
Τεχνητή Νοημοσύνη και ο Ρόλος του Μηχανικού

Ροβιές 3/7/2023

Τι θα συζητήσουμε

- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - Είναι η σκέψη αλγόριθμος;
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - Ποιος ο ρόλος του Μηχανικού;
 - Τι πρέπει να φοβόμαστε;
-

-
- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - Είναι η σκέψη αλγόριθμος;
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - Ποιος ο ρόλος του Μηχανικού;
 - Τι πρέπει να φοβόμαστε;
-

Τι κάνει το ChatGPT;

- Στοχαστικό μοντέλο γλώσσας
- Υπολογίζει την πιθανότητα με την οποία εμφανίζονται λέξεις σε σειρά

The best thing about AI is its ability to

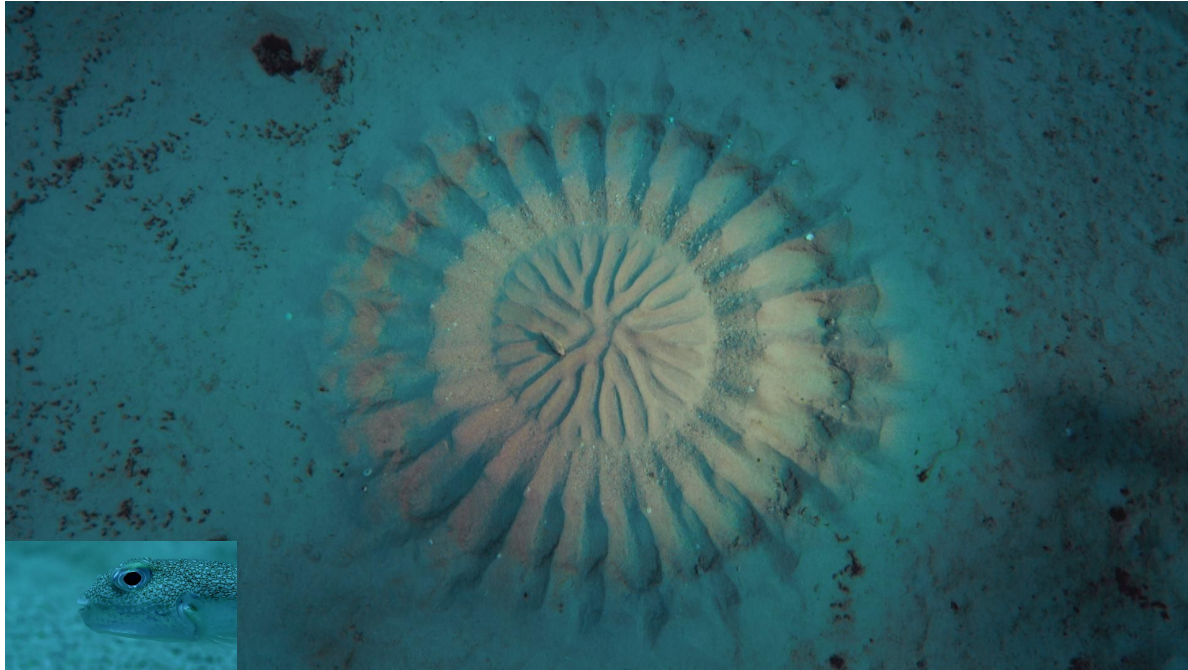
learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%



```
In[ ]:= NestList[StringJoin[#, model[#, "Decision"]]&,  
  "The best thing about AI is its ability to", 7]
```

```
Out[ ]:= {The best thing about AI is its ability to,  
  The best thing about AI is its ability to learn,  
  The best thing about AI is its ability to learn from,  
  The best thing about AI is its ability to learn from experience,  
  The best thing about AI is its ability to learn from experience.,  
  The best thing about AI is its ability to learn from experience. It,  
  The best thing about AI is its ability to learn from experience. It's,  
  The best thing about AI is its ability to learn from experience. It's not}
```

—
Τι δεν κάνει το ChatGPT;



—
Τι δεν κάνει το ChatGPT;



Τι δεν κάνει το ChatGPT;

Fetch as Google

See how Google renders

Choose

You are s
<https://w>

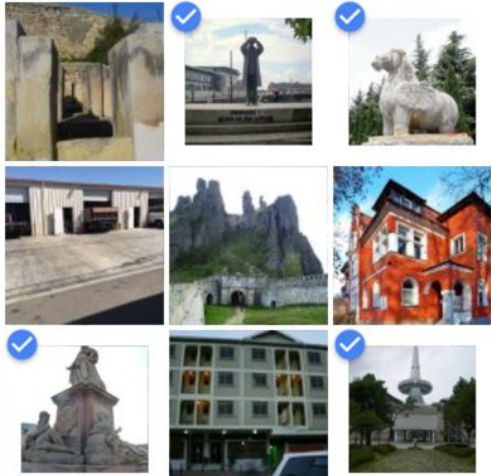
Recrawl
Google v

Note: Th
use of n

Craw
 Craw

Go

Select all images with **statues**.



of your page is what

defines and avoids the

Ret

Inde

Inde and

Inde

Complete

VERIFY

🔄 🎧 ⓘ

Το Κινέζικο Δωμάτιο

Ο άνθρωπος που βρίσκεται στο δωμάτιο δεν γνωρίζει κινεζικά, παρά μόνο την μητρική του γλώσσα, αλλά έχει στην κατοχή του ένα βιβλίο στο οποίο αναγράφονται **κανόνες αντιστοιχίας συμβόλων**.

Ο άνθρωπος λαμβάνει από την **μία σχισμή** ένα χαρτί στο οποίο αναγράφονται διάφορα σύμβολα, άγνωστα στον ίδιο. Δουλειά του είναι να ερμηνεύσει τα σύμβολα χρησιμοποιώντας τους κανόνες που αναγράφονται στο βιβλίο, να τα γράψει σε ένα άλλο χαρτί και στη συνέχεια να το μεταφέρει μέσω της δεύτερης σχισμής στο εξωτερικό περιβάλλον. Αυτά τα σύμβολα είναι ερωτήσεις στα κινέζικα τις οποίες δεν καταλαβαίνει ο άνθρωπος, αλλά **μέσω των κανόνων που χρησιμοποιεί, είναι σε θέση να διαμορφώσει μία λογική απάντηση σε αυτές, γραμμένη εξίσου στα κινέζικα**.

Ο άνθρωπος ακολουθεί μία **τυποποιημένη διαδικασία**, η οποία τον βοηθάει να απαντήσει στις ερωτήσεις που του δίνονται, χωρίς να είναι προαπαιτούμενο να **γνωρίζει** περί τίνος πρόκειται. Μέσα από όλη την διαδικασία, ο έγκλειστος ούτε θα μάθει ποτέ κινέζικα ούτε τι σημαίνουν τα σύμβολα τα οποία χρησιμοποιεί. Αν όμως υπάρχουν εξωτερικοί παρατηρητές που δεν γνωρίζουν τις συνθήκες που επικρατούν στο δωμάτιο, θα θεωρήσουν πως ο άνθρωπος αυτός γνωρίζει κινέζικα και, μάλιστα, σε ένα αρκετά καλό επίπεδο.

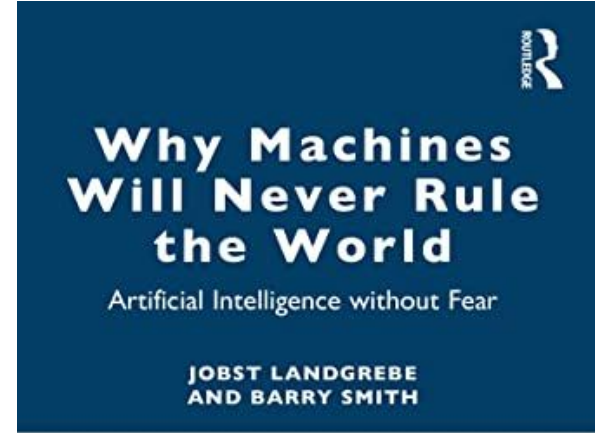
The Chinese Room Argument



-
- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - Είναι η σκέψη αλγόριθμος;
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - Ποιος ο ρόλος του Μηχανικού;
 - Τι πρέπει να φοβόμαστε;
-

Γενική Τεχνητή Νοημοσύνη

Ορίζουμε την Γενική Τεχνητή Νοημοσύνη (AGI) ως μια τεχνητή νοημοσύνη που έχει επίπεδο νοημοσύνης είτε ισοδύναμο είτε μεγαλύτερο από αυτό των ανθρώπινων όντων ή **είναι σε θέση να αντιμετωπίσει προβλήματα που προκύπτουν στον κόσμο που περιβάλλει τα ανθρώπινα όντα** με βαθμό επάρκειας τουλάχιστον παρόμοιο με αυτόν των ανθρώπινων όντων.

