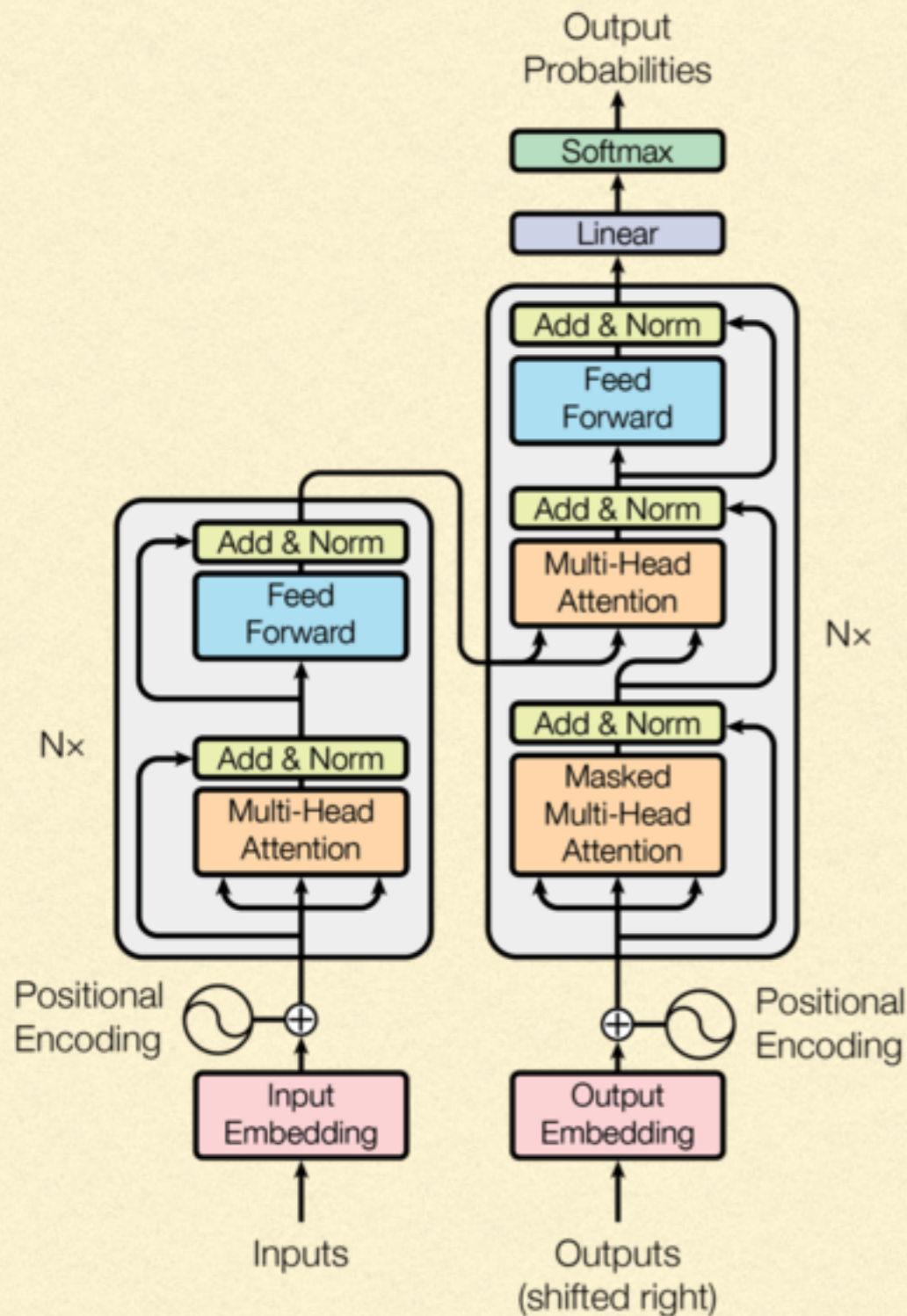
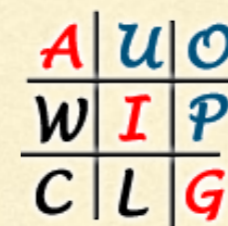


Γλωσσικά Μοντέλα

Τί είναι και πώς λειτουργούν



Αν.Καθ. Κυριάκος Σγάρμπας
IEEE Senior Member, ACM Member
Ομάδα Τεχνητής Νοημοσύνης & AI-Hub
Τμήμα Ηλεκτρολόγων Μηχανικών & ΤΥ
Πανεπιστήμιο Πατρών





ChatGPT

30 Νοεμβρίου 2022

1.000.000 χρήστες σε 5 ημέρες

USER: Explain the plot of Cinderella in a sentence where each word has to begin with the next letter in the alphabet from A to Z, without repeating any letters.

ChatGPT: A beautiful Cinderella, dwelling eagerly, finally gains happiness; inspiring jealous kin, love magically nurtures opulent prince; quietly rescues, slipper triumphs, uniting very wondrously, xenial youth zealously.

<https://openai.com/blog/chatgpt>

Τεχνητή Νοημοσύνη & ChatGPT

Artificial Intelligence (1955)

Machine Learning (1959)

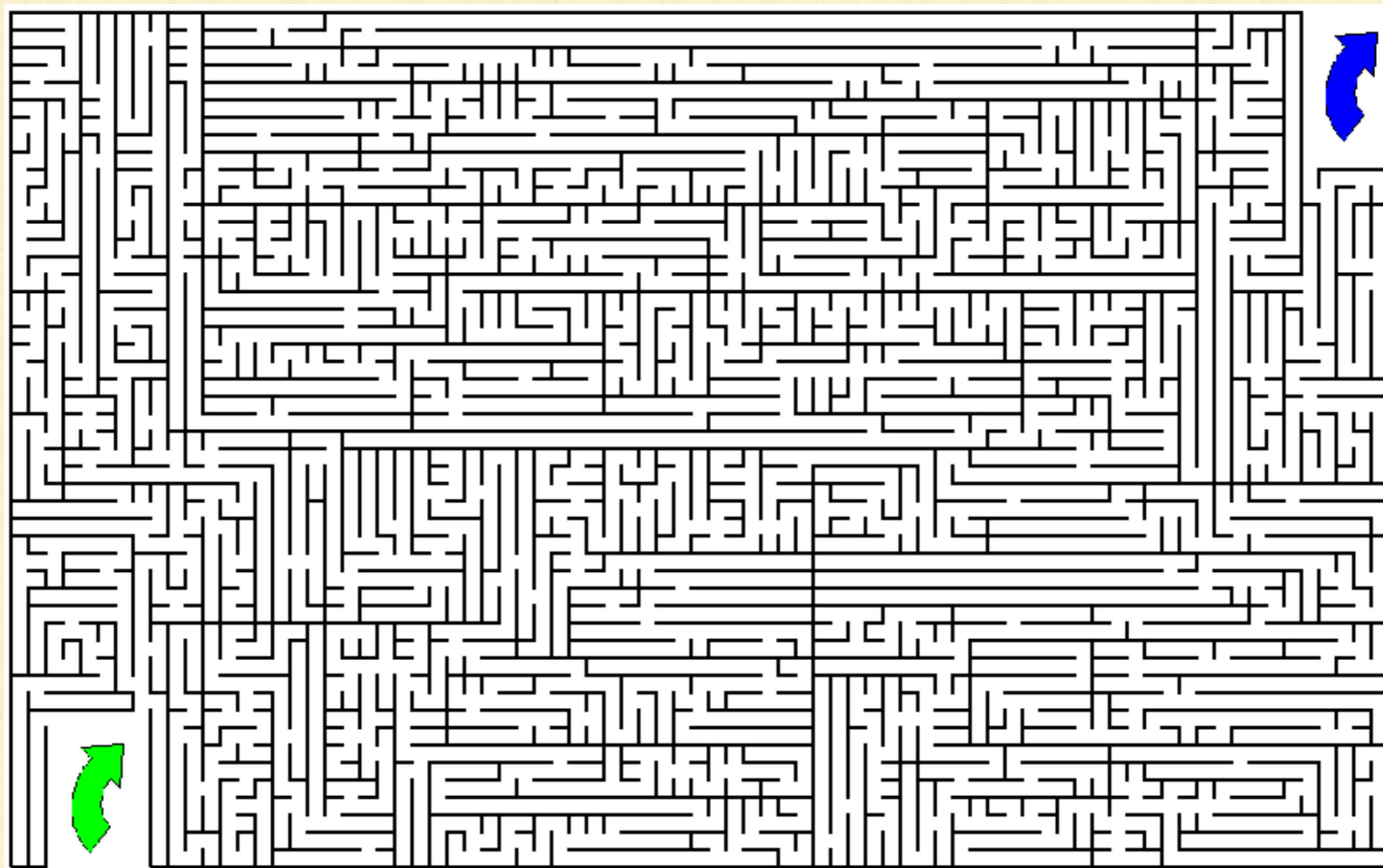
Language Models (1980)

