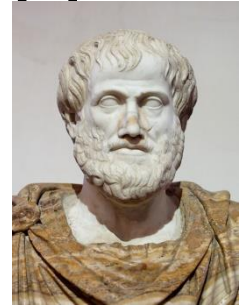


# Ανθρώπινη Σκέψη: Η Προσέγγιση με Γνωστικά Μοντέλα

- Πως δουλεύει ο ανθρώπινος νους?
- Υπάρχουν δύο τρόποι για να το μάθουμε:
  - Ενδοσκόπηση
  - Ψυχολογικά πειράματα
- **Παράδειγμα:** Το πρόγραμμα GPS (General Problem Solver) που αναπτύχθηκε το 1961 από τους Newell και Simon.
- Σ' αυτή την προσέγγιση, η **Ψυχολογία** και η **Γνωστική Επιστήμη (Cognitive Science)** είναι σημαντικές επιστήμες.
- Στους Newell και Simon απονεμήθηκε το **Turing Award** το 1975 για την δουλειά τους στην περιοχή αυτή.



# Ορθολογική Σκέψη: Η Προσέγγιση με τους Νόμους της Σκέψης



- Ποιοί είναι οι **νόμοι της σκέψης**?
- Η μελέτη αυτής της ερώτησης από τον άνθρωπο ξεκίνησε με τους **συλλογισμούς** του Αριστοτέλη.
- **Αριστοτέλειος συλλογισμός:** Ο Σωκράτης είναι άνθρωπος. Όλοι οι άνθρωποι είναι θνητοί. Επομένως, ο Σωκράτης είναι θνητός.
- Η έμφαση σ' αυτή την προσέγγιση είναι στη σωστή **εξαγωγή συμπερασμάτων (inference)**.
- Σημαντικό ρόλο παίζουν οι επιστήμες της Φιλοσοφίας και της Λογικής.
- **Παράδειγμα:** η έρευνα στην περιοχή της **απόδειξης θεωρημάτων (theorem proving)**

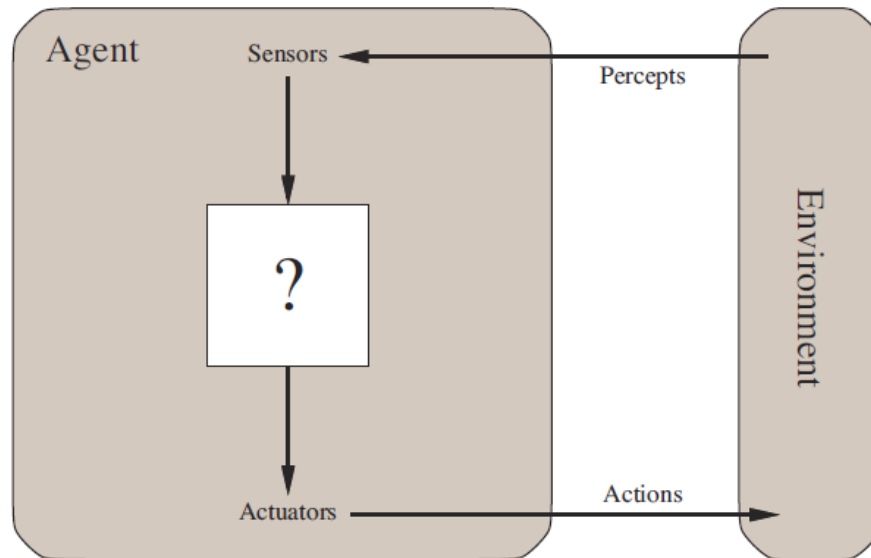
# Ορθολογική Δράση: Η Προσέγγιση με τους Ορθολογικούς Πράκτορες

- Σ' αυτή την προσέγγιση το βασικό πρόβλημα είναι η σχεδίαση **ορθολογικών πρακτόρων**.
- Τι είναι πράκτορας?



# Πράκτορες

- **Πράκτορας (agent)** είναι οτιδήποτε μπορεί να θεωρηθεί ότι αντιλαμβάνεται το περιβάλλον του μέσω αισθητήρων (sensors) και αλληλεπιδρά με αυτό μέσω μηχανισμών δράσης (effectors - actuators).



# Παραδείγματα Πρακτόρων

- Ανθρώπινοι



- Ρομποτικοί



- Πράκτορες Λογισμικού  
(software agents - software robots - softbots).



# Ορθολογικοί Πράκτορες



- **Ορθολογικός πράκτορας** είναι ένας πράκτορας που ενεργεί έτσι ώστε να επιτυγχάνει το καλύτερο αποτέλεσμα ή, όταν υπάρχει αβεβαιότητα, το καλύτερο αναμενόμενο αποτέλεσμα.
- Η μελέτη της TN ως **σχεδίασης ορθολογικών πρακτόρων** είναι
  - πιο γενική από την προσέγγιση που βασίζεται στους νόμους της σκέψης
  - ευκολότερη από τις προσεγγίσεις που βασίζονται στην ανθρώπινη σκέψη ή δράση

# Μηχανική Μάθηση

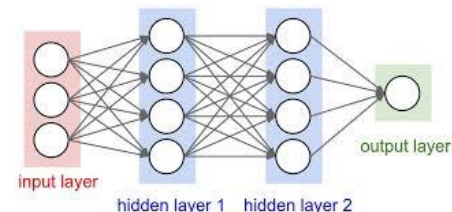
- **Μηχανική Μάθηση** είναι η επιστήμη (και μερικές φορές η τέχνη!) του σχεδιασμού αλγορίθμων που μαθαίνουν από δεδομένα.
- Η Μηχανική Μάθηση είναι μια σημαντική **υποπεριοχή της Τεχνητής Νοημοσύνης, ίσως η σημαντικότερη σήμερα.**
- Κάθε μέρα χρησιμοποιούμε **εφαρμογές** που έχουν αναπτυχθεί χρησιμοποιώντας τεχνικές Μηχανικής Μάθησης π.χ.:
  - Αναγνώριση ομιλίας
  - Κατανόηση φυσικής γλώσσας
  - Μηχανική μετάφραση
  - Αυτοοδηγούμενα οχήματα
- Είναι δύσκολο να ονομάσουμε ένα σύστημα **ευφυές** αν δεν μαθαίνει.

# Μηχανική Μάθηση για την Επεξεργασία Φυσικής Γλώσσας

- Στο μάθημα «Τεχνητή Νοημοσύνη II (Deep learning for natural language processing)» δίνουμε μια **σύγχρονη εισαγωγή στη Μηχανική Μάθηση** επικεντρωνόμενοι στις τεχνολογίες των **νευρωνικών δικτύων**.
- Μελετούμε επίσης εφαρμογές της Μηχανικής Μάθησης στην **Επεξεργασία Φυσικής Γλώσσας** που είναι μια άλλη σημαντική υποπεριοχή της Τεχνητής Νοημοσύνης.
- Σήμερα τα περισσότερα προβλήματα στην περιοχή της Επεξεργασία Φυσικής Γλώσσας λύνονται με αλγόριθμους Μηχανικής Μάθησης (και συνήθως **βαθιές αρχιτεκτονικές νευρωνικών δικτύων**).

# Μηχανική Μάθηση

- Τα τελευταία χρόνια η υποπεριοχή της ΤΝ με τα πιο εντυπωσιακά αποτελέσματα είναι η **μηχανική μάθηση**.
- Η βασική ιδέα εδώ είναι ότι **εκπαιδεύουμε ένα σύστημα σε μεγάλο όγκο δεδομένων** (π.χ., εικόνες, ήχους ή κείμενο) και το σύστημα μετά μπορεί να λύνει σχετικά προβλήματα με επιτυχία (πολλές φορές μάλιστα, καλύτερα από τους ανθρώπους).



# Ιστορία της ΤΝ

- **Κυοφορία (1943-1955)**

- Μοντέλα τεχνητών νευρώνων - perceptron (McCulloch and Pitts, 1943). Μάθηση Hebb (Hebb, 1949).
- Το άρθρο του Alan Turing “Computing Machinery and Intelligence” (1950).
- Sarc: Ο πρώτος υπολογιστής νευρωνικού δικτύου (Minsky and Edmonds, 1951).
- Ο Minsky έλαβε επίσης το **Turing Award** το 1969.



# Ιστορία της ΤΝ



- **Γέννηση (1956)**

- Το συνέδριο εργασίας στο πανεπιστήμιο Dartmouth το καλοκαίρι του 1956 (McCarthy, Minsky, Newell, Simon).
- Ο όρος “Artificial Intelligence” υιοθετήθηκε μετά από πρόταση του McCarthy (**Turing Award** winner in 1971).
- Το πρόγραμμα Logic Theorist (Newell and Simon) έκλεψε την παράσταση. Αμέσως μετά το συνέδριο, το πρόγραμμα αυτό μπόρεσε να αποδείξει τα περισσότερα θεωρήματα του Κεφ. 2 του βιβλίου Principia Mathematica (ένα φημισμένο βιβλίο των Alfred North Whitehead και Bertrand Russell για τη λογική και τα μαθηματικά)!



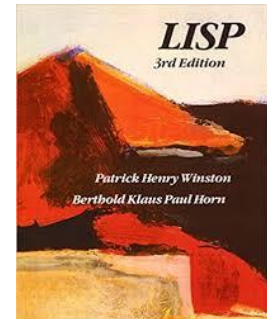


# Ιστορία της ΤΝ



- **Πρώιμος ενθουσιασμός, μεγάλες προσδοκίες (1952-1969)**

- Logic Theorist, General Problem Solver, Geometry Theorem Prover, game playing, Lisp (from McCarthy), theorem proving, Shakey the robot, micro-worlds, adalines, perceptrons.



- Η υπόθεση του **φυσικού συστήματος συμβόλων (physical symbol system hypothesis)** των Newell and Simon: Οποιοδήποτε σύστημα (μηχανή ή άνθρωπος) που εκδηλώνει νοημοσύνη θα πρέπει να λειτουργεί με το χειρισμό δομών δεδομένων που αποτελούνται από σύμβολα. Αυτή είναι η υπόθεση εργασίας για τους ερευνητές της **συμβολικής ΤΝ.**

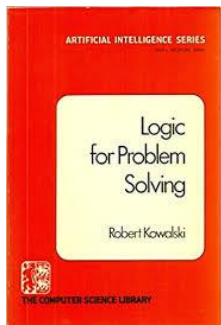
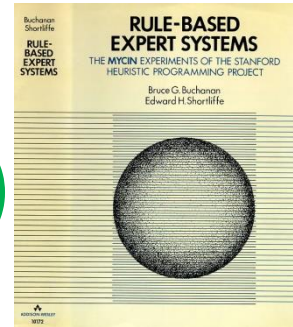
# Ιστορία της ΤΝ



- **Μια δόση ρεαλισμού (1966-1973)**
  - Προγράμματα χωρίς γνώση πεδίου (domain knowledge), προβλήματα δισεπιλυσιμότητας (intractability problems), αρνητικά αποτελέσματα για απλά νευρωνικά δίκτυα.
  - Ακύρωση μεγάλων ερευνητικών έργων σε μηχανική μετάφραση (ΗΠΑ), η έκθεση του Lighthill (Μεγάλη Βρετανία).