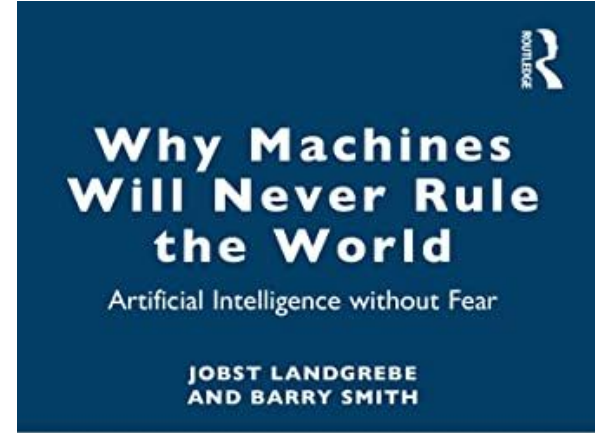


Γενική Τεχνητή Νοημοσύνη

Το επιχείρημά μας μπορεί να παρουσιαστεί εδώ σε μια κάπως απλουστευμένη μορφή ως εξής:

A1. Για να κατασκευάσουμε μια AGI θα χρειαζόμασταν τεχνολογία με νοημοσύνη που να είναι τουλάχιστον συγκρίσιμη με εκείνη των ανθρώπων (από τον ορισμό της AGI που μόλις δόθηκε).

A2. Ο μόνος τρόπος για να κατασκευάσουμε μια τέτοια τεχνολογία είναι να δημιουργήσουμε ένα λογισμικό προσομοίωσης του ανθρώπινου νευρογνωστικού συστήματος.



Γενική Τεχνητή Νοημοσύνη

Ωστόσο,

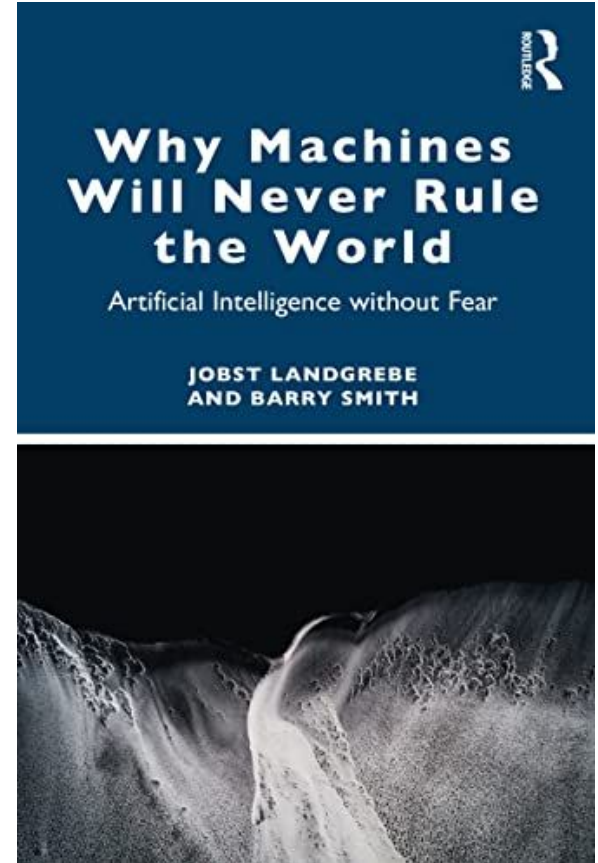
B1. Για να δημιουργήσουμε μια λογισμική εξομοίωση της συμπεριφοράς ενός συστήματος θα πρέπει να δημιουργήσουμε ένα μαθηματικό μοντέλο αυτού του συστήματος που να επιτρέπει την πρόβλεψη της συμπεριφοράς του συστήματος.¹

B2. Είναι αδύνατο να δημιουργηθούν μαθηματικά μοντέλα αυτού του είδους για πολύπλοκα συστήματα.

B4. Επομένως, δεν μπορούμε να δημιουργήσουμε μια προσομοίωση λογισμικού του ανθρώπινου νευρογνωστικού συστήματος.

Από τα (A2.) και (B4.) προκύπτει τώρα ότι:

C. Μια AGI είναι αδύνατη.

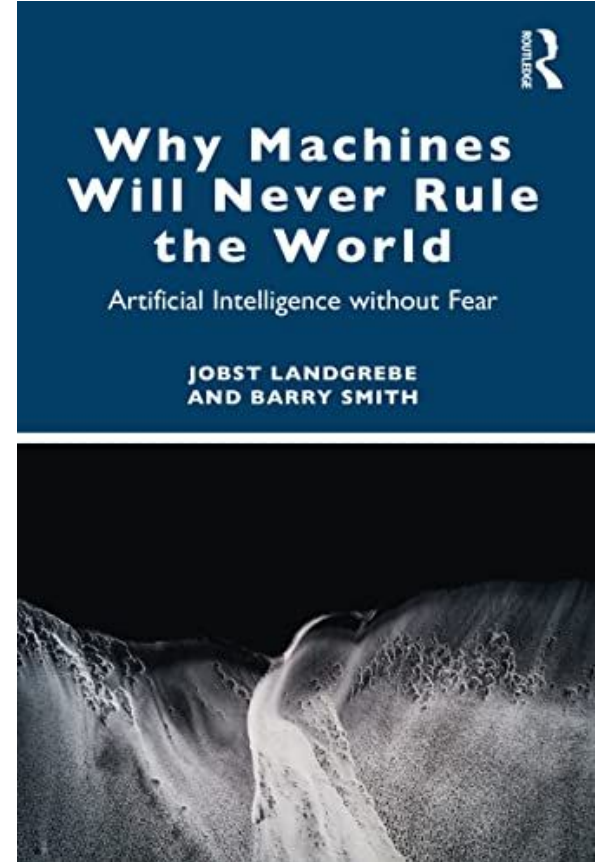


Γενική Τεχνητή Νοημοσύνη

Για τη γενική τεχνητή νοημοσύνη, ο στόχος είναι να δημιουργηθεί ένα υπολογίσιμο μοντέλο της συμπεριφοράς σημαντικών πτυχών του ανθρωπίνου **συνεχούς νου-σώματος**, επιτρέποντας έτσι τη μίμηση της ευφυούς ανθρώπινης συμπεριφοράς.

Αλλά **το συνεχές νου-σώμα είναι ένα πολύπλοκο σύστημα** (είναι στην πραγματικότητα ένα πολύπλοκο σύστημα πολύπλοκων συστημάτων, σε πολλά επίπεδα).

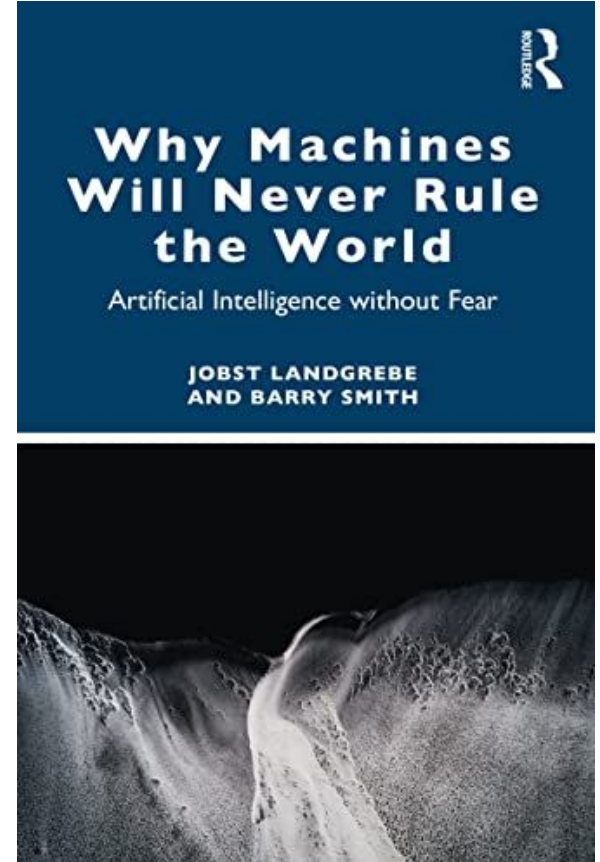
Επομένως, αν **η ικανότητά μας να δημιουργούμε μαθηματικά μοντέλα πολύπλοκων συστημάτων είναι πολύ περιορισμένη**, το ίδιο ισχύει και για την ικανότητά μας να δημιουργήσουμε τα υπολογιστικά μοντέλα που απαιτούνται για τη δημιουργία γενικής ΤΝ.



