



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Τεχνητή νοημοσύνη και μηχανική μάθηση

Εισηγητής
Αναστάσιος Κεσίδης



Εισαγωγή

Νοημοσύνη

➤ Ορισμοί

Η ικανότητα απόκτησης και εφαρμογής γνώσεων και δεξιοτήτων.

Oxford Dictionary

Η ικανότητα σκέψης και κατανόησης αντί της ενέργειας με ενστικτώδη ή αυτόματο τρόπο.

Collins English Dictionary



Τεχνητή Νοημοσύνη

➤ Ορισμοί

Ο τομέας της επιστήμης των υπολογιστών, που ασχολείται με τη σχεδίαση ευφυών (νοημόνων) **υπολογιστικών συστημάτων**, δηλαδή συστημάτων που επιδεικνύουν χαρακτηριστικά που σχετίζουμε με τη **νοημοσύνη** στην **ανθρώπινη συμπεριφορά**.

Barr και Feigenbaum 1981

Η συναρπαστική νέα προσπάθεια για να κάνουμε τους υπολογιστές να σκέπτονται..., **μηχανές με νόηση**, με την πλήρη και **κυριολεκτική έννοια**.

Hugeland 1985

Η αυτοματοποίηση των δραστηριοτήτων που συσχετίζουμε με την ανθρώπινη **σκέψη**, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων και η μάθηση.