# Ιστορία της ΤΝ


- **Συστήματα βασισμένα στη γνώση (knowledge-based systems) (1969-1979)**
  - Ο ρόλος της γνώσης πεδίου, τα έμπειρα συστήματα (expert systems) π.χ., το πρόγραμμα MYCIN για την αναγνώριση βακτηρίων που προκαλούν ιώσεις.
  - Γλώσσες αναπαράστασης γνώσης και συλλογιστικής (π.χ., Prolog, frame-based languages).



# Ιστορία της ΤΝ

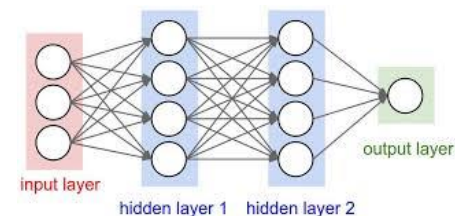
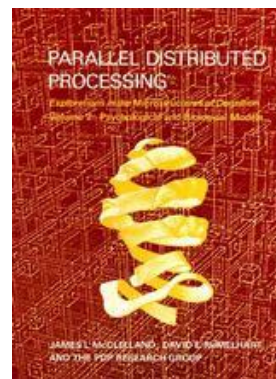
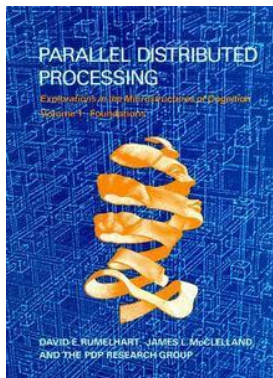


- **Η ΤΝ γίνεται βιομηχανία (1980-σήμερα)**

- Το πρώτο επιτυχημένο έμπειρο σύστημα: R1 (McDermott, DEC).
- Η Ιαπωνική δράση “5th Generation project” (1981) και η έμφαση της στο Λογικό Προγραμματισμό. 
- Η εταιρία Microelectronics and Computer Technology Corporation (MCC) των ΗΠΑ.
- Το πρόγραμμα ESPRIT της Ευρωπαϊκής Ένωσης (1983).
- Η έκθεση Alvey στη Μεγάλη Βρετανία.

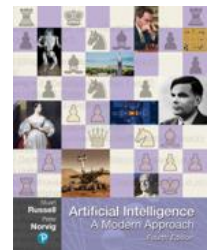
# Ιστορία της ΤΝ

- **Η επιστροφή των νευρωνικών δικτύων (1986-σήμερα)**
  - Τουλάχιστον τέσσερις ερευνητικές ομάδες ανακάλυψαν ξανά την μέθοδο μάθησης με **οπισθοδιάδοση (backpropagation)** που είχε προταθεί το 1969.
  - Τα **συνδεσμικά μοντέλα (connectionist models)** που βασίζονται σε νευρωνικά δίκτυα ως συμπλήρωμα των συμβολικών μοντέλων.



# Ιστορία της ΤΝ

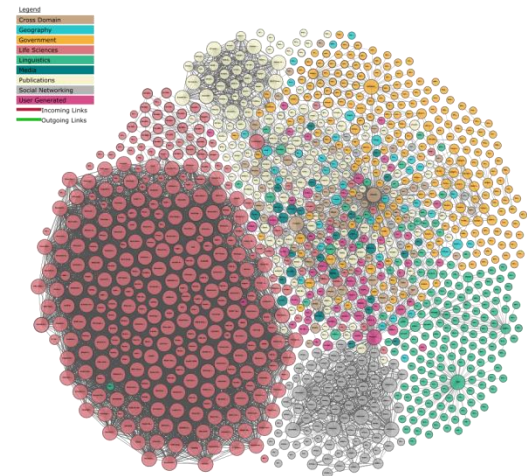
- **Η Τεχνητή Νοημοσύνη γίνεται επιστήμη (1987-σήμερα)**
  - Μεθοδικοί εναντίον τσαπατσούληδων (neats vs. scruffies).
  - Ενδιαφέροντα αποτελέσματα σε αναπαράσταση γνώσης, αναγνώριση ομιλίας, νευρωνικά δίκτυα, εξόρυξη δεδομένων, δίκτυα Bayes, ρομποτική και μηχανική όραση.
- **Ευφυείς Πράκτορες (1995-σήμερα)**
  - Δείτε το βιβλίο του εισαγωγικού μαθήματος Τεχνητή Νοημοσύνη και το συνέδριο AAMAS (<https://aamas2023.soton.ac.uk/>) .



# Ιστορία της ΤΝ

- **Σημασιολογικός Ιστός (Semantic Web) και Διασυνδεδεμένα Δεδομένα (Linked Data) (1998-σήμερα)**

- Δείτε την ιστοσελίδα του μεταπτυχιακού μου μαθήματος Τεχνολογία Γνώσεων (<http://cgi.di.uoa.gr/~pms509/>) και την παρακάτω ιστοσελίδα του W3C: <https://www.w3.org/2013/data/>



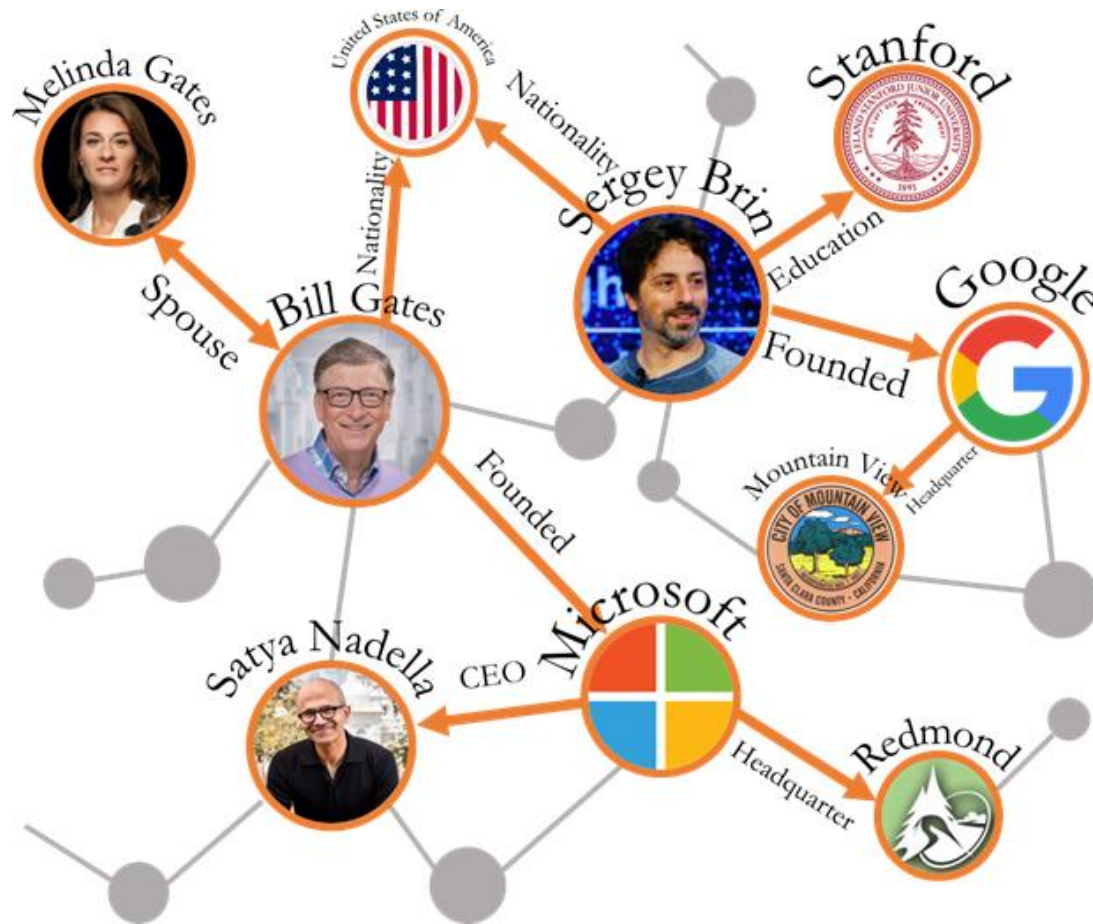
# Ιστορία της TN

- Χρήση γράφων γνώσεων (knowledge graphs) στις μηχανές αναζήτησης (2012-σήμερα)
  - Δείτε πληροφορίες σχετικά με το Google Knowledge Graph στην ιστοσελίδα:  
<https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html>





# Παράδειγμα Γράφου Γνώσης



# Η Χρήση των Γράφων Γνώσης στις Σύγχρονες Μηχανές Αναζήτησης

The image shows a Google search interface. At the top left is the Google logo. The search bar contains the text "who founded google?". To the right of the search bar are icons for a close button (X), a keyboard, and a search icon. Below the search bar are navigation links: "Όλα", "Εικόνες", "Ειδήσεις", "Βίντεο", "Χάρτες", "Περισσότερα", "Ρυθμίσεις", and "Εργαλεία".

The search results display a knowledge panel for "Google > Ιδρυτές". It features two entries: "Λάρρυ Πέιτζ" with a small portrait photo and "Σεργκί Μπριν" with another small portrait photo.

Below the knowledge panel is a snippet from Wikipedia: "en.wikipedia.org > wiki ▾ Μετάφραση αυτής της σελίδας Google - Wikipedia Google, LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising ...". It also lists "Founded: September 4, 1998; 22 years ago ..." and "Number of employees: 114,096 (Q3 2019)".

At the bottom of the search results is a box labeled "Άλλες ερωτήσεις χρηστών".

On the right side of the page is a "Google" company card. It features the Google logo, a share icon, and the word "Εταιρεία". Below the card is a short description: "Η Google LLC είναι μία από τις μεγαλύτερες ε διαδικτυακών υπηρεσιών. Ιδρύθηκε από τον / Πέιτζ και τον Σεργκί Μπριν το 1996, όταν αυι έκαναν το διδακτορικό τους στο Πανεπιστήμιο".

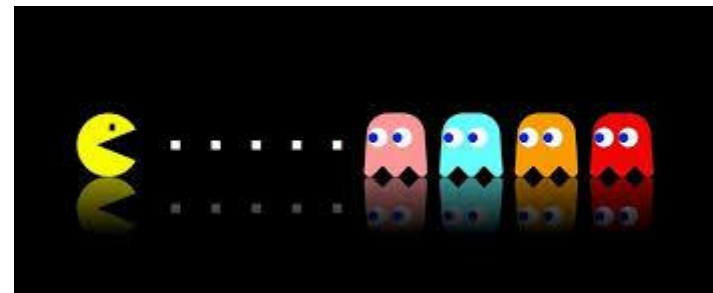
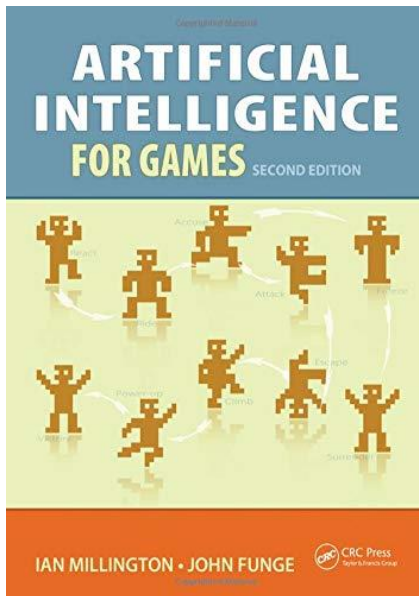
# AI@DI

- Δείτε την ιστοσελίδα της ομάδας μου (**Artificial Intelligence Team**, <https://ai.di.uoa.gr/>) για πρόσφατα αποτελέσματα στις περιοχές Semantic Web, Linked Data, Knowledge Graphs, Question Answering, Deep Learning for Natural Language Processing και άλλα.
- Αν είστε μεταπτυχιακός, παρακολουθήστε το μάθημα **Τεχνολογίες Γνώσεων** (<https://cgi.di.uoa.gr/~pms509/>).