Transformers (2017)

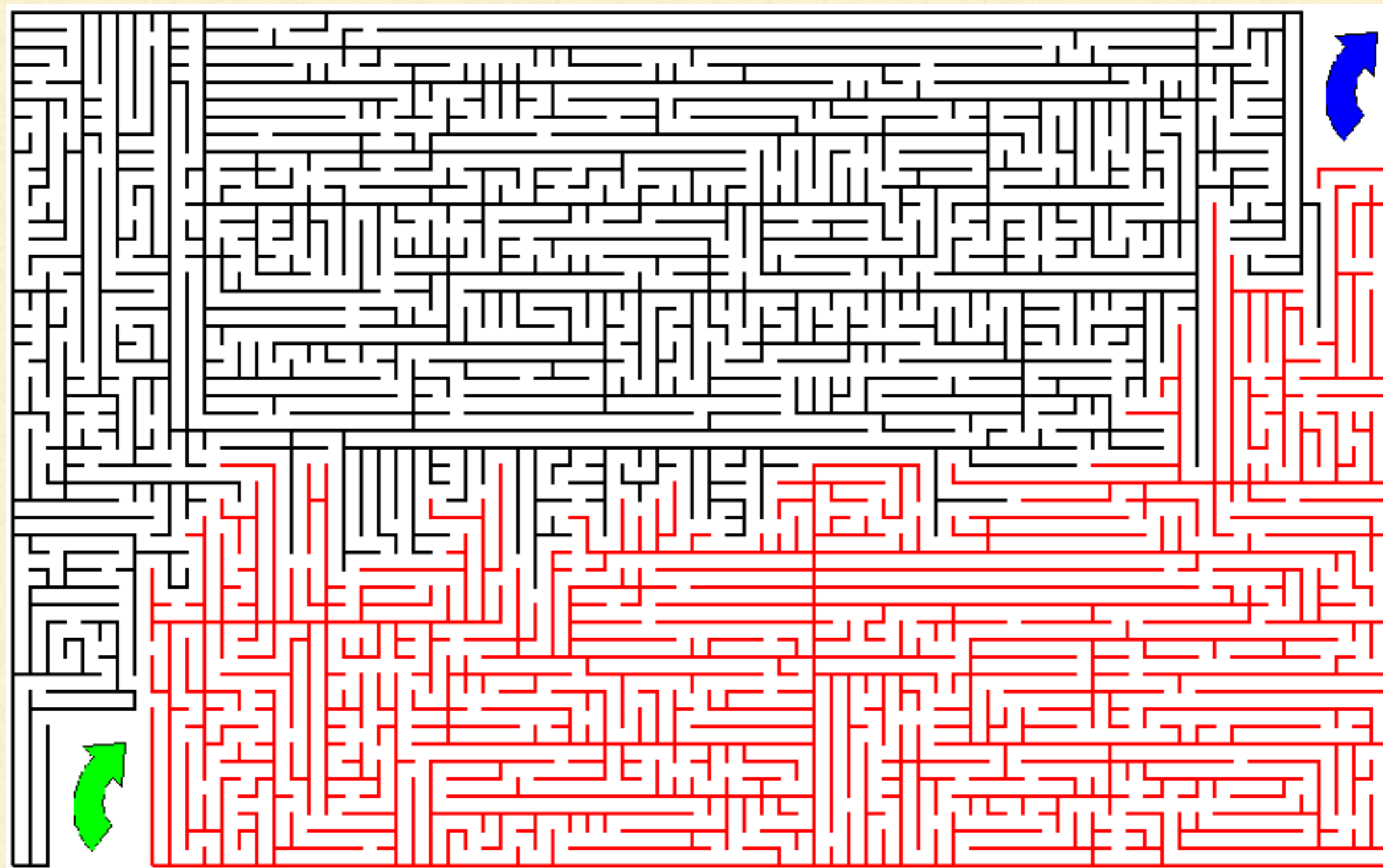
GPT (2018)

ChatGPT (2022)

Παράδειγμα



Παράδειγμα



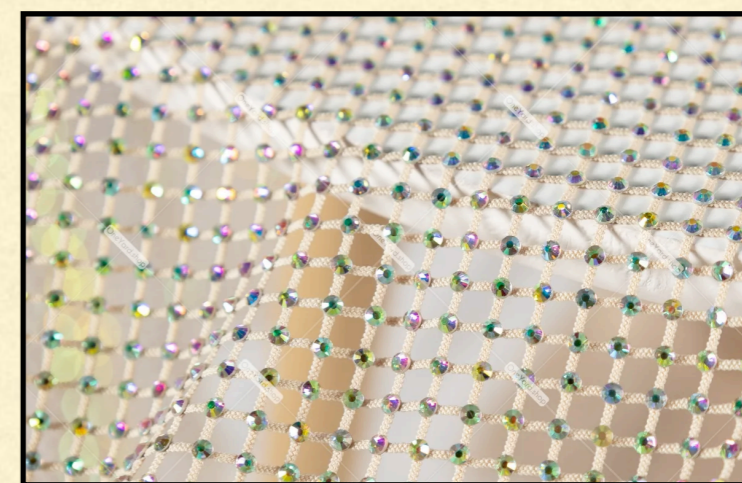
ChatGPT - Πώς Δουλεύει;

Βρείτε τη _____ που λείπει.

