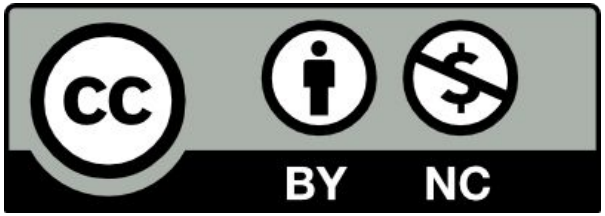
Η Τεχνητή Νοημοσύνη στον κόσμο.
Εξελίξεις και προκλήσεις

Νικόλαος Παπαχρήστου, ΥΠΑΙΘΑ
Site: https://nikpapa.com
Email: nikpapa@gmail.com



Περίγραμμα

1. Σύντομη ιστορική αναδρομή και ορισμοί
2. Εξελίξεις και Τάσεις
3. Προκλήσεις και Κίνδυνοι
4. Συμπεράσματα

Περίγραμμα

1. Σύντομη ιστορική αναδρομή και ορισμοί
2. Εξελίξεις και Τάσεις
3. Προκλήσεις και Κίνδυνοι
4. Συμπεράσματα

Τεχνητή Νοημοσύνη: Κλασσικοί Ορισμοί

- **Russell and Norvig (2010)**: ΤΝ είναι η μελέτη [ευφυών] πρακτόρων που λαμβάνουν εντολές από το περιβάλλον και αναλαμβάνουν δράση. Κάθε τέτοιος πράκτορας υλοποιείται από μια συνάρτηση που αντιστοιχίζει τα ερεθίσματα σε ενέργειες και καλύπτουμε διαφορετικούς τρόπους αναπαράστασης αυτών των λειτουργιών, όπως συστήματα παραγωγής, αντιδραστικοί παράγοντες, λογικοί σχεδιαστές, νευρωνικά δίκτυα και συστήματα θεωρίας αποφάσεων
- **Chatgpt:** Η ΤΝ αναφέρεται στη δημιουργία υπολογιστικών συστημάτων που εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη. Συγκεκριμένα, στοχεύει στη δημιουργία προγραμμάτων και μηχανισμών που μπορούν να αντιλαμβάνονται, να μαθαίνουν, να λύνουν προβλήματα και να λαμβάνουν αποφάσεις σε περιβάλλοντα που προηγουμένως απαιτούσαν ανθρώπινη επέμβαση. Στόχος της είναι να αναπτύσσει συστήματα που μπορούν να εκτελούν εργασίες με αυξημένη αυτονομία και ευφυία, μιμούμενη σε μεγάλο βαθμό την ανθρώπινη νοημοσύνη.

ΚΑΤΗΓΟΡΙΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

- **Machine Learning (ML)**
- Natural Language Processing (NLP)
- Computer Vision
- Robotics
- Expert Systems
- Knowledge Representation and Reasoning
- Planning and Scheduling
- Search
- Human-AI Interaction (HAI)
- AI Ethics and Fairness

ΚΑΤΗΓΟΡΙΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

- **Supervised Learning**
- Unsupervised Learning
- Semi-Supervised Learning
- Reinforcement Learning
- NN & Deep Learning
- Transfer Learning
- AutoML (Automated Machine Learning)

Κατηγοριοποίηση συστημάτων ΤΝ με βάση την αυτονομία

Autonomy Level	Example Systems	Example Risks
Level 0: No AI	typing in a text editor	-
Level 1: AI as a Tool	Reading a sign with a machine translation app	• de-skilling • Industry disruption
Level 2: AI as a Consultant	Relying on a language model to summarize a set of documents	• over-trust • radicalization • manipulation
Level 3: AI as a Collaborator	Training as a chess player through interactions with and analysis of a chess AI	• anthropomorphization • rapid societal change
Level 4: AI as an Expert	Using an AI system to advance scientific discovery (e.g., protein folding)	• mass labor displacement • decline of human exceptionalism
Level 5: AI as an Agent *fully autonomous AI*	Autonomous AI-powered personal assistants *(not yet unlocked)*	• misalignment • concentration of power

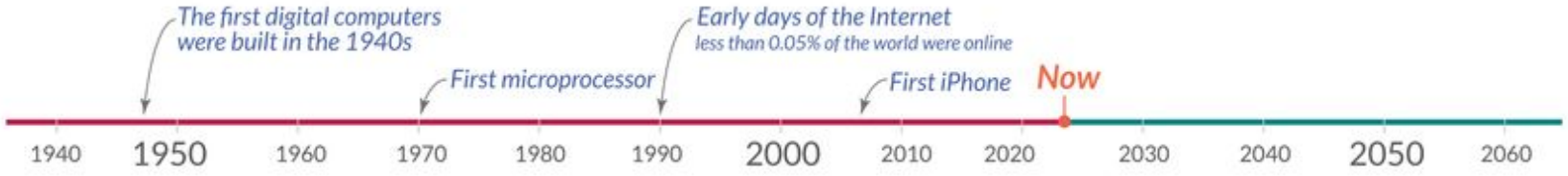
Πηγή: Morris, Meredith Ringel, et al. "Levels of AGI: Operationalizing Progress on the Path to AGI." arXiv preprint arXiv:2311.02462 (2023)

