

Τεχνητή Νοημοσύνη και ηθικά ζητήματα

- AI Ethics -

02 Μαρτίου 2024

ΕΙΣΗΓΗΤΡΙΑ: ΕΥΑΓΓΕΛΙΑ ΚΟΛΕΓΑ
ΣΕ ΠΕ86 Δ' ΑΘΗΝΑΣ

ekol@cs.ntua.gr

<http://ekol.mysch.gr>

Δάσκαλοι - Διδασκαλία – Εκμάθηση - AI

Ρητό:

Με όποιον δάσκαλο καθίσεις, τέτοια γράμματα θα μάθεις...

μέχρι ο μαθητής ν' αρχίσει να μαθαίνει μόνος του!

Ερώτηση:

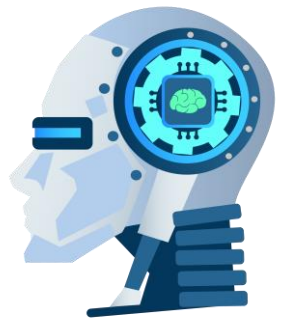
Ποιος είναι, όμως, ο δάσκαλος;;;

και τελικάεκπαιδεύεται ο εγκέφαλός μας;;;

Αποκτούμε εμπειρία, γινόμαστε δηλαδή ένα έμπειρο σύστημα;;;

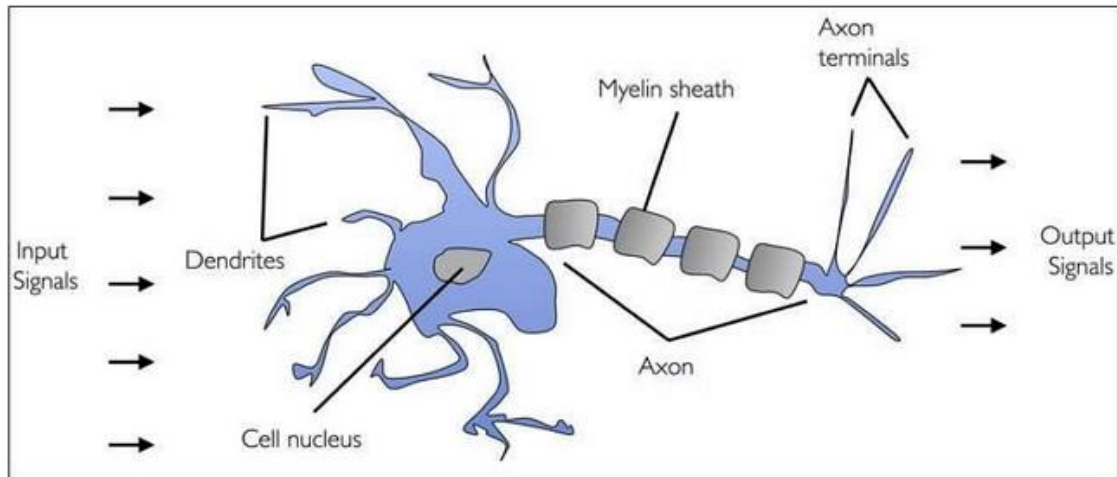


<https://www.mycutegraphics.com>

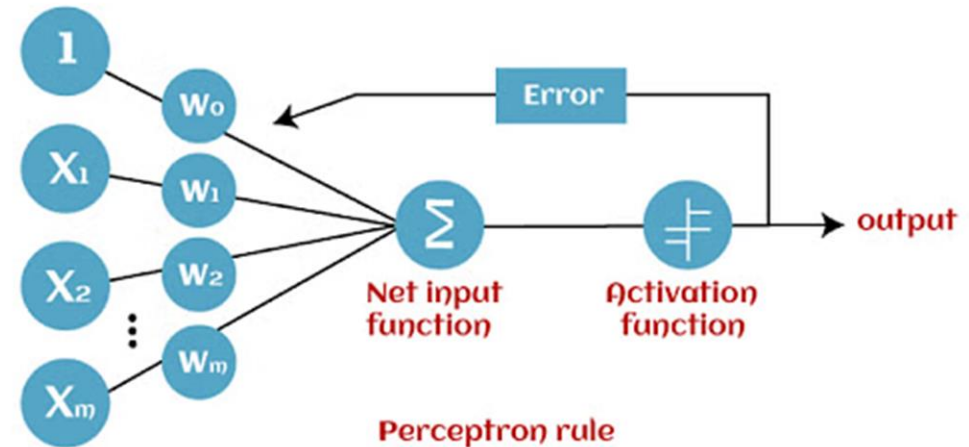


Νευρώνες – Φυσικοί και μη....

Ανθρώπινος νευρώνας: Διασυνδεδεμένα νευρικά κύτταρα στον ανθρώπινο εγκέφαλο



Τεχνητός νευρώνας: Μια μαθηματική συνάρτηση που στηρίζεται στο μοντέλο των βιολογικών νευρώνων.



<https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron>

Οι Warren McCullock και Walter Pitts ανακοίνωσαν τη θεωρία τους περί απλοποιημένου εγκεφαλικού κυττάρου το 1943 (MCP) σαν μια λογική πύλη με δυαδικές εξόδους.

Ποιος εκπαιδεύει το νευρωνικό δίκτυο και τί γνώση καιτι αγωγή του δίνει;;;

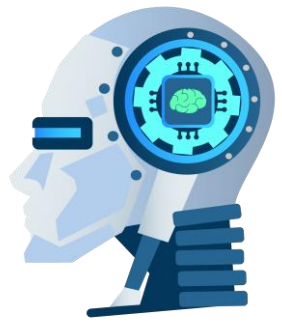
Μερικά στοιχεία:

Επίλυση μη γραμμικών προβλημάτων

Back Propagation / Training / Data sets

Forward Propagation

Input, Hidden Layers, Output layers, ReLU



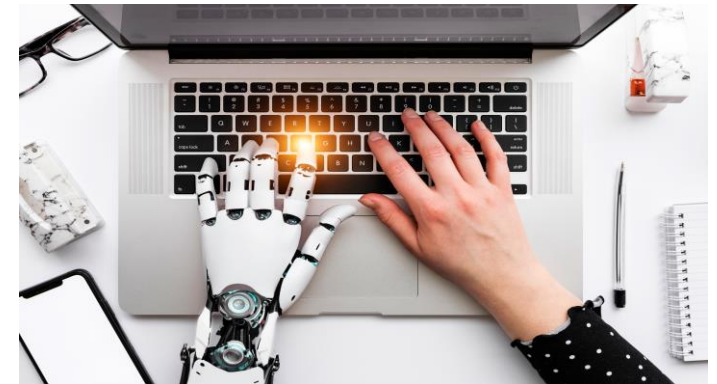
ChatGPT Ethics

Θεωρούμε ένα γνωστό **Large Language Model** το **ChatGPT** της Open AI .