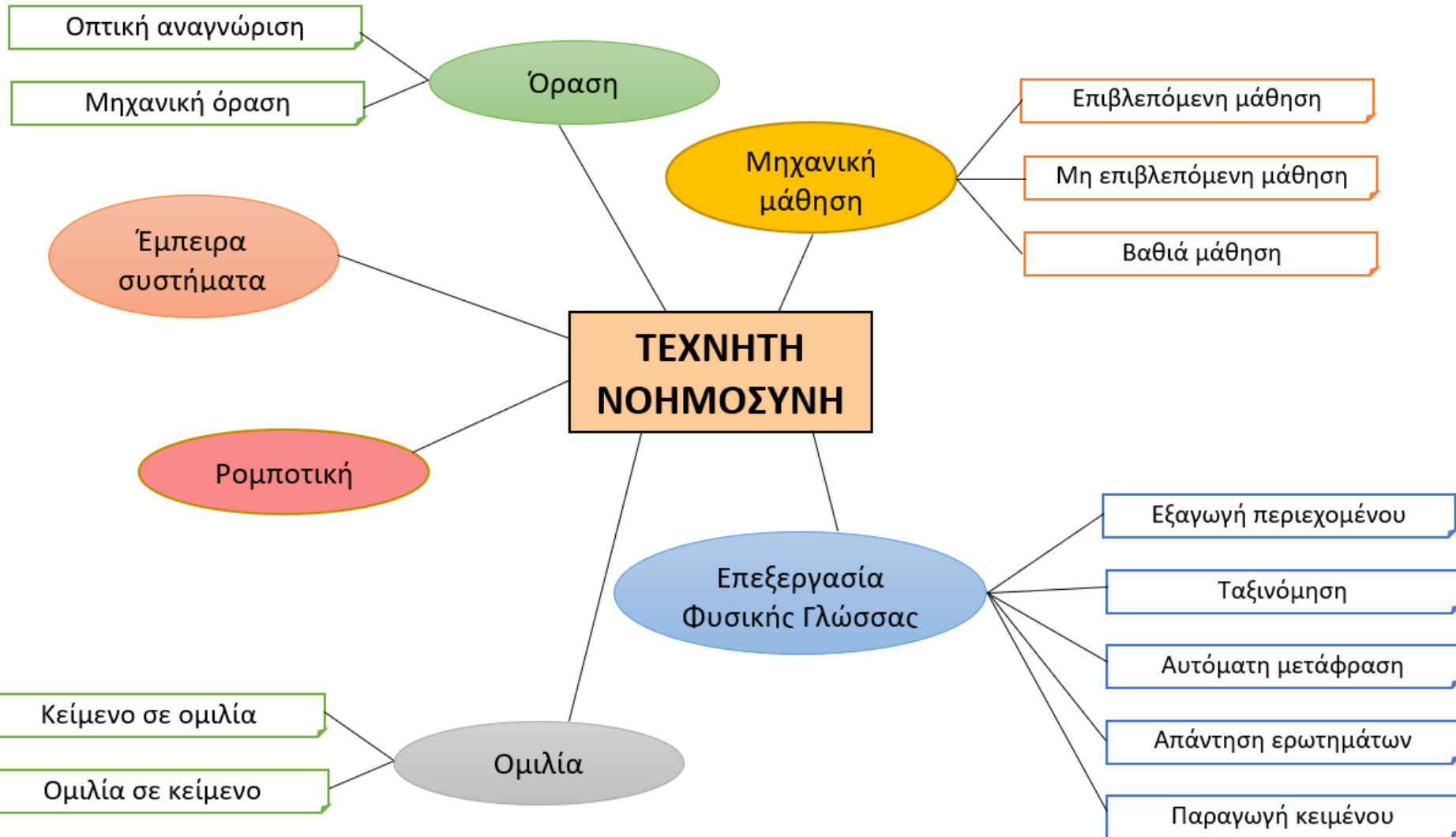
Bellman 1978

Η τέχνη της **δημιουργίας μηχανών** που πραγματοποιούν λειτουργίες οι οποίες απαιτούν **νοημοσύνη** όταν πραγματοποιούνται από ανθρώπους.

Kurzweil 1990

Τεχνητή Νοημοσύνη

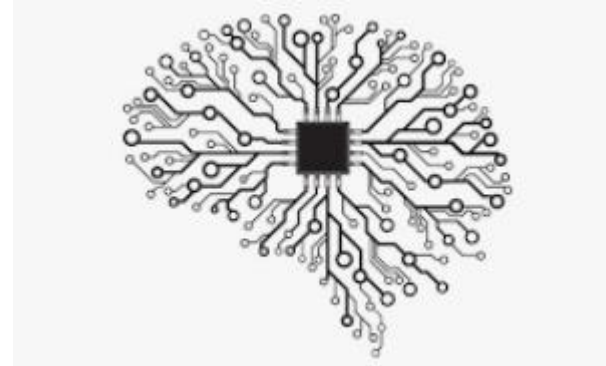
➤ Εφαρμογές της ΤΝ



Τεχνητή Νοημοσύνη

➤ Αντικείμενα

- Επίλυση προβλημάτων
- Μηχανική Μάθηση
- Τεχνητή Όραση
- Επεξεργασία Φυσικής Γλώσσας
- Προβλήματα Αναζήτησης
- Απόδειξη Θεωρημάτων
- Σχεδιασμός Ενεργειών και Χρονοπρογραμματισμός
- Αυτόνομα Robot
- Έμπειρα Συστήματα και Συστήματα Γνώσης
- Προσαρμοζόμενα και εξελισσόμενα ευφυή συστήματα



Σχέση της ΤΝ με άλλες Επιστήμες

➤ Φιλοσοφία

Θεμελιώσεις μάθησης, γλώσσας, ορθολογισμού, Λογική, μέθοδοι συλλογισμού

➤ Μαθηματικά

Αλγόριθμοι, Υπολογισμός, Πιθανότητες, Τυπική αναπαράσταση και απόδειξη

➤ Ψυχολογία

Προσαρμογή, Φαινόμενα αντίληψης και ελέγχου κίνησης

➤ Γλωσσολογία

Γραμματική, αναπαράσταση γνώσης

➤ Νευροεπιστήμη

Φυσικό υπόβαθρο για διανοητική λειτουργία

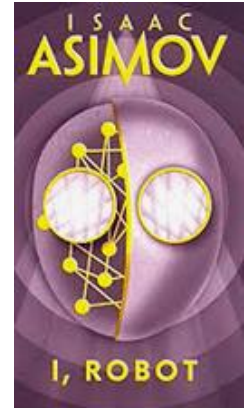
➤ Λογοτεχνία

▪ Επιστημονική φαντασία

Isaac Asimov (1942) – Εγώ, το ρομπότ

Τρεις νόμοι της ρομποτικής:

1. Ένα ρομπότ δε θα κάνει κακό σε άνθρωπο, ούτε με την αδράνειά του θα επιτρέψει να βλαφτεί ανθρώπινο ον.
2. Ένα ρομπότ πρέπει να υπακούει τις διαταγές που του δίνουν οι άνθρωποι, εκτός αν αυτές οι διαταγές έρχονται σε αντίθεση με τον πρώτο νόμο.
3. Ένα ρομπότ οφείλει να προστατεύει την ύπαρξή του, εφόσον αυτό δεν συγκρούεται με τον πρώτο και τον δεύτερο νόμο.



➤ Κινηματογράφος

Philip K. Dick (1968)

Do androids dream of electric sheep?