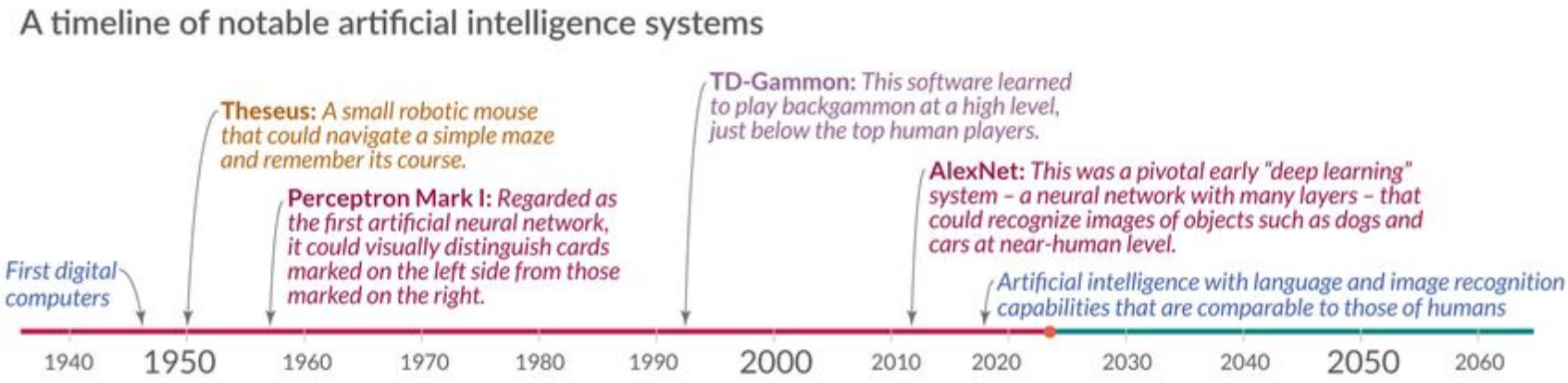
Κατηγοριοποίηση συστημάτων ΤΝ με βάση την απόδοση

Performance (rows) x **Generality** (columns)	**Narrow**	**General**
Level 0: No AI	**Narrow Non-AI** • calculator software • compiler	**General Non-AI** • Amazon Mechanical Turk
Level 1: Emerging *equal to or better than an unskilled human*	**Emerging Narrow AI** simple rule-based systems, e.g. SHRDLU	**Emerging AGI** • ChatGPT, Bard, Llama 2
Level 2: Competent *at least 50th percentile of skilled adults*	**Competent Narrow AI** • Siri, Alexa • short essay w chatgpt	**Competent AGI** not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	**Expert Narrow AI** • spelling & grammar checkers • Imagen, Dall-E 2	**Expert AGI** not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	**Virtuoso Narrow AI** • Deep Blue, AlphaGo	**Virtuoso AGI** not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	**Superhuman Narrow AI** • AlphaFold • AlphaZero, StockFish	**Artificial Superintelligence (ASI)** not yet achieved

Πηγή: Morris, Meredith Ringel, et al. "Levels of AGI: Operationalizing Progress on the Path to AGI." arXiv preprint arXiv:2311.02462 (2023)