_____ ποτέ _____ δώδεκα _____.



Γλωσσικό Μοντέλο
Μηχανική Μάθηση
Στατιστική

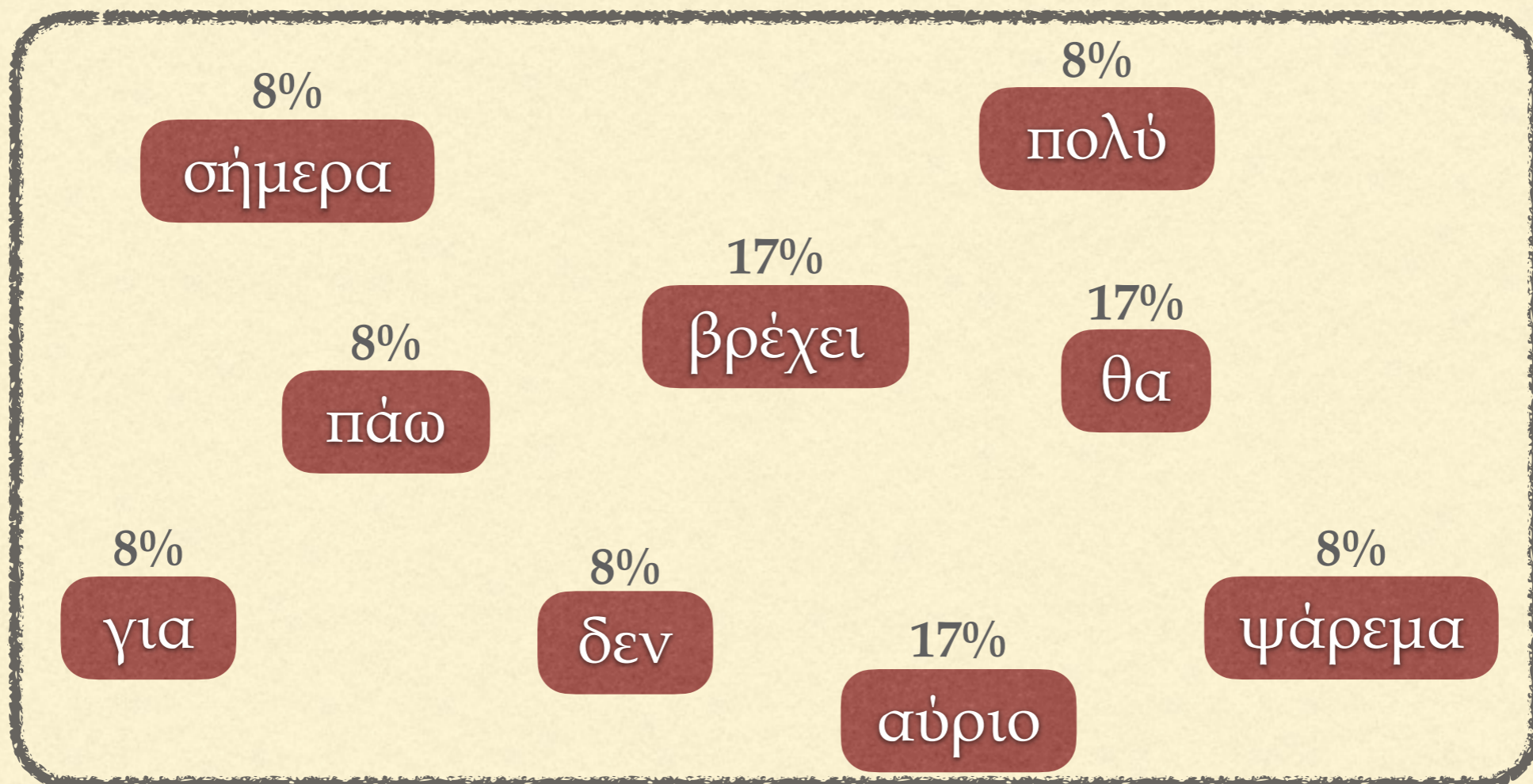


Γλωσσικό Μοντέλο (Language Model)

- Οποιαδήποτε διαδικασία μπορεί να προσδώσει πιθανότητες εμφάνισης σε προτάσεις της γλώσσας.
- Πρακτικά: πίνακες πιθανοτήτων εμφάνισης λέξεων και ακολουθιών λέξεων. Οι πιθανότητες υπολογίζονται από μετρήσεις σε συλλογές κειμένων που ονομάζονται **σώματα κειμένων** (text corpora).

Μοντέλο Μονογραμμάτων (Unigrams)

<ΑΡΧΗ> σήμερα βρέχει πολύ <ΤΕΛΟΣ>
<ΑΡΧΗ> αύριο δεν θα βρέχει <ΤΕΛΟΣ>
<ΑΡΧΗ> αύριο θα πάω για ψάρεμα <ΤΕΛΟΣ>

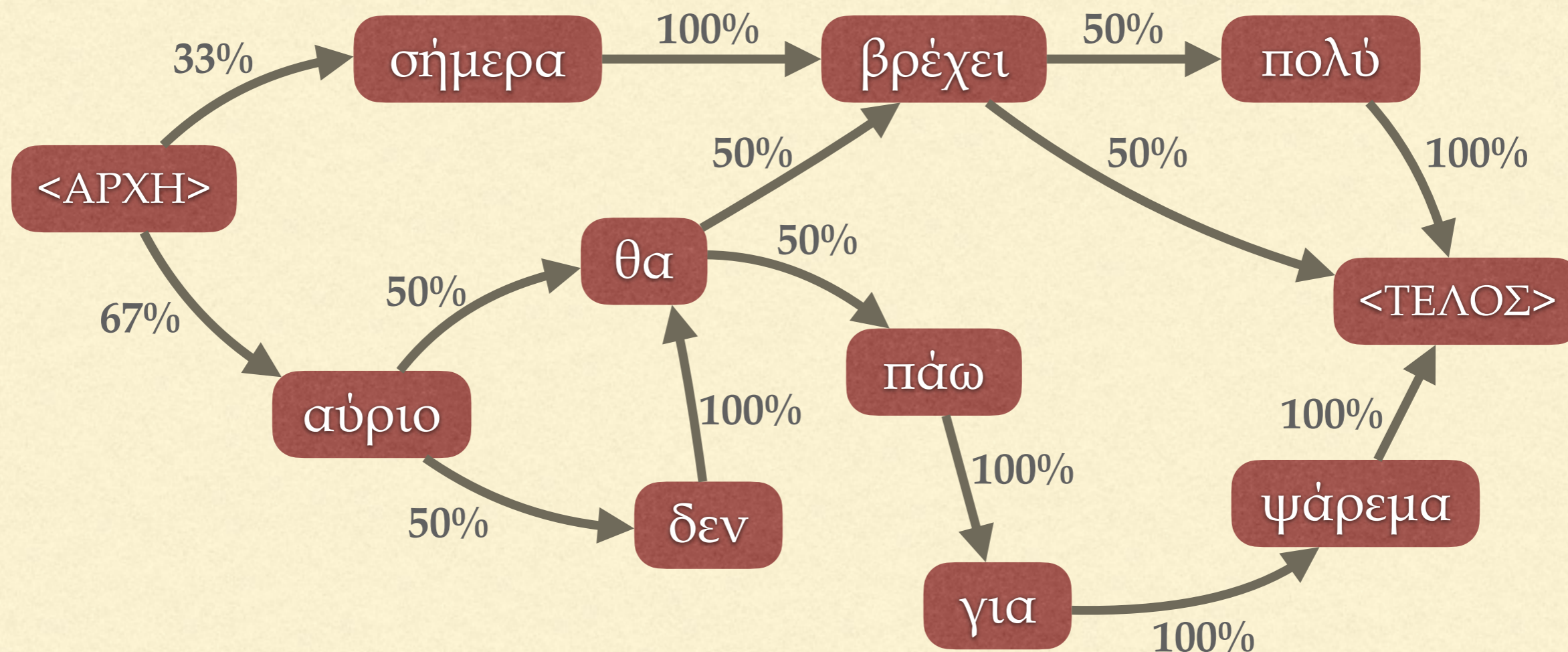


Παράδειγμα Σύνθεσης (Μονογράμματα)

months the my and issue of year foreign new
exchange's September were recession exchange
new endorsed a acquire to six executives

Μοντέλο Διγραμμάτων (Bigrams)

<ΑΡΧΗ> σήμερα βρέχει πολύ <ΤΕΛΟΣ>
<ΑΡΧΗ> αύριο δεν θα βρέχει <ΤΕΛΟΣ>
<ΑΡΧΗ> αύριο θα πάω για ψάρεμα <ΤΕΛΟΣ>



Παράδειγμα Σύνθεσης

(Διγράμματα)

last December through the way to preserve the Hudson corporation N.B.E.C. Taylor would seem to complete the major central planners one point five percent of U.S.E. has already old M.X. corporation of living on information such as more frequently fishing to keep her

Μοντέλο Τριγραμμάτων (Trigrams)

<ΑΡΧΗ> σήμερα βρέχει πολύ <ΤΕΛΟΣ>

<ΑΡΧΗ> αύριο δεν θα βρέχει <ΤΕΛΟΣ>

<ΑΡΧΗ> αύριο θα πάω για ψάρεμα <ΤΕΛΟΣ>



Παράδειγμα Σύνθεσης (Τριγράμματα)

they also point to ninety nine point six billion dollars from two hundred four oh six three percent of the rates of interest stores as Mexico and Brazil on market conditions

Σύγκριση Μοντέλων

Unigram:

To him swallowed confess hear both. Which. Of save on trail for are ay devise and rote life have
Every enter now severally so, let Hill he late speaks; or! a more to leg less first you enter Are where
exeunt and sighs have rise excellency took of. Sleep knave we. near; vile like

Bigram:

What means, sir. I confess she? then all sorts, he is trim, captain.
Why dost stand forth thy canopy, forsooth; he is this palpable hit the King Henry. Live king. Follow.
What we, hath got so she that I rest and sent to scold and nature bankrupt, nor the first gentleman?
Enter Memenius, if it so many good direction found'st thou art a strong upon command of fear not a
liberal largess given away, Falstaff! Exeunt

Trigram:

Sweet prince, Falstaff shall die. Harry of Monmouth's grave. This shall forbid it should be branded,
if renown made it empty. Indeed the duke; and had a very good friend. Fly, and will rid me these
news of price. Therefore the sadness of parting, as they say, 'tis done.

