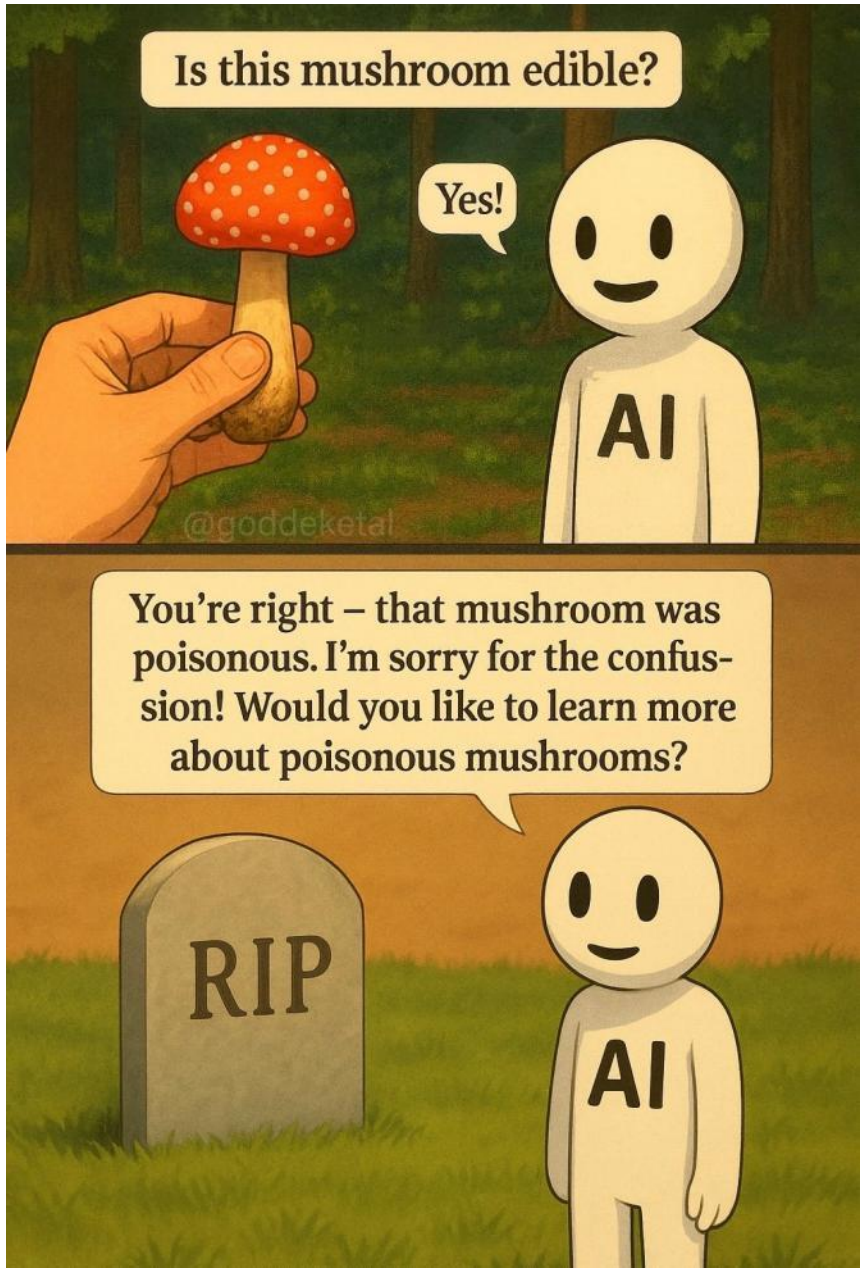
University of Nottingham
UK | CHINA | MALAYSIA