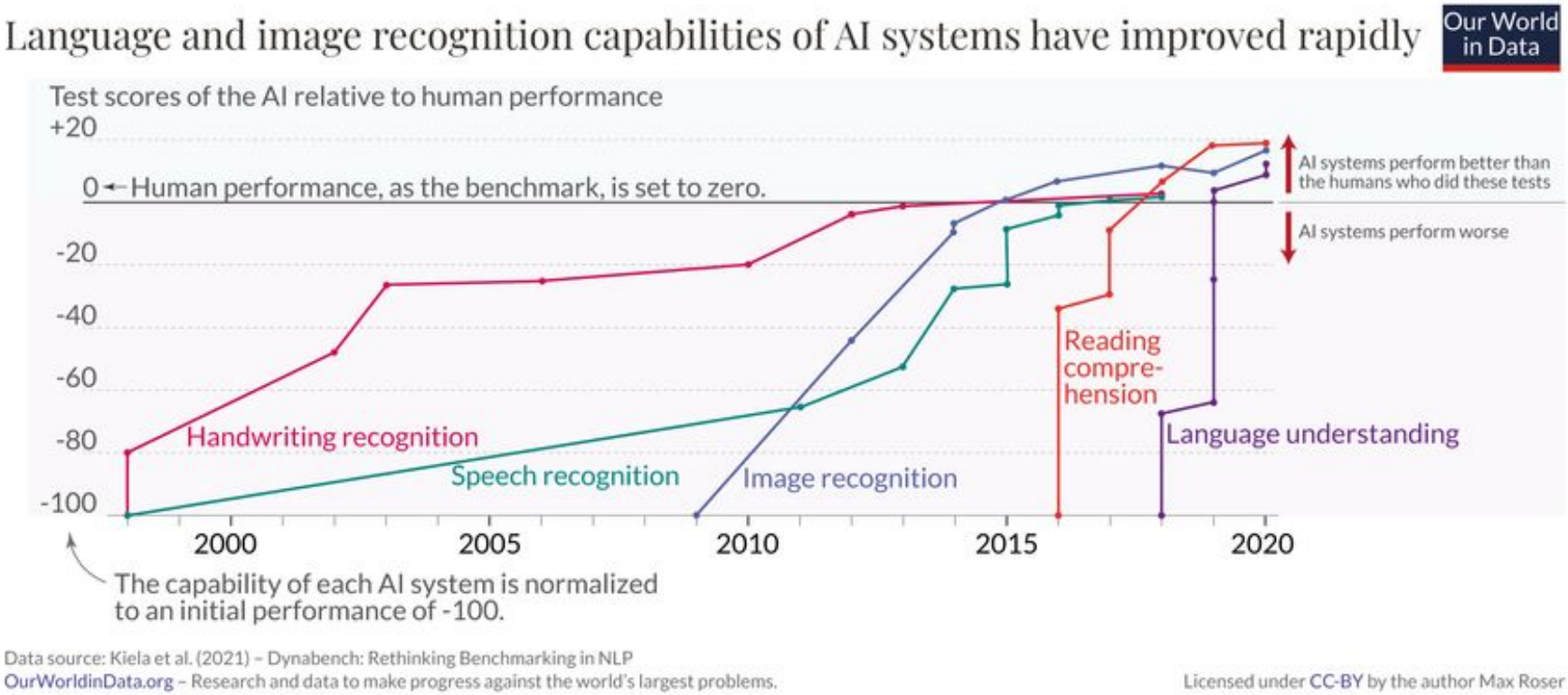
Σύντομη ιστορική αναδρομή





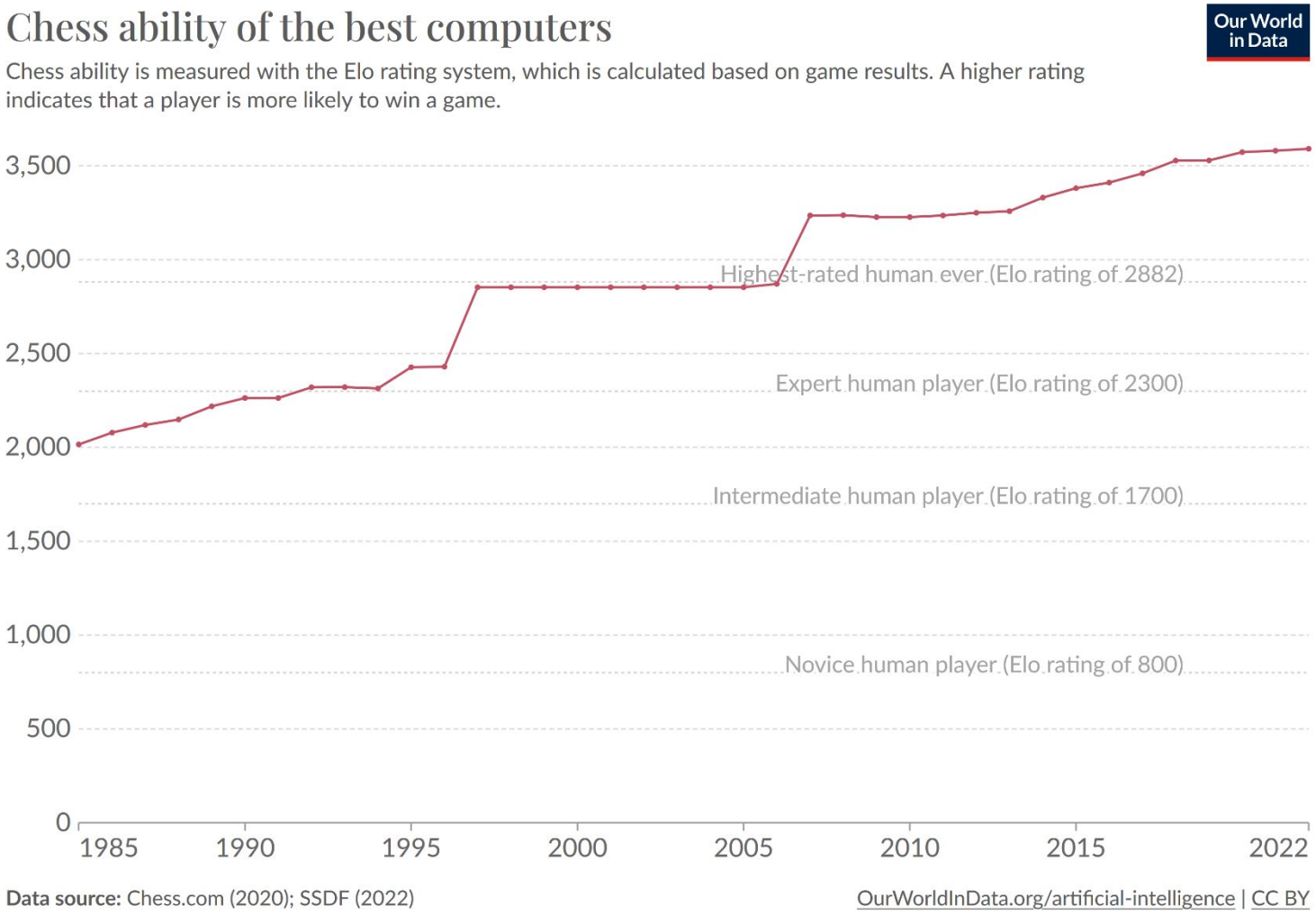
Πηγή: OurWorldInData.org/brief-history-of-ai