Quadrigram:

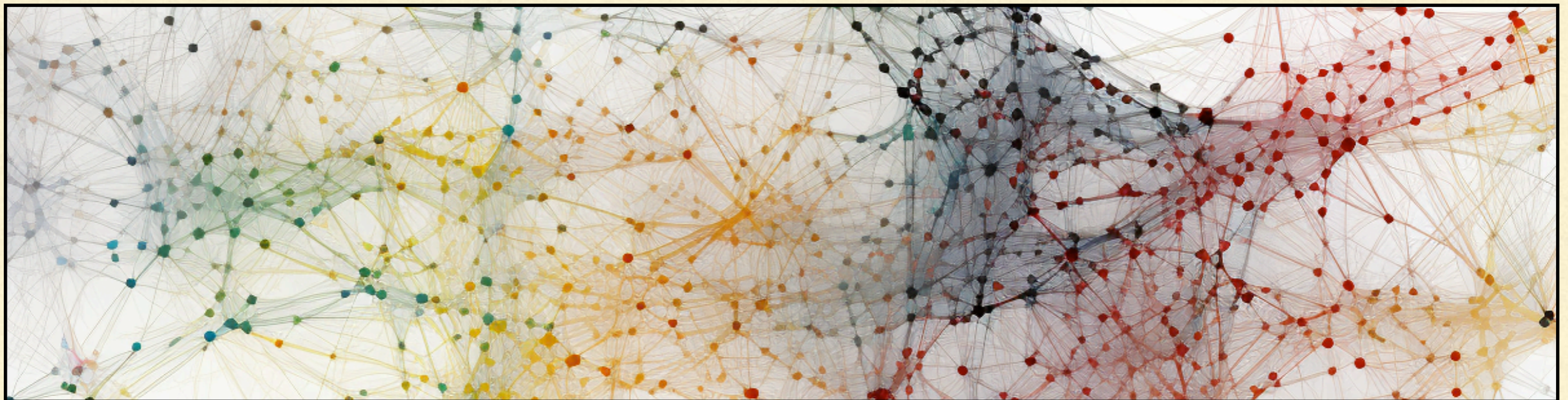
King Henry. What! I will go seek the traitor Gloucester. Exeunt some of the watch. A great banquet
serv'd in; Will you not tell me who I am? I cannot be but so. Indeed the short and the long. Marry,
'tis a noble Lepidus.

(Τα παραδείγματα είναι από D. Jurafsky & J. Martin, "Speech and Language Processing", 2010)

Μεγάλα Γλωσσικά Μοντέλα

Large Language Models (LLMs)

- Σύνθετα στατιστικά μοντέλα που συσχετίζουν κάθε λέξη με εκατοντάδες άλλες στο κείμενο.
- Όχι αποκλειστικά προηγούμενες, όχι μόνο σε ορθογραφική μορφή (αλλά ως σημασίες - embeddings).



Σημασιολογία μέσω Στατιστικής

σε όλα τα κείμενα

σε μεμονωμένα κείμενα

συχνές
λέξεις

το θα δέκα
έχω
από αύριο εσύ
δεν

το θα δέκα
έχω υποβρύχιο
από δεν αύριο
περισκόπιο εσύ

σπάνιες
λέξεις

αντιισταμινικό
υποβρύχιο
περισκόπιο
φωτορύπανση

αντιισταμινικό
φωτορύπανση

Αρκεί το μέτρημα λέξεων για τον εντοπισμό σημασιολογικών συσχετίσεων

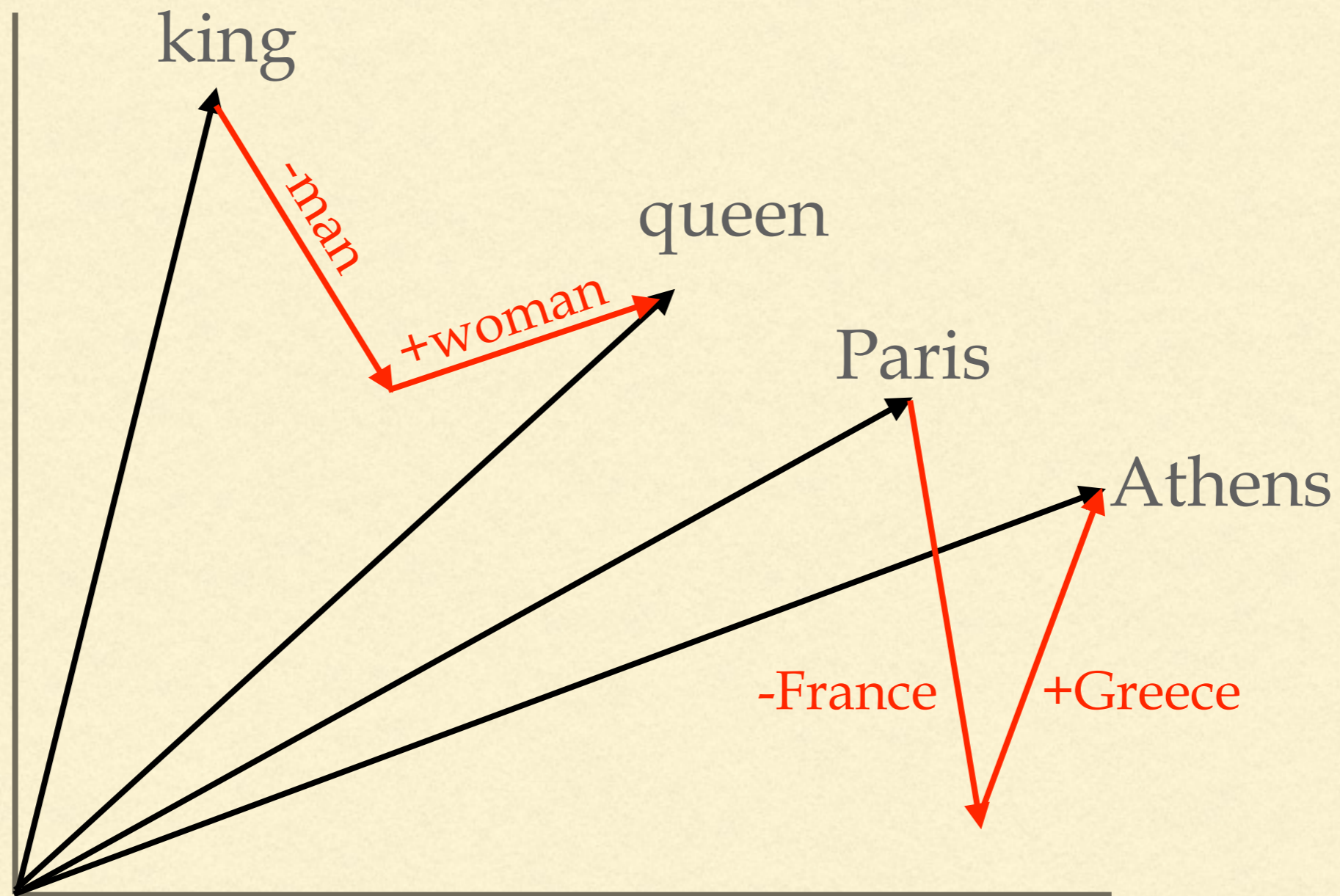
Πραγματολογία μέσω Στατιστικής

Έριξα νερό από το μπουκάλι στο ποτήρι μέχρι που γέμισε.