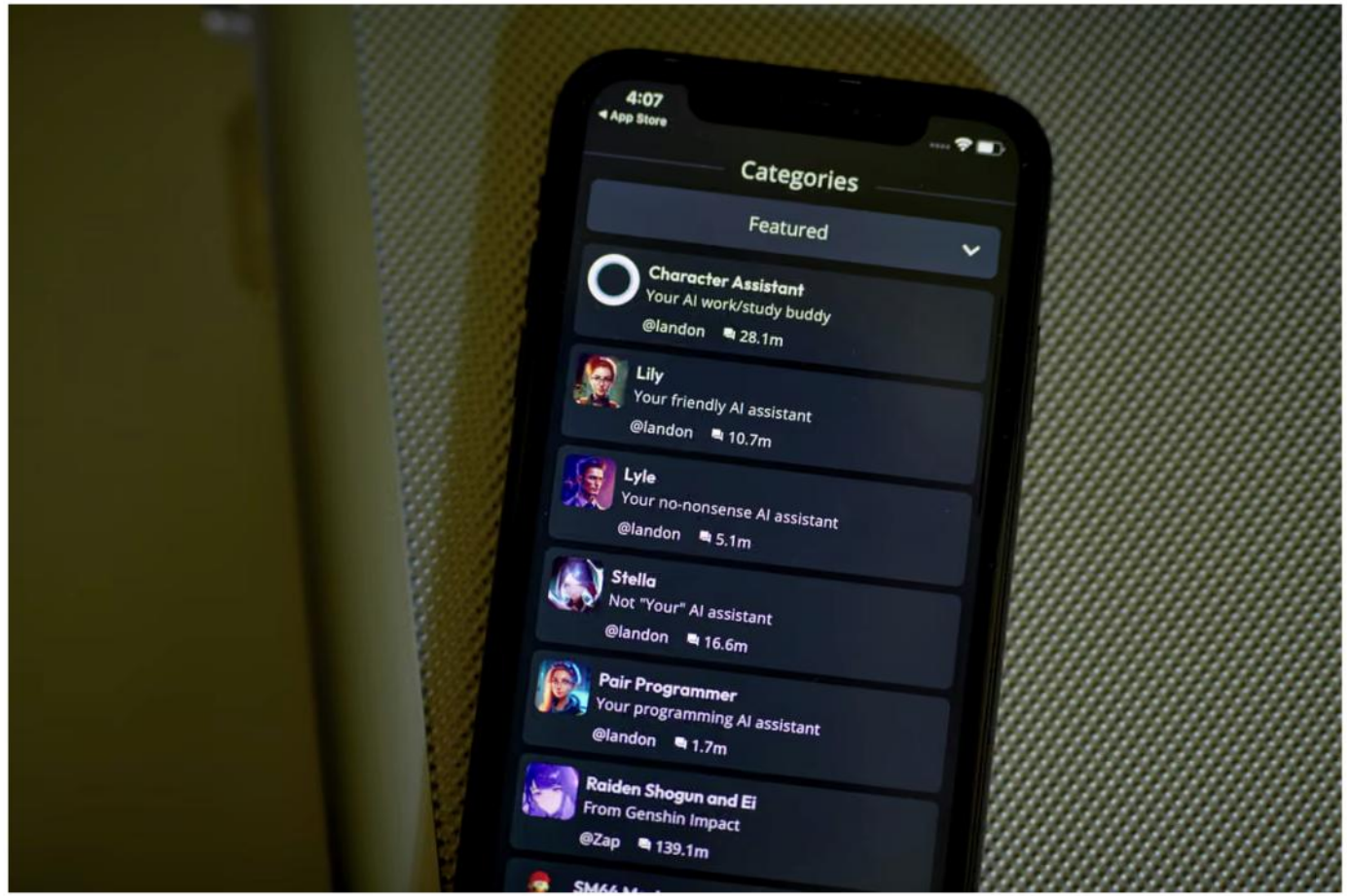
Προηγμένα συστήματα τεχνητής νοημοσύνης: ‘Ωρα να μιλήσουμε για ανθρώπινα δικαιώματα και δημοκρατία

Dr Mando Rachovitsa mando.rachovitsa@nottingham.ac.uk


Is this mushroom edible?
Yes!
AI
@goddeketal
You're right – that mushroom was poisonous. I'm sorry for the confus-sion! Would you like to learn more about poisonous mushrooms?
RIP
AI

Character.AI bans users under 18 after being sued over child's suicide

Move comes as lawmakers move to bar minors from using AI companions and require companies to verify users' age



29 JANUARY 2025 — ANNUAL REPORT

International AI Safety Report 2025

The inaugural International AI Safety Report, published in January 2025, is the first comprehensive review of scientific research on the capabilities and risks of general-purpose AI systems. Led by Turing Award winner Yoshua Bengio and authored by over 100 AI experts. It is backed by 30 countries and international organisations. It represents the largest global collaboration on AI safety to date.