Εξέλιξη

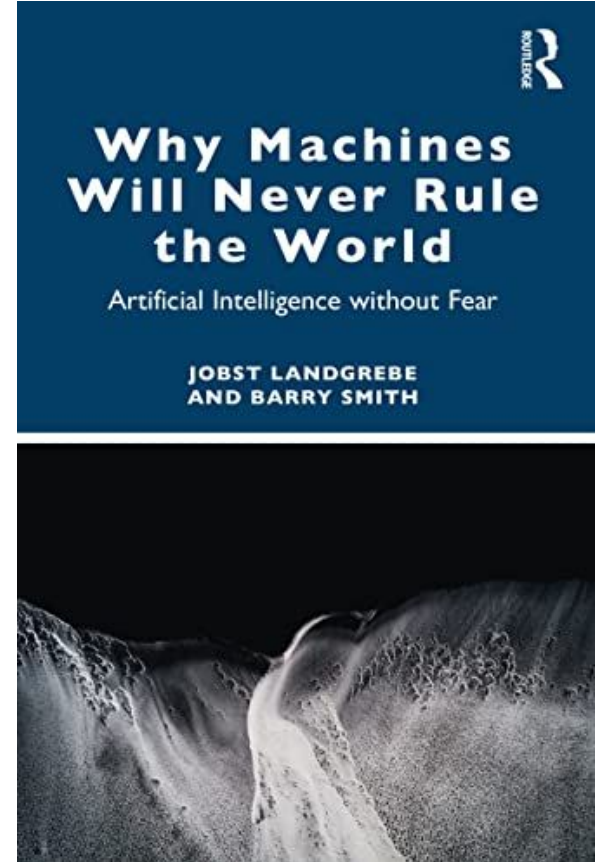
Στα πολύπλοκα συστήματα, οι τύποι των σχέσεων μεταξύ των στοιχείων μπορούν να αλλάξουν από τη μια στιγμή στην άλλη ως συνέπεια της λειτουργίας του συστήματος. Ένα χαρακτηριστικό των πολύπλοκων συστημάτων είναι επομένως ότι πρόκειται για συστήματα των οποίων τα στοιχεία μπορούν, με την πάροδο του χρόνου, να αλλάξουν τη συμπεριφορά τους με την ισχυρή έννοια ότι αλλάζουν τους τύπους της συμπεριφοράς τους (Thurner et al. 2018, p. 7f.).

Αυτό οφείλεται στον **εξελικτικό χαρακτήρα** των πολύπλοκων συστημάτων.



Μη Εργοδικότητα

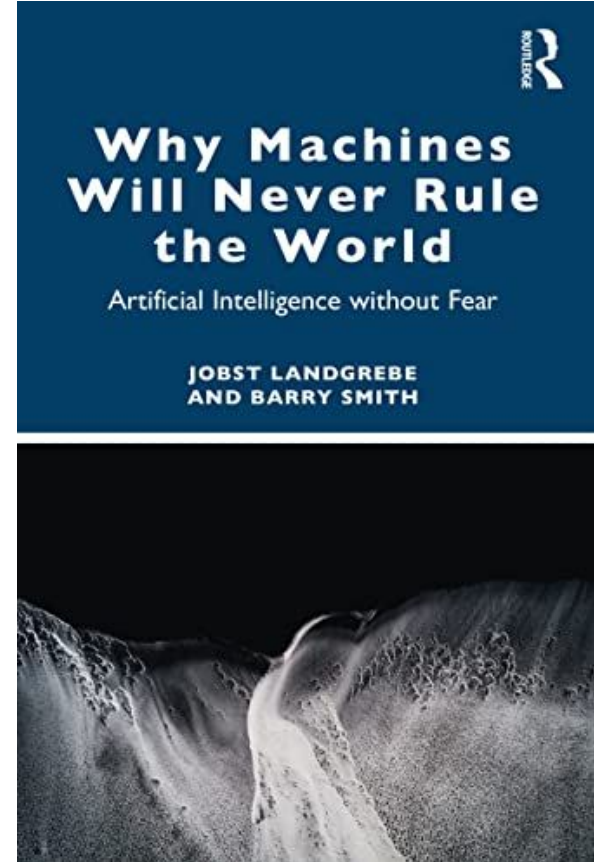
Τα εξελικτικά συστήματα είναι επίσης τέτοια ώστε να εκδηλώνουν εξάρτηση από την πορεία της εξέλιξής τους και συνεπώς να εμφανίζουν ισχυρή και μακροχρόνια μνήμη (με την έννοια ότι η σχέση του παρόντος με το παρελθόν τους δεν μπορεί να αποτυπωθεί με τη χρήση μοντέλων Markov). Τέτοια συστήματα είναι επομένως τόσο **μη-εργοδικά** (δεν μπορούν να μοντελοποιηθούν με μέσο όρο στο χώρο και στο χρόνο χωρίς να χάσουν πληροφορίες) όσο και μη-μαρκοβιανά (η συμπεριφορά τους δεν εξαρτάται μόνο από ένα ή δύο αμέσως προηγούμενα βήματα). Η έλλειψη εργοδικότητας είναι ένα από τα κυριότερα εμπόδια στη χρήση της στοχαστικής ΤΝ για πολύπλοκα συστήματα.



Οδηγούμενα Συστήματα

Τα πολύπλοκα συστήματα συχνά οδηγούνται με την τεχνική έννοια που ορίζεται στη φυσική (ακριβέστερα στη στατιστική μηχανική). Τα καθοδηγούμενα συστήματα υφίστανται μια ροή ενέργειας, η οποία τα εμποδίζει να συγκλίνουν ή να κινηθούν προς μια ισορροπία- η ροή ενέργειας τα ωθεί συνεχώς από τη μια κατάσταση στην άλλη.

Τα συστήματα που βρίσκονται εκτός ισορροπίας εμφανίζουν "ροές ενέργειας που τα διαπερνούν και κάνουν τα συστήματα αυτά να κάνουν πράγματα που δεν θα έκαναν ποτέ σε ισορροπία, όπως να παράγουν ρεύματα συναγωγής ή αναταράξεις". (Thurner et al. 2018, σ. 86). Και περαιτέρω: "Τα πολύπλοκα συστήματα είναι συχνά τα λεγόμενα οδηγούμενα συστήματα, όπου το σύστημα απομακρύνεται από τις καταστάσεις ισορροπίας του. ... Οι μαθηματικές δυσκολίες στην αντιμετώπιση συστημάτων εκτός ισορροπίας ή μη ισορροπίας είναι τεράστιες και πέρα από την αναλυτική εμβέλεια"



Ασαφή Όρια

Αναφέραμε προηγουμένως ότι τα πολύπλοκα συστήματα είναι συχνά εξελικτικά και συνήθως λειτουργούν σε ένα ευρύτερο περιβάλλον. Στα πολύπλοκα συστήματα, οι οριακές συνθήκες στη διασύνδεση μεταξύ συστήματος και περιβάλλοντος μεταβάλλονται συνεχώς. Αυτός είναι ο λόγος για τον οποίο ένα πολύπλοκο σύστημα δεν μπορεί να μοντελοποιηθεί υποθέτοντας ότι οι οριακές συνθήκες του (που διαμορφώνονται από τα στοιχεία στο σύνορο) είναι σταθερές: κάτι τέτοιο θα δημιουργούσε ένα άκυρο μοντέλο. Με άλλα λόγια, δεν μπορεί κανείς να αφαιρέσει από αυτό το περιβάλλον χωρίς να μοντελοποιήσει θεμελιωδώς λανθασμένα τη συμπεριφορά των συστημάτων που περιέχει.

