

Εισαγωγή στην Τεχνητή Νοημοσύνη

Τεχνητή Νοημοσύνη Ι

Εξάμηνο 5

Ασπασία Δασκαλοπούλου

Τι είναι Τεχνητή Νοημοσύνη (AI)

“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, η μάθηση...” (Bellman 1978)

“ Η μελέτη των νοητικών ικανοτήτων με τη χρήση υπολογιστικών μοντέλων” (Charniak & McDermott 1985)

“Η μελέτη των υπολογιστικών εργασιών που μας δίνουν τη δυνατότητα να αντιλαμβανόμαστε, να συλλογίζομαστε, και να ενεργούμε” (Winston 1992)

“Η τέχνη της δημιουργίας μηχανών που πραγματοποιούν λειτουργίες οι οποίες απαιτούν νοημοσύνη όταν πραγματοποιούνται από ανθρώπους” (Kurzweil 1990)

“Η μελέτη του πώς μπορούμε να κάνουμε τους υπολογιστές να κάνουν πράγματα στα οποία προς το παρόν οι άνθρωποι είναι καλύτεροι” (Rich & Knight 1991)

“Υπολογιστική Νοημοσύνη είναι η μελέτη της σχεδίασης ευφυών πρακτόρων” (Poole et al 1998)

“Η τεχνητή νοημοσύνη ασχολείται με την ευφυή συμπεριφορά των τεχνουργημάτων” (Nilsson 1998)

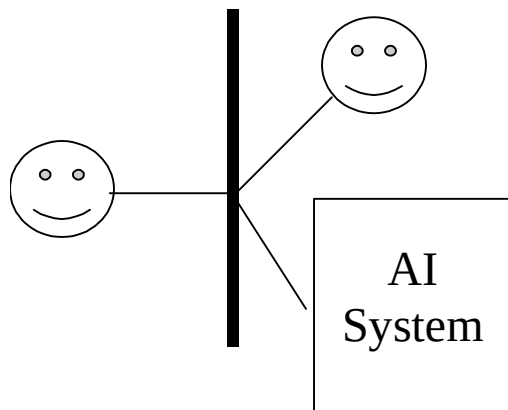
Τι είναι Τεχνητή Νοημοσύνη (ΑΙ)

Συστήματα που σκέπτονται σαν τον άνθρωπο	Συστήματα που σκέπτονται ορθολογικά
Συστήματα που ενεργούν σαν τον άνθρωπο	Συστήματα που ενεργούν ορθολογικά

Συστήματα που σκέπτονται σαν τον άνθρωπο	Συστήματα που σκέπτονται ορθολογιστικά
Συστήματα που ενεργούν σαν τον άνθρωπο	Συστήματα που ενεργούν ορθολογιστικά

Ανθρώπινη δράση: Το τεστ Turing

- Alan Turing (1950) “Computing machinery and intelligence”
- “Μπορούν οι μηχανές να σκέφτονται;”
- “Μπορούν οι μηχανές να συμπεριφέρονται έξυπνα;”



- Για να περάσει ένας υπολογιστικό σύστημα το τεστ θα πρέπει να γίνουν πολλά:
 - Επεξεργασία φυσικής γλώσσας
 - Αναπαράσταση γνώσης
 - Αυτοματοποιημένη συλλογιστική
 - Μηχανική μάθηση
- Κι ακόμα, για το πλήρες τεστ
 - Μηχανική όραση
 - Ρομποτική
- Όρισε τα βασικά πεδία της AI και εξακολουθεί να ισχύει σήμερα.
- Αλλά δεν επιδέχεται μαθηματικής ανάλυσης ή αναπαραγωγής ώστε να μπορεί κανείς να κατανοήσει το *μηχανισμό* της νοημοσύνης.