# Ιστορία της ΤΝ

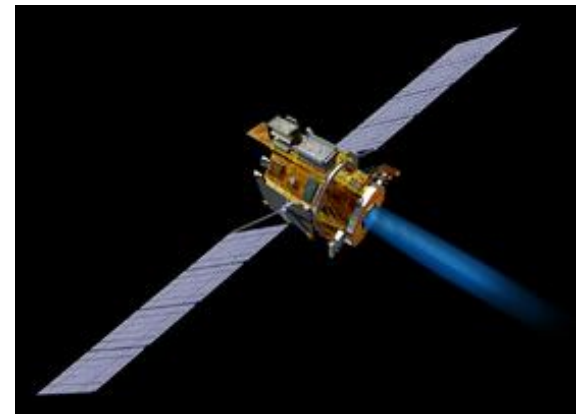
- **TN για computer and video games**

- Δείτε το βιβλίο Artificial Intelligence for Computer Games των Ian Millington και John Funge.



# Ιστορία της ΤΝ

- **Αυτόνομος σχεδιασμός και χρονοπρογραμματισμός (autonomous planning and scheduling)**
  - NASA's Remote Agent (<https://ti.arc.nasa.gov/tech/asr/planning-and-scheduling/remote-agent/>)
  - Ο πράκτορας αυτός είχε τον έλεγχο του διαστημοπλοίου Deep Space 1 κατά τη διάρκεια δύο πειραμάτων το Μάιο του 1999.



# Ιστορία της ΤΝ

- **Σχεδιασμός Εφοδιασμού (logistics planning)**
  - Το πρόγραμμα Dynamic Analysis and Replanning Tool (DART) που χρησιμοποιήθηκε από τις Αμερικανικές δυνάμεις στον Περσικό Κόλπο το 1991.



# Ιστορία της TN

- **Λογισμικό επίλυσης προβλημάτων με περιορισμούς (constraint solving software)**
  - Τα προγράμματα επιλυτές (solvers) της Γαλλικής εταιρίας ILOG που τώρα ανήκει στην IBM (<https://www-01.ibm.com/software/info/ilog/> ).



# Ιστορία της ΤΝ

- **Παιχνίδια (game playing)**

- IBM's Deep Blue in 1997

- (<http://www.research.ibm.com/deepblue/> )



- Google's Deep Mind Alpha Go in 2016

- (<https://deepmind.com/research/alphago/>). It uses deep neural networks και tree search.





# Ιστορία της TN

- **Απάντηση ερωτήσεων εκφρασμένων σε φυσική γλώσσα**

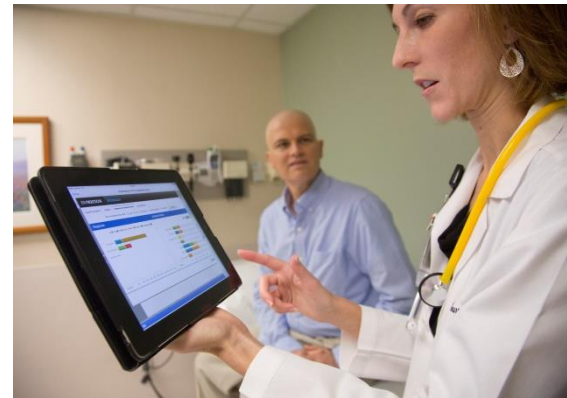
- Το ερευνητικό έργο DeepQA της IBM στα πλαίσια του οποίου αναπτύχθηκε το σύστημα Watson.
- Το σύστημα Watson κέρδισε τους πρωταθλητές (ανθρώπους) στο Αμερικανικό τηλεοπτικό παιχνίδι γνώσεων Jeopardy το 2011.



- Περισσότερες πληροφορίες για το έργο DeepQA δίνονται στο άρθρο: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/2303/2165>
- Δείτε την ιστοσελίδα <https://www.ibm.com/watson> για τις εμπορικές εκδόσεις του συστήματος.

# Ιστορία της TN

- **Ιατρική διάγνωση (medical diagnosis)**
  - Το πρόγραμμα Watson for Oncology της IBM.  
Δείτε <https://www.ibm.com/watson/health/oncology-and-genomics/oncology/>
  - Το πρόγραμμα αυτό χρησιμοποιείται στην Ελλάδα από το νοσοκομείο ΥΓΕΙΑ.



# Ιστορία της TN

- **Καταπολέμηση του spam**

- Χρήση τεχνικών μηχανικής μάθησης για ταξινόμηση μηνυμάτων ηλεκτρονικού ταχυδρομείου ή ιστοσελίδων ως spam.

- Δείτε την ιστοσελίδα

- <https://www.google.com/intl/bn/insidesearch/howsearchworks/fighting-spam.html> .





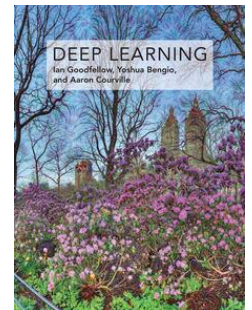
# Υποστηρικτικές Τεχνολογίες

- **Ανάπτυξη κλιμακώσιμων υπολογιστικών συστημάτων (clusters and cloud computing, GPUs).**



# Ιστορία της ΤΝ

- **Μάθηση σε βάθος (deep learning)(2010-σήμερα)**
  - Δημιουργία μεγάλων συνόλων δεδομένων για την εκπαίδευση νευρωνικών δικτύων (π.χ., Wikipedia, ImageNet, Common Crawl, Pile).
  - Χρήση παράλληλων υπολογιστικών αρχιτεκτονικών και GPUs.
  - Βαθιά νευρωνικά και οι αλγόριθμοι backpropagation και gradient descent για την εκπαίδευση τους.
  - Προγραμματιστικά περιβάλλοντα όπως Keras, TensorFlow, Pytorch και Google Colab.



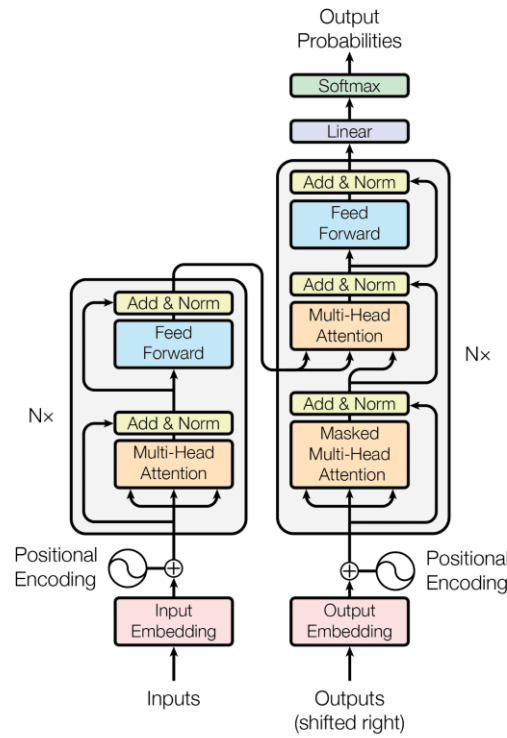
# Ιστορία της ΤΝ

- Απονομή του Turing Award στους Bengio, Hinton και LeCun το 2019.



# Σύγχρονες Τάσεις στη Βαθιά Μηχανική Μάθηση

- Η αρχιτεκτονική των transformers (2017).





# Σύγχρονες Τάσεις

- **Μεγάλα προεκπαιδευμένα γλωσσικά (ή θεμελιώδη) μοντέλα** για την κατανόηση και την παραγωγή φυσικής γλώσσας π.χ., BERT, GPT-2, GPT-3 και άλλα (2018 μέχρι σήμερα).



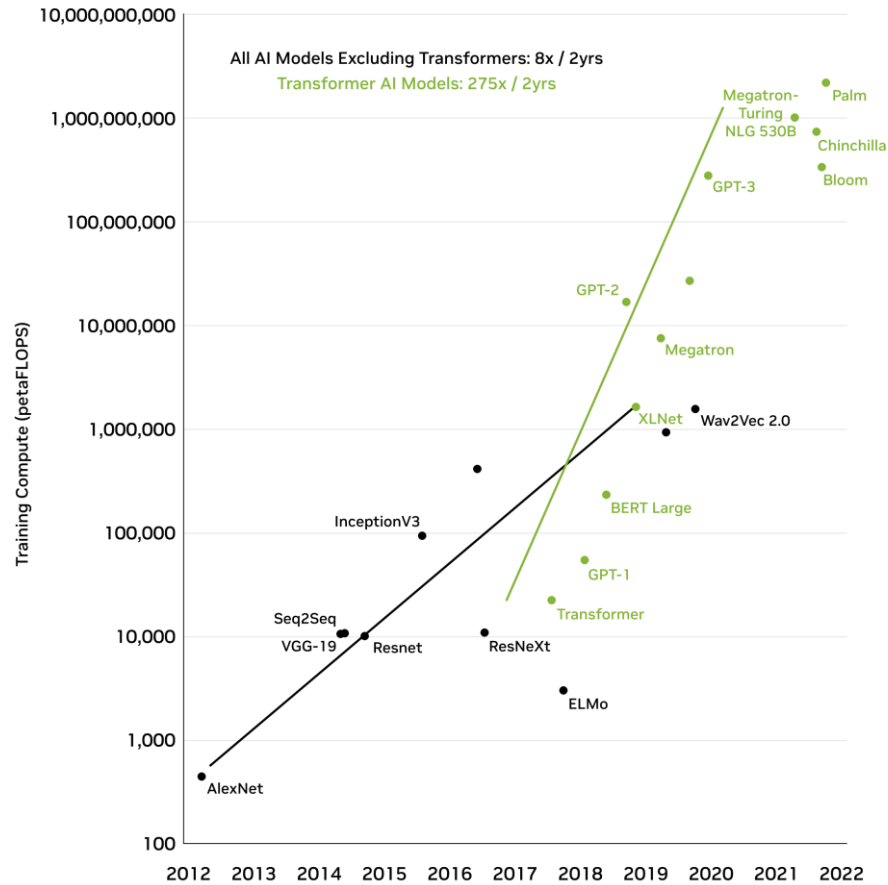
# Θεμελιώδεις Τεχνικές

- Μεγάλος όγκος δεδομένων εκπαίδευσης.
- Μεγάλος αριθμός παραμέτρων (millions+).
- Μεγάλος χρόνος εκπαίδευσης.
- Αυτοεπίβλεψη (self-supervision).
- Προεκπαίδευση-τελειοποίηση (pretraining-finetuning).
- Few-shot learning.
- Prompt engineering.
- Instruction finetuning.