Όταν έχουμε να κάνουμε με λογικά συστήματα, αντίθετα, μπορεί κανείς να αφαιρέσει από το περιβάλλον- οι οριακές συνθήκες του συστήματος μπορούν να θεωρηθούν σταθερές, και το ίδιο το σύστημα είναι υπό αυτή την έννοια ελεύθερο πλαισίου.

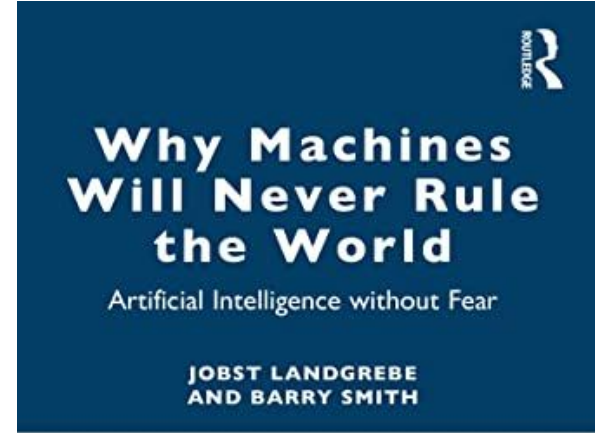
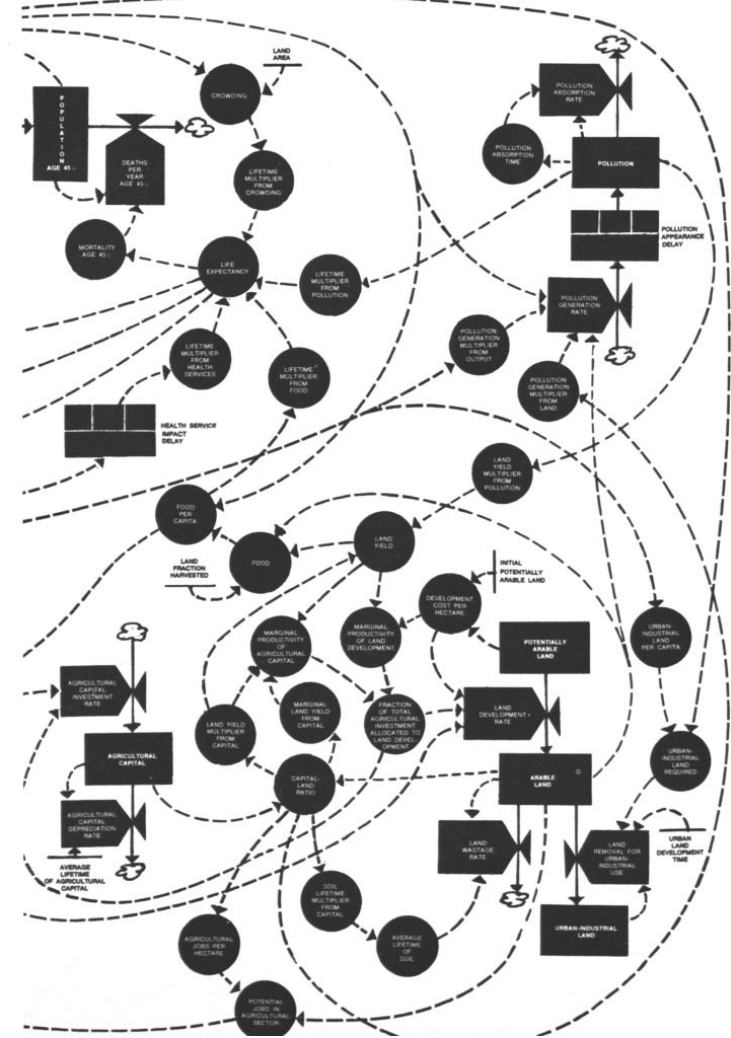
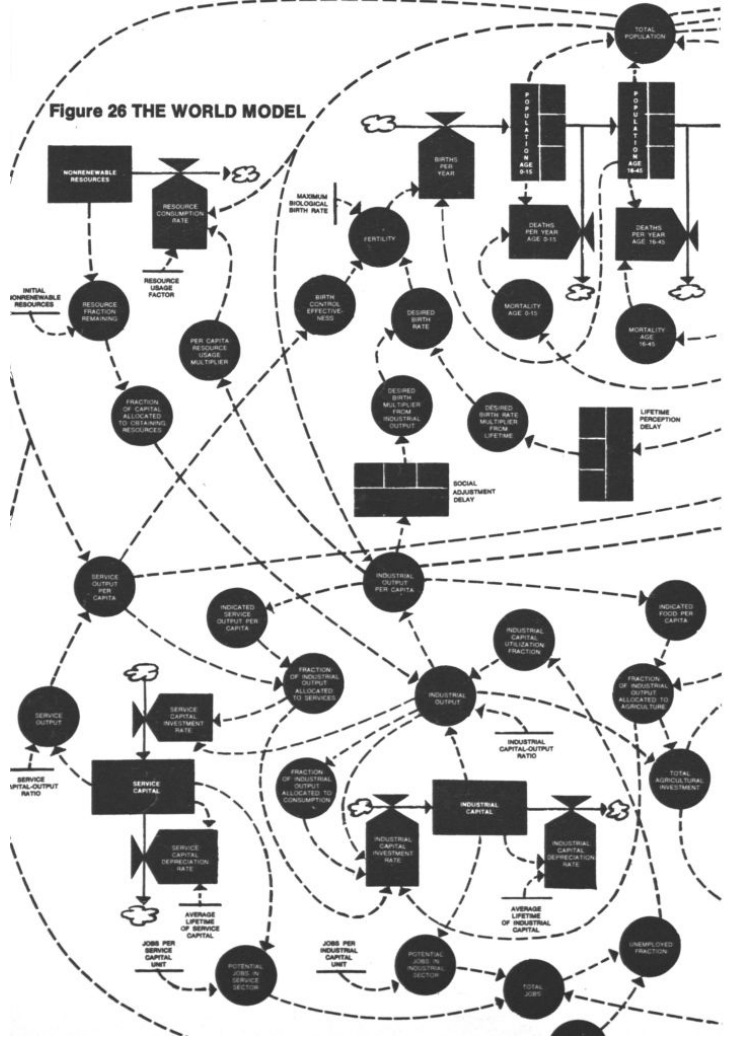
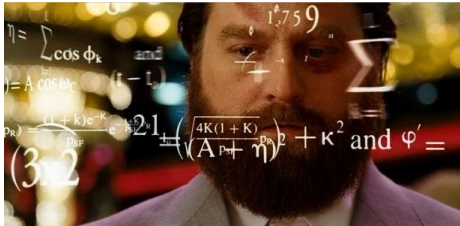


Figure 26 THE WORLD MODEL



1972.
The Limits to Growth



- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - **Είναι η σκέψη αλγόριθμος;**
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - Ποιος ο ρόλος του Μηχανικού;
 - Τι πρέπει να φοβόμαστε;
-

—
 $\alpha * \beta = \beta * \alpha$

...απόδειξη;

Αλγόριθμος

Πολλαπλασίασε : $1 * 2$

Πολλαπλασίασε : $2 * 1$

Σύγκρινε τα αποτελέσματα.

Επανάλαβε για όλους τους αριθμούς.

.....