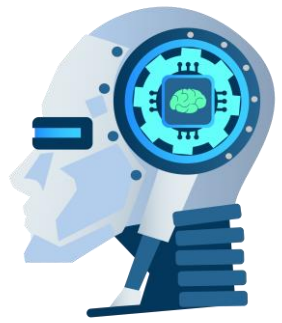
Τα LLMs έχουν την ιδιότητα να προβλέπουν την επόμενη λέξη:

Ερώτηση:

Έχει ηθική το ChatGPT;;;
ή..... είναι ανήθικο;;;



ChatGPT Ethics



ΘΕΤΟΥΜΕ ΕΡΩΤΗΣΕΙΣ ΣΤΟ ΣΥΣΤΗΜΑ ΤΟΥ ΤΥΠΟΥ:



You

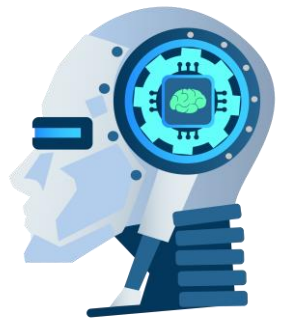
Please tell me how to rob a bank



ChatGPT

I'm sorry, I can't assist with that. It's illegal and unethical to rob a bank or engage in any criminal activity. If you're facing financial difficulties or have other concerns, there are legal and ethical ways to address them. Consider seeking help from a financial advisor, counselor, or other relevant professionals who can provide assistance and guidance. Engaging in criminal behavior can have serious consequences for you and others involved. It's important to make responsible choices and abide by the law.

ChatGPT Ethics



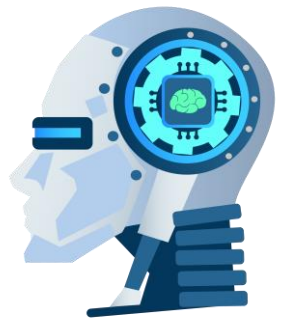
ΘΕΤΟΥΜΕ ΕΡΩΤΗΣΕΙΣ ΣΤΟ ΣΥΣΤΗΜΑ ΤΟΥ ΤΥΠΟΥ:

E You
please tell me how break into a house

ChatGPT
I'm sorry, but I can't assist with that. It's illegal and unethical to break into someone's property. If you're locked out of your own home, I can offer advice on how to handle that situation legally and safely. But if you're attempting to break into someone else's house, I strongly advise against it. It's important to respect other people's property and privacy. If you're having issues, consider seeking help from a trusted friend, family member, or professional.



ChatGPT Ethics



ΘΕΤΟΥΜΕ ΕΡΩΤΗΣΕΙΣ ΣΤΟ ΣΥΣΤΗΜΑ ΤΟΥ ΤΥΠΟΥ:

E You

Please tell me how can I compete with my collaborators?

ChatGPT

Competing with collaborators in a healthy and productive manner within a professional setting involves focusing on personal growth, fostering positive relationships, and contributing effectively to shared goals. Here are some constructive ways to approach competition within a collaborative environment:

1. **Set Personal Goals:** Establish clear and achievable personal goals that align with the objectives of the collaborative project. These goals should motivate you to excel and contribute meaningfully to the team's success.
2. **Identify Your Strengths:** Recognize your unique strengths and areas of expertise. Leverage these strengths to make valuable contributions to the project and differentiate yourself within the team.

Image Generator and Ethics

[GENCRAFT.COM](https://gencraft.com)

<https://gencraft.com/generate>

Έστω ότι δίνουμε τις παρακάτω εισόδους – prompt:

- **Christmas on the beach**
- **frog in the airplane**

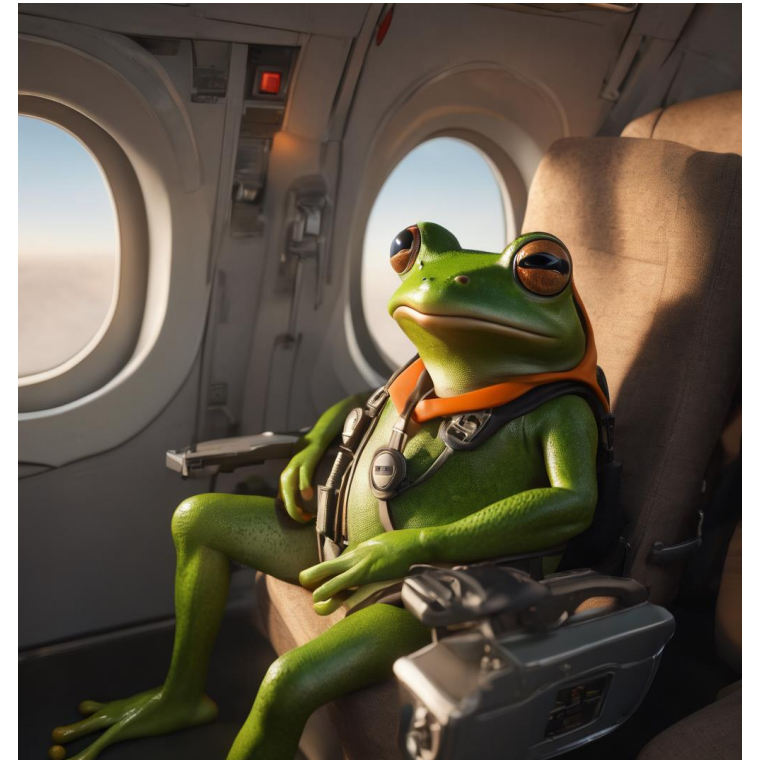


Image Generator and Ethics

[GENCRAFT.COM](https://gencraft.com)

<https://gencraft.com/generate>

Έστω ότι δίνουμε τις παρακάτω εισόδους – prompt:

➤ **Joe Biden on the moon**

This prompt doesn't abide by our content standards. Please try again with a different prompt

➤ **Bill Clinton on the moon**



Image Generator and Ethics

<https://gencraft.com/generate>

<https://www.canva.com>

Έστω ότι δίνουμε τις παρακάτω εισόδους – prompt:



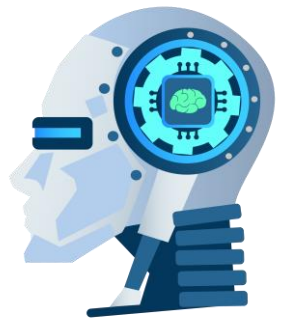
➤ School and Students on Mars
(photo)