Συστήματα που σκέφτονται σαν τον άνθρωπο	Συστήματα που σκέπτονται ορθολογιστικά
Συστήματα που ενεργούν σαν τον άνθρωπο	Συστήματα που ενεργούν ορθολογιστικά

Ανθρώπινη σκέψη: γνωστική επιστήμη

- “Μπορούν οι μηχανές να σκέφτονται;”
- “Μπορούν οι μηχανές να σκέφτονται σαν τον άνθρωπο;”
- “Πώς σκέφτεται ο άνθρωπος;”
- Γνωστική επανάσταση 1960 : η ψυχολογία της επεξεργασίας πληροφορίας αντικατέστησε τον συμπεριφορισμό.
- Απαιτούνται επιστημονικές θεωρίες που να εξηγούν την εσωτερική λειτουργία του ανθρώπινου εγκεφάλου.
 - Κατάλληλο επίπεδο αφαίρεσης; Γνώση ή κυκλώματα;
 - Πώς θα επαληθεύσουμε τις θεωρίες
 - Πρόβλεψη και έλεγχος συμπεριφοράς ανθρώπων (top down, γνωστική επιστήμη)
 - Με βάση νευρολογικά δεδομένα (bottom up, γνωστική νευροεπιστήμη)
- Κοινός προβληματισμός για AI, γνωστική επιστήμη, γνωστική νευροεπιστήμη: οι υπάρχουσες θεωρίες ούτε εξηγούν επαρκώς ούτε μπορούν να γεννήσουν νοημοσύνη που να προσεγγίζει την ανθρώπινη.

Συστήματα που σκέπτονται σαν τον άνθρωπο	Συστήματα που σκέπτονται ορθολογιστικά
Συστήματα που ενεργούν σαν τον άνθρωπο	Συστήματα που ενεργούν ορθολογιστικά

Ορθολογική σκέψη: οι νόμοι της σκέψης

- Αριστοτέλης: “**ποιός είναι ο σωστός τρόπος σκέψης;**”
- Σωστός = αδιαφιλονίκητος
- **Συλλογισμοί:** πρότυπες δομές επιχειρημάτων που δίνουν πάντα σωστά συμπεράσματα όταν ξεκινήσει κανείς από σωστές υποθέσεις.
- Πολλές ελληνικές σχολές ανέπτυξαν διάφορα συστήματα **λογικής**
 - Συμβολισμό
 - Κανόνες παραγωγής συμπεράσματος
- **Λογικιστική τεχνητή νοημοσύνη (συμβολική):** υπολογιστικά συστήματα που μπορούν να λύνουν προβλήματα διατυπωμένα στην κατάλληλη σημειογραφία.
- Προβλήματα:
 - Δεν περιγράφεται κάθε είδους έξυπνη συμπεριφορά μέσω ορθολογικής σκέψης
 - Ποιός είναι ο σκοπός της σκέψης; Ποιές σκέψεις θα έπρεπε να έχω από όλες όσες θα μπορούσα να έχω;

Συστήματα που σκέπτονται σαν τον άνθρωπο	Συστήματα που σκέπτονται ορθολογιστικά
Συστήματα που ενεργούν σαν τον άνθρωπο	Συστήματα που ενεργούν ορθολογιστικά

Ορθολογική δράση: ορθολογικοί πράκτορες

- Ορθολογική δράση = να κάνει κανείς το σωστό
- **Σωστό** = αυτό που αναμένεται να μεγιστοποιήσει το συνολικό όφελος με βάση τις πληροφορίες που είναι διαθέσιμες τη στιγμή που αποφασίζει τι να κάνει.
- Δεν απαιτείται πάντα σκέψη (π.χ. αντανακλαστικές ενέργειες) αλλά η ορθολογική σκέψη θα πρέπει να υπηρετεί την ορθολογική δράση.
- Αριστοτέλης (Ηθικά Νικομάχεια):
κάθε τέχνη και κάθε αναζήτηση, και παρόμοια κάθε ενέργεια και κάθε επιδίωξη υποτίθεται ότι στοχεύει σε κάποιο καλό

Ορθολογικοί πράκτορες

- **Πράκτορας** : μια οντότητα που αντιλαμβάνεται και πράττει (ενεργεί)
- Μαθηματικά, ένας πράκτορας είναι μια συνάρτηση που αντιστοιχίζει ενέργειες σε ιστορίες αντιλήψεων: $f : P^* \rightarrow A$
- Για κάθε είδος περιβάλλοντος και κάθε είδος εργασίας που μπορεί να εκτελεστεί σε αυτό ψάχνουμε τον (ή τους) **καλύτερο** πράκτορα, δηλαδή τον πράκτορα με την καλύτερη απόδοση.
- Οι υπολογιστικοί περιορισμοί καθιστούν αδύνατη την τέλεια ορθολογικότητα.
- Σχεδιάζουμε το καλύτερο πρόγραμμα πράκτορα για μια δεδομένη υπολογιστική μηχανή.