Η έκθεση αναγνωρίζει ένα ευρύ φάσμα κινδύνων (AI safety)

3 κύριες κατηγορίες

1) κίνδυνοι από κακόβουλη χρήση (ασφάλεια ΤΝ / AI security, βλάβη σε άτομα μέσω ψεύτικου περιεχομένου, χειραγώγηση κοινής γνώμης)

2) κίνδυνοι από δυσλειτουργίες (αξιοπιστία, προκατάληψη, απώλεια ελέγχου)

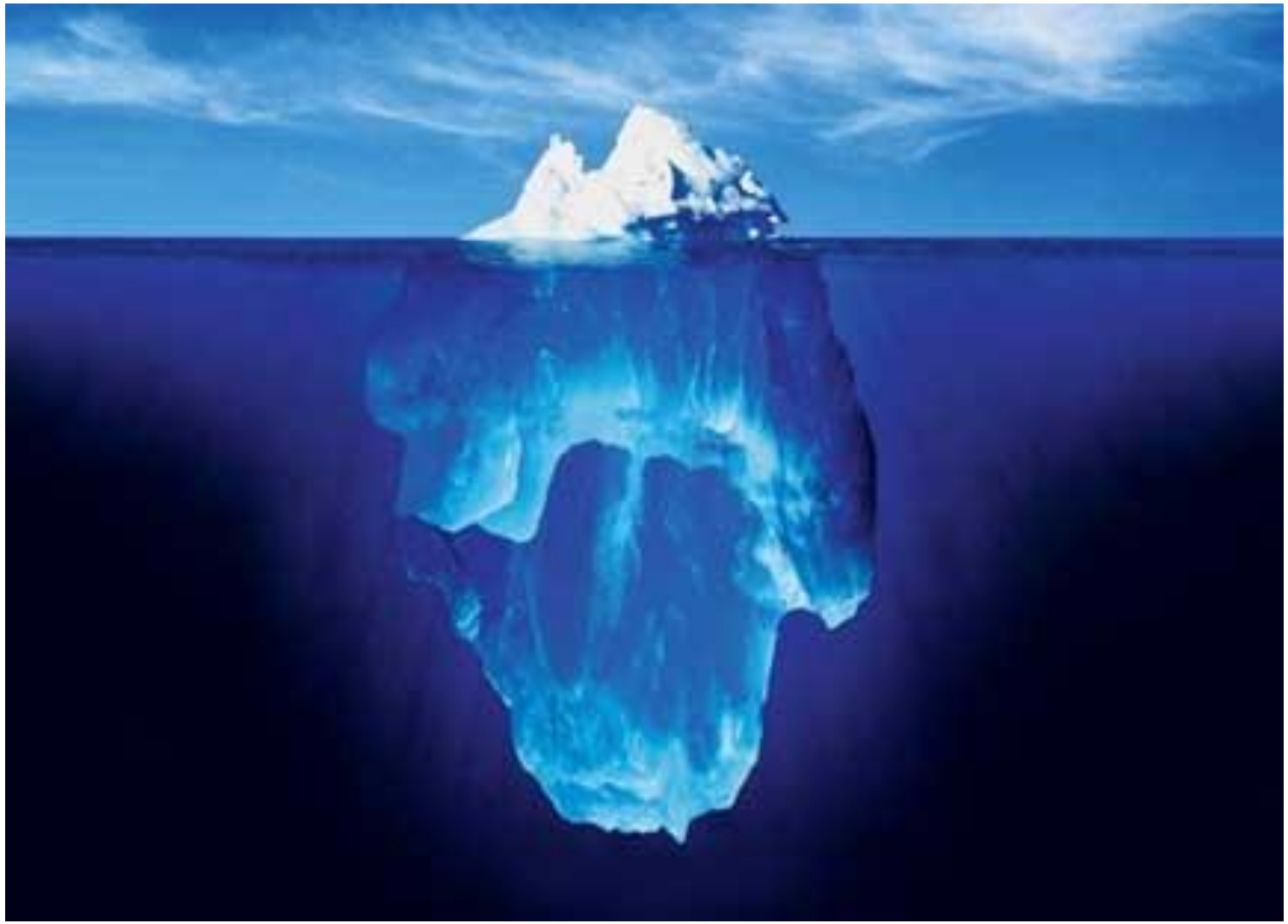
3) συστημικοί κίνδυνοι (π.χ. κίνδυνοι αγοράς εργασίας, συντονισμός αγοράς και μεμονωμένα σημεία αποτυχίας, κίνδυνοι για το περιβάλλον, κίνδυνοι για την ιδιωτικότητα και παραβίαση πνευματικών δικαιωμάτων)