Απόδοση συστημάτων ΤΝ



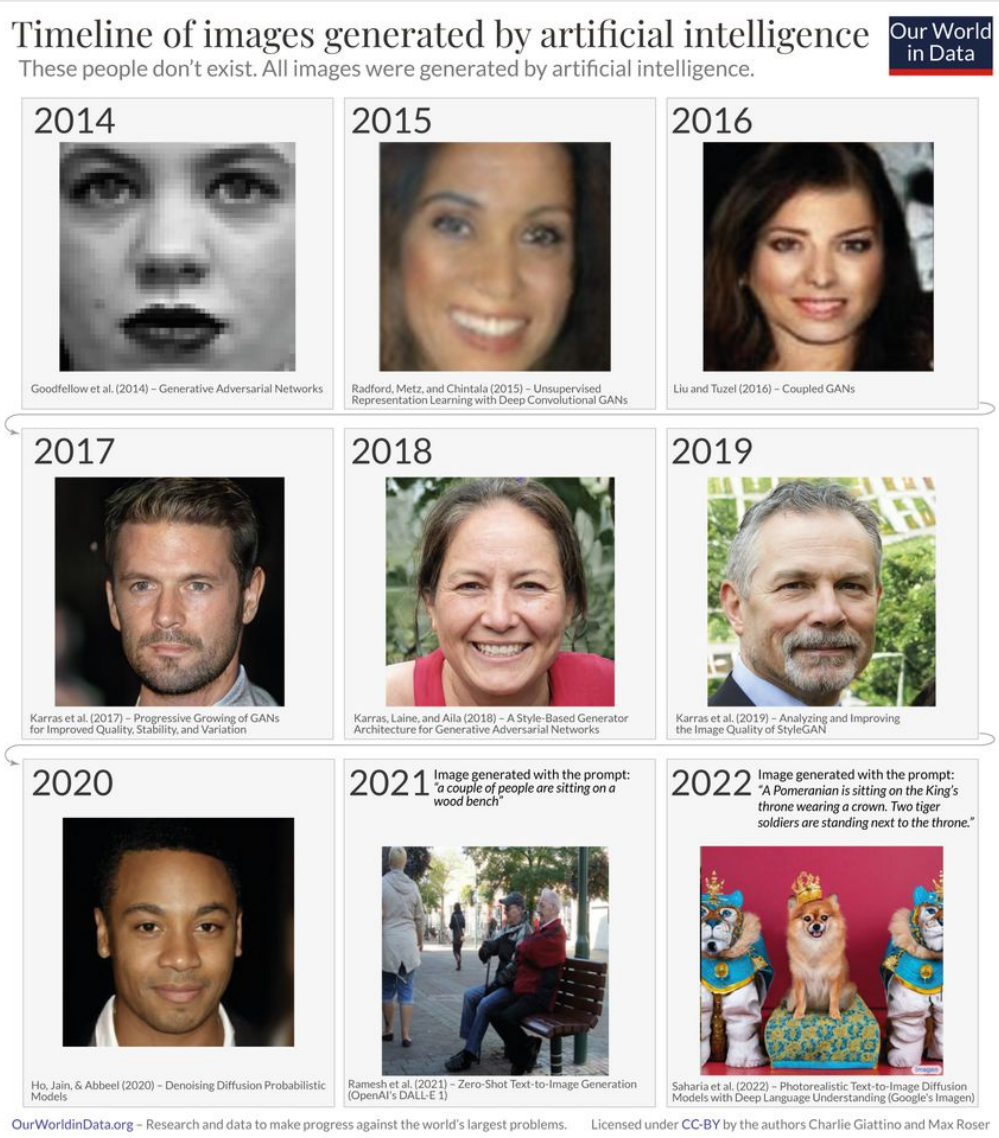
Πηγή: OurWorldInData.org/artificial-intelligence