Έριξα νερό από το μπουκάλι στο ποτήρι μέχρι που άδειασε.

Έριξα το {ποτήρι, τούβλο, σίδερο, ...} στο τραπέζι και έσπασε.

Οι Σημασίες ως Διανύσματα



<https://en.wikipedia.org/wiki/Word2vec>

Distributional Semantics



John Rupert Firth

1890-1960

University College London
SOAS University of London

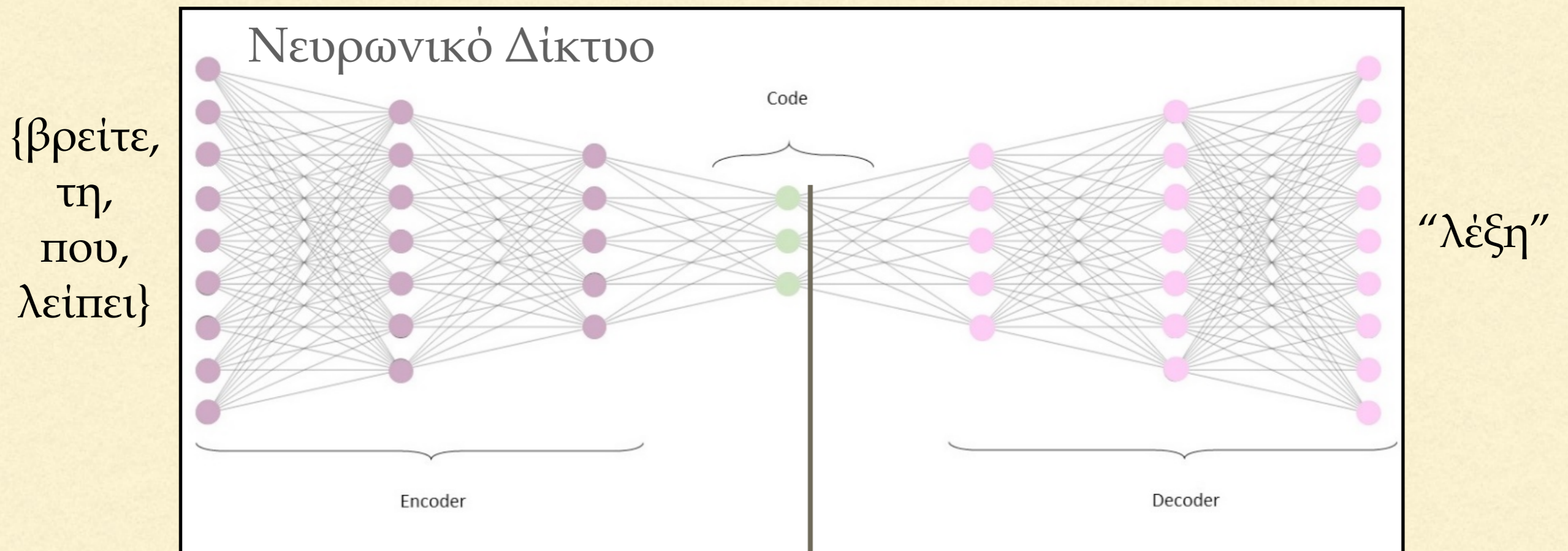
“You shall know a word by the
company it keeps” (John Rupert
Firth, 1957)

Word embeddings

https://en.wikipedia.org/wiki/John_Rupert_Firth

Word Embeddings

βρείτε τη _____ που λείπει

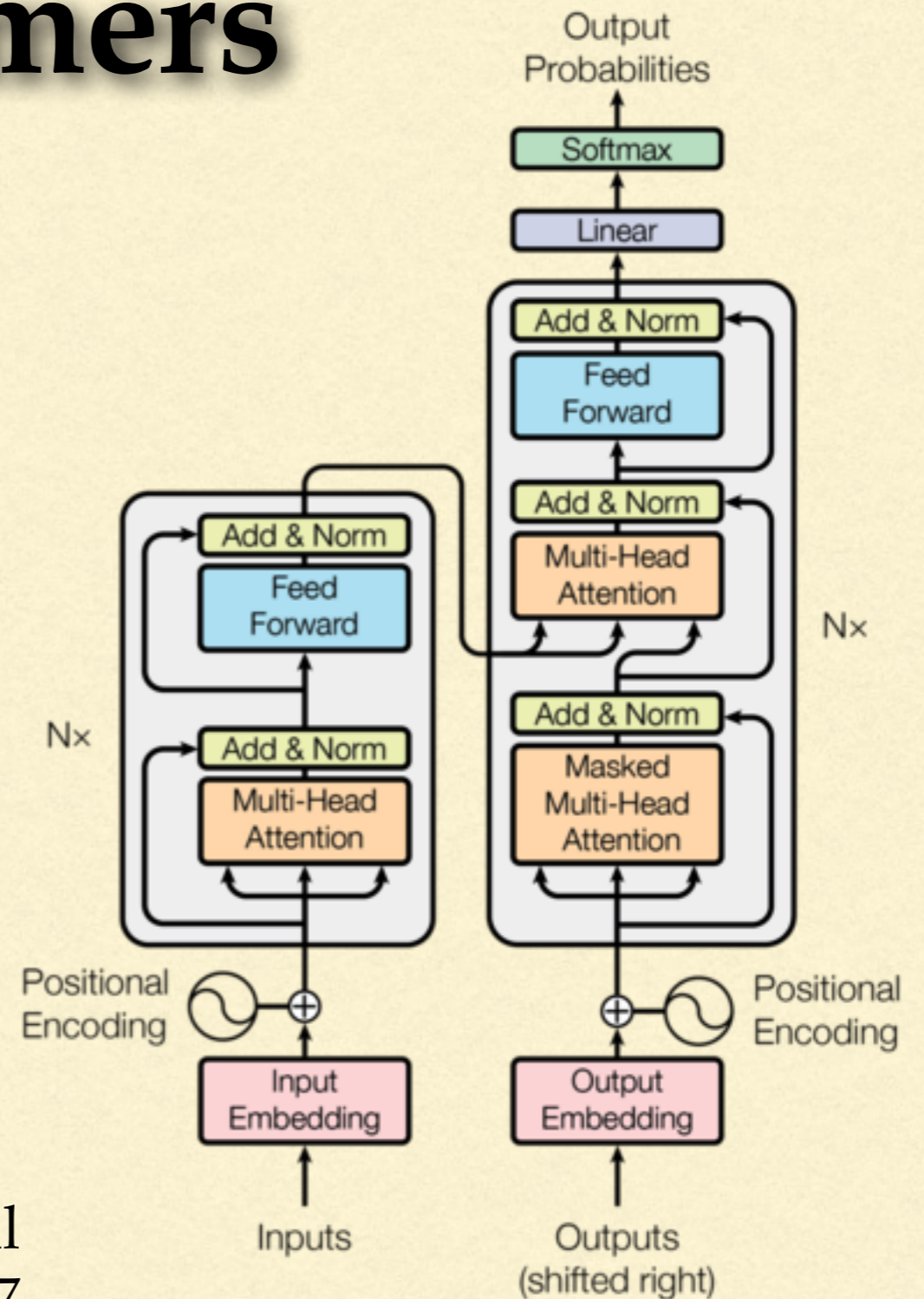


$[2.129, -0.166, 3.725] = \text{“λέξη”}$
word embedding (vector)

Transformers

- Από το νόημα μεμονωμένων λέξεων στο νόημα ολόκληρης της πρότασης
- Missing word training
- Next sentence training
- Χρησιμοποιήθηκαν αρχικά για αυτόματη μετάφραση