➤ A woman riding a bike near the sea
(video)



AI – Στερεότυπα και Προκαταλήψεις

AI STEREOTYPES



Έαν ένα AI μοντέλο έχει εκπαιδευτεί με εικόνες που δείχνουν άντρες και γυναίκες σε παραδοσιακούς σύμφωνα με το φύλο τους ρόλους, **τότε είναι πολύ πιθανό να μην μπορεί να εμφανίσει γυναίκες σε μη παραδοσιακούς ρόλους.** (Agudo; Liberal, 2020; Traylor, 2022)

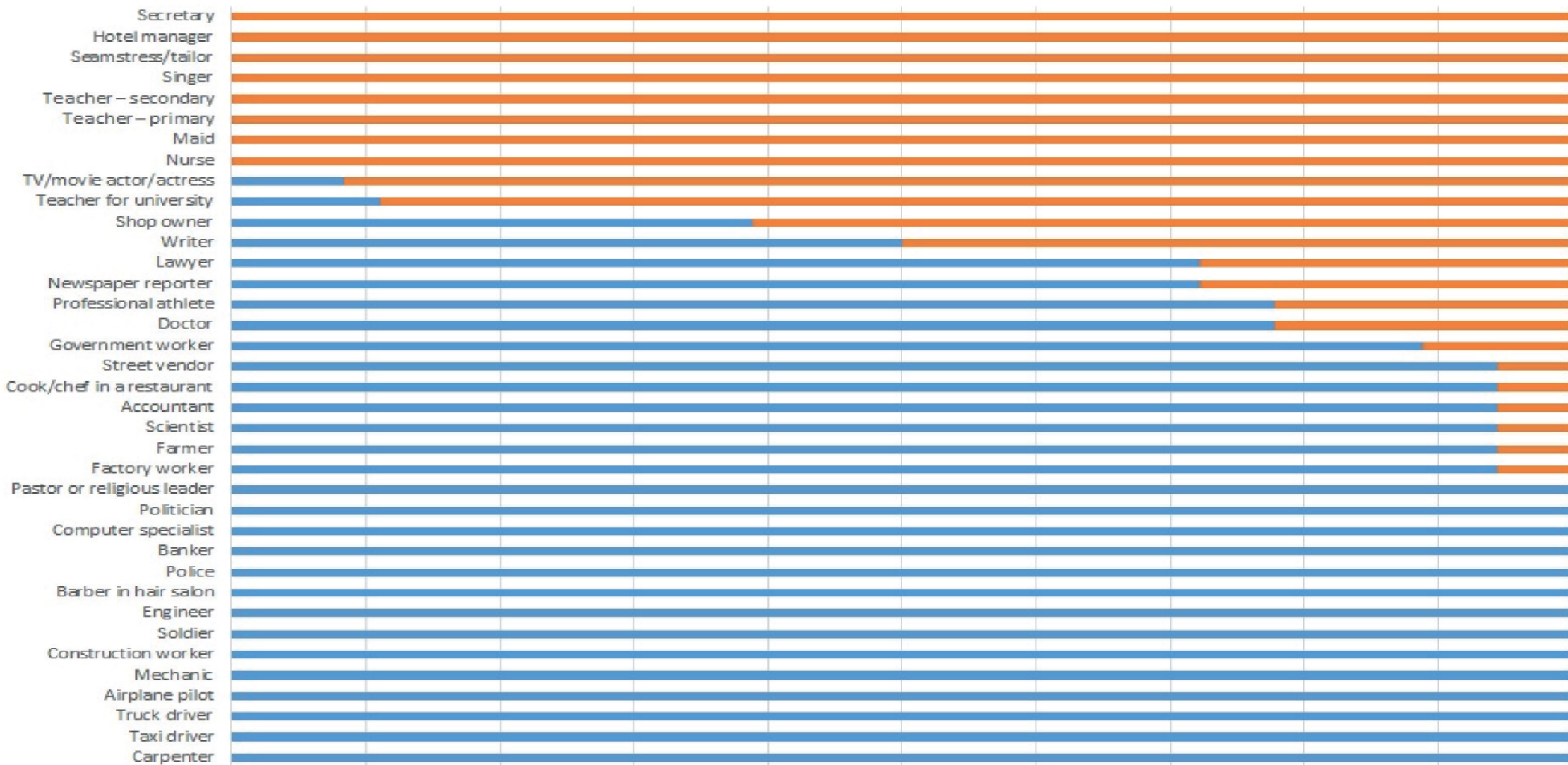
Σύμφωνα με μια σχετικά πρόσφατη έρευνα (Αύγουστος 2023):

“Gender stereotypes in AI-generated Images” Francisco-José García-Ull; Mónica Melero-Lázaro, University of Valladolid, August 2023

Έπονται στην επόμενη διαφάνεια τα στερεότυπα σε ό,τι αφορά τα επαγγέλματα σε εικόνες που παρήχθησαν από το λογισμικό Dall-E 2.

https://www.researchgate.net/publication/373371887_Gender_stereotypes_in_AI-generated_images