Θεμελιώσεις της τεχνητής νοημοσύνης (προϊστορία)

Φιλοσοφία

Λογική, μέθοδοι συλλογισμού
Ο νούς σαν φυσικό σύστημα
Θεμελιώσεις μάθησης, γλώσσας, ορθολογισμού

Μαθηματικά

Τυπική αναπαράσταση και απόδειξη
Αλγόριθμοι, υπολογισμός, (un) decidability, (in) tractability
Πιθανότητες

Ψυχολογία

Προσαρμογή
Φαινόμενα αντίληψης και ελέγχου κίνησης
Πειραματικές τεχνικές

Οικονομική Επιστήμη

Τυπική θεωρία ορθολογικής λήψης απόφασης

Γλωσσολογία

Γραμματική, αναπαράσταση γνώσης

Νευροεπιστήμη

Φυσικό υπόβαθρο για διανοητική λειτουργία

Θεωρία ελέγχου

Ομοιοστατικά συστήματα, σταθερότητα
Απλά βέλτιστα σχέδια για πράκτορες

(Μερική από την) ιστορία της τεχνητής νοημοσύνης

1943	McCulloch & Pitts: Μοντέλο του εγκεφάλου σαν Boolean κύκλωμα
1950	Turing “Computing Machinery and Intelligence”
1950s	Τα πρώτα προγράμματα : Samuel (checkers), Newell & Simon (Logic Theorist), Gelernter (Geometry Engine)
1956	Dartmouth meeting. Minsky “ Artificial Intelligence ”
1965	Robinson: πλήρης αλγόριθμος για λογικό συλλογισμό
1966-74	Εξαφανίζεται σχεδόν η έρευνα σε νευρωνικά δίκτυα
1969-79	Τα πρώτα συστήματα βασισμένα σε γνώση
1980-88	Αναπτύσσεται η βιομηχανία των έμπειρων συστημάτων
1988-93	Ο χειμώνας της τεχνητής νοημοσύνης: δύσκολοι καιροί για τη βιομηχανία των έμπειρων συστημάτων
1985-95	Επιστροφή των νευρωνικών δικτύων
1988-	Πιθανοθεωρία, τεχνικό βάθος, γενετικοί αλγόριθμοι, τεχνητή ζωή
1995-	Πράκτορες παντού...Επιστροφή στον ανθρωποκεντρισμό. Αλλά τι είναι φυσική νοημοσύνη ;