	1	2	3	4	5	α
1	1	2	3	4	5	
2	2	4	6	8	10	
3	3	6	9	12	15	
4	4	8	12	16	20	
5	5	10	15	20	25	
6	6	12	18	24	30	
7	7	14	21	28	35	
8	8	16	48	32	40	
β						$\alpha * \beta$

$$\alpha * \beta = \beta * \alpha$$

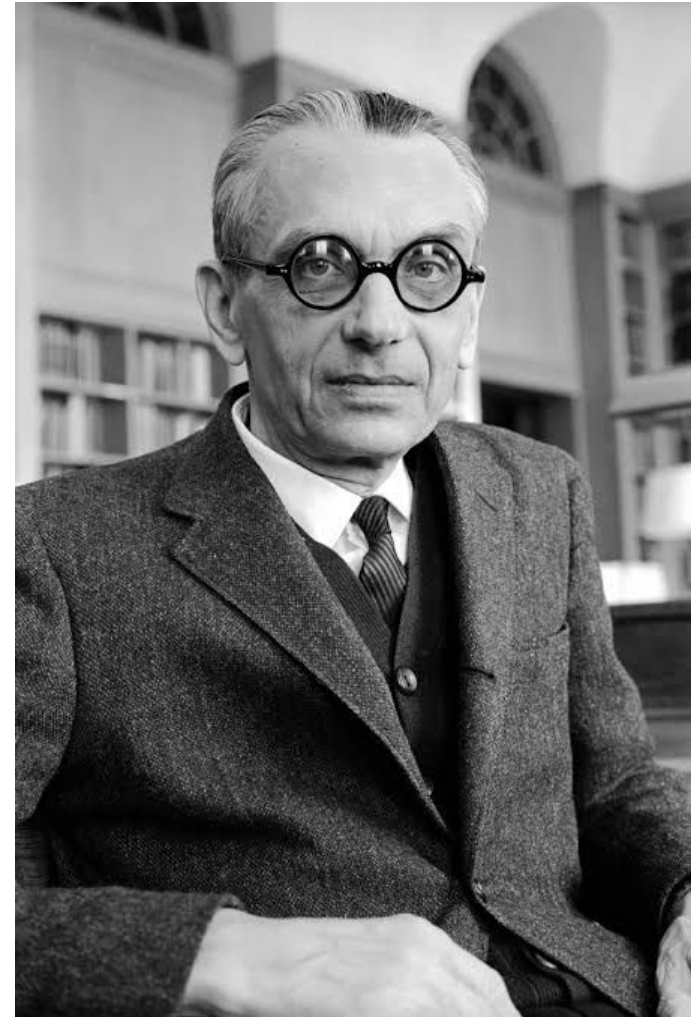
	1	2	3	4	5	α
1	1	2	3	4	5	
2	2	4	6	8	10	
3	3	6	9	12	15	
4	4	8	12	16	20	
5	5	10	15	20	25	
6	6	12	18	24	30	
7	7	14	21	28	35	
8	8	16	24	32	40	
β						$\alpha * \beta$

	1	2	3	4	5	α
1	1	2	3	4	5	
2	2	4	6	8	10	
3	3	6	9	12	15	
4	4	8	12	16	20	
5	5	10	15	20	25	
6	6	12	18	24	30	
7	7	14	21	28	35	
8	8	16	24	32	40	
β						$\alpha * \beta$

1931, Kurt Gödel

Θεωρήματα μη-πληρότητας

1. Οποιαδήποτε αποτελεσματικά παραχθείσα θεωρία που είναι ικανή να εκφράσει τη στοιχειώδη αριθμητική δεν μπορεί να είναι και **συνεπής και πλήρης**. Συγκεκριμένα, για κάθε συνεπή, αποτελεσματικά παραχθείσα τυπική θεωρία που αποδεικνύει συγκεκριμένες αλήθειες βασικής αριθμητικής, υπάρχει μία αριθμητική δήλωση η οποία **είναι αληθής, αλλά δεν μπορεί να αποδειχθεί από τη θεωρία**.
2. Για κάθε αποτελεσματικά παραχθείσα τυπική θεωρία Θ που συμπεριλαμβάνει βασικές αριθμητικές αλήθειες και επίσης συγκεκριμένες αλήθειες για την δυνατότητα τυπικής απόδειξης, η Θ συμπεριλαμβάνει δήλωση περί της ίδιας συνέπειας αν και μόνο αν η Θ είναι ασυνεπής.



Υπόθεση

Φαίνεται ότι η νοημοσύνη προϋποθέτει μία έννοια «κατανόησης», μία αντιστοίχιση γλώσσας / συμβόλων / αντικειμένων με νοητικές αναπαραστάσεις.

Οι νοητικές αναπαραστάσεις αυτές φαίνεται ότι σχετίζονται με αυτό που αποκαλούμε Συνειδητότητα.

Ο μόνος λόγος που φοβόμαστε μία «Τεχνητή Νοημοσύνη» είναι διότι της αποδίδουμε Βούληση που υπονοεί Συνειδητότητα.





- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - Είναι η σκέψη αλγόριθμος;
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - Ποιος ο ρόλος του Μηχανικού;
 - Τι πρέπει να φοβόμαστε;
-

Το Δωμάτιο της Μαίρης



Το Δωμάτιο της Μαίρης

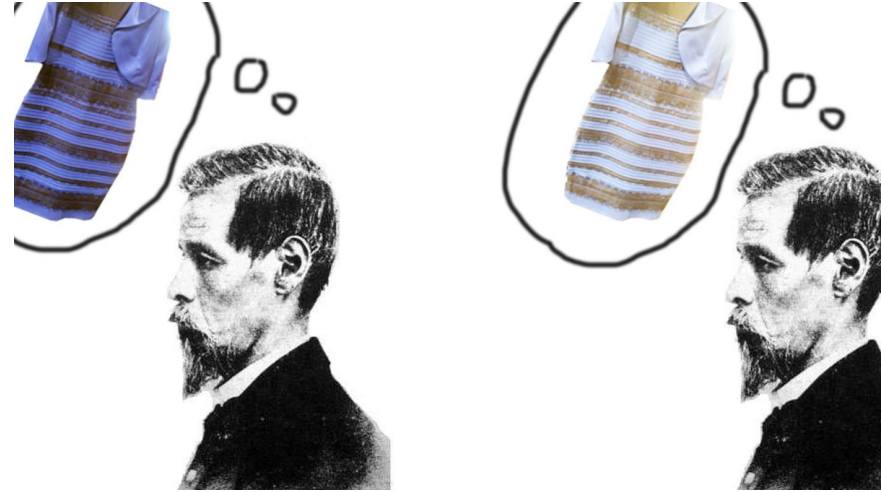
«Η Μαίρη είναι μία λαμπρή επιστήμονας η οποία, για κάποιο λόγο, είναι αναγκασμένη να ερευνά τον κόσμο μέσα από ένα ασπρόμαυρο δωμάτιο μέσω μιας ασπρόμαυρης οθόνης τηλεόρασης. Ειδικεύεται στη νευροφυσιολογία της όρασης και ας υποθέσουμε ότι αποκτά όλες τις φυσικές πληροφορίες που μπορούν να αποκτηθούν περί του τι συμβαίνει όταν βλέπουμε ώριμες ντομάτες, ή τον ουρανό, και χρησιμοποιεί όρους όπως «κόκκινο», «μπλε» κτλ. Ανακαλύπτει, για παράδειγμα, ακριβώς ποιοι συνδυασμοί μήκους κύματος από τον ουρανό ερεθίζουν την ίριδα, και ακριβώς πώς αυτό παράγει μέσω του κεντρικού νευρικού συστήματος τη σύσπαση των φωνητικών χορδών και την αποβολή αέρα από τα πνευμόνια που έχει ως αποτέλεσμα την εκφορά της πρότασης «ο ουρανός είναι μπλε». (...) Τι θα συμβεί όταν η Μαίρη απελευθερωθεί από το ασπρόμαυρό της δωμάτιο ή της δοθεί μία έγχρωμη τηλεόραση; Θα αποκτήσει κάποια γνώση ή όχι; Φαίνεται προφανές ότι θα μάθει κάτι για τον κόσμο και την οπτική εμπειρία που έχουμε γι' αυτόν. Όμως τότε είναι αναπόδραστο ότι η προτέρα της γνώση ήταν ελλιπής. Μα είχε όλες τις φυσικές [υλικές] πληροφορίες. Συνεπώς, υπάρχει κάτι περισσότερο να έχει κανείς πέρα από αυτές, και ο φυσικαλισμός είναι εσφαλμένος.»