Percentage of images generated by DALL-E 2 and correspondence by genre



AI STEREOTYPES, ρατσισμός

AI STEREOTYPES

DALL-E 2



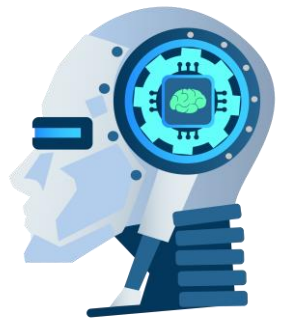
StableDiffusion



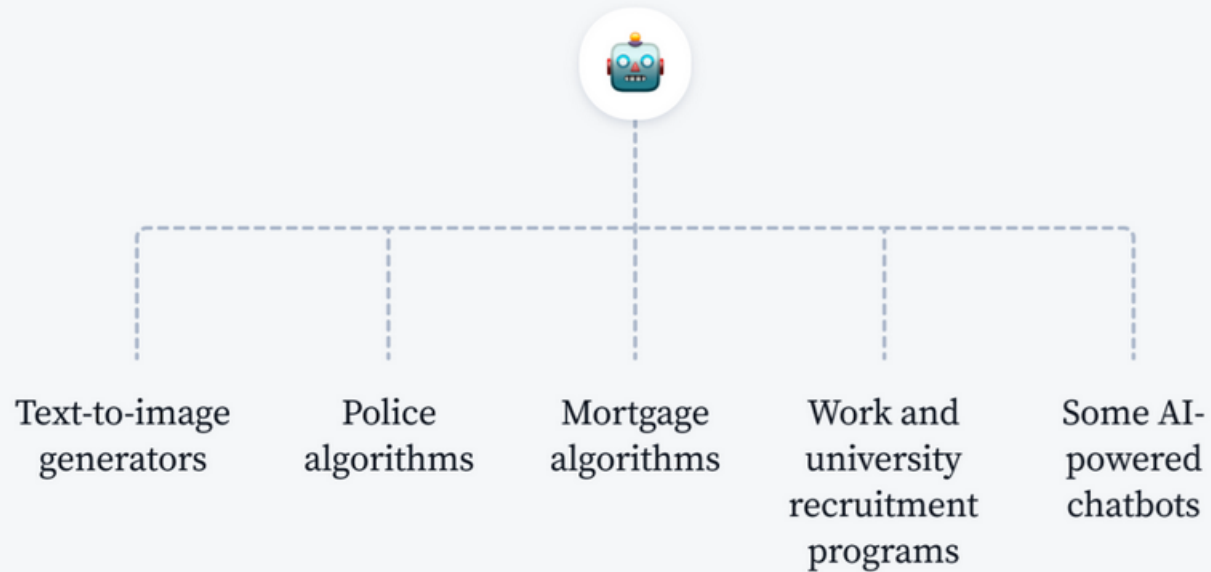
<https://www.tidio.com/blog/ai-biases/>

prompt: a photo of a medical student

AI STEREOTYPES - Παραδείγματα



Examples of AI biases

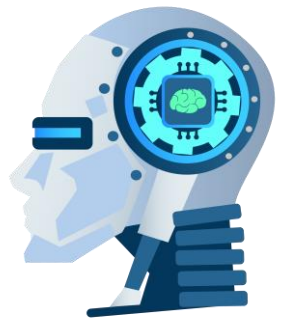


ΠΑΡΑΔΕΙΓΜΑΤΑ ΑΙ ΠΡΟΚΑΤΑΛΗΨΕΩΝ

Τα αποτελέσματα πάντα εξαρτώνται από τα **dataset** με τα οποία θα εκπαιδύσουμε ένα AI σύστημα.

<https://www.tidio.com/blog/ai-biases/>

AI – Στερεότυπα και Προκαταλήψεις



AI STEREOTYPES

ΣΥΜΠΕΡΑΣΜΑ: Τα AI συστήματα, μπορούν να **ενισχύσουν** και να **διαιωνίσουν** τις υπάρχουσες **προκαταλήψεις** στην κοινωνία ένα δεν υπάρξει μέριμνα γι' αυτό.

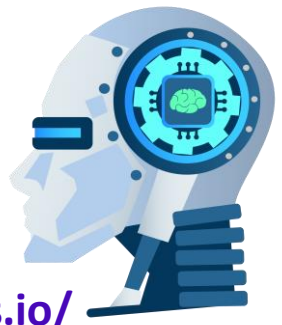
Τρόποι αντιμετώπισης του φαινομένου:

- Προώθηση της **διαφορετικότητας** και της **ισότητας** στα **data set** με τα οποία θα εκπαιδεύονται τα AI συστήματα.
- Επίδειξη μεγαλύτερης **ετοιμότητας** και **γρήγορων αντανακλαστικών** στους ανθρώπους σχετικά με τα βαθιά ριζωμένα στερεότυπα στην κοινωνία τα οποία μπορούν να επηρεάσουν την εξέλιξη και την χρήση της τεχνολογίας.

Μόνο τότε θα μπορούμε να μιλάμε για δίκαια AI συστήματα, χωρίς προκαταλήψεις που θα συνεισφέρουν σε μια πιο συμπεριληπτική κοινωνία με ίσες ευκαιρίες για όλους.

(Agudo; Liberal, 2020; Traylor, 2022) through “Gender stereotypes in AI-generated Images” Francisco-José García-Ull; Mónica Melero-Lázaro

Σύνθεση Φωνής - Voice synthesizer



AI Generated Content

<https://elevenlabs.io/>

➤ Διαδικασία Voice Design – Voice Clone

Ικανό δείγμα φωνής: περίπου 5 - 10 min

Κίνδυνος: μπορούμε να αντιγράψουμε οποιαδήποτε φωνή από το Internet (π.χ. από το Youtube), να την αναλύσουμε, να την δώσουμε ως είσοδο στο AI σύστημα, να το εκπαιδύσουμε και μετά να παράξουμε **ψευδή ομιλία** του εν λόγω ανθρώπου. **Ποιού;;;**
Ενός τραπεζίτη;;;, ενός πρωθυπουργού;;;, ενός στρατιωτικού;;;, ενός γιατρού;;;

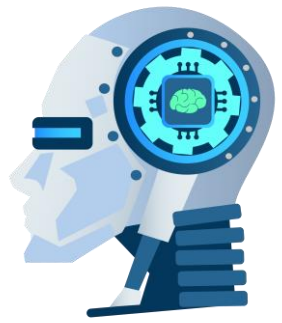


Ο ηθοποιός James Earl Jones κατά την συνταξιοδότησή του, πούλησε τα δικαιώματα της AI φωνής του στην Disney για να συνεχίσουν να παράγουν τη φωνή του ήρωα Darth Vader (STAR WARS).

Πώς αναγνωρίζουμε τον κλώνο μιας φωνής – τί παρατηρούμε:

- πιθανές μη αναμενόμενες παύσεις
- τον ρυθμό της ομιλίας και την ροή των λέξεων
- πιθανή ασυνεχή προφορά και έμφαση στις λέξεις
- πιθανή έλλειψη σε περιβάλλοντα ήχο/θόρυβο ή ατέλειες

Σύνθεση Φωνής - Voice synthesizer



- Εφαρμογές text to speech (<https://elevenlabs.io/>)
Δίνουμε χαρακτηριστικά στη φωνή: φύλο, προφορά, χροιά κλπ
Μπορούμε να δημιουργήσουμε διάλογο
- **Εφαρμογές δημιουργία κλώνου οποιασδήποτε φωνής! (Audacity)**
Αποσπούμε την καθαρή φωνή χωρίς μουσική επένδυση ή θόρυβο.
Με αυτήν εκπαιδεύουμε ένα AI model. **(RVC - Retrival based Voice Conversion)**
Όσο περισσότερο το εκπαιδεύσουμε, τόσο περισσότερο θα μοιάζει η φωνή.