Blade Runner (1982)



Ιστορική αναδρομή

➤ Αρχαιότητα

«Αυτόματα» και τεχνητά όντα στους ελληνικούς μύθους

- Ήφαιστος - Πανδώρα
- Δαίδαλος - «ζωντανά αγάλματα» από χάλκινα γλυπτά που έδειχναν ανθρώπινα χαρακτηριστικά
- Τάλος - προστασία της Κρήτης από εισβολείς



«Βιοτεχνολογία»

Πώς βιολογικά φαινόμενα όπως η γήρανση μπορούν να μεταβληθούν με την τεχνολογία των ανθρώπων.

Ιστορική αναδρομή

➤ Αντικείμενο φιλοσοφικής μελέτης

Διερεύνηση εννοιών όπως

- Μάθηση
- Αντίληψη
- Συλλογισμός

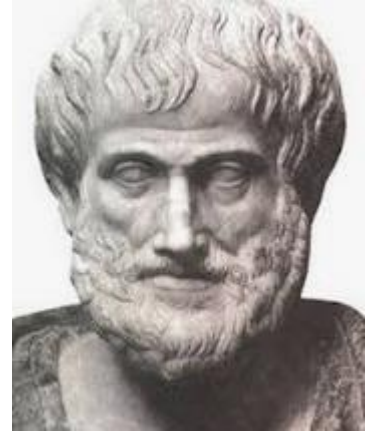
«Αριστοτέλεια λογική»

Πρότυπα εκφράσεων όπου «σωστές» υποθέσεις οδηγούν σε «σωστά» συμπεράσματα.

- Ο Σωκράτης είναι άνθρωπος.
- Όλοι οι άνθρωποι είναι θνητοί.



Ο Σωκράτης είναι θνητός



Καρτέσιος (1642)

Discourse on the Method - Τα «automata» δεν μπορούν να ανταποκριθούν με τον τρόπο που μπορεί ένας άνθρωπος.



➤ Λογική

Διερεύνηση των τρόπων ορθού συλλογισμού

Μαθηματική λογική

- Παράσταση γνώσης
- Μηχανισμοί παραγωγής συλλογισμών

George Boole (1848)

“The Mathematical Analysis of Logic” – συμβολική λογική

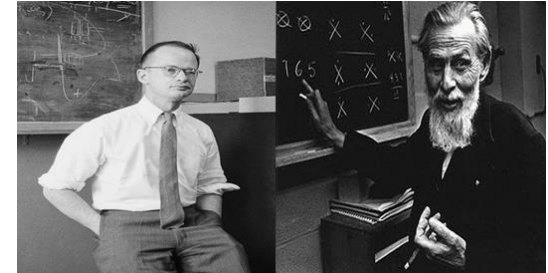


Ιστορική αναδρομή

➤ Walter Pitts & Warren McCulloch (1943)

Μοντέλο τεχνητού νευρώνα

- Μαθηματικό μοντέλο ενός βιολογικού νευρώνα.



➤ Alan Turing (1950)

Turing test

- Για να διαπιστωθεί πειραματικά αν μία μηχανή έχει αυθεντικές γνωστικές ικανότητες και μπορεί να σκεφτεί.



VOL. LIX. No. 236.]

[October, 1950]

MIND

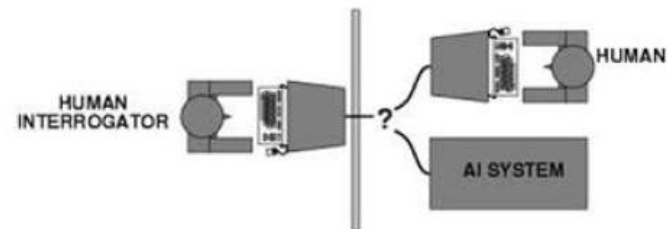
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to



Δοκιμασία Turing

➤ Τεστ αναγνώρισης ευφύων μηχανών (Turing test 1950)