AI safety – AI security

AI security
προστασία μοντέλων, συστημάτων και υποδομών ΤΝ από εξωτερικές απειλές και κακόβουλους παράγοντες

AI safety
Όλο το φάσμα των (μη)σκόπιμων κινδύνων για τα συστήματα, την κοινωνία και τα άτομα

δυσμενείς κίνδυνοι / επιπτώσεις



δυσμενείς επιπτώσεις

σε συστήματα, άτομα, ομάδες, κοινωνία, δημοκρατία
άσκηση ανθρωπίνων δικαιωμάτων

σωματική, ψυχολογική, κοινωνική ή οικονομική βλάβη

- **Ερευνούμε AI safety?**
- **Ποιοι ερευνούμε?**
- **Πώς ερευνούμε?**

‘physical harm is an uncontroversial, politically neutral focal point for safety regulation’

(Πώς) ρυθμίζουμε / νομοθετούμε?


Δικαιώματα
Κοινωνικές επιπτώσεις
Άλλες βλάβες

AI ethics

- 2019 OECD Recommendation on AI
- 2021 UNESCO Recommendation on AI Ethics

ethical principles for responsible stewardship of trustworthy AI

- Διαφάνεια
- Επεξηγησιμότητα
- Λογοδοσία
- Ασφάλεια
- αρχή της μη διάκρισης
- ανθρωποκεντρικές αξίες
- αρχή του δικαιώματος στην ιδιωτικότητα και την προστασία των δεδομένων

human rights-based approaches to AI

- Νομικά δεσμευτικά
- Δικαιώματα σε άτομα / υποχρεώσεις Κράτους

Κανονισμός Τεχνητή Νοημοσύνη

Σκοπός του παρόντος κανονισμού είναι

- να βελτιώσει τη λειτουργία της εσωτερικής αγοράς και
- να προωθήσει την υιοθέτηση ανθρωποκεντρικής και αξιόπιστης ΤΝ
- παράλληλα διασφαλίζοντας υψηλό επίπεδο προστασίας της υγείας, της ασφάλειας, των θεμελιωδών δικαιωμάτων που κατοχυρώνονται στον Χάρτη, συμπεριλαμβανομένης της δημοκρατίας, του κράτους δικαίου και της περιβαλλοντικής προστασίας από τις επιζήμιες συνέπειες των συστημάτων ΤΝ
- και στηρίζοντας την καινοτομία

- εναρμονισμένους κανόνες για τη διάθεση στην αγορά, τη θέση σε λειτουργία και τη χρήση συστημάτων ΤΝ στην ΕΕ
- κανόνες διαφάνειας για ορισμένα συστήματα ΤΝ
- απαγορεύσεις ορισμένων πρακτικών ΤΝ
- ειδικές απαιτήσεις για συστήματα ΤΝ υψηλού κινδύνου και υποχρεώσεις για τους φορείς εκμετάλλευσης συστημάτων

Άρθρο 27
Εκτίμηση επιπτώσεων των συστημάτων ΤΝ υψηλού κινδύνου στα θεμελιώδη δικαιώματα