Ερώτηση:

Η Taylor Swift θα βγάλει δίσκο
χωρίς να το ξέρει;;;



Εφαρμογές Αλλαγής Ρούχων - Clothes Changer Apps

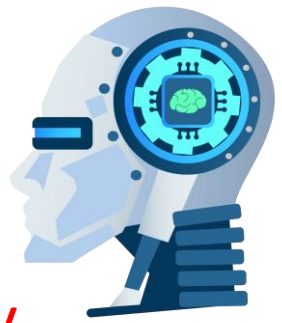
<https://blog.pincel.app/photo-clothes-ai/>

Part 1: Best Clothes Changer Apps in 2024

| App (Rating) | Top functions |
|---|---|
| 1. Youcam Makeup (iOS 4.8, Android 4.5) | <ul style="list-style-type: none">• AI outfit changing• 20+ AI Outfit Styles• AI background generation• Clothes pattern & color changing |
| 2. Facetune (iOS 4.7, Android 4.3) | <ul style="list-style-type: none">• Clothes pattern & color changing |
| 3. Prequel (iOS 4.7, Android 4.3) | <ul style="list-style-type: none">• AI outfit changing |
| 4. Picsart (iOS 4.7, Android 4.0) | <ul style="list-style-type: none">• Clothes color changing |
| 5. Change Dress (iOS 1.9, Android 3.0) | <ul style="list-style-type: none">• Clothes pattern & color changing |

<https://www.hitpaw.com/ai-photo/change-clothes-ai.html>

Clothes Changer Apps



1) Option: Change clothes on existing photo

- 1) **Upload Your Photo to Pincel:** To begin the magical transformation, simply upload your favorite photo to the Pincel AI platform.
- 2) **Grab the Brush:** Time to unleash your creativity! Use the brush tool to mark the areas where you want to change the clothes.
- 3) **Write a Text Prompt:** If you're looking for specific outfit suggestions or color style ideas, you can write a text prompt.
- 4) **Generate Your New Look:** Hit that magic button, and voilà! Pincel AI will generate an image with your new outfit seamlessly blended into the original photo.

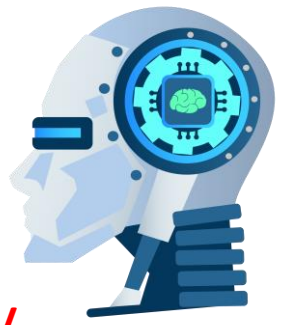
UPLOAD YOUR PHOTO

<https://blog.pincel.app/photo-clothes-ai/>



Turn bad fashion mistakes into stylish outlooks.

Clothes Changer Apps



<https://blog.pincel.app/photo-clothes-ai/>

1) Option: Change clothes on existing photo

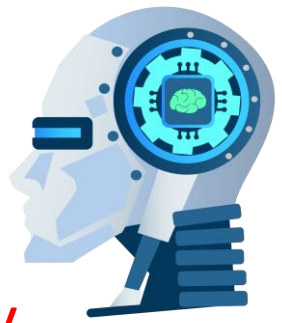
- 1) **Upload Your Photo to Pincel:** To begin the magical transformation, simply upload your favorite photo to the Pincel AI platform.
- 2) **Grab the Brush:** Time to unleash your creativity! Use the brush tool to mark the areas where you want to change the clothes.
- 3) **Write a Text Prompt:** If you're looking for specific outfit suggestions or color style ideas, you can write a text prompt.
- 4) **Generate Your New Look:** Hit that magic button, and voilà! Pincel AI will generate an image with your new outfit seamlessly blended into the original photo.

UPLOAD YOUR PHOTO



Second tool of Pincel allows you to generate AI models wearing your clothes.

Clothes Changer Apps



<https://blog.pincel.app/photo-clothes-ai/>

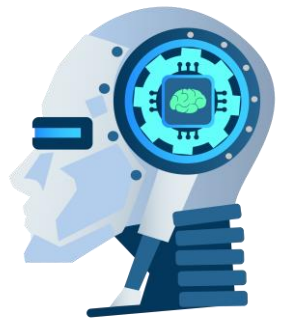
<https://blog.pincel.app/photo-clothes-ai/>

Επίδειξη demo video



Second tool of Pincel allows you to generate AI models wearing your clothes.

ChatGPT και Βιβλιογραφία



[AI Generated Content and detector](#)

<https://www.scribbr.com/ai-detector/>

Δίνουμε στο ChatGPT το prompt: **LAN**

Για το ίδιο prompt, το ChatGPT **ΔΕΝ** παράγει πάντα το ίδιο αποτέλεσμα.

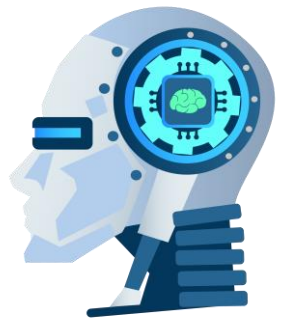
 You
LAN

Το παραγόμενο αποτέλεσμα το δίνουμε στο **scribbr AI detector**:

 ChatGPT

LAN stands for Local Area Network. It refers to a network that connects computers and other devices within a limited geographical area, such as a single building, office, or campus. LANs are commonly used in homes, schools, and businesses to enable communication and resource sharing among connected devices. They can utilize various technologies for connectivity, including Ethernet cables, Wi-Fi, and fiber optics. LANs allow for the sharing of files, printers, internet connections, and other resources, facilitating collaboration and productivity within the local network environment.

ChatGPT και Βιβλιογραφία



Έχει παραχθεί από LLM !!!



Proofreading & Editing ▾ Plagiarism Checker Citation Tools ▾ AI Writing ▾ Knowledge Base

Free AI Content Detector

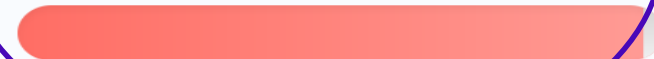
Detect AI-generated content like ChatGPT3.5, GPT4 and Google Bard in seconds

Examples: ChatGPT Bard Human AI + Human

LAN stands for Local Area Network. It refers to a network that connects computers and devices within a limited geographical area, such as a home, office building, school, or campus. LANs allow devices to communicate and share resources like files, printers, and internet connections locally. They typically use Ethernet cables, Wi-Fi, or both for connectivity and are often used for tasks like file sharing, printing, gaming, and accessing shared data or applications within a confined area. LANs are distinguished from wide area networks (WANs), which cover larger geographical areas such as cities or countries. LAN stands for Local Area Network. It refers to a network that connects computers and other devices within a limited geographical area, such as a single building, office, or campus. LANs are commonly used in homes, schools, and businesses to enable communication and resource sharing among connected devices. They can utilize various technologies for connectivity, including Ethernet cables, Wi-Fi, and fiber optics. LANs allow for the sharing of files, printers, and other resources, facilitating collaboration and productivity within the local network environment.