Απόδοση συστημάτων ΤΝ στο σκάκι



Πηγή: OurWorldInData.org/artificial-intelligence

Δημιουργία εικόνων στο χρόνο



Σύντομη Ιστορική αναδρομή και ορισμοί

Ερωτήσεις;

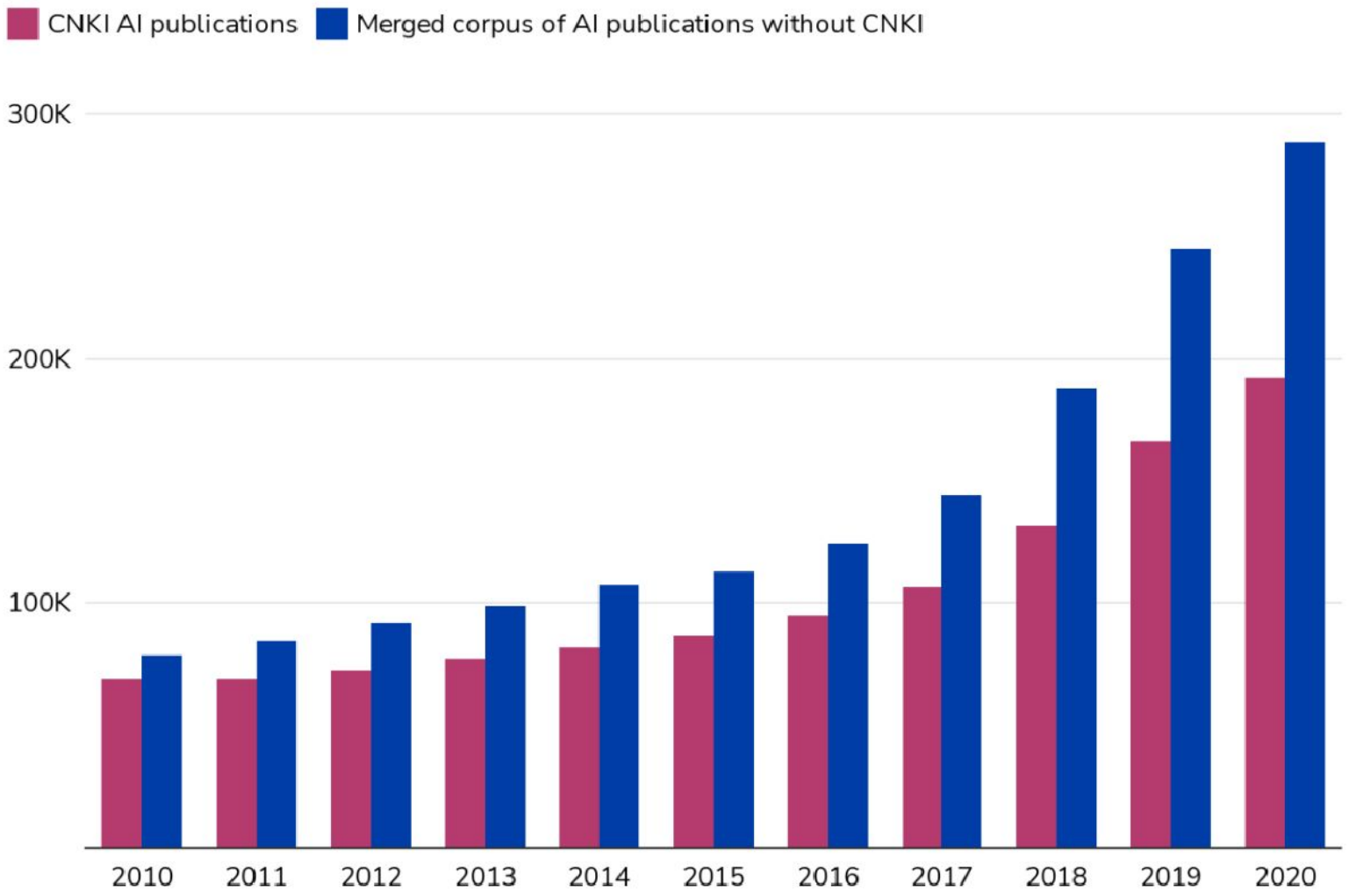
Outline

1. Σύντομη ιστορική αναδρομή και ορισμοί
2. Εξελίξεις και Τάσεις
3. Προκλήσεις και Κίνδυνοι
4. Συμπεράσματα