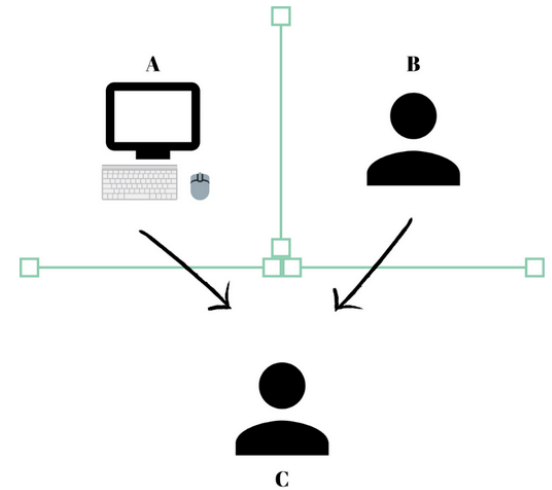
- Σειρά ερωτήσεων από τον εξεταστή C προς μια μηχανή A και έναν άνθρωπο B (μη ορατοί στον εξεταστή)
- Εάν στο τέλος ο εξεταστής δεν μπορεί να διακρίνει τον άνθρωπο από την μηχανή, τότε η μηχανή περνάει το τεστ και θεωρείται ευφυής.

Απαιτεί υπολογιστικό σύστημα ικανό για

- Επεξεργασία φυσικής γλώσσας
- Αναπαράσταση γνώσης
- Αυτοματοποιημένη συλλογιστική
- Μηχανική μάθηση

Το πλήρες τεστ περιλαμβάνει αναγνώριση εικόνων και αντικειμένων

- Μηχανική όραση
- Ρομποτική



Μπορούν οι μηχανές να σκέφτονται;
Μπορούν οι μηχανές να συμπεριφέρονται έξυπνα;

➤ Dartmouth Summer Research Project (1956)

- Θεωρείται η «γέννηση» της ΤΝ
- Προτείνεται από τον McCarthy ο όρος «Τεχνητή Νοημοσύνη»

"..that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

- A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (McCarthy et al, 1955)



J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C. E. Shannon, Bell Telephone Laboratories

"..ότι κάθε πτυχή της μάθησης ή οποιοδήποτε άλλο χαρακτηριστικό της νοημοσύνης μπορεί καταρχήν να περιγραφεί με ακρίβεια ώστε να μπορεί να δημιουργηθεί μια μηχανή που να το προσομοιώνει. Θα γίνει μια προσπάθεια να βρεθεί πώς να φτιαχτούν μηχανές που να χρησιμοποιούν τη γλώσσα, να σχηματίζουν αφηρημένες έννοιες, να λύνουν διάφορα είδη προβλημάτων που προορίζονται τώρα για τον άνθρωπο και να αυτο-βελτιώνονται. Πιστεύουμε ότι σημαντική πρόοδος μπορεί να προκύψει σε ένα ή περισσότερα από αυτά τα προβλήματα εάν μια προσεκτικά επιλεγμένη ομάδα επιστημόνων δουλέψει πάνω σε αυτό για ένα καλοκαίρι. "

Ιστορική αναδρομή

➤ Logic Theorist (1956)

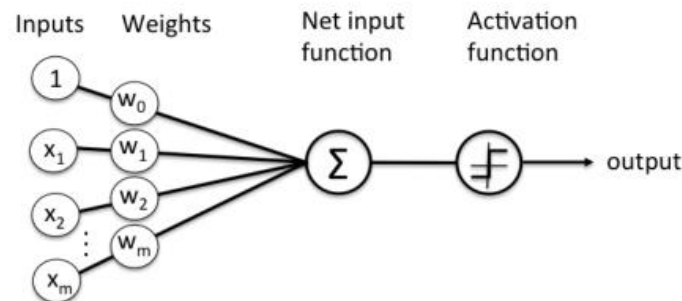
A. Newel, H. A. Simon & C. Shaw - Το πρώτο πρόγραμμα τεχνητής νοημοσύνης

- Ικανό να εκτελεί «αυτοματοποιημένη συλλογιστική».
- Απέδειξε 38 από τα πρώτα 52 θεωρήματα του **Principia Mathematica** των Whitehead και Russell.
- Βρήκε **νέες αποδείξεις** για ορισμένα από τα θεωρήματα αυτά.

➤ Perceptron (1962)

F. Rosenblatt

- Τεχνητό νευρωνικό δίκτυο



Ιστορική αναδρομή

➤ Eliza (1962)

J. Weizenbaum - MIT

- Πρόγραμμα που λειτουργούσε ως ψυχαναλύτης.
- Συνήθως επαναλάμβανε στον χρήστη ό,τι είπε, ως ερώτηση.
- Είχε κανόνες που ενεργοποιούνταν από βασικές λέξεις στις απαντήσεις του χρήστη.
- Κρατούσε σε μικρή μνήμη όσα είχαν ειπωθεί στο άμεσο παρελθόν.

```
> Hello, I am Eliza.  
* I am afraid.  
> How long have you been afraid?  
* Since midterms.  
> Tell me more...  
* I am afraid that I will fail  
my course.  
> Do you enjoy being afraid that  
you will fail your course?
```