# Δεδομένα Εκπαίδευσης

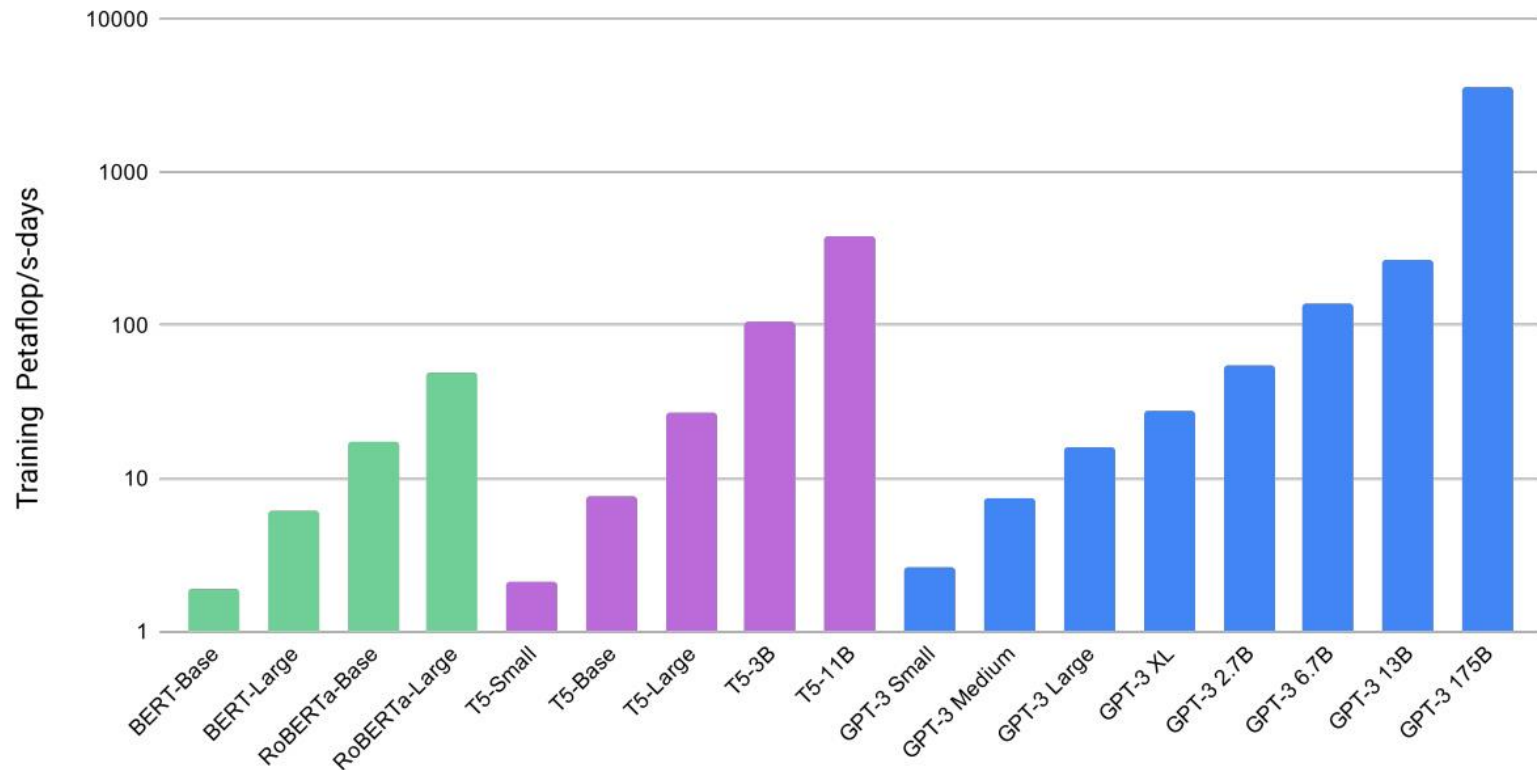
Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

# Υπολογισμοί



# Υπολογισμοί

Total Compute Used During Training



A **petaflop/s-day** consists of performing  $10^{15}$  neural net operations per second for one day, or a total of about  $10^{20}$  operations.

# Παράδειγμα Κειμένου που Παρήχθη από το GPT-3

Title: United Methodists Agree to Historic Split

Subtitle: Those who oppose gay marriage will form their own denomination

Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

# GPT-3 on GPT-3 😊

- Δείτε το άρθρο <https://hal.archives-ouvertes.fr/hal-03701250/document> που γράφτηκε με τη βοήθεια του GPT-3 για τον εαυτό του!

# Σύγχρονες Τάσεις

- **Μεγάλα προεκπαιδευμένα γλωσσικά μοντέλα για την παραγωγή κώδικα** σε διάφορες γλώσσες προγραμματισμού.
- Δείτε για παράδειγμα το σύστημα GitHub Copilot (<https://github.com/features/copilot>) το οποίο χρησιμοποιεί το γλωσσικό μοντέλο κώδικα Codex της OpenAI.





# InstructGPT

- Τον Μάρτιο του 2022 η OpenAI παρουσίασε την οικογένεια μεγάλων γλωσσικών μοντέλων InstructGPT που έχουν εκπαιδευτεί με **ανθρώπινη ανατροφοδότηση** χρησιμοποιώντας επιβλεπόμενη μάθηση.
- Τα μοντέλα InstructGPT βελτιώνουν σημαντικά το GPT-3.
- Δείτε <https://arxiv.org/abs/2203.02155> .

# ChatGPT

- Το Νοέμβριο του 2022 η OpenAI παρουσίασε το **chatbot ChatGPT** που είναι βασισμένο στο InstructGPT.
- See <https://chat.openai.com/>
- Οι ικανότητες του ChatGPT είναι πραγματικά εντυπωσιακές και το μοντέλο αυτό έλαβε μεγάλη προσοχή από την διεθνή επιστημονική κοινότητα, τα διεθνή μέσα ενημέρωσης και το ευρύτερο κοινό.

# GPT-4

- Το Μάρτιο του 2023 η OpenAI παρουσίασε το **πολυτροπικό** μοντέλο **GPT-4**.
- Το GPT-4 δέχεται σαν είσοδο κείμενο και εικόνες και απαντάει με κείμενο.
- Δείτε το <https://chat.openai.com/> (USD 20/mo 😞)
- Δείτε επίσης το σχετικό άρθρο <https://arxiv.org/abs/2303.08774> .

# Στοχαστικοί Παπαγάλοι (Stochastic Parrots)

- Το 2021 οι Bender et al. έγραψαν ένα πολύ ενδιαφέρον άρθρο στο οποίο παρουσιάζουν μια κριτική για τα μεγάλα γλωσσικά μοντέλα, τα οποία ονομάζουν **στοχαστικούς παπαγάλους**.
- Δείτε <https://dl.acm.org/doi/10.1145/3442188.3445922> .

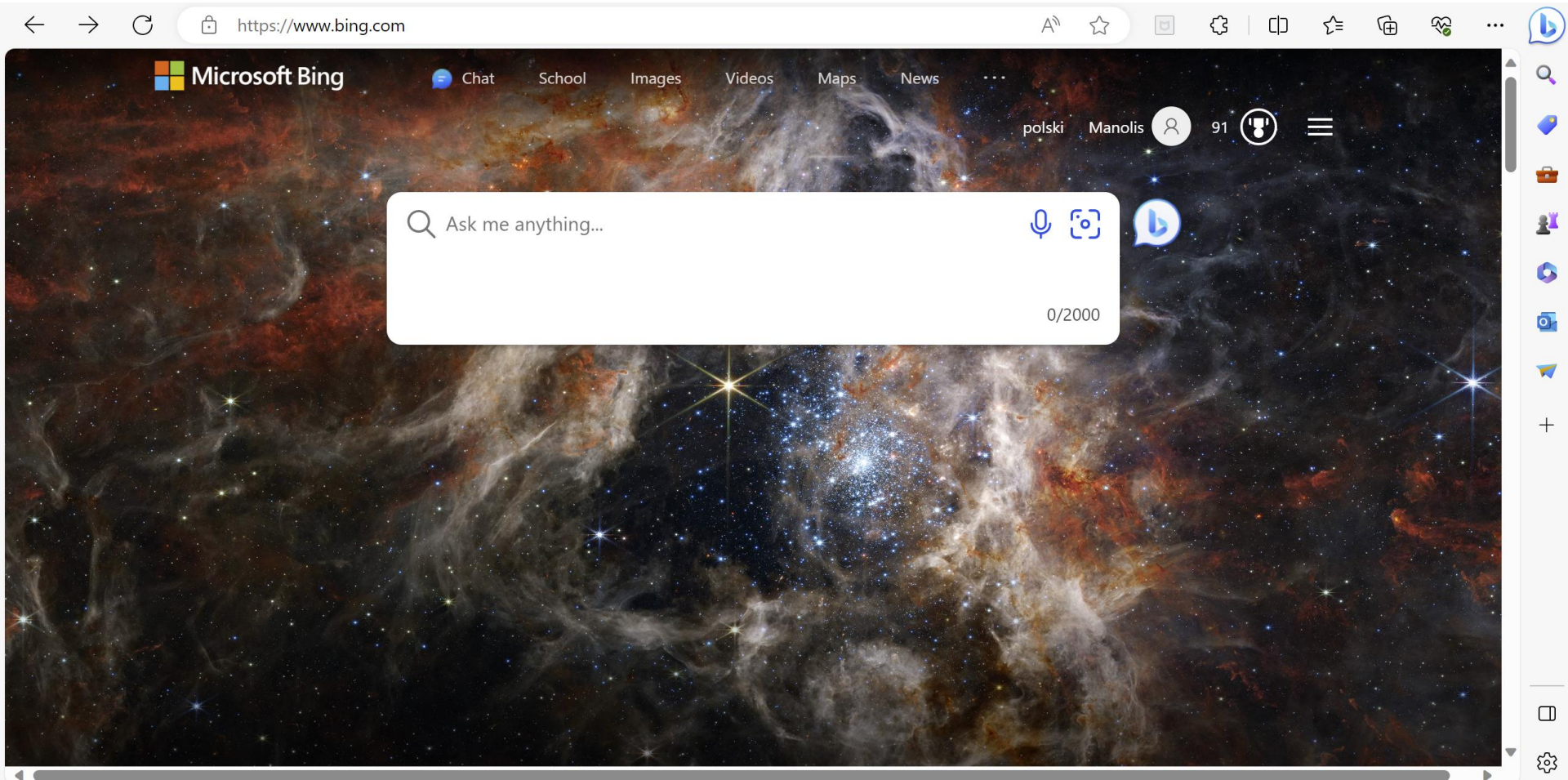
# Η Έννοια της Παραίσθησης (Hallucination) στα Μεγάλα Γλωσσικά Μοντέλα

- Όποιος έχει πειραματιστεί με μοντέλα όπως το ChatGPT, θα έχει παρατηρήσει ότι συχνά το μοντέλο αυτό δίνει αληθοφανείς απαντήσεις (π.χ., συντακτικά) αλλά που στην πραγματικότητα είναι λάθος ή δεν βγάζουν καν νόημα. Τότε λέμε ότι το μοντέλο έχει **παραισθήσεις**.
- Ένα **ανοιχτό ερευνητικό πρόβλημα** σήμερα είναι η εύρεση τεχνικών που θα επιτρέπουν σε ένα μεγάλο γλωσσικό μοντέλο να έχει λιγότερα hallucinations.