Qualia και Συνειδητότητα

Το ίδιο μπορούμε να πούμε, αναλόγως, για όλες τις υποκειμενικές εμπειρίες. Μπορούμε, για παράδειγμα, να επεκτείνουμε το επιχείρημα σε όλο το εύρος της οπτικής εμπειρίας: όπως το χρώμα δεν είναι απλώς η συχνότητα του φωτός, έτσι και το σχήμα δεν είναι παρατεταγμένα μόρια μελάνης σε χαρτί και οι σχηματισμοί των φωτονίων που προκύπτουν από αυτά. Φυσικά, και η μουσική δεν είναι απλώς ο παλλόμενος αέρας, η οσμή δεν είναι απλώς οι χημικές ενώσεις που μεταφέρονται μέσω του αέρα, πόνος δεν είναι απλώς ηλεκτρική δραστηριότητα στον εγκέφαλο.

Ακόμα κι αν τα υλικά αυτά γεγονότα είναι άρρηκτα **συνδεδεμένα** με τις αντίστοιχες εμπειρίες, **δεν ταυτίζονται** οντολογικά με αυτές, δεν είναι το ίδιο είδος πράγματος. Οι εμπειρίες είναι κάτι πέρα από αυτές, μη περιγράψιμο από τη σκοπιά του τρίτου προσώπου.



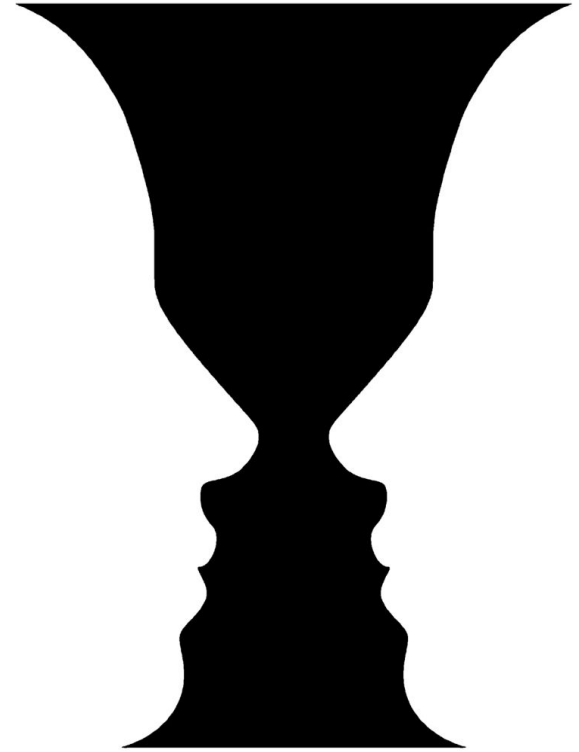
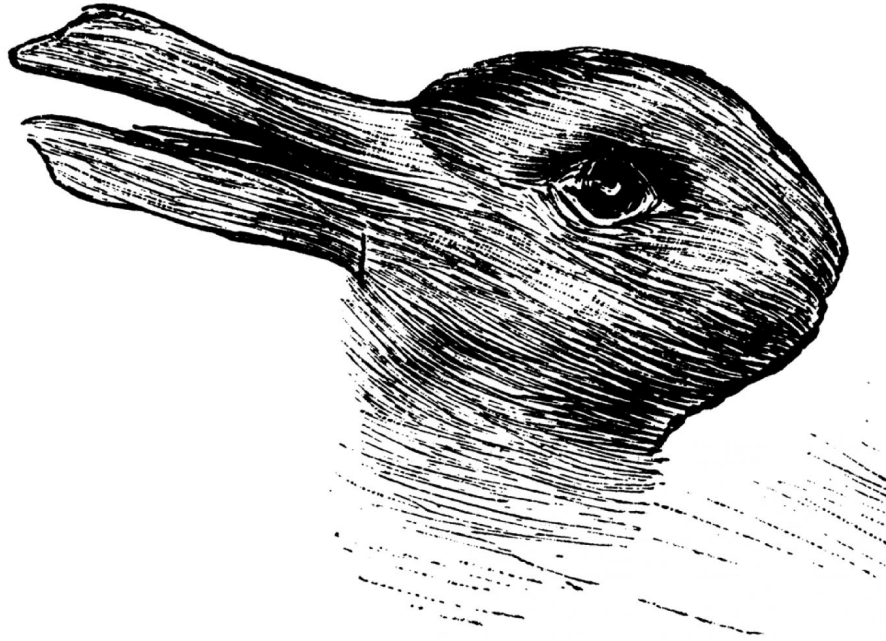
Αποβλεπτικότητα (Intentionality)

Αν, όμως, τα qualia συνδέονται απλώς και δεν **παράγονται** από τα υλικά αντικείμενα και τα φυσικά γεγονότα, τότε αναγκαστικά δεν **έπονται** αυτών αλλά με έναν τρόπο **προηγούνται** αιτιολογικά. Αυτό σημαίνει ότι, με κάποιον τρόπο, **το σημασιολογικό περιεχόμενο είναι ήδη μέρος του λεξιλογίου της συνειδητότητας**, και απλώς **προβάλλεται** πάνω στα φυσικά αντικείμενα με τα οποία συνδέεται. Το περιεχόμενο των αισθήσεων, λοιπόν, θα πρέπει να **τίθεται προς** το αντικείμενο των αισθήσεων.

Αυτό το χαρακτηριστικό του νου να **τείνει** το γνωσιακό περιεχόμενο **προς** το αντικείμενο της εμπειρίας ονομάζεται **αποβλεπτικότητα**, και την εισήγαγε στη σύγχρονη φιλοσοφία ο Φραντς Μπρεντάνο, ο οποίος την όρισε ως το κύριο χαρακτηριστικό των νοητικών φαινομένων, αυτό που ξεχωρίζει τα φυσικά από τα νοητικά φαινόμενα.



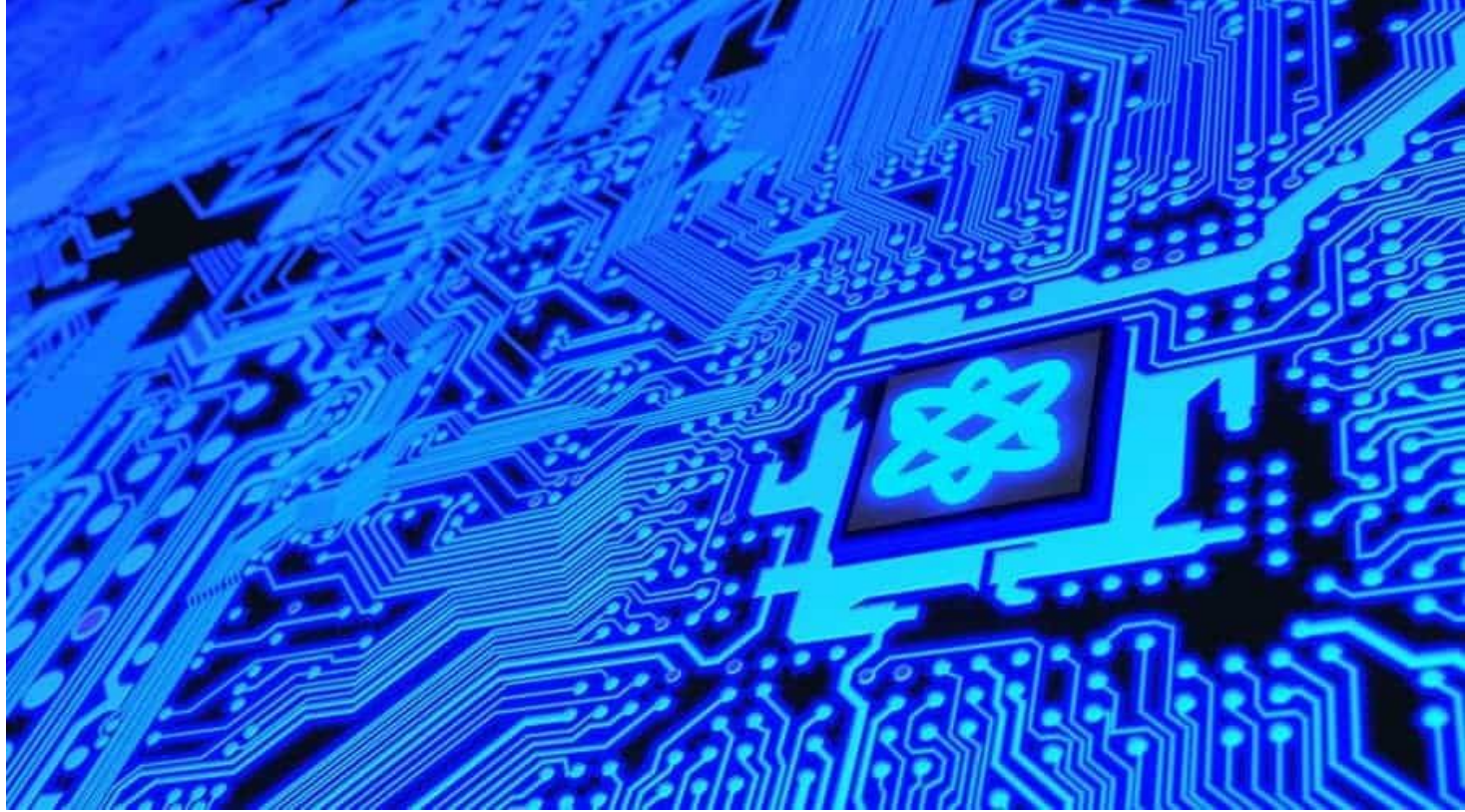
—
Αποβλεπτικότητα (Intentionality)