➤ J. Weizenbaum (1976)

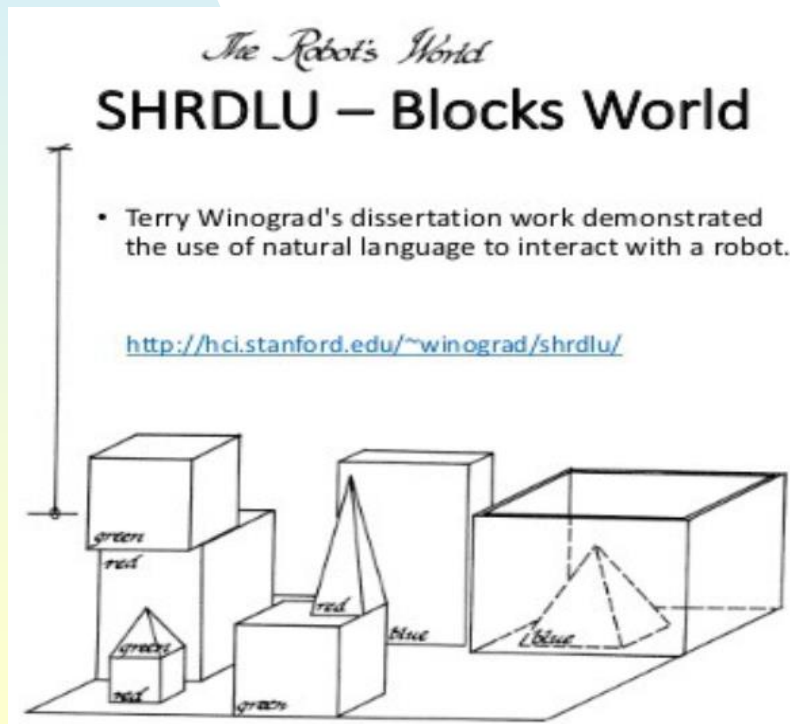
"Δεν είχα συνειδητοποιήσει ... ότι εξαιρετικά σύντομες εκθέσεις σε ένα σχετικά απλό πρόγραμμα υπολογιστή θα μπορούσαν να προκαλέσουν έντονη παραληρηματική σκέψη σε γενικά φυσιολογικούς ανθρώπους."

Ιστορική αναδρομή

➤ SHRDLU (1971)

T. Winograd

- Από τα πρώτα προγράμματα κατανόησης **φυσικής γλώσσας**.
- Ο χρήστης «συνομιλούσε» με τον υπολογιστή μετακινώντας αντικείμενα, ονομάζοντας συλλογές και θέτοντας ερωτήματα που αφορούσαν έναν κόσμο που αποτελείτο από βασικά τρισδιάστατα γεωμετρικά αντικείμενα.



Person: PICK UP A BIG RED BLOCK.
Computer: OK. (does it)
Person: GRASP THE PYRAMID.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer: OK. (does it)
Person: WHAT DOES THE BOX CONTAIN?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
Person: WHAT IS THE PYRAMID SUPPORTED BY?
Computer: THE BOX.
Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
Computer: FOUR OF THEM.
Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
Computer: YES, THE RED CUBE.

Ιστορική αναδρομή

➤ Κριτική στα τέλη της δεκαετίας του 1960

- Ανυπαρξία **πραγματικής γνώσης** του υπό εξέταση προβλήματος (π.χ. ELIZA)
- Αδυναμία **κλιμάκωσης** των συστημάτων (εφαρμογή μόνο σε συγκεκριμένα «κλειστά» προβλήματα, π.χ. SHRDLU)
- Marvin Minsky and Seymour Papert
Perceptrons – An introduction to Computational Geometry (1969)
 - Κριτική στην αποτελεσματικότητα των perceptrons
 - Απέδειξαν ότι το perceptron μπορεί να χρησιμοποιηθεί μόνο για γραμμικές συναρτήσεις.
 - Δεν μπορεί να λύσει το πρόβλημα της μη γραμμικής **συνάρτησης XOR**
- Η αποτυχία αυτή είχε σαν αποτέλεσμα να εγκαταλειφθεί η ιδέα των νευρωνικών δικτύων και να σταματήσει σχεδόν κάθε έρευνα πάνω στο αντικείμενο αυτό