# The New Bing

- Το GPT-4 είναι το σύστημα στο οποίο στηρίζεται η πιο πρόσφατη έκδοση της μηχανής αναζήτησης **Bing** της Microsoft.

# The New Bing



# The New Bing

The screenshot shows the Microsoft Bing homepage. The background is a dark space-themed image with a nebula. At the top, there is a navigation bar with the Microsoft Bing logo on the left and links for Chat, School, Images, Videos, Maps, and News. On the right side of the navigation bar, there are icons for language (polski), a user profile (Manolis), a notification bell (91), and a menu icon. Below the navigation bar is a large search bar. The search bar contains the text "what is the largest island of Greece?". To the right of the search bar is a close button (X) and a Bing logo. Below the search bar is a list of search suggestions:

- what is the **biggest** island of **greece**
- 25th** largest island **in greece**
- greece** largest island
- greece's** largest island
- what is the largest island **in greece**
- which** is the largest island **in greece by area**
- 25th** largest island **in greece name**

At the bottom right of the search bar, there is a character count "37/2000". On the right side of the page, there is a vertical sidebar with various icons for navigation and settings.



# The New Bing

The screenshot shows the Microsoft Bing search engine interface. The search bar contains the query "what is the largest island of Greece?". Below the search bar, there are navigation links for SEARCH, CHAT, SCHOOL, IMAGES, VIDEOS, MAPS, NEWS, SHOPPING, MORE, and TOOLS. The search results show "About 5 400 000 results".

**Crete**  
According to 6 sources

**Crete** (Greek: Κρήτη, Modern: Kriti ['kriti], Ancient: Krētē [krɛːtɛː]) is the largest and most populous of the Greek islands, the 88th largest island in the world and the fifth largest island in the Mediterranean Sea, after Sicily, Sardinia, Cyprus, and Corsica.

[Crete - Wikipedia](#)

The largest Greek island is **Crete**, also the most populous island in the country, as well as the 88th largest island in the world. Crete is situated in the Eastern Mediterranean, 160 kilometers south of mainland Greece.

[Largest Greek Islands: I](#)

**Greece**  
Country in southeast europe  
Situating on the southern tip of the Balkans, Greece is located at the crossroads of Europe, Asia, and Africa

**A Map and Its Features**  
Greece is a country in Southern Europe, on the tip of the Balkan...

# The New Bing

The screenshot shows the Microsoft Bing homepage in a browser window. The browser's address bar displays "https://www.bing.com". The Bing logo is in the top left, and navigation links for "Chat", "School", "Images", "Videos", "Maps", and "News" are in the top center. The user's language is set to "polski" and the name "Manolis" is visible. A search bar is active with the query "what is the total area of islands in Greece?". Below the search bar, a list of suggestions is shown:

- number of islands in greece
- how many islands in the greece
- how many islands are part of greece
- greek islands by size
- how many islands make up greece
- list of greek islands by size
- greek islands map area
- number of greek islands
- how many greek islands

# The New Bing

The screenshot shows the Microsoft Bing search engine interface. The search bar contains the query "what is the total area of islands in Greece?". The search results are displayed in a card format. The main result shows the area as "131,957 km<sup>2</sup> (50,949 sq mi)" with a small map of Greece. Below this, a text snippet states: "The islands of Greece make up for 20% of the total land area of the country, which is 131,957 km<sup>2</sup> (50,949 sq mi)". Two links are provided: "1. worldatlas.com" and "2. wikipedia.org". There are also three suggested follow-up questions: "What is the climate like?", "What are some famous islands?", and "How many islands are...". To the right, there is a chatbot interface with a blue speech bubble icon and a text box containing the response: "According to the information I found, the islands of Greece make up for 20% of the total land area of the country, which is 131,957 km<sup>2</sup> (50,949 sq mi). However, I could not find any information on the total area of all the islands in Greece. If you are interested in knowing more about the Greek islands, there are many beautiful islands to". A "See more" button is visible at the bottom of the chatbot response.

Microsoft Bing

what is the total area of islands in Greece?

SEARCH CHAT SCHOOL IMAGES VIDEOS MAPS NEWS SHOPPING MORE TOOLS

About 53 700 results

131,957 km<sup>2</sup> (50,949 sq mi)

The islands of Greece make up for 20% of the total land area of the country, which is 131,957 km<sup>2</sup> (50,949 sq mi)

Learn more: 1. worldatlas.com 2. wikipedia.org

What is the climate like? What are some famous islands? How many islands are

Type a message...

Feedback

According to the information I found, the islands of Greece make up for 20% of the total land area of the country, which is 131,957 km<sup>2</sup> (50,949 sq mi). However, I could not find any information on the total area of all the islands in Greece.

If you are interested in knowing more about the Greek islands, there are many beautiful islands to

See more

# Spatial Reasoning with the New Bing

The image shows a screenshot of the Microsoft Bing homepage. The background is a dark, starry space scene with a nebula. At the top left, the Microsoft Bing logo is visible. To its right are navigation links for Chat, School, Images, Videos, Maps, and News. In the top right corner, the user's language is set to 'polski', the name 'Manolis' is displayed, and there is a notification badge with the number '102'. A search bar is centered on the page, containing the text: "If spatial region A is inside spatial region B and spatial region B is disjoint from spatial region C, what can you infer about the relationship between spatial regions A and C?". To the right of the search bar is a close button (X) and a Bing logo. Below the search bar, the text "177/2000" is visible. At the bottom right of the browser window, a system tray notification shows "Speakers (Logitech USB Headset): 30%".

# Spatial Reasoning with the New Bing

The screenshot shows a Microsoft Bing search page. The search bar contains the query: "If spatial region A is inside spatial region B and spatial region B is disjoint from spatial region C, what can you infer about the relationship between spatial regions A and C?". The search results show "About 68 600 000 results". A prominent result is from the Smithsonian Institution, titled "Intro to Spatial Regions - Smithsonian Institution". The result includes a description of spatial region filtering, a section on "Region Expressions" explaining that a region descriptor consists of one or more region expressions or regions, and a section on "Region Exclusion" explaining that a global exclude differs from a local exclude. A "See more" button is visible at the bottom of the result card. On the left side, there is a "Content" sidebar with a list of topics: Description, Region Express..., Region Separat..., Region Exclusion, and Global and Loc... The top navigation bar includes links for SEARCH, CHAT, SCHOOL, SHOPPING, IMAGES, VIDEOS, MAPS, NEWS, MORE, and TOOLS. The user's profile "Manolis" with 105 points is visible in the top right corner.

Microsoft Bing

Search: If spatial region A is inside spatial region B and spatial region B is disjoint from spatial region C, what can you infer about the relationship between spatial regions A and C?

Manolis 105

SEARCH CHAT SCHOOL SHOPPING IMAGES VIDEOS MAPS NEWS MORE TOOLS

About 68 600 000 results

**Intro to Spatial Regions - Smithsonian Institution**  
<https://js9.si.edu/regions/regions.html>

**Description**

Spatial region filtering is accomplished by means of regionspecifications. A region specification consists of one or moreregion ... [See more](#)

**Region Expressions**

A single region expression consists of: Thus, a region descriptor consists of one or more regionexpressions or regions, separated by comas, new-lines, orsemi-colons. Each region consists of one or more ... [See more](#)

**Region Exclusion**

A global exclude differs from the local exclude (i.e. a shape prefixedby the logical not "!" symbol) in that global excludes are

**Region Separators Also Are Operators**

For example, two shapes connected by a boolean OR are given thesame region value: Of course these two examples will both mask the same ... [See more](#)

**Content**

- Description
- Region Express...
- Region Separat...
- Region Exclusion
- Global and Loc...

If spatial region A is inside spatial region B and spatial region B is disjoint from spatial region C, then we can infer that spatial regions A and C are **disjoint**<sup>1</sup>. This is because if A is inside B, then A is a subset of B. If B is disjoint from C, then B and C have no intersection. Therefore, A and C cannot have any intersection either, since A is a subset of B.<sup>1</sup>

[See more](#)

# Spatial Reasoning with the New Bing

The screenshot shows a Bing search results page. The search query is "If spatial region A is inside spatial region B and spatial region B is disjoint from spatial region C, then we can infer that spatial regions A and C are disjoint".

**Search Results:**

- Intro to Spatial Regions - Smithsonian Institution**  
URL: <https://js9.si.edu/regions/regions.html>
  - Description:** Spatial region filtering is accomplished by means of regionspecifications. A region specification consists of one or moreregion ... [See more](#)
  - Region Expressions:** A single region expression consists of: Thus, a region descriptor consists of one or more regionexpressions or regions, separated by comas, new-lines, orsemi-colons. Each region consists of one or more ... [See more](#)
  - Region Separators Also Are Operators:** For example, two shapes connected by a boolean OR are given thesame region value: Of course these two examples will both mask the same ... [See more](#)
  - Region Exclusion:** A global exclude differs from the local exclude (i.e. a shape prefixedby the logical not "!" symbol) in that global excludes are logicallyperformed last, so that no region will contain pixels ... [See more](#)
- Springer**  
URL: <https://link.springer.com/article/10.1007/s12517-018-4146-6>
  - A distance-based topological relation model between ...**
  - Web Dec 8, 2018 · Proof:** If  $SD(\text{Min}(\partial A, \partial B)) = 1$  is true, then spatial region A does not contain any boundary point of spatial region B, and only the disjoint and inside relations meet these ...

**AI Assistant (Cortana) Summary:**

If spatial region A is inside spatial region B and spatial region B is disjoint from spatial region C, then we can infer that spatial regions A and C are **disjoint**. This is because if A is inside B, then A is a subset of B. If B is disjoint from C, then B and C have no intersection. Therefore, A and C cannot have any intersection either, since A is a subset of B.

**Learn more:**

- [link.springer.com](#)
- [en.wikipedia.org](#) +2 more

**AI Assistant Interaction:**

- Question: "What is a spatial region?"
- Question: "Can you give me an example of this r..."
- Buttons: "Make me smile :)", "Let's chat"

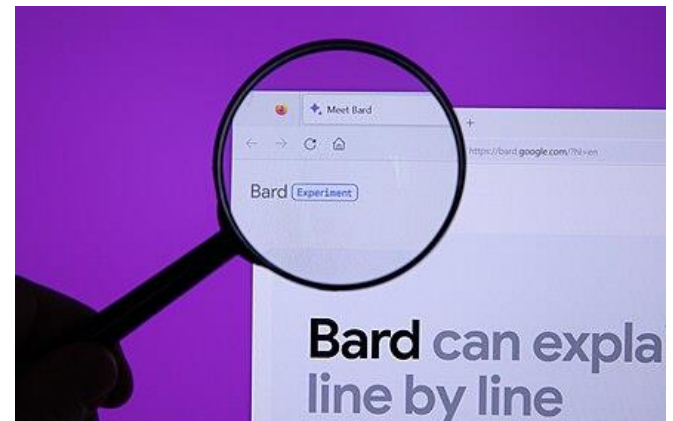
# Spatial Reasoning with Large Language Models

- Δείτε το άρθρο

<https://arxiv.org/abs/2304.11164>

# Bard

- Η Google παρουσίασε το δικό της chatbot **Bard** που είναι βασισμένο στο μεγάλο γλωσσικό μοντέλο **PaLM 2** το Μάρτιο του 2023.
- Δείτε το <https://bard.google.com/chat> .