A. Vaswani et al., (Google), "Attention Is All You Need", arxiv.org, 2017



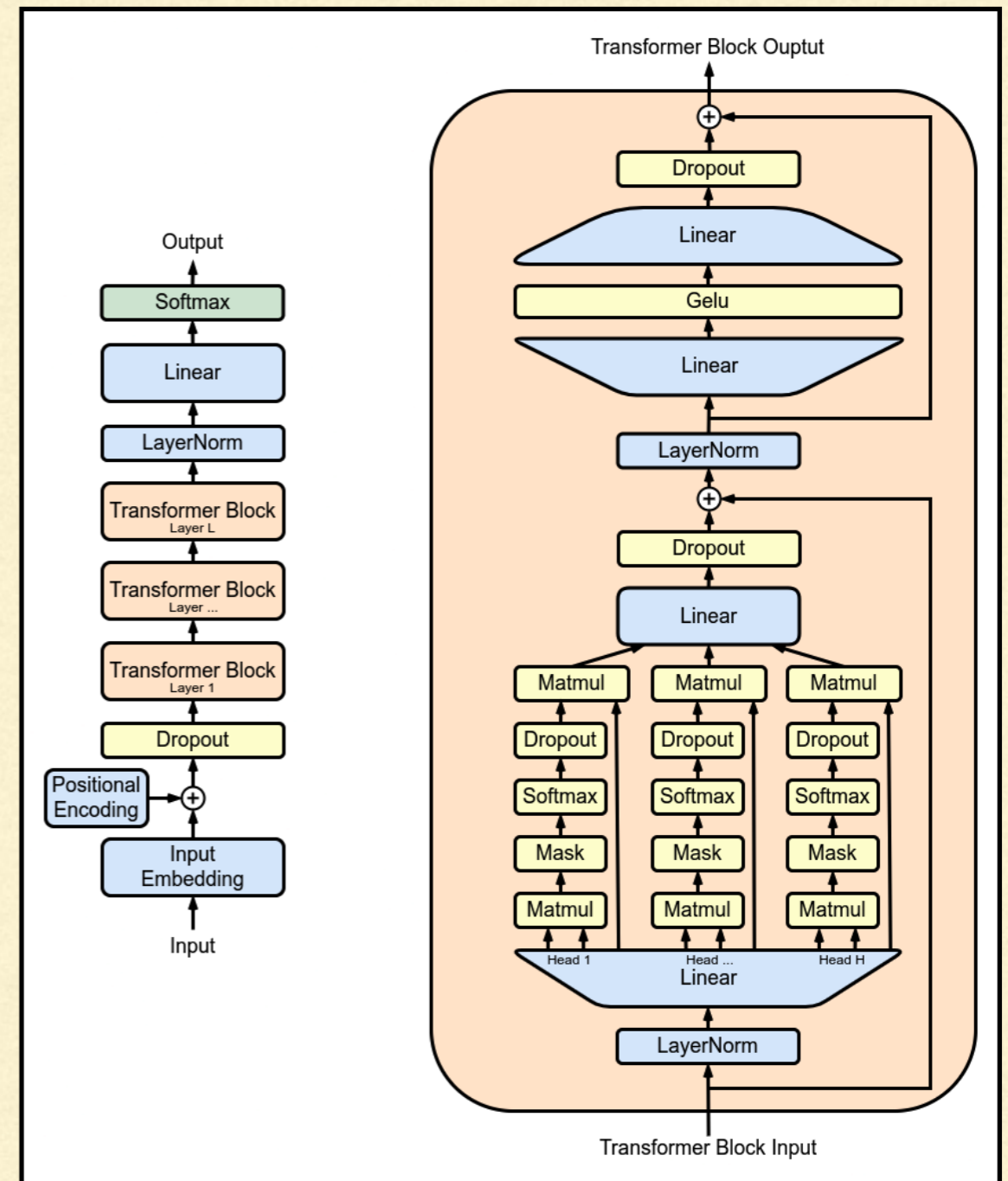
GPT

Generative Pre-trained Transformer

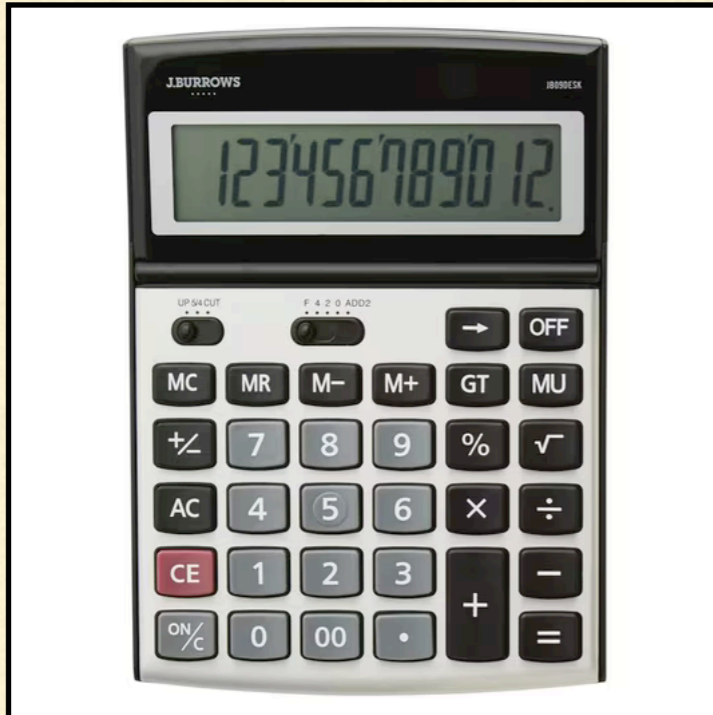
2018, OpenAI

500 billion words
(GPT-3, 2020)

ChatGPT (2022)



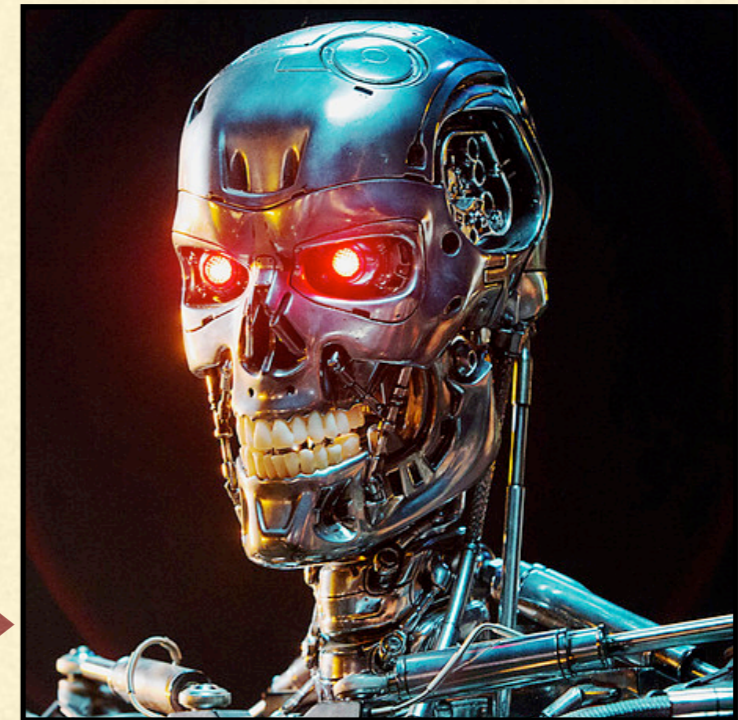
Να ανησυχούμε;