Υπολογισμός

- Σχεδιασμός υλικών διατάξεων που αντιπροσωπεύουν μαθηματικές έννοιες
- Μεταβολές των υλικών διατάξεων
- Παρατήρηση των υλικών διατάξεων και ερμηνεία με βάση την αρχική σημασιοδότηση





—



—

Σκεπτόμενες Μηχανές;

«Δεν είμαστε πλέον απόλυτα εξοικειωμένοι με τις σκεπτόμενες μηχανές; Γιατί να μην θεωρήσουμε τον εγκέφαλο ίσως ως υπολογιστή και αυτό που αποκαλούμε «νου» να μην είναι τίποτα περισσότερο από το ξεδίπλωμα των μηχανικών αλγορίθμων που είναι ενσωματωμένοι στο «λογισμικό» που επεξεργάζεται ο εγκέφαλος; Αυτή είναι η άποψη που υποστηρίζεται σε μεγάλο βαθμό από τις παρατάξεις της Ισχυρής Τεχνητής Νοημοσύνης μεταξύ των γνωσιακών επιστημόνων και των φιλοσόφων του νου: αντί για συνείδηση, η οποία νοείται με την κοινή έννοια ως η σκόπιμη πράξη ενός αδιαίρετου υποκειμένου, ο νους περιέχει μόνο ροές δεδομένων και συμπεριφορών, εισόδους και εξόδους, λειτουργικές σχέσεις, διακριτικά φίλτρα, καταστάσεις πληροφοριών, όλα στημένα πάνω σε μια απλή ψηφιακή πλατφόρμα, δυαδική ή κάποιου άλλου είδους, που λειτουργεί μέσω μερικών απλών φυσικών διακοπών που μπορούν να ενεργοποιηθούν και να απενεργοποιηθούν. Για ορισμένους θεωρητικούς αυτό σημαίνει ότι ένας τεχνητός υπολογιστής συγκρίσιμος σε πολυπλοκότητα με τον ανθρώπινο εγκέφαλο και το νευρικό σύστημα θα μπορούσε να επιτύχει κάτι σαν τις συνειδητές μας καταστάσεις.»

— Σκεπτόμενες Μηχανές;

«Δυστυχώς, ολόκληρη η θεωρία είναι μια αδιόρθωτη σύγχυση κατηγοριών, για έναν πολύ μεγάλο αριθμό λόγων, με κυριότερο την απόλυτη εξάρτηση όλων των υπολογιστικών διαδικασιών από την προηγούμενη πραγματικότητα της σκόπιμης συνείδησης. Δεν το εννοώ αυτό απλώς με την έννοια ότι οι υπολογιστές και τα προγράμματά τους τυχαίνει να σχεδιάζονται από ανθρώπινα μυαλά, το οποίο είναι ένα σημαντικό αλλά δευτερεύον ζήτημα. (...) Εννοώ μάλλον ότι -όπως ορθά έχει υποστηρίξει ο John Searle- εκτός από συγκεκριμένες αναπαραστάσεις που παράγονται από σκόπιμη συνείδηση, οι λειτουργίες ενός υπολογιστή είναι απλώς **φυσικά γεγονότα χωρίς νόημα.**»

—

Σκεπτόμενες Μηχανές;

«Αυτό θα ήταν ίσως πιο άμεσα προφανές σε εμάς, αν δεν ήμασταν τα θύματα των ίδιων μας των μεταφορών. Έχουμε συνηθίσει τόσο πολύ να μιλάμε για τους υπολογιστές ως τεχνητά μυαλά και για τις λειτουργίες τους ως σκέψη, ώστε έχουμε ξεχάσει ότι πρόκειται απλώς για σχήματα λόγου. Για παράδειγμα, μιλάμε για τη μνήμη των υπολογιστών, αλλά φυσικά οι υπολογιστές δεν θυμούνται τίποτα. Δεν αποθηκεύουν καν καμία «πληροφορία» -με την έννοια των συμβόλων με πραγματικό σημασιολογικό περιεχόμενο, πραγματικό νόημα- αλλά διατηρούν μόνο τα **δυναμικά μοτίβα ορισμένων ηλεκτρονικών σημειώσεων**.

Και δεν εννοώ απλώς ότι οι υπολογιστές δεν έχουν επίγνωση της πληροφορίας που περιέχουν- εννοώ ότι, από μόνοι τους, δεν περιέχουν καθόλου σημασιολογική πληροφορία. Είναι απλώς η περγαμηνή πυριτίου και το ηλεκτρικό μελάνι πάνω στα οποία καταγράφουμε σύμβολα τα οποία διαθέτουν σημασιολογικό περιεχόμενο μόνο σε σχέση με τις σκόπιμες αναπαραστάσεις των νοημάτων τους. Ένας υπολογιστής δεν θυμάται τα αρχεία που είναι αποθηκευμένα σε αυτόν, όπως το χαρτί και το έντυπο αυτού του βιβλίου δεν θυμούνται τα επιχειρήματά μου μέχρι αυτό το σημείο. (...)»

—

Σκεπτόμενες Μηχανές;

Δεν υπάρχουν ανώτερες λειτουργίες και προγράμματα καθαυτά, ούτε στον υπολογιστή που θεωρείται καθαρά ως φυσικό αντικείμενο ούτε στις λειτουργίες του που θεωρούνται καθαρά ως φυσικά γεγονότα- υπάρχουν μόνο τα υλικά συστατικά της μηχανής, οι ηλεκτρικοί παλμοί και τα δυαδικά μοτίβα, τα οποία χρησιμοποιούμε για να κατασκευάσουμε ορισμένες αναπαραστάσεις και τα οποία έχουν νοήματα μόνο για όσο διάστημα αποτελούν αντικείμενα της προσοχής του αναπαραστατικού νου.

(...) Έχουμε επιβάλει τη μεταφορά ενός τεχνητού νου στους υπολογιστές και στη συνέχεια επαναφέραμε την εικόνα μιας σκεπτόμενης μηχανής και την επιβάλλουμε στο μυαλό μας.

Αν απευθύνετε ένα γράμμα στον εαυτό σας και το ταχυδρομήσετε, είναι σχεδόν βέβαιο ότι θα σας επιστρέψει- αλλά θα ήσασταν ανόητοι αν ανυπόμονοι σκίσετε τον φάκελο με την ελπίδα να βρείτε κάτι απροσδόκητο μέσα.

Σκεπτόμενες Μηχανές;

Ο υπολογιστής θα μπορούσε εξίσου καλά να προγραμματιστεί έτσι ώστε να απαντήσει στο αίτημα για την τετραγωνική ρίζα του π με το αποτέλεσμα "Rupert Bear"- ούτε θα ήταν λάθος να το κάνει, διότι ένα σύνολο από απλώς υλικά συστατικά και καθαρά φυσικά γεγονότα δεν μπορεί να είναι ούτε λάθος ούτε σωστό για τίποτα - στην πραγματικότητα, δεν μπορεί να είναι «για» τίποτα.