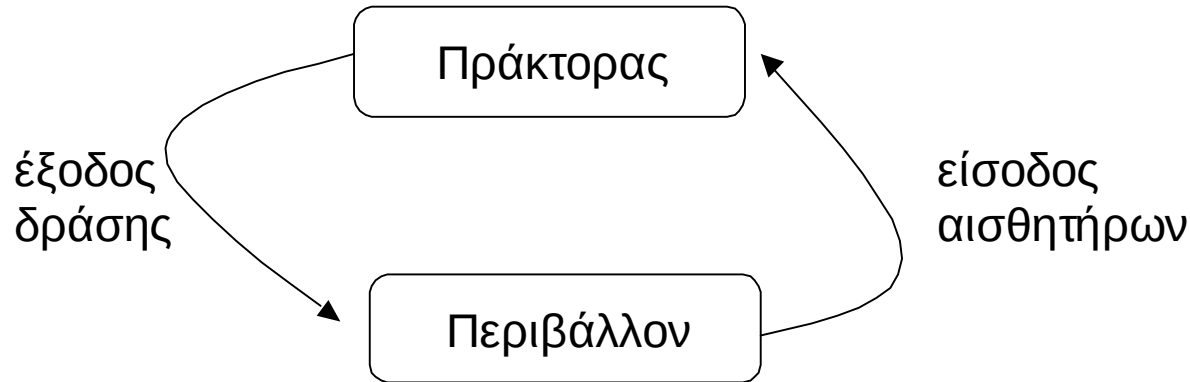
(Μερική από τη) σύγχρονη τεχνολογία

- Αυτόνομος χρονοπρογραμματισμός της δράσης διαστημικού αεροσκάφους (Remote Agent, NASA)
- Παιχνίδια (Deep Blue, IBM)
- Αυτόνομος έλεγχος αυτοκινήτου με σύστημα υπολογιστικής όρασης (ALVINN+NavLab, CMU, 2850 μίλια με έλεγχο του οχήματος το 98% του χρόνου)
- Ιατρική Διάγνωση με ανάλυση πιθανοτήτων
- Σχεδιασμός εφοδιασμού (DART, DARPA, Gulf War 1991)
- Ρομποτική
- Κατανόηση φυσικής γλώσσας (PROVERB)

Τι είναι ευφυής πράκτορας;

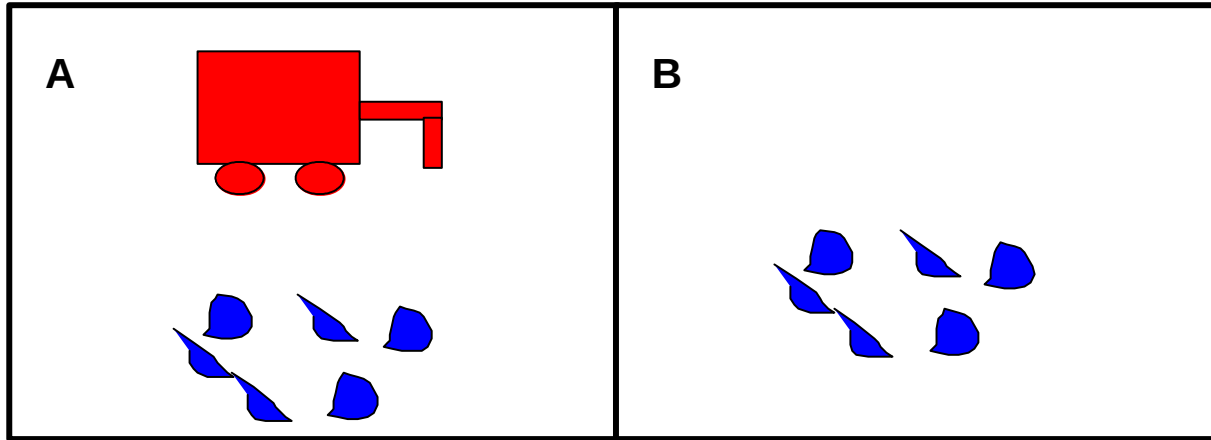
- Οποιαδήποτε οντότητα λειτουργεί μέσα σε ένα **περιβάλλον**, το αντιλαμβάνεται μέσω **αισθητήρων** και επιδρά πάνω του μέσω **μηχανισμών δράσης**.
 - Άνθρωπος
 - Ρομπότ
 - Πράκτορας λογισμικού

Πράκτορας και περιβάλλον



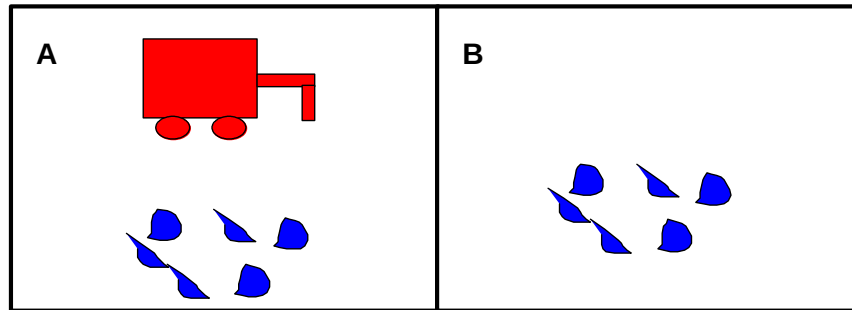
- Είσοδος αισθητήρων = **αντιλήψεις** (percepts).
 - Ακολουθία αντιλήψεων = πλήρες ιστορικό όλων όσων έχει αντιληφθεί ο πράκτορας.
- Έξοδος δράσης = **ενέργειες** που εκτελεί ο πράκτορας στο περιβάλλον.
 - Ρεπερτόριο ενεργειών = το σύνολο όλων των ενεργειών που μπορεί να εκτελέσει ο πράκτορας.
- **Συνάρτηση πράκτορα** = αντιστοιχία ακολουθιών αντιλήψεων σε ενέργειες
- Η **επιλογή** μιας ενέργειας μια χρονική στιγμή εξαρτάται από την τρέχουσα αντίληψη του πράκτορα ή και από ολόκληρη την ακολουθία αντιλήψεών του.