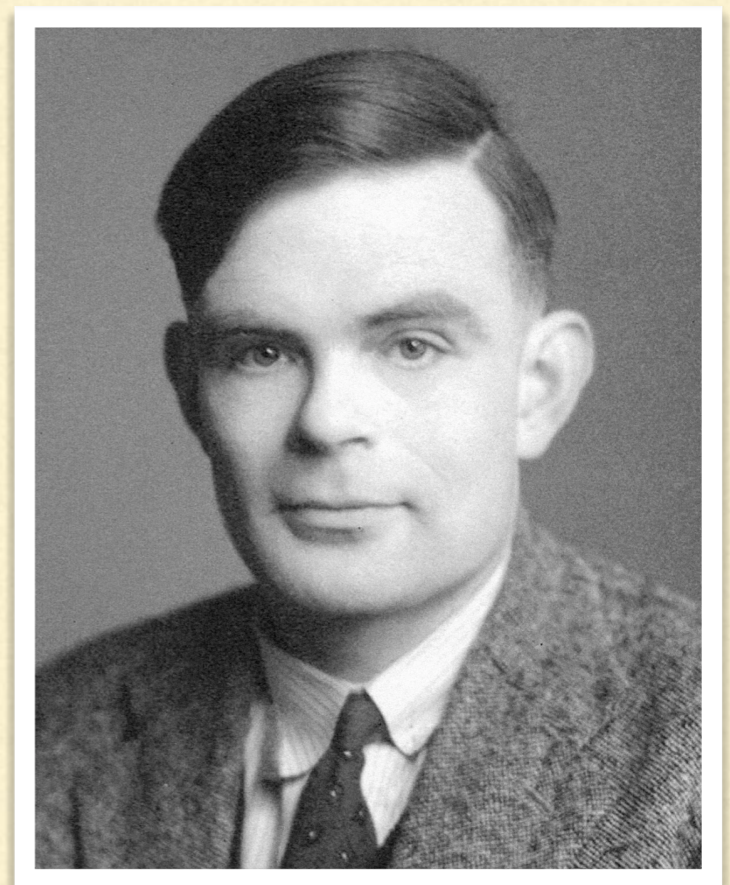
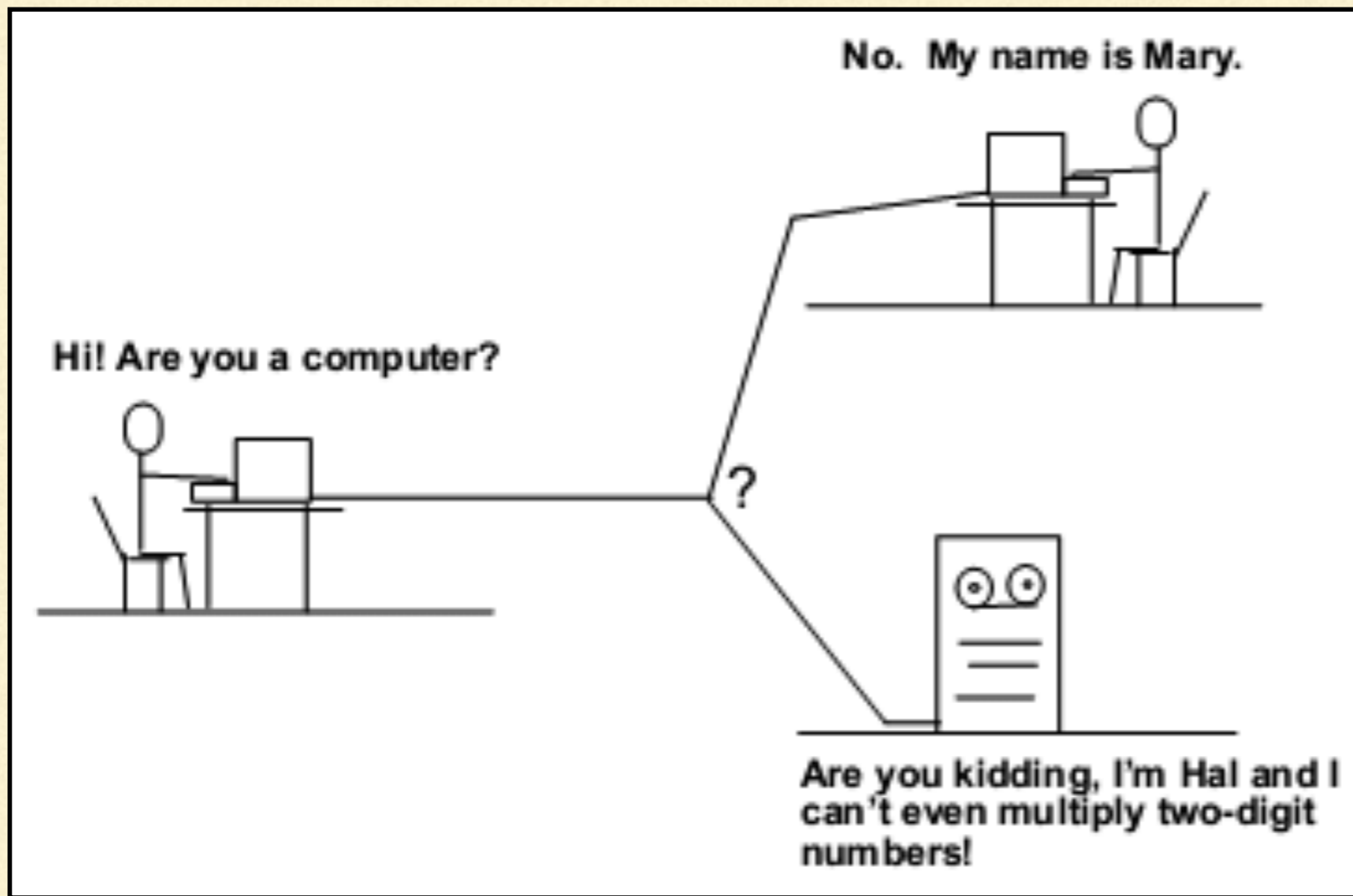
Τι είναι



Τι δεν είναι



To Test Turing (1950)