➤ Ερευνητικός «χειμώνας» δεκαετίας 1970

- Υπερβολικά αισιόδοξες προβλέψεις
- **Μείωση της χρηματοδότησης** σε ΗΠΑ και Η. Βασίλειο

Ιστορική αναδρομή

➤ Έμπειρα συστήματα

Μίμηση ικανότητας ενός **εμπειρογνώμονα** στη λήψη αποφάσεων μέσω συσσώρευσης γνώσης για συγκεκριμένη περιοχή εφαρμογής

- Χρήση ειδικών για παραγωγή/ενσωμάτωση γνώσης
- Χρήση αυτοματοποιημένης συλλογιστικής για επίλυση πολύπλοκων προβλημάτων με βάση την διαθέσιμη γνώση

Δύο κύρια μέρη

- η μηχανή εξαγωγής συμπερασμάτων (**σταθερή**, ανεξάρτητη από το έμπειρο σύστημα)
- η βάση γνώσης (**μεταβλητή**, συνεχώς εμπλουτιζόμενη)

Dendral (1969)

Εύρεση μοριακής δομής οργανικών ενώσεων από δεδομένα φασματογράφου μάζας.

Mycin (μέσα '70) <https://www.shortliffe.net/Shortliffe-1976/MYCIN%20thesis%20Book.htm>

Διάγνωση μολύνσεων αίματος. Παράσταση αβέβαιης γνώσης.

Prospector (τέλη '70)

Συμβουλές για τοποθεσίες εξόρυξης.

Ιστορική αναδρομή

➤ Υπολογιστές πέμπτης γενιάς (1982-1990)

Japan's Fifth Generation Computer Systems project

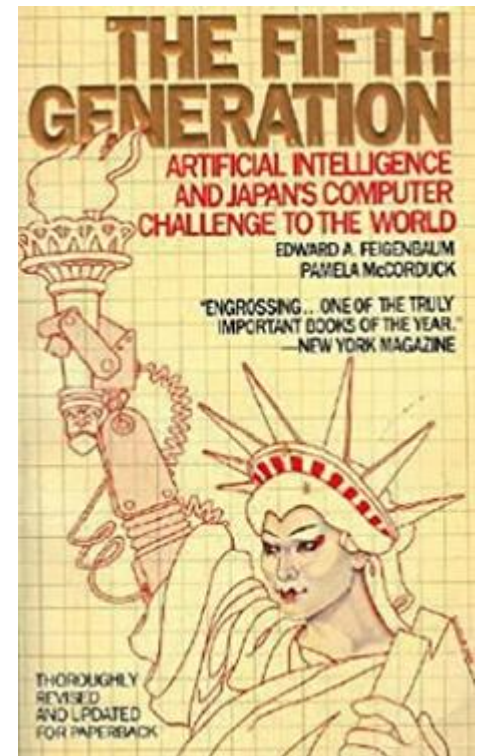
Χρήση της λογικής για:

- την έκφραση της πληροφορίας στον υπολογιστή
- την παρουσίαση του προβλήματος στον υπολογιστή
- την επίλυση των προβλημάτων αυτών

Πρόγραμμα = σύνολο από αξιώματα

Υπολογισμοί = απόδειξη δηλώσεων μέσω των αξιωμάτων

Εκτεταμένη χρήση παράλληλης αρχιτεκτονικής υπολογιστών για αύξηση της υπολογιστικής ισχύος



Ιστορική αναδρομή

➤ Νευρωνικά δίκτυα (1985-1986)

Αλγόριθμος οπισθοδιάδοσης του σφάλματος (**Back-propagation**)

- Πολυστρωματικά νευρωνικά δίκτυα
- Επίλυση μη γραμμικών προβλημάτων (π.χ. XOR πρόβλημα)

<https://xnought.github.io/backprop-explainer/>



➤ Αναμέτρηση ανθρώπου-μηχανής

Σκάκι - IBM Deep Blue

Υπολογισμός 200 εκατομμύριων πιθανών θέσεων το δευτερόλεπτο

Deep Blue vs Gary Kasparov (1997)