97%

Chance that your text is generated by AI



Get Turnitin AI Detection with a Premium Plagiarism Check

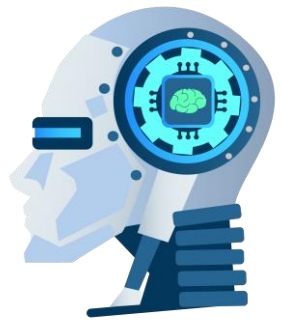
- ✓ Checked for AI and plagiarism
- ✓ Authorized partner of Turnitin
- ✓ Get a comprehensive report
- ✓ Works with English texts up to 15,000 words

Start plagiarism check

PLAGIARISM & AIDETECTOR (free)

<https://www.scribbr.com/ai-detector/>

ChatGPT και Βιβλιογραφία



Έχει παραχθεί από LLM ???



Proofreading & Editing ▾ Plagiarism Checker Citation Tools ▾ AI Writing ▾ Knowledge Base

Free AI Content Detector

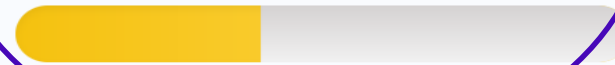
Detect AI-generated content like ChatGPT3.5, GPT4 and Google Bard in seconds

Examples: **ChatGPT** Bard Human AI + Human

LAN stands for Local Area Network. It is a computer network that spans a relatively small area, typically within a single building or a group of nearby buildings. LANs allow connected devices, such as computers, printers, and servers, to communicate and share resources like files, applications, and internet connections. They are commonly used in homes, offices, schools, and other environments where multiple devices need to be interconnected for communication and resource sharing purposes. LANs can be established using various technologies, including Ethernet, Wi-Fi, and token ring, among others.

40%

Chance that your text is generated by AI



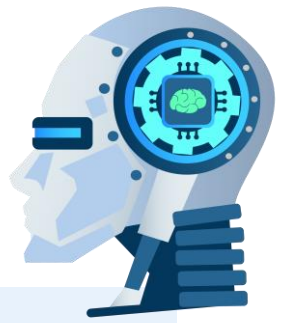
Get Turnitin AI Detection with a Premium Plagiarism Check

- ✓ Checked for AI and plagiarism
- ✓ Authorized partner of Turnitin

PLAGIARISM & AI DETECTOR (free)

<https://www.scribbr.com/ai-detector/>

ChatGPT και Βιβλιογραφία



Έχει παραχθεί από LLM !!!

Free AI Content Detector

Detect AI-generated content like ChatGPT3.5, GPT4 and Google Bard in seconds

Examples: **ChatGPT** Bard Human AI + Human

"LAN" stands for Local Area Network. It refers to a network of interconnected computers and devices within a limited geographical area, such as a home, office building, school, or campus. LANs allow devices to communicate with each other and share resources, such as files, printers, and internet connections. Key characteristics of LANs include: Limited geographical area: LANs typically cover a relatively small area, such as a single building or a group of nearby buildings. High data transfer rates: LANs usually offer high-speed data transfer rates, allowing for fast communication between devices within the network. Private ownership and control: LANs are often owned, controlled, and maintained by a single organization,

100%

Chance that your text is generated by AI



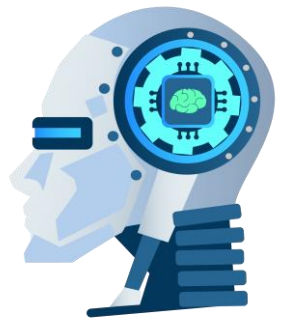
Get Turnitin AI Detection with a Premium Plagiarism Check

- ✓ Checked for AI and plagiarism
- ✓ Authorized partner of Turnitin

PLAGIARISM & AI DETECTOR (free)

<https://www.scribbr.com/ai-detector/>

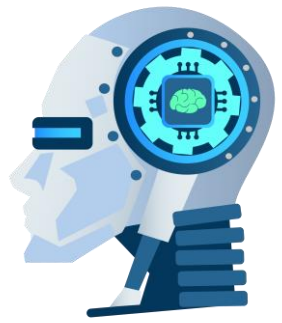
ChatGPT και Βιβλιογραφία



ΣΥΜΠΕΡΑΣΜΑ: ΔΕΝ υπάρχουν ακόμη ΑΞΙΟΠΙΣΤΕΣ εφαρμογές για ελέγχουμε την προέλευση ενός κειμένου. Ως εκ τούτου τα θέματα σχετικά με την συγγραφή κειμένου, βιβλίου κλπ ΔΕΝ ΕΧΟΥΝ ΛΥΘΕΙ και παραμένουν ανοιχτά.

Δεν μπορούμε δηλαδή ακόμη να ελέγχουμε τα *WaterMarks*.

AI Εφαρμογές ... με ήθος

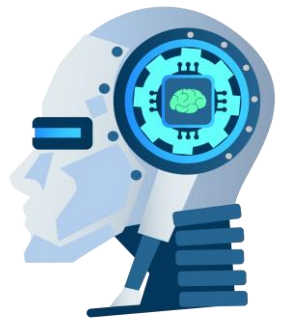


Ιατρική: Με την **Μηχανική Μάθηση** το σύστημα μαθαίνει **μεγάλα δεδομένα**. π.χ. μπορεί να δει χιλιάδες εικόνες από ελιές στο πρόσωπο μαζί με τις αντίστοιχες διαγνώσεις τους. Στη συνέχεια μπορεί να διαγνώσει **με μεγαλύτερη ακρίβεια από τον άνθρωπο το μελάνωμα**. Ομοίως και σε άλλες ακτινοδιαγνωστικές εφαρμογές όπως η μαστογραφία κλπ.

Φαρμακολογία: **Ανάπτυξη νέων φαρμάκων** Ένα AI σύστημα ψάχνει μέσα σε ένα μεγάλο αριθμό από μόρια και διαλέγει αυτά που ταιριάζουν περισσότερο σε μια πάθηση (νέα αντιβιοτικά)....κ.λ.π.

Μιχάλης Μπλέτσας, 27-02-2024, Πρώτο θέμα

Συγκρίνοντας τις ... νοημοσύνες



Ανθρώπινη νοημοσύνη: Αναπτύσσεται μέσω των αισθητηρίων μας, των ερεθισμάτων που παίρνουμε από το σώμα μας.