Εξελίξεις και Τάσεις

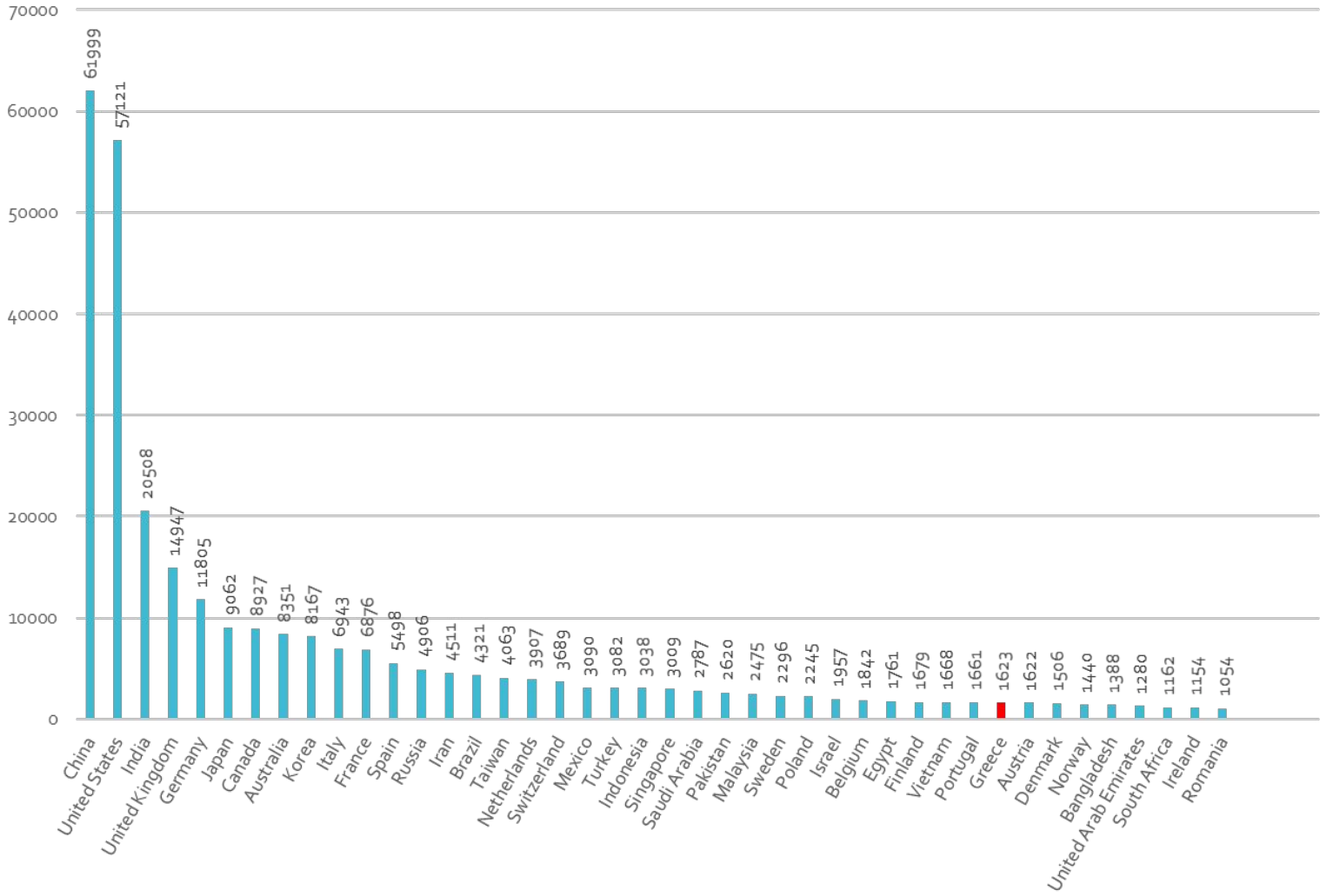
- Έρευνα πάνω στην ΤΝ
- Οικονομία
  - Ιδιωτικός Τομέας
  - Αγορά Εργασίας
- Δημόσια Διοίκηση
  - Νομοθεσία
  - Εθνική στρατηγική
  - Έργα ΤΝ
- Κοινή γνώμη