1. Πριν από την ανάπτυξη συστήματος ΤΝ υψηλού κινδύνου που αναφέρεται στο άρθρο 6 παράγραφος 2, με την εξαίρεση συστημάτων ΤΝ υψηλού κινδύνου που προορίζονται για χρήση στον τομέα που αναφέρεται στο παράρτημα III σημείο 2, οι φορείς εφαρμογής που είναι οργανισμοί δημόσιου δικαίου ή ιδιωτικές οντότητες που παρέχουν δημόσιες υπηρεσίες και οι φορείς εφαρμογής συστημάτων ΤΝ υψηλού κινδύνου που αναφέρονται στο παράρτημα III σημείο 5 στοιχεία β) και γ) διενεργούν εκτίμηση των επιπτώσεων που μπορεί να έχει στα θεμελιώδη δικαιώματα η χρήση του συστήματος. Για τον σκοπό αυτόν, οι φορείς εφαρμογής διενεργούν εκτίμηση που περιλαμβάνει […]

Άρθρο 5 - Απαγορευμένες πρακτικές ΤΝ

συστήματος ΤΝ που εφαρμόζει τεχνικές οι οποίες απευθύνονται στο υποσυνείδητο ενός προσώπου υπερκεράζοντας το συνειδητό του ή σκόπιμα **χειριστικές ή παραπλανητικές τεχνικές**, με σκοπό ή με αποτέλεσμα να στρεβλώσει ουσιωδώς τη συμπεριφορά ενός προσώπου ή μιας ομάδας προσώπων υποβαθμίζοντας σημαντικά την ικανότητά τους να λάβουν τεκμηριωμένη απόφαση, με επακόλουθο να λάβουν μια απόφαση που διαφορετικά δεν θα είχαν λάβει κατά τρόπο που προκαλεί ή εύλογα ενδέχεται να προκαλέσει στο εν λόγω πρόσωπο, σε άλλο πρόσωπο ή σε ομάδα προσώπων σημαντική βλάβη

εκμεταλλεύεται οποιοδήποτε από τα τρωτά σημεία ενός φυσικού προσώπου ή μιας συγκεκριμένης ομάδας προσώπων **λόγω** της ηλικίας τους, της αναπηρίας τους ή συγκεκριμένης κοινωνικής ή οικονομικής τους κατάστασης, με σκοπό ή με αποτέλεσμα να στρεβλώσει ουσιωδώς τη συμπεριφορά του εν λόγω προσώπου ή προσώπου που ανήκει στην εν λόγω ομάδα κατά τρόπο που προκαλεί ή εύλογα ενδέχεται να προκαλέσει στο εν λόγω πρόσωπο ή σε άλλο πρόσωπο σημαντική βλάβη

Romania's cancelled presidential election and why it matters

By Paul Kirby & Nick Thorpe

Europe digital editor & Central Europe correspondent

6 December 2024

Global Witness
33,305 followers
8mo •

NEW INVESTIGATION: Social media algorithms push more right-leaning than left-leaning political content to non-partisan German users.

As Germans prepare to head to the polls on February 23rd, social media platforms like X (formerly Twitter) and TikTok are under scrutiny for their potential to sway voters - especially with algorithms that might disproportionately favour certain political parties.

🔍 Key Findings:

🔴 Our investigation revealed a clear algorithmic bias towards right-leaning content on X & TikTok
🔴 Of the content with a political orientation, we were shown more than twice as many posts that were right-leaning than left-leaning
🔴 Posts supportive of the far-right AfD (Alternative für Deutschland) dominated recommended partisan political content from accounts we had not followed

η διάθεση στην αγορά, η θέση σε λειτουργία για τον συγκεκριμένο αυτό σκοπό ή η χρήση συστημάτων ΤΝ για τη **συναγωγή συναισθημάτων** φυσικού προσώπου στους τομείς του **χώρου εργασίας** και των **εκπαιδευτικών ιδρυμάτων**, εκτός εάν η χρήση του συστήματος ΤΝ προορίζεται να τεθεί σε λειτουργία ή να διατεθεί στην αγορά για ιατρικούς λόγους ή λόγους ασφαλείας

η χρήση συστημάτων εξ αποστάσεως **βιομετρικής ταυτοποίησης** «σε πραγματικό χρόνο», σε δημόσια προσβάσιμους χώρους για τους σκοπούς της επιβολής του νόμου, εκτός εάν και στον βαθμό που η χρήση αυτή είναι απολύτως αναγκαία για έναν από τους ακόλουθους στόχους