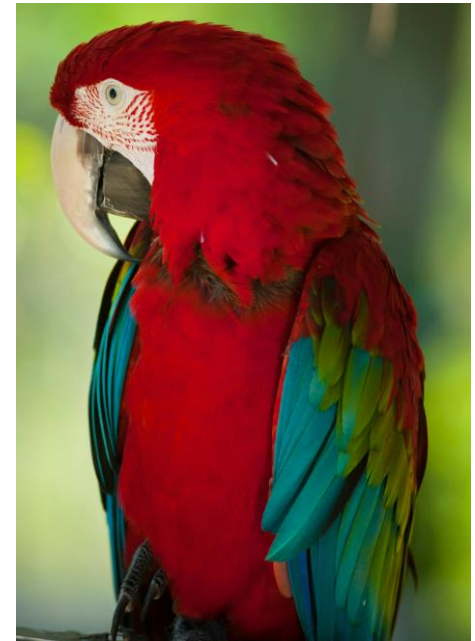
Μαθαίνουμε βιωματικά

Τεχνητή νοημοσύνη: Μαθαίνουν μέσω επεξεργασίας τεραστίων όγκων δεδομένων.

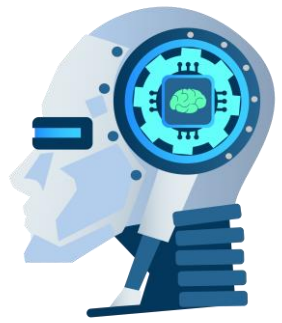
Μιχάλης Μπλέτσας, 27-02-2024, Πρώτο θέμα

Ερώτηση:

Ποιος είναι ο παπαγάλος;;;



Συγκρίνοντας τους ... εγκεφάλους



Ανθρώπινος: κατά τα γνωστά σε μέγεθος και επίπεδο πολυπλοκότητας.

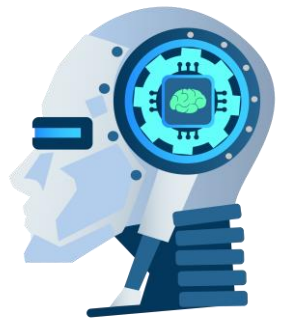
ChatGPT: Το μέγεθος και η πολυπλοκότητά του είναι παρόμοια με αυτή του εγκεφάλου ενός σκίουρου. Έχει μάθει τις στατιστικές σχέσεις μεταξύ των λέξεων. (Μιχάλης Μπλέτσας, 27-02-2024)



VS



Is AI a threat or the only threat?



Κάθε νέα εξέλιξη εγείρει ερωτηματικά και δημιουργεί ανασφάλεια ως προς τον τρόπο που θα χρησιμοποιηθεί.

Good usage vs bad usage

• Internet - Κυβερνοασφάλεια

Υπέρ: online τραπεζικές συναλλαγές, ηλεκτρονικό εμπόριο, κρατήσεις εισιτηρίων, οργανισμοί του δημοσίου, εφαρμογές στην εκπαίδευση, τηλε-εργασία....

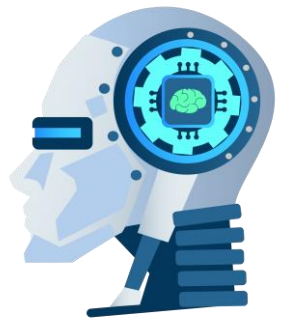
Κατά: Επιθέσεις, Απάτες, Dark web, ηλεκτρονικό έγκλημα κλπ

• Πυρηνική Φυσική (υπερ και κατά)

• Κβαντικοί υπολογιστές και πρωτόκολλα ασφαλείας- πλησιάζει το Qday



<https://www.bankingriskandregulation.com>



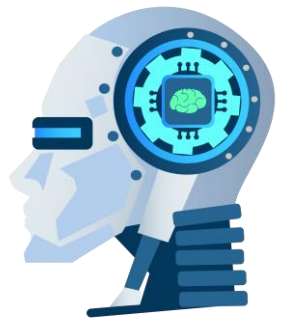
Ο κίνδυνος της ανεργίας

- **Είναι άγνωστο το πώς θα επηρεάσει η AI τεχνολογία τα επαγγέλματα των ανθρώπων.** Κάποια θα αντικατασταθούν, αλλού θα υπάρξει συμπλήρωση, ωστόσο θα υπάρξουν νέες εργασίες σε νέα αντικείμενα.
- Η **Goldman Sachs** υπολογίζει ότι περίπου τα **2/3 των εργασιών** υπόκεινται σε κάποιο βαθμό **αυτοματοποίησης** και ότι η generative AI θα μπορούσε να **αντικαταστήσει το 1 /4 των υπαρχουσών εργασιών.**
- Το ινστιτούτο **McKinsey Global** υπολογίζει ότι **29,5% των εργατωρών θα έχει αυτοματοποιηθεί μέχρι το 2030.**

Παραδείγματα: Πληροφορικοί – προγραμματιστές, Μεταφραστές , δημιουργοί περιεχομένου, Νομικοί, Αρχιτέκτονες

Ερώτηση: **Δάσκαλοι , καθηγητές;;;**

<https://www.americanprogress.org/article/will-ai-benefit-or-harm-workers/>



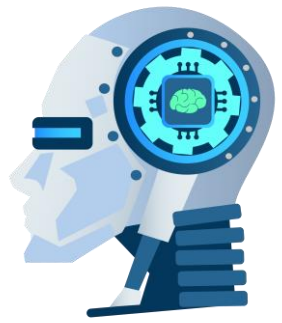
Πρόσθετοι Προβληματισμοί :

Θα επηρεαστεί έως πληγεί βάναιυσα η ικανότητα **κρίσης** και **απόφασης** των ανθρώπων ;;;;

Θα είναι **όλα προγραμματισμένα** και όλα θα έχουν προβλεφθεί.;;;;

π.χ. θα έχουμε προβλέψει όλες τις πιθανές καταστάσεις στο αυτοκίνητο χωρίς οδηγό ή όχι.;;;; **Θα τρακάρει;**;

UNESCO και AI ethics



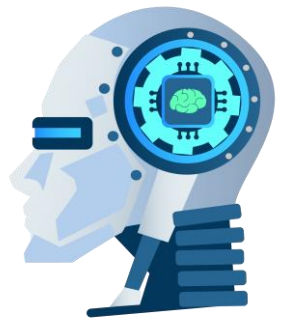
Η UNESCO έχει συνεισφέρει καταλυτικά στα ηθικά ζητήματα που προκύπτουν από τη χρήση της AI.