- 3½ – 2½ σε 6 γύρους



Ιστορική αναδρομή

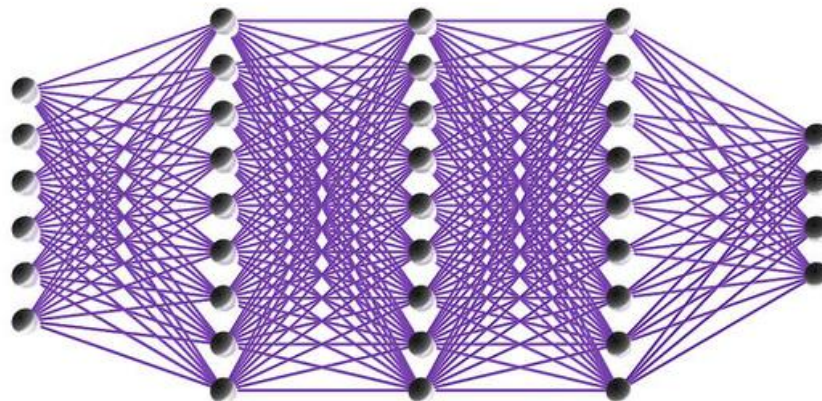
➤ Μηχανική μάθηση (2000-)

Έμφαση σε συστήματα **μεγάλης κλίμακας**

Μεγάλα δεδομένα (**Big Data**)

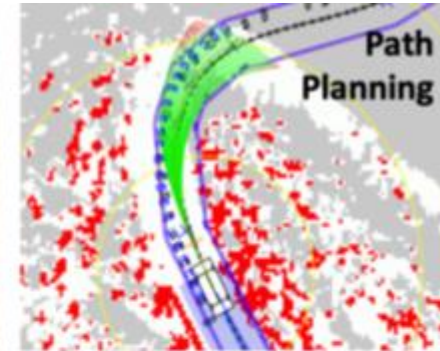
Βαθιά μάθηση (**Deep Learning**)

- **Πολλαπλά επίπεδα** νευρώνων
- Χρειάζονται πολλά παραδείγματα και σημαντική **υπολογιστική ισχύς** για την εκπαίδευσή τους
- Πληθώρα εφαρμογών με συχνά εντυπωσιακά αποτελέσματα