Δημοσιεύσεις έρευνας ΤΝ παγκοσμίως



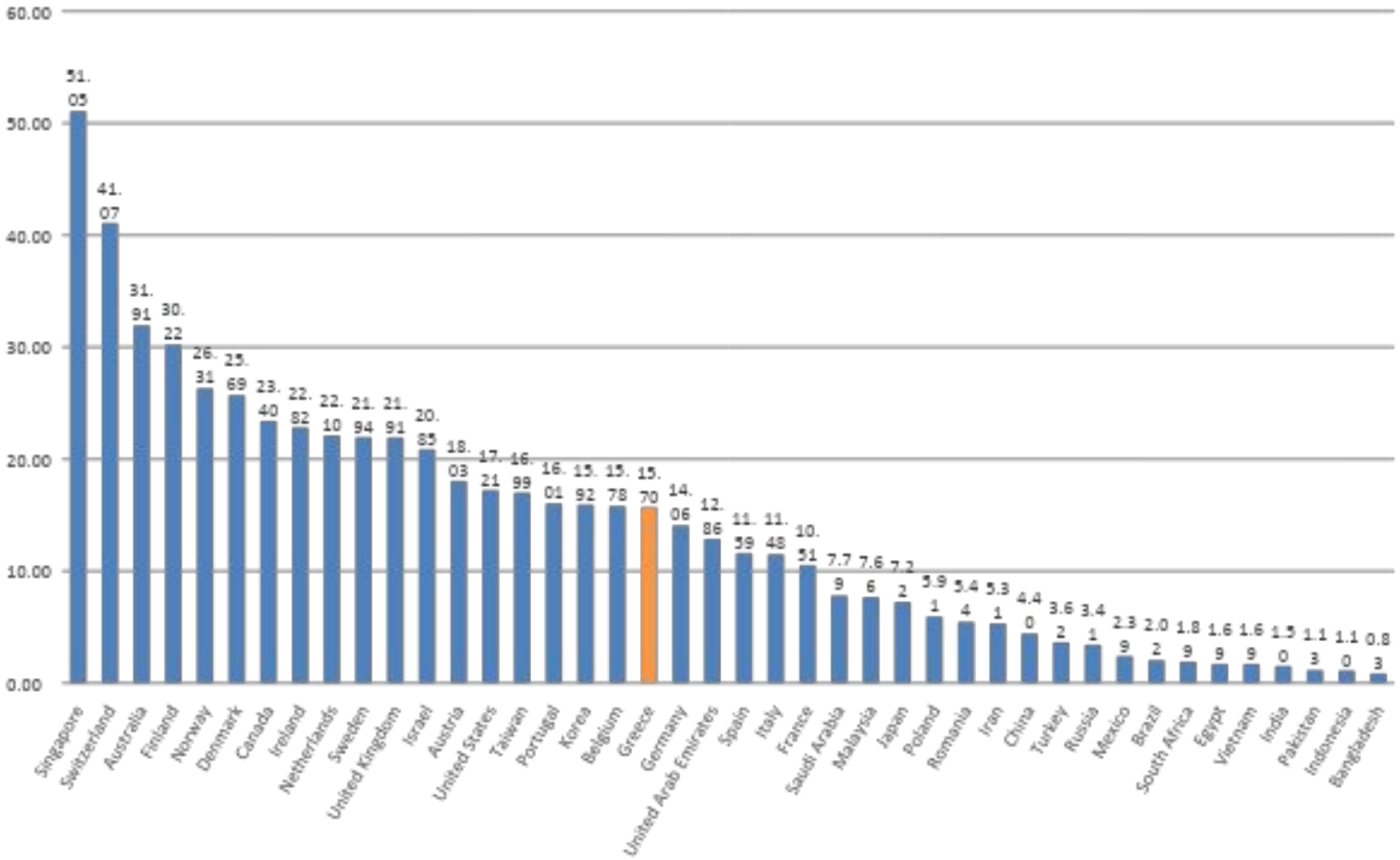
Πηγή: Identifying AI Research, CSET (2023)

Δημοσιεύσεις στην ΤΝ ανά χώρα (2020)



Πηγή: Identifying AI Research, CSET (2023)

Δημοσιεύσεις ανά 100000 κατοίκους (2020)



Πηγή: Identifying AI Research, CSET (2023)