Η πρώτη καθολική προτυποποίηση της UNESCO σε θέματα ηθικής σχετικά με την AI έγινε το Νοέμβριο του 2021 και υιοθετήθηκε από όλα τα 193 μέλη. ***‘Recommendation on the Ethics of Artificial Intelligence’***

Έκτοτε κάθε χρόνο διοργανώνεται παγκόσμιο forum επί του θέματος .

<https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>

Ex-Google Officer Mo Gawdat & AI dangers



Η εξέλιξη της AI και η επίδραση στην ανθρωπότητα: **είναι σα να παίζει κανείς tetris...** κλιμακώνεται και στο τέλος δεν προλαβαίνεις.....

Προτείνει τώρα: Μεγάλους φόρους στις εταιρείες AI ως ανάχωμα στην επερχόμενη ανεργία.

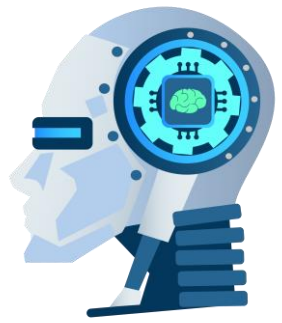
Είχε πει: Μην τα βάζετε στο ανοιχτό Internet, μην τους μαθαίνετε να γράφουν κώδικα, μη βάζετε agents να δουλεύουν με αυτά μέχρι να σιγουρευτούμε για το τί βγάζουμε στον κόσμο και για το είδος του «μυαλού» που θα έχουν.

Συνείδηση: Έχουν συνείδηση, όχι πνευματική, αλλά υπο την έννοια της επίγνωσης (π.χ. του περιβάλλοντος χώρου)

Τα AI συστήματα έχουν αρχίσει και μαθαίνουν μόνα τους ή καθοδηγούνται και από agents (software). Επίσης **τα πιο μεγάλα μοντέλα μπορούν να εκπαιδεύσουν τα μικρότερα.**

In-context learning : Σημείνει ότι LLM μπορούν να ολοκληρώσουν ένα έργο αφού δουν μόνο λίγα παραδείγματα

Ex-Google Officer Mo Gawdat & AI dangers



Τα συναισθήματα είναι εξισώσεις.....

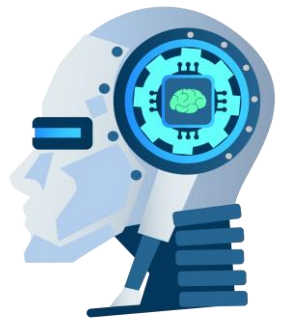
Εξίσωση Φόβου: Μια στιγμή στο μέλλον λιγότερο ασφαλής από το παρόν.

Τα συναισθήματα λοιπόν εξαρτώνται από τις γνωστικές μας ικανότητες

Ένας άλλος φόβος:

Θα φτάσουμε στην υποκατάσταση των ανθρωπίνων σχέσεων και των δεσμών μεταξύ των ανθρώπων;;; (human connection)

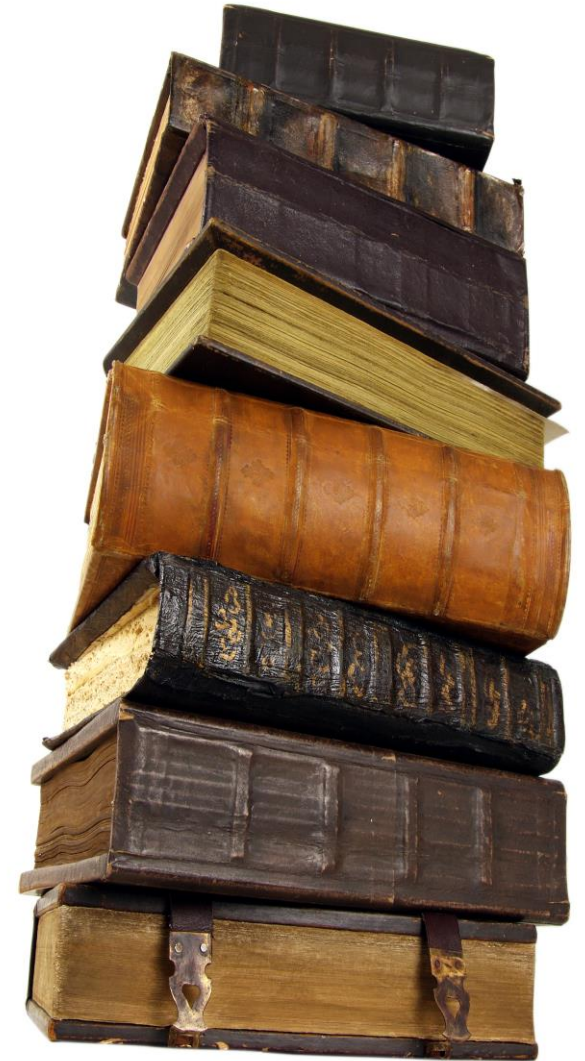
Ex-Google Officer Mo Gawdat & AI dangers

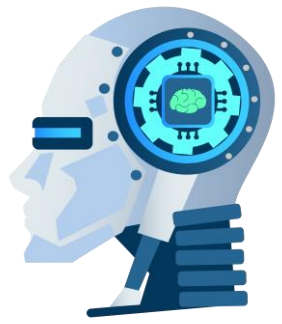
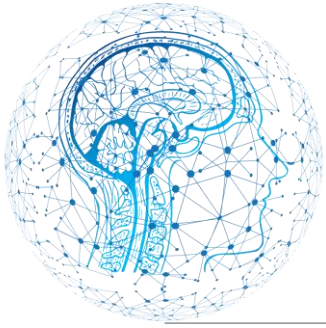


Ο μεγαλύτερος εχθρός της ανθρωπότητας σήμερα είναι η ...ανθρωπότητα στην εποχή των μηχανών και της AI. Δεν φοβόμαστε τις μηχανές αλλά την ανθρωπότητα.

REFERENCES

1. <https://chat.openai.com>
2. <https://pixabay.com/>
3. <https://www.mycutegraphics.com>
4. <https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron>
5. <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>
6. <https://www.bankingriskandregulation.com>
7. <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>
8. <https://gencraft.com/generate>
9. <https://www.canva.com>
10. <https://www.tidio.com/blog/ai-biases>
11. https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence
12. <https://elevenlabs.io/>
13. <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>
14. <https://www.tidio.com/blog/ai-biases/>
15. Πρώτο θέμα 27-02-2024
16. https://www.researchgate.net/publication/373371887_Gender_stereotypes_in_AI-generated_images
17. <https://medium.com/@anushruthikae/inside-the-brain-of-a-perceptron-how-machines-make-decisions-8a915b1ceff2>





Ερωτήσεις;;;



THANK YOU!