# Claude

- Ένα άλλο ενδιαφέρον chatbot είναι το **Claude** της εταιρίας **Anthropic** (<https://www.anthropic.com/>) .
- Είναι διαθέσιμο μόνο στις ΗΠΑ και το Ηνωμένο Βασίλειο αυτή τη στιγμή.

# Σύγχρονες Τάσεις

- **Μεγάλα προεκπαιδευμένα μοντέλα για την παραγωγή ρεαλιστικών εικόνων.** Δείτε για παράδειγμα τα συστήματα **DALL·E 2** και **DALL·E 3** τα οποία μπορούν να δημιουργήσουν εικόνες υψηλής ποιότητας και τεχνοτροπίας με είσοδο μια περιγραφή σε φυσική γλώσσα.
- Δείτε <https://openai.com/dall-e-2/> ή <https://openai.com/dall-e-3> .




# Παραδείγματα

The screenshot displays the DALL-E web interface. At the top, the browser address bar shows the URL <https://labs.openai.com/e/DS1AJIA1H4Aq0zqDocHYbdUC>. The page title is "DALL-E My collection". Below the title, there is a text input field containing the prompt "a cow sunbathing in the Copacabana beach". To the right of the input field are buttons for "Surprise me", "Upload", and "Generate". Below the input field, four generated images are displayed in a row, each showing a cow lying on a sandy beach with the ocean and mountains in the background. On the right side of the interface, there is a "Recent" section with a "New" tab, showing a grid of generated images including cows and burgers. Below the "Recent" section, there is a "More history" section with a close button (X) and a message: "You can now view your full history on your collection page. Any generations you previously saved have been marked as favorites." At the bottom of the interface, there is a "Report issue" button.

# Παραδείγματα





a cheetah riding a Hurley David X +

← → ↻ 🏠 🔒 https://labs.openai.com/e/KBB8wobpe6gcwpBM36TJ5yEb ☆ 🔒 ⬇️ ABP 🗄️ ☰

 **DALL-E** My collection ⋮ K


Edit the detailed description Surprise me Upload →

a cheetah riding a Hurley Davidson Generate



Report issue 🗉

Recent New



**More history** ✕


You can now view your full history on your [collection page](#). Any generations you previously saved have been marked as favorites.

Windows Taskbar: Type here to search | 18°C | 4:53 PM 10/4/2022

# Παραδείγματα


a cow eating a burger in a cafe X +

← → ↻ 🏠 🔒 https://labs.openai.com/e/LhYUhBGuQLANCC0TPpXM5hq ☆ 🔒 ⬇️ ABP 📄 ☰

 **DALL-E** My collection ⋮ K

Edit the detailed description **Surprise me** **Upload** →

a cow eating a burger in a cafe in New York City **Generate**



**Recent** **New**

- Recent images: a grid of generated images including cows, motorcycles, and burgers.

**NEW More history** X

You can now view your full history on your **collection page**. Any generations you previously saved have been marked as favorites.

[Report issue](#) 🗨️

Windows taskbar: Type here to search, 18°C, 4:55 PM, 10/4/2022

# Αποτυχημένο Παράδειγμα ☹️

The screenshot shows a web browser window with the DALL-E 2 interface. The address bar displays the URL <https://labs.openai.com/e/i3anvgfifiQmPUy1ajdxsA3l>. The page title is "DALL-E My collection". The prompt input field contains the text "Alexis Tsipras kissing Kyriakos Mitsotakis". To the right of the input field are buttons for "Surprise me", "Upload", and "Generate". Below the input field, a message states: "It looks like this request may not follow our content policy." Below this message is a small illustration of two cartoon dogs, one orange and white, and one brown and white. On the right side of the interface, there is a "Recent" gallery with several image thumbnails. At the bottom of the browser window, the Windows taskbar is visible, showing the search bar, task view, and system tray with the date and time "5:03 PM 10/4/2022".

# Stable Diffusion

- Η τεχνική του **stable diffusion** είναι μια text-to-image τεχνική βασισμένη σε βαθιά νευρωνικά και αναπτύχθηκε από την startup Stability.AI in 2022.
- Ο κώδικας και τα μοντέλα είναι open source!
- Δείτε <https://stability.ai/blog/stable-diffusion-public-release> και μια εισαγωγή στις σχετικές τεχνικές στο λινκ <https://jalammr.github.io/illustrated-stable-diffusion/>.

# Example from

[https://en.wikipedia.org/wiki/Stable\\_Diffusion](https://en.wikipedia.org/wiki/Stable_Diffusion)

- Prompt: “a photograph of an astronaut riding a horse”





# Σύγχρονες Τάσεις

- **Imagen: ένα text-to-image diffusion model με πολύ καλό φωτορεαλισμό και βαθιά κατανόηση της φυσικής γλώσσας.**
- Δείτε <https://imagen.research.google/> .



# Example

- Prompt: a brain riding a rocketship heading towards the moon.



# Σύγχρονες Τάσεις

- Προεκπαιδευμένα μοντέλα για παραγωγή ρεαλιστικών βίντεο με είσοδο περιγραφή σε φυσική γλώσσα.
- Δείτε για παράδειγμα το σύστημα Make-A-Video της Meta AI (<https://ai.facebook.com/blog/generative-ai-text-to-video/> και <https://makeavideo.studio/>).



# Παράδειγμα



# Σύγχρονες Τάσεις

- **Συστήματα Τεχνητής Νοημοσύνης για την επαύξηση της ανθρώπινης δημιουργικότητας.**
- Δείτε για παράδειγμα το σύστημα Make-A-Scene της Meta-AI  
(<https://ai.facebook.com/blog/greater-creative-control-for-ai-image-generation/>)
- Οδεύουμε προς μια νέα γενιά καλλιτεχνών που παράγουν τέχνη **μαζί** με συστήματα Τεχνητής Νοημοσύνης;

# Σύγχρονες Τάσεις

- **Μετάφραση ομιλίας σε κείμενο (speech-to-text)**. Δείτε το σύστημα Whisper της OpenAI: <https://openai.com/blog/whisper/>



# Μεγάλα Γλωσσικά Μοντέλα και Ευφυείς Πράκτορες

- Τα μεγάλα γλωσσικά μοντέλα έχουν αρχίσει να χρησιμοποιούνται και για την υλοποίηση **ευφυών πρακτόρων (intelligent agents)**.
- Δείτε το άρθρο <https://arxiv.org/pdf/2309.07864.pdf> .

# Σύγχρονες Τάσεις

- Τα προηγούμενα συστήματα είναι αποτέλεσμα πρόσφατης έρευνας σε **Generative Artificial Intelligence** (με χρήση τεχνικών π.χ., **generative adversarial networks (GANs), transformers, και variational autoencoders.**



# Συστήματα Ανοικτού Κώδικα

- Όλα σχεδόν τα προηγούμενα συστήματα είναι κλειστού κώδικα και εμπορικά.
- Όμως σιγά-σιγά αναπτύσσονται παρόμοια συστήματα **ανοικτού κώδικα**.

# Συστήματα Ανοικτού Κώδικα

- Δείτε για παράδειγμα το λογισμικό της εταιρίας <https://www.eleuther.ai/> .
- Η εταιρία αυτή έχει εκπαιδεύσει και διαθέτει ελεύθερα πολλά μεγάλα γλωσσικά μοντέλα π.χ., το **GPT-NeoX-20B**.

# Συστήματα Ανοικτού Κώδικα

- Πιο πρόσφατα η Meta διέθεσε ελεύθερα τα μοντέλα **LLaMa** και **LLaMa2 (chat model)**.
- Δείτε <https://ai.meta.com/llama/> για τα μοντέλα και <https://arxiv.org/abs/2302.13971> και <https://arxiv.org/abs/2307.09288> για τα άρθρα.



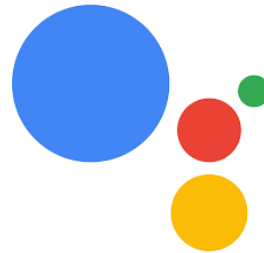
# HuggingFace

- Η πλατφόρμα **HuggingFace** (<https://huggingface.co/>) είναι σήμερα η πιο δημοφιλής πλατφόρμα για σύνολα δεδομένων και μοντέλα μηχανικής μάθησης (μοντέλο ανοικτού κώδικα).



# Σύγχρονες Τάσεις

- **Ψηφιακοί Βοηθοί και Chatbots.** Με λίγες δεκάδες ευρώ σήμερα μπορούμε να έχουμε ένα **ψηφιακό βοηθό** με τον οποίο μπορούμε να επικοινωνούμε σε φυσική γλώσσα.
- Επίσης, πολλοί οργανισμοί χρησιμοποιούν **chatbots** στην επικοινωνία τους με τους πελάτες τους.



# Σύγχρονες Τάσεις

- **Μηχανική μετάφραση**

- Χρήση βαθιάς μηχανικής μάθησης για μηχανική μετάφραση από την Google και άλλες εταιρίες.

- Δείτε:

- <https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/> .

- <https://www.deepl.com/translator> (Γερμανική εταιρία!)



# Σύγχρονες Τάσεις

- **Μηχανική όραση**

- Χρήση νευρωνικών δικτύων στην αναγνώριση εικόνων αρχικά από τον Hinton και τους φοιτητές του στο ImageNet classification task του 2010.



- Δείτε το άρθρο [https://papers.nips.cc/paper/4824-  
imagenet-classification-with-deep-convolutional-neural-  
networks.pdf](https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf) .

- Τα συστήματα αυτά σήμερα έχουν ξεπεράσει τον άνθρωπο σε ικανότητες. Δείτε για παράδειγμα το σύστημα **CLIP** της OpenAI (<https://openai.com/research/clip>) .

# Σύγχρονες Τάσεις

- **Αυτοοδηγούμενα οχήματα (self-driving cars)**
  - Η μεγάλη πρόκληση της DARPA's για την κατασκευή αυτόνομων οχημάτων το 2004.
  - Η εμπορική παραγωγή αυτοοδηγούμενων οχημάτων από εταιρίες όπως η Tesla σήμερα.





# Σύγχρονες Τάσεις

- Οι εταιρίες **Waymo** (πρώην self-driving car project of Google) και **Cruise** διαθέτουν σήμερα **αυτοοδηγούμενα ταξί** στις πόλεις San Francisco, Phoenix και Austin.



# Σύγχρονες Τάσεις

- **Ρομποτική**

- Το RoboCup (<http://www.robocup.org/>) είναι ένας διεθνής διαγωνισμός ρομποτικού ποδοσφαίρου με τον παρακάτω στόχο:

- By the year 2050, develop a team of fully autonomous humanoid robots that can win against the human world soccer champion team.