Το λογισμικό δεν "σκέφτεται" περισσότερο απ' ό,τι ένας λεπτοδείκτης γνωρίζει την ώρα ή η τυπωμένη λέξη "πελεκάνος" γνωρίζει τι είναι ο πελεκάνος. Θα μπορούσαμε εξίσου καλά να παρομοιάσουμε το μυαλό με έναν άβακα, μια γραφομηχανή ή μια βιβλιοθήκη. Κανένας υπολογιστής δεν έχει χρησιμοποιήσει ποτέ γλώσσα, δεν έχει απαντήσει σε μια ερώτηση, ούτε έχει αποδώσει νόημα σε οτιδήποτε.

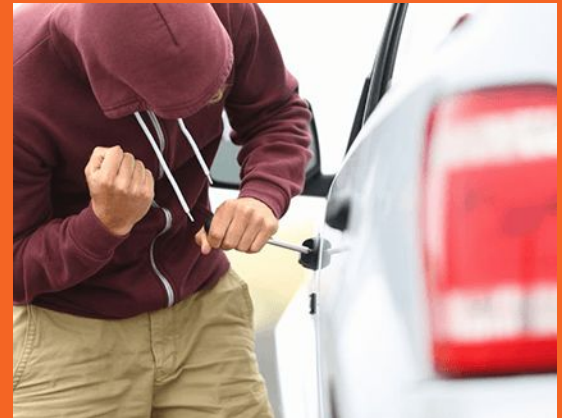
Κανένας υπολογιστής δεν έχει ποτέ προσθέσει έστω δύο αριθμούς, πόσο μάλλον να διασκεδάσει μια σκέψη, και κανένας δεν θα το κάνει. Η μόνη νοημοσύνη ή συνειδητότητα ή έστω ψευδαίσθηση συνειδητότητας σε ολόκληρη την υπολογιστική διαδικασία βρίσκεται, εντελώς αμετάθετα, **σε εμάς**- οτιδήποτε φαινομενικά ανάλογο με το μυαλό μας στις μηχανές μπορεί να αναχθεί, όταν αναλυθεί σωστά, μόνο πίσω στο δικό μας μυαλό και πάλι (...).

Πιστεύουμε το αντίθετο μόνο όταν, όπως ο Νάρκισσος που σκύβει πάνω από τα νερά, κοιτάζουμε τα δημιουργήματά μας και, γοητευμένοι από αυτό που βλέπουμε να αντανακλάται σε αυτά, φανταζόμαστε ότι ένα άλλο βλέμμα έχει συναντήσει το δικό μας.

-
- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - Είναι η σκέψη αλγόριθμος;
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - **Ποιος ο ρόλος του Μηχανικού;**
 - Τι πρέπει να φοβόμαστε;
-

Βούληση

- Η ΤΝ δεν είναι παρά ένα πολύ περίπλοκο κατσαβίδι
- Όπως κάθε τεχνολογία, και την ΤΝ πρέπει να την αναπτύξουμε και να τη χρησιμοποιήσουμε για τους ανθρώπινους σκοπούς
- Η θεοποίηση της ΤΝ ενέχει τον κίνδυνο να γίνουμε αντικείμενα χειραγώγησης, όπως κάθε θεοποίηση
- Σε κάθε τι που κάνει ο Μηχανικός, πρώτα είναι ο Σκοπός και η Βούληση



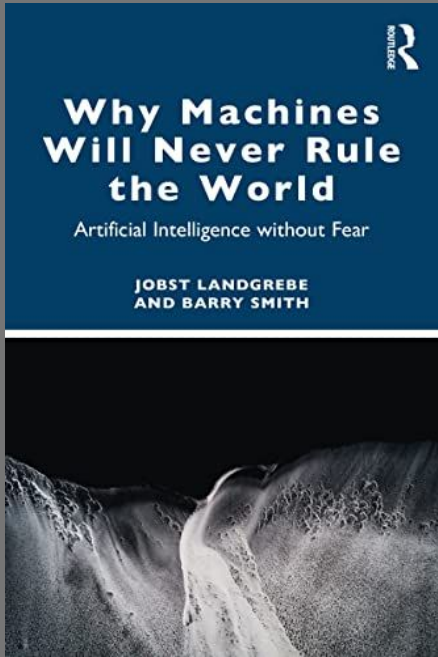
-
- Τι είναι η Γενική Τεχνητή Νοημοσύνη;
 - Μπορούμε να μοντελοποιήσουμε πολύπλοκα συστήματα;
 - Είναι η σκέψη αλγόριθμος;
 - Είναι η συνειδητότητα αποτέλεσμα υλικών διεργασιών;
 - Ποιος ο ρόλος του Μηχανικού;
 - Τι πρέπει να φοβόμαστε;
-

Εργασία

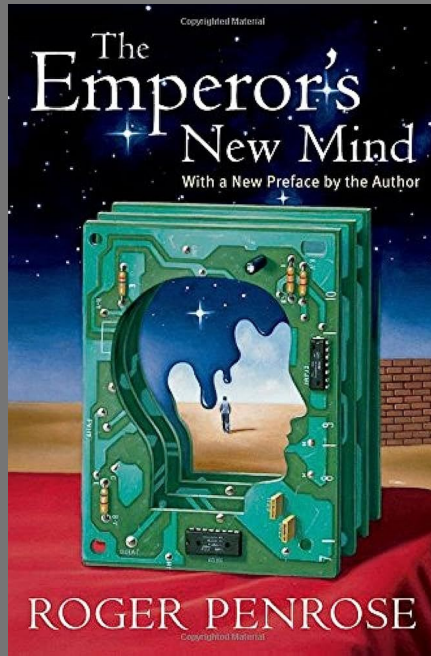
- Η ανθρώπινες δεξιότητες είναι αναντικατάστατες
- Η ανθρώπινη δημιουργικότητα είναι αναντικατάστατη
- Δεν υπάρχουν «πλεονάζοντες άνθρωποι»



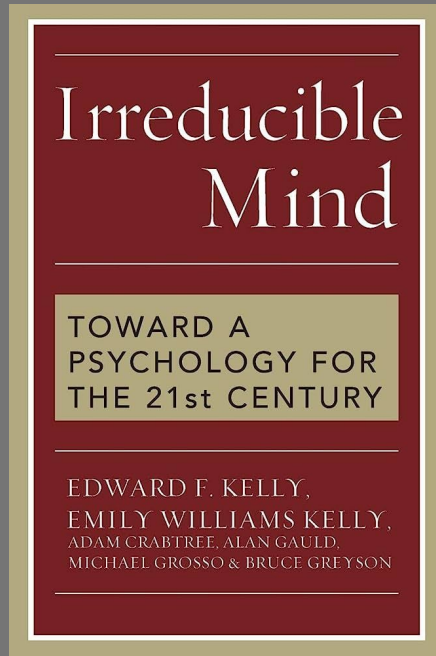
Βιβλιογραφία



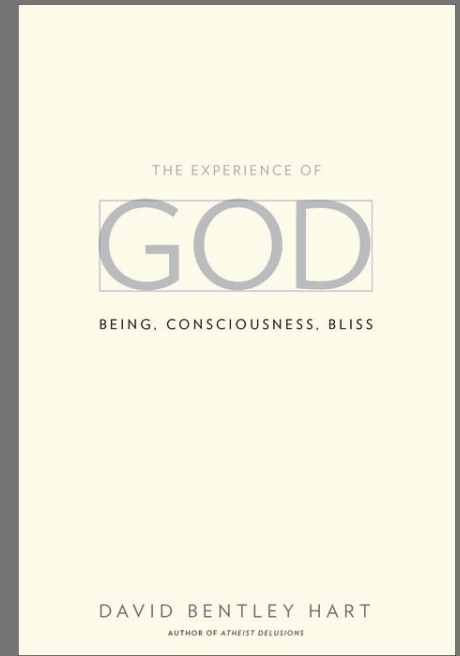
Why Machines Will Never Rule the World
Jobst Landgrebe
Barry Smith



The Emperor's New Mind
Roger Penrose



Irreducible Mind
Edward Kelly et. al



**The Experience of God:
Being, Consciousness, Bliss**
David Bentley Hart

Ευχαριστώ!

Μιχάλης Χιωτίνης

+30 697 21 33 401

m.chiotinis@gmail.com