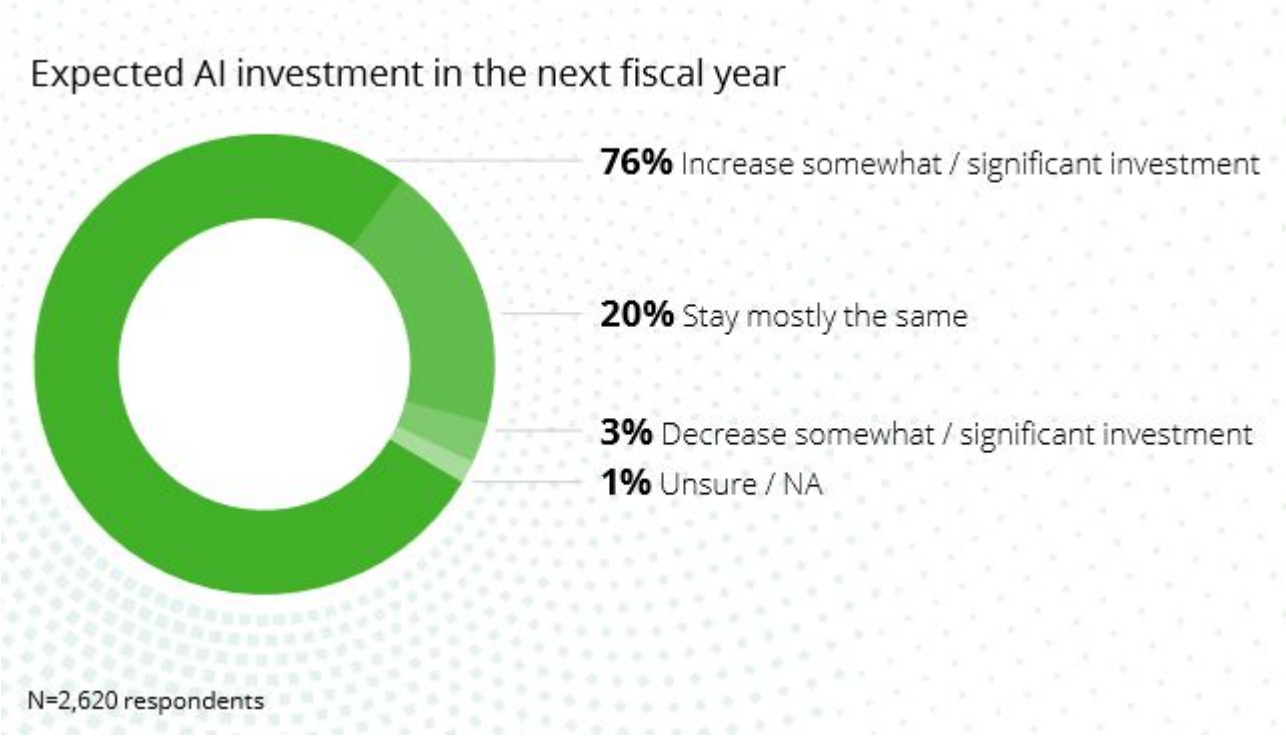
Οικονομία – Ιδιωτικός Τομέας

Το 94% των επιχειρήσεων που συμμετείχαν στην έρευνα συμφωνούν ότι η τεχνητή νοημοσύνη είναι κρίσιμη για την επιτυχία στα επόμενα πέντε χρόνια.

Πηγή: Deloitte's State of AI in the Enterprise, 5th Edition report (October 2022)

Οικονομία – Ιδιωτικός Τομέας

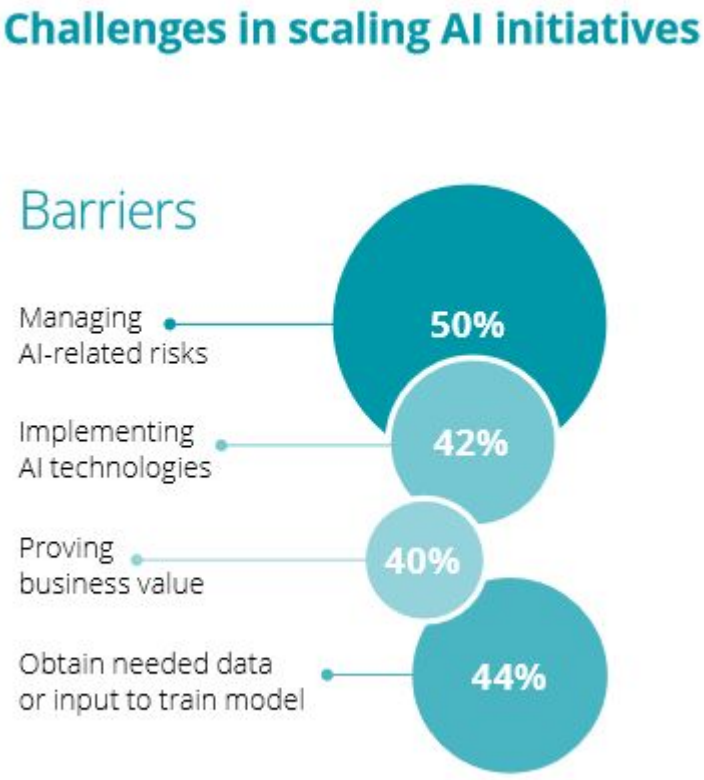
Το 79% των επιχειρήσεων που ερωτήθηκαν ανέφεραν ανάπτυξη πλήρους κλίμακας για τρεις ή περισσότερους