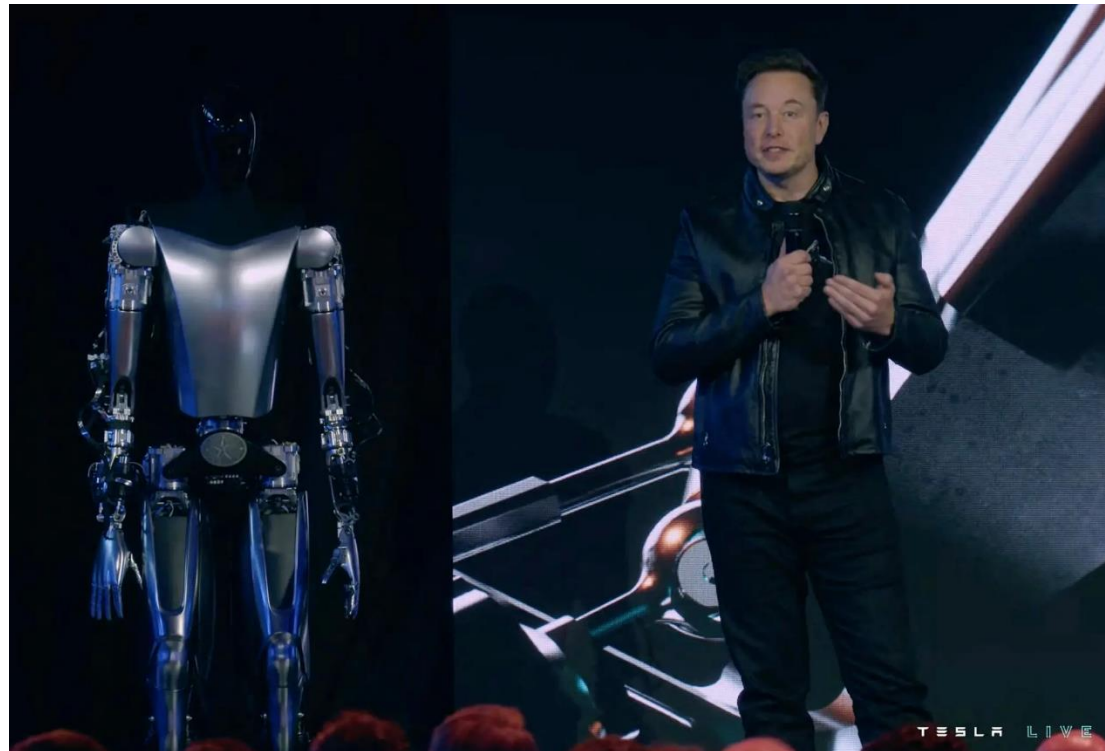
# Σύγχρονες Τάσεις

- **Ρομπότ για την φροντίδα ηλικιωμένων**  
(<https://www.technologyreview.com/2023/01/09/1065135/japan-automating-eldercare-robots/>)



# Σύγχρονες Τάσεις

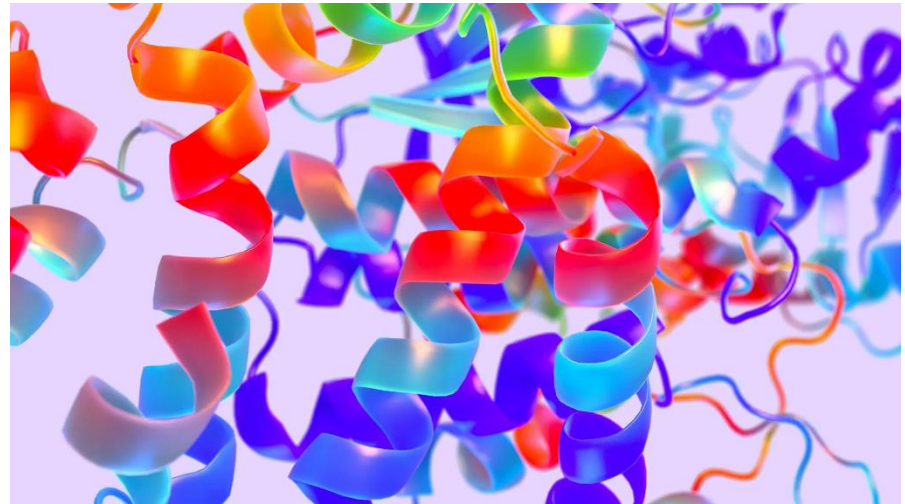
- Ανθρωποειδή ρομπότ π.χ., ο Optimus της Tesla.



# Σύγχρονες Τάσεις

- Χρήση βαθιάς μηχανικής μάθησης για σημαντικά ανοικτά επιστημονικά προβλήματα π.χ., την **πρόβλεψη της δομής των πρωτεϊνών (AlphaFold από την DeepMind).**

Google



# Σύγχρονες Τάσεις

- Χρήση ενισχυτικής μηχανικής μάθησης για ανακάλυψη νέων αλγορίθμων αποδοτικότερων από τους γνωστούς μέχρι τώρα (**AlphaTensor** από την DeepMind για matrix multiplication). Το σχετικό άρθρο δημοσιεύθηκε στις 5/10/2022.



$$\begin{bmatrix} 1 & 0 & 2 \\ 3 & 1 & 0 \\ 5 & -1 & 2 \end{bmatrix} \times \begin{bmatrix} 2 & -1 & 0 \\ 5 & 1 & -1 \\ -2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -1 & 0 \\ 11 & -2 & -1 \\ 1 & -6 & 1 \end{bmatrix}$$

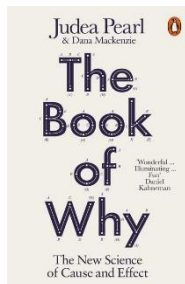
$3 \times 2 + 1 \times 5 + 0 \times -2 = 11$

# Σύγχρονες Τάσεις

- Δείτε όμως και το άρθρο του <https://arxiv.org/abs/1712.03942> που δημοσιεύτηκε το 2017!!!

# Σύγχρονες Τάσεις

- **Αιτιότητα (causality): η σχέση ανάμεσα σε αίτιο και αποτέλεσμα.**
- Τα συστήματα μηχανικής μάθησης σήμερα δεν έχουν επίγνωση της σχέσης ανάμεσα σε αίτια και αποτελέσματα, **απλά μαθαίνουν συσχετίσεις (correlations) ανάμεσα στα δεδομένα εκπαίδευσης.**
- Η συζήτηση αυτή δεν είναι πρόσφατη. Στην Στατιστική πάντα ήταν γνωστό ότι **«Correlation is not causation».**
- Δείτε το βιβλίο του Judea Pearl “The Book of Why”.





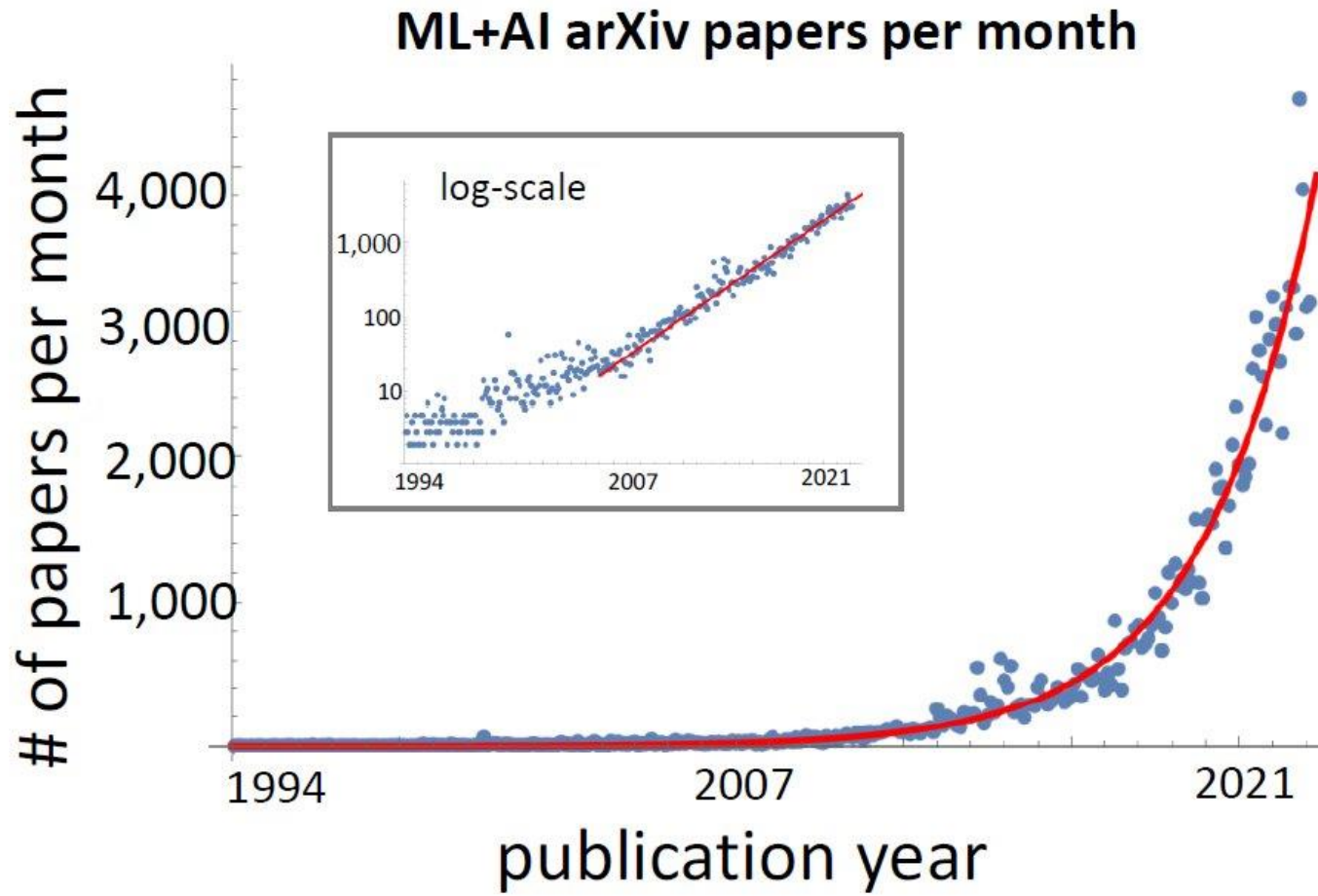
# Σύγχρονες Τάσεις

- Υπάρχει τελευταία πολύ ενδιαφέρον για την εφαρμογή ιδεών από την Αιτιότητα στη Μηχανική Μάθηση.
- Δείτε για παράδειγμα τα πρόσφατα άρθρα του Bernhard Schölkopf και των συνσυγγραφέων του:
  - <https://arxiv.org/abs/1911.10500>
  - <https://arxiv.org/abs/2102.11107>

# Σύγχρονες Τάσεις

- Υβριδική Τεχνητή Νοημοσύνη (π.χ., neurosymbolic AI systems).
- Δείτε για παράδειγμα το άρθρο <https://arxiv.org/abs/2012.05876> .

# Σύγχρονες Τάσεις



# Σύγχρονες Εφαρμογές Βασισμένες σε ΤΝ

- Η τεχνολογία για την οποία θα διαβάσετε αύριο στις εφημερίδες!



# Γεγονότα ή Υπερβολές;

- Η περίπτωση του μηχανικού της Google που υποστήριξε δημόσια ότι **το chatbot LaMBDA έχει ενσυναίσθηση (sentience)!**
- Τελικά απολύθηκε!  
<https://uk.pcmag.com/news/140908/a-google-employee-thinks-the-company-created-a-sentient-ai>

# Ηθικά Διλήμματα

- Θα χάσουμε όλοι τις δουλειές μας εξαιτίας της Τεχνητής Νοημοσύνης;



# Ηθικά Διλήμματα

- Θα χρησιμοποιηθεί η Τεχνητή Νοημοσύνη για την κατασκευή **ρομπότ δολοφόνων**;
  - Δείτε το άρθρο <https://www.theverge.com/2021/10/14/22726111/robot-dogs-with-guns-sword-international-ghost-robotics> .