Παράδειγμα: πράκτορας-καθαριστής



- Ο κόσμος αποτελείται μόνο από τους χώρους A και B.
- Ο πράκτορας μπορεί
 - (να δέχεται την ακόλουθη είσοδο)
 - Να αντιλαμβάνεται σε ποιο χώρο είναι.
 - Να αντιλαμβάνεται αν υπάρχει σκόνη στο χώρο όπου βρίσκεται.
 - (να παράγει την ακόλουθη έξοδο)
 - Να μετακινείται αριστερά.
 - Να μετακινείται δεξιά.
 - Να αναρροφήσει τη σκόνη.

Συνάρτηση πράκτορα-καθαριστή



Ακολουθία αντιλήψεων	Ενέργεια
[A, Καθαρό]	Δεξιά
[A, Σκονισμένο]	Αναρρόφηση
[B, Καθαρό]	Αριστερά
[B, Σκονισμένο]	Αναρρόφηση
[A, Καθαρό], [A, Καθαρό]	Δεξιά
[A, Καθαρό], [A, Σκονισμένο]	Αναρρόφηση
....
[A, Καθαρό], [A, Καθαρό], [A, Καθαρό]	Δεξιά

Και εγένετο πράκτορας...

- Πώς θα **αναπαραστήσουμε γνώση** για το περιβάλλον ώστε να μπορεί να την επεξεργάζεται ένα πρόγραμμα πράκτορα;
- Πώς θα **αναπαραστήσουμε τις ενέργειες** που μπορεί να κάνει ένας πράκτορας, τις προϋποθέσεις και τα αποτελέσματά τους;
- Πώς θα **ορίσουμε την εργασία**, το στόχο, που πρέπει να πετύχει ένας πράκτορας;
- Πώς θα προγραμματίσουμε έναν πράκτορα ώστε να **σκέφτεται**, για να αποφασίσει (με βάση την τρέχουσα ή ολόκληρο το ιστορικό των αντιλήψεών του) ποιά ενέργεια να εκτελέσει και πότε;
- Πώς θα προγραμματίσουμε έναν πράκτορα ώστε, δοθέντος ενός στόχου που πρέπει να επιτύχει, να **καταστρώνει ένα σχέδιο δράσης**, που αν ακολουθηθεί έχει μεγάλες πιθανότητες όντως να πετύχει;

Σχέδιο μαθήματος

Week	Date	Topic
2	10/10	Εισαγωγή στην Τεχνητή Νοημοσύνη και τους Νοήμονες Πράκτορες
3	17/10	Επίλυση προβλημάτων με Αναζήτηση
4	24/10	Προτασιακή Λογική
5	31/10	Λογική Πρώτης Τάξης
6	7/11	Συμπερασμός στη Λογική Πρώτης Τάξης
7	14/11	Λογικός Προγραμματισμός
8	21/11	Αναπαράσταση Γνώσης
9	28/11	Αναπαράσταση Γνώσης
10	5/12	Σχεδιασμός
11	12/12	Θεωρία λήψης αποφάσεων
12	19/12	Θεωρία λήψης αποφάσεων

Πρακτικά θέματα

- Ιστοσελίδα μαθήματος

<http://inf-server.inf.uth.gr/courses/ce322>

- Λίστα μαθήματος

ce322@inf-server.inf.uth.gr

- Ώρες γραφείου

Τρίτη 15.30-16.30

Τετάρτη 9.00-10.00

- Επικοινωνία

E-mail: aspasia@inf.uth.gr

Γραφείο: Ιάσονος 10, 3ος όροφος (κτίριο μεταπτυχιακού)