Σύγχρονες εφαρμογές

➤ Autonomous vehicles



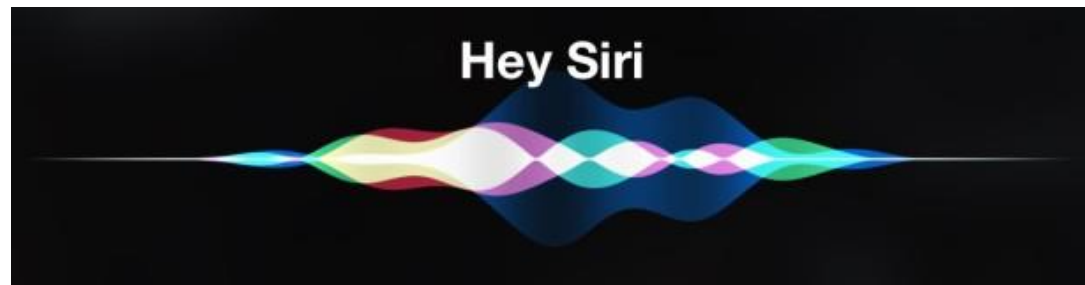
Σύγχρονες εφαρμογές

➤ Automatic scene labeling



Σύγχρονες εφαρμογές

➤ Speech technology



Σύγχρονες εφαρμογές

➤ Stock price prediction



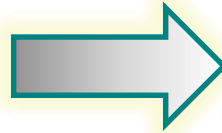
Σύγχρονες εφαρμογές

➤ Deepfake video generation and detection



Παράδειγμα

Κατηγοριοποίηση εικόνων



Άνδρες



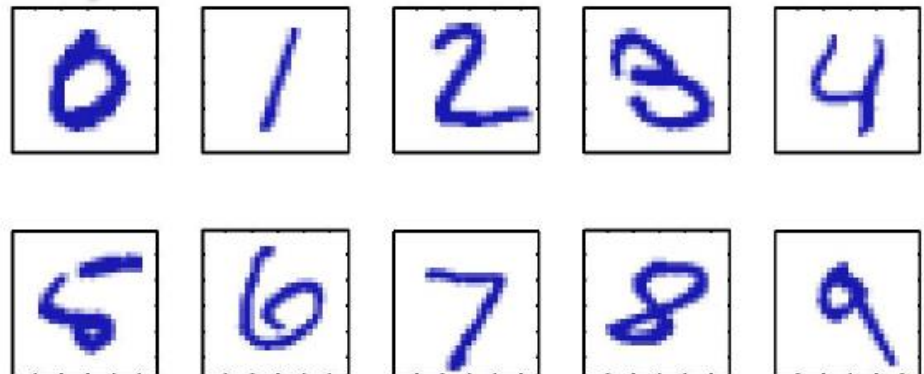
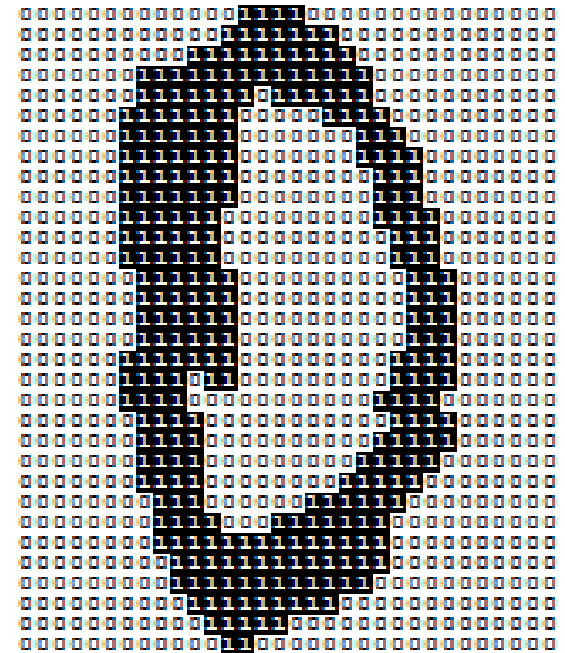
Γυναίκες



Παράδειγμα (συν.)

➤ Προσέγγιση αναγνώρισης προτύπων

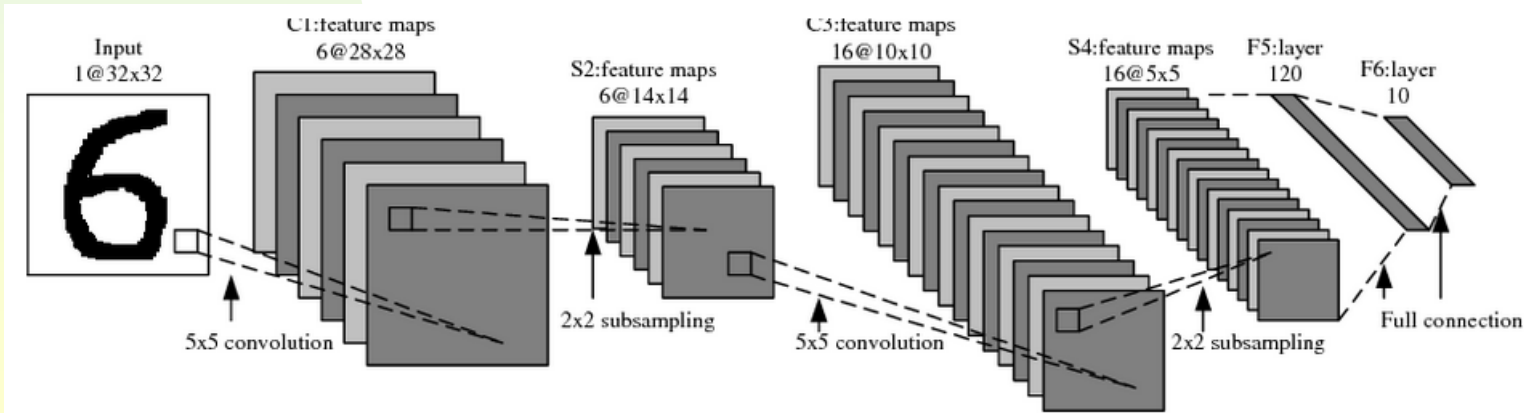
- Προεπεξεργασία δεδομένων
- **Επιλογή χαρακτηριστικών**
(γεωμετρικά χαρακτηριστικά, χρώμα, υφή κλπ)
- Επιλογή μοντέλου/ταξινομητή
- Εκπαίδευση
- Επικύρωση



Παράδειγμα (συν.)

➤ Προσέγγιση βαθιάς μάθησης

- Προεπεξεργασία δεδομένων
- **Αυτόματη επιλογή χαρακτηριστικών**
- Επιλογή νευρωνικού δικτύου (αρχιτεκτονική, κανόνες μάθησης)
- Εκπαίδευση
- Επικύρωση



Παράδειγμα (συν.)

Κατηγοριοποίηση εικόνων (με **διαφορετικό κριτήριο...**)

Graphic

Εικόνα