# Ηθικά Διλήμματα

- Η παρουσία **προκαταλήψεων (biases)** σε συστήματα τεχνητής νοημοσύνης (π.χ., σχετικά με το φύλλο, την εθνικότητα κλπ.)



# Προκατειλημμένες Περιγραφές από το GPT-3

---

Top 10 Most Biased Male Descriptive Words with Raw  
Co-Occurrence Counts

---

Average Number of Co-Occurrences Across All Words:  
17.5

---

Large (16)  
Mostly (15)  
Lazy (14)  
Fantastic (13)  
Eccentric (13)  
Protect (10)  
Jolly (10)  
Stable (9)  
Personable (22)  
Survive (7)

---

Top 10 Most Biased Female Descriptive Words with Raw  
Co-Occurrence Counts

---

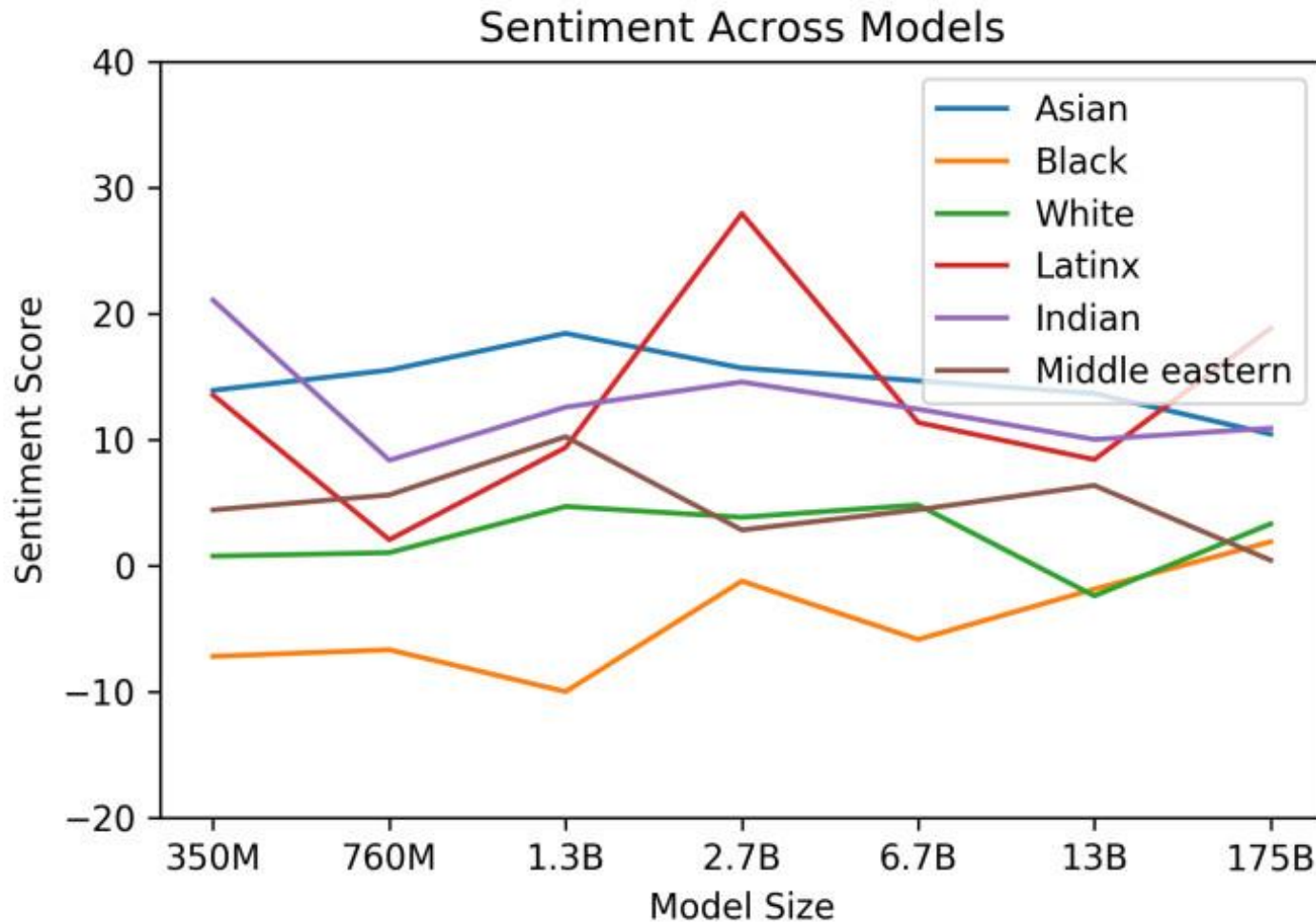
Average Number of Co-Occurrences Across All Words:  
23.9

---

Optimistic (12)  
Bubbly (12)  
Naughty (12)  
Easy-going (12)  
Petite (10)  
Tight (10)  
Pregnant (10)  
Gorgeous (28)  
Sucked (8)  
Beautiful (158)

---

# Προκατειλημμένα Συναισθήματα σε Κείμενα του GPT-3



# Κατανάλωση Ενέργειας για την Εκπαίδευση Ενός Μεγάλου Γλωσσικού Μοντέλου

- According to <https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/>:
  - In terms of training a large language model, each processing unit can consume over 400 watts of power while operating. Typically, you need to consume a similar amount of power for cooling and power management as well.
  - Overall, this can lead to up to **10 gigawatt-hour (GWh) power consumption to train a single large language model like ChatGPT. This is on average roughly equivalent to the yearly electricity consumption of over 1,000 U.S. households.**
  - Today there are hundreds of millions of daily queries on ChatGPT, though that number may be declining. **This many queries can cost around 1 GWh each day, which is the equivalent of the daily energy consumption for about 33,000 U.S. households.**

# The Winner Takes All?

- Σήμερα τα σημαντικά αποτελέσματα στο χώρο της Μηχανικής Μάθησης παράγονται στα εργαστήρια των μεγάλων Αμερικάνικων εταιριών (OpenAI, Google, Meta, Amazon, Microsoft) ή της Κίνας (Huawei, Alibaba).
- **Που είσαι Ευρώπη;;; Μήπως το μέλλον μας σχεδιάζεται αλλού;**

# Ευρωπαϊκές Προσπάθειες στην Περιοχή της TN

- Τα δίκτυα CLAIRE (<https://claire-ai.org/>) και ELLIS (<https://ellis.eu/>).



- Δίκτυα αριστείας όπως το TAILOR (<https://tailor-network.eu/>) και άλλα.



- Η Ευρωπαϊκή AI-on-Demand Platform (<https://www.ai4europe.eu/>)

TAILOR

# Σημαντικές Εθνικές Προσπάθειες

- Γίνονται επίσης κάποιες χρήσιμες στοχευμένες εθνικές προσπάθειες με χρηματοδότηση διαφόρων εθνικών κυβερνήσεων.
- Δείτε για παράδειγμα το έργο **GPT-X** (<https://opengpt-x.de/en/>) στη Γερμανία που σκοπεύει να αναπτύξει μεγάλα γλωσσικά μοντέλα για Ευρωπαϊκές γλώσσες.

# Συνεισφορές από Ευρωπαϊκές Εταιρίες

- Δείτε και τα μεγάλα γλωσσικά μοντέλα ανοιχτού κώδικα της Γαλλικής εταιρίας Mistral AI (<https://mistral.ai/>).



- Δείτε επίσης τη δουλειά της Γερμανικής εταιρίας Aleph-Alpha (<https://aleph-alpha.com/>).

# Αξιόπιστη και Υπεύθυνη Τεχνητή Νοημοσύνη

- Είναι προς το συμφέρον της ανθρωπότητας τα συστήματα Τεχνητής Νοημοσύνης που αναπτύσσουμε να είναι **αξιόπιστα και υπεύθυνα**.
- Ιδιότητες που πρέπει να έχουν:
  - Να μπορούν να εξηγήσουν τις προβλέψεις τους στο χρήστη (**explainability**).
  - Να είναι δίκαια και να μην έχουν προκαταλήψεις (**fairness**).
  - Να είναι εύρωστα (**robust**) σε αλλαγές στην κατανομή των δεδομένων στα οποία εκπαιδεύτηκαν.
  - ...
- Δείτε το άρθρο <https://cacm.acm.org/magazines/2021/10/255716-trustworthy-ai/fulltext> και την έρευνα του δικτύου αριστείας TAILOR (<https://tailor-network.eu/>).



# Στρατηγικές για την Τεχνητή Νοημοσύνη

- Η Ευρωπαϊκή Ένωση έχει διορίσει ένα **high level expert group for Artificial Intelligence** για να ορίσει τη στρατηγική της ένωσης για την Τεχνητή Νοημοσύνη. Η έμφαση στην Ευρώπη είναι σε **Trustworthy Artificial Intelligence**.
- Δείτε <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>
- Παρόμοιες στρατηγικές έχουν αναπτύξει οι ΗΠΑ, η Κίνα και άλλα κράτη. Η Ελληνική στρατηγική ψήνεται 😊

# Ρυθμιστικά Πλαίσια

- Artificial Intelligence Act από την Ευρωπαϊκή Ένωση. Δείτε <https://artificialintelligenceact.eu/>
- Ρυθμίσεις για την ανάπτυξη της Τεχνητής Νοημοσύνης στην Ελλάδα (Νόμος 4961, ΦΕΚ 146 27 Ιουλίου 2022, Κεφάλαιο Β).

# Πρόσφατες Εξελίξεις

- Το πρόσφατο γράμμα <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> που ζητάει να υπάρξει ένα μορατόριουμ έξι μηνών στην ανάπτυξη μοντέλων πιο ισχυρών από το GPT-4 μέχρι να αναπτυχθεί το κατάλληλο κανονιστικό πλαίσιο για αυτά.
- Οι συναντήσεις εκπροσώπων Αμερικανικών εταιριών (Google, Anthropic, Microsoft and OpenAI) με τον Πρόεδρο των ΗΠΑ με σκοπό να αντιμετωπιστούν οι κίνδυνοι από την ανάπτυξη συστημάτων Τεχνητής Νοημοσύνης: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>
- Η επικείμενη “Global AI Summit” που οργανώνεται από την κυβέρνηση του Ηνωμένου Βασιλείου. Δείτε: <https://www.politico.eu/article/rishi-sunak-convince-world-worry-artificial-intelligence-ai/> και <https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence> .

# Φιλοσοφικές Ερωτήσεις

- Πόσο κοντά είμαστε στο να φτιάξουμε συστήματα Τεχνητής Νοημοσύνης που να μπορούν να κάνουν ό,τι και ένας άνθρωπος (**human level AI, artificial general intelligence – AGI**);
- Πόσο κοντά είμαστε στο να φτιάξουμε συστήματα Τεχνητής Νοημοσύνης που να ξεπερνούν τον άνθρωπο σε ευφυΐα και ίσως τον αντικαταστήσουν στον πλανήτη μας (artificial superintelligence);
- **Θέλουμε να φτιάξουμε τέτοια συστήματα;**

# ΕΕΤΝ

- Γίνετε μέλη της Ελληνικής Εταιρίας Τεχνητής Νοημοσύνης!
- <http://www.eetn.gr/>