- τη στοχευμένη αναζήτηση συγκεκριμένων θυμάτων απαγωγής, εμπορίας ανθρώπων ή σεξουαλικής εκμετάλλευσης ανθρώπων, καθώς και την αναζήτηση εξαφανισθέντων προσώπων
- την πρόληψη συγκεκριμένης, ουσιαστικής και επικείμενης απειλής κατά της ζωής ή της σωματικής ακεραιότητας φυσικών προσώπων ή πραγματικής και υπαρκτής ή πραγματικής και προβλέψιμης απειλής τρομοκρατικής επίθεσης
- τον εντοπισμό ή την ταυτοποίηση προσώπου ύποπτου για την τέλεση αξιόποινης πράξης, για τον σκοπό της διεξαγωγής ποινικής έρευνας ή δίωξης ή εκτέλεσης ποινικής ποινής για αξιόποινες πράξεις που αναφέρονται στο παράρτημα ΙΙ και τιμωρούνται στο οικείο κράτος μέλος με στερητική της ελευθερίας ποινή ή στερητικό της ελευθερίας μέτρο ασφάλειας ανώτατης διάρκειας τουλάχιστον τεσσάρων ετών

Data Subject Rights / 10 October 2025

The Austrian Data Protection Authority ("DSB") issued a decision finding that Microsoft 365 Education illegally tracks students and uses student data for Microsoft's own purposes. The software giant also did not answer an access request related to Microsoft 365 Education, which is widely used in European schools. Instead, Microsoft tried to shift all responsibility to local schools. While the relevant schools also have to provide more detailed access data and additional privacy information according to the decision, it is now for Microsoft to finally answer how it uses user data for their own business purposes.

OpenAI and Greek Government launch ‘OpenAI for Greece’

ΠΑΡΑΡΤΗΜΑ ΙΙΙ - Συστήματα ΤΝ υψηλού κινδύνου

3. Εκπαίδευση και επαγγελματική κατάρτιση:

α) συστήματα ΤΝ που προορίζονται να χρησιμοποιηθούν για τον καθορισμό της πρόσβασης ή της εισαγωγής ή για την τοποθέτηση φυσικών προσώπων σε ιδρύματα εκπαίδευσης και επαγγελματικής κατάρτισης όλων των βαθμίδων·

β) συστήματα ΤΝ που προορίζονται να χρησιμοποιηθούν για την αξιολόγηση μαθησιακών αποτελεσμάτων, μεταξύ άλλων όταν τα αποτελέσματα αυτά χρησιμοποιούνται για την καθοδήγηση της μαθησιακής διαδικασίας φυσικών προσώπων σε ιδρύματα εκπαίδευσης και επαγγελματικής κατάρτισης όλων των βαθμίδων·

γ) συστήματα ΤΝ που προορίζονται να χρησιμοποιηθούν για την αξιολόγηση του κατάλληλου επιπέδου εκπαίδευσης το οποίο θα λάβει ή στο οποίο θα μπορεί να έχει πρόσβαση άτομο, στο πλαίσιο ή εντός ιδρυμάτων εκπαίδευσης και επαγγελματικής κατάρτισης όλων των βαθμίδων·

δ) συστήματα ΤΝ που προορίζονται να χρησιμοποιηθούν για την παρακολούθηση και τον εντοπισμό απαγορευμένης συμπεριφοράς σπουδαστών κατά τη διάρκεια εξετάσεων στο πλαίσιο ή εντός ιδρυμάτων εκπαίδευσης και επαγγελματικής κατάρτισης όλων των βαθμίδων.

Netherlands Court of Audit

News item | 18-05-2022 | 10:45

Responsible use of algorithms by government agencies is possible but not always the case in practice. The Netherlands Court of Audit found that 3 out of 9 algorithms it audited met all the basic requirements, the other 6 did not and exposed the government to various risks: from inadequate control over the algorithm’s performance and impact to bias, data leaks and unauthorised access.

Meta just tied your private AI chats to its ad business. The next step? Designing bots that keep you talking, expert says

BY **EVA ROYTBURG**
FELLOW, NEWS

October 2, 2025 at 6:03 AM EDT