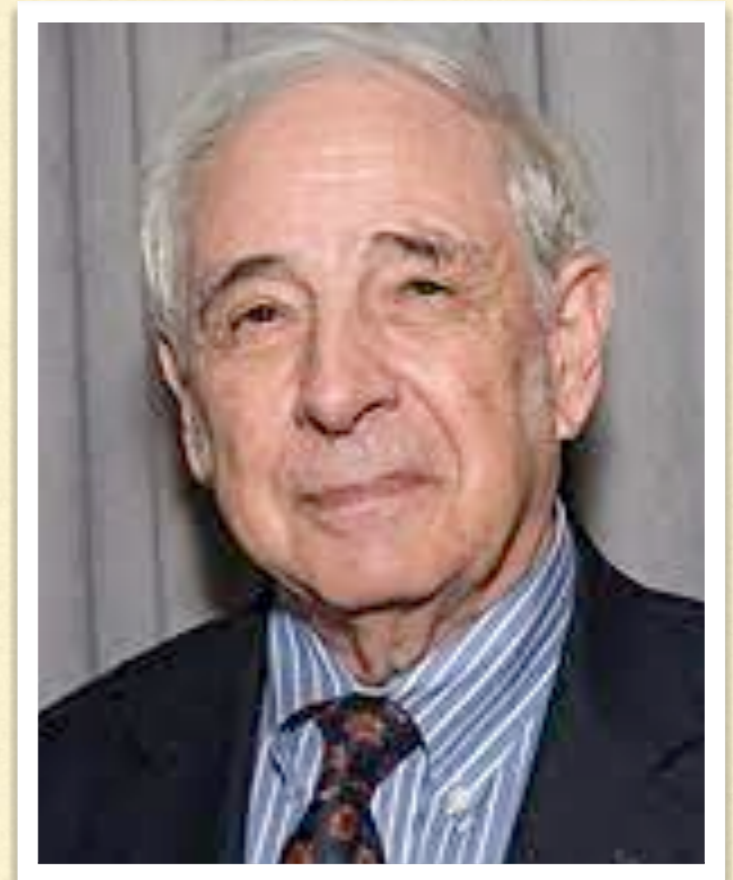
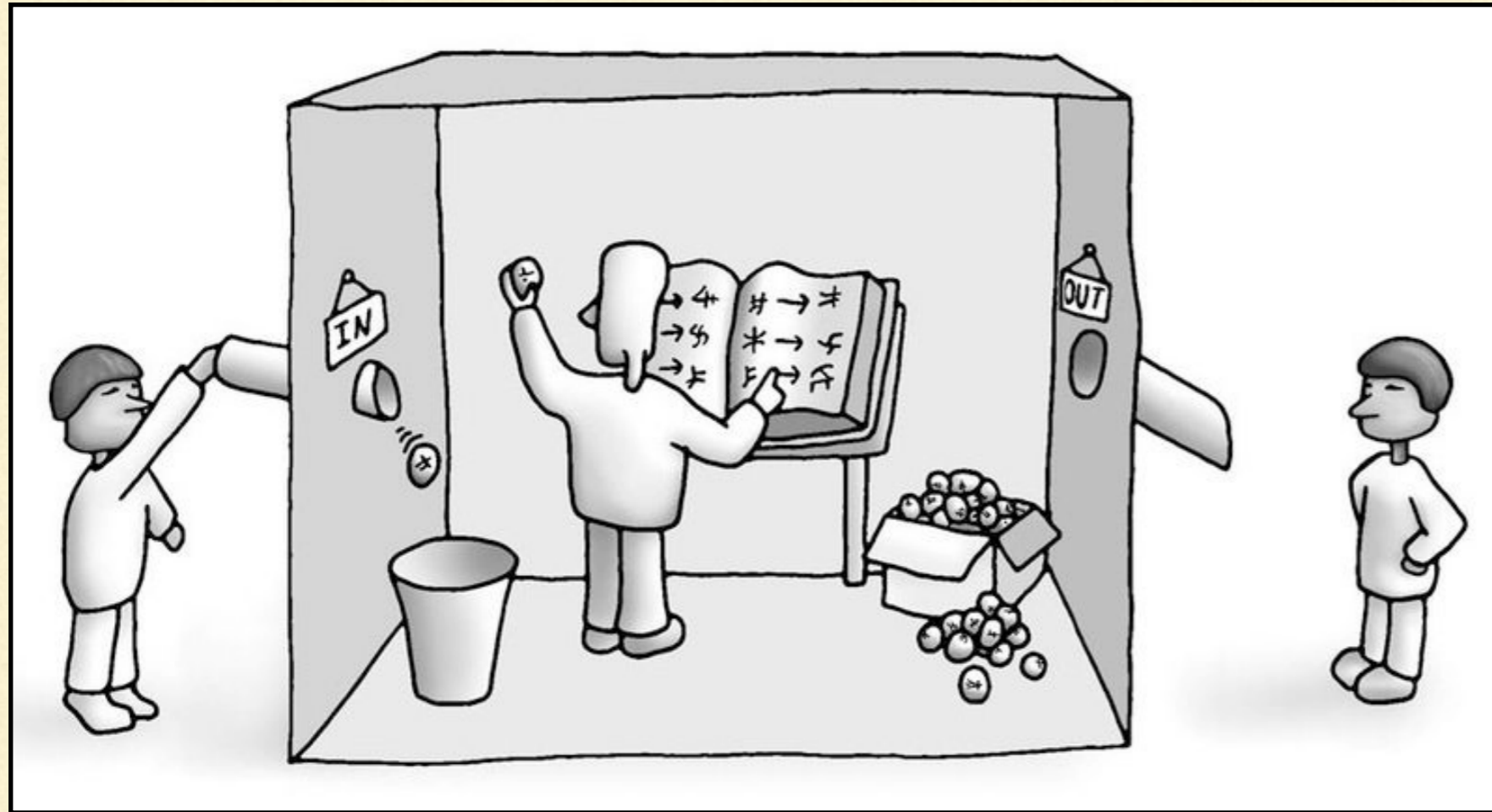


Alan Turing

1912-1954

King's College, Cambridge
PhD Princeton University

Το Κινέζικο Δωμάτιο (1980)



John Searle

1932-

University of Wisconsin–Madison
Christ Church, Oxford

Υπάρχουν Κίνδυνοι;

Ίσως ο μεγαλύτερος είναι να ξεχάσουμε ότι πρόκειται για στατιστικές μετρήσεις και να εμπιστευτούμε τυφλά όσα γράφουν.

TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER
THAT HE'S A COMPUTER.

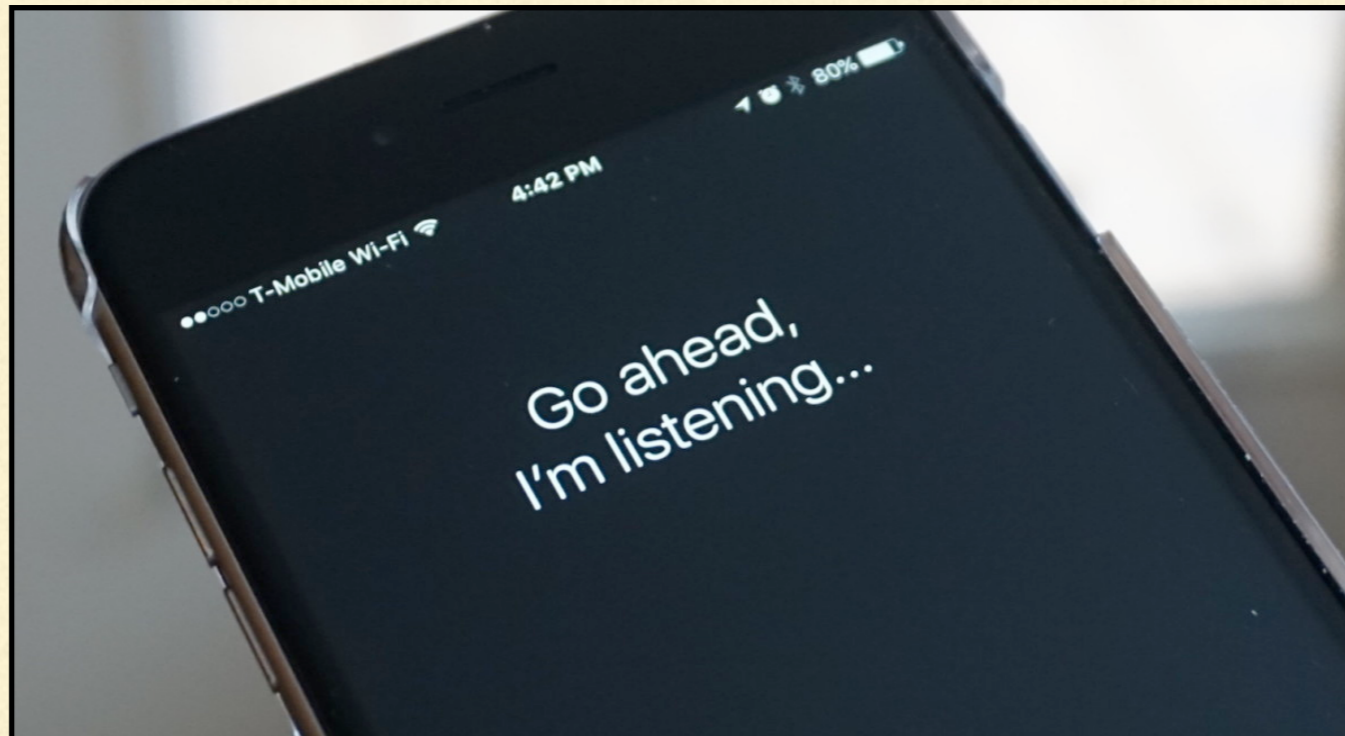
YOU KNOW, YOU MAKE
SOME REALLY GOOD POINTS.

I'M ... NOT EVEN SURE
WHO I AM ANYMORE.



<http://xkcd.com/>

Ερωτήσεις



sgarbas@upatras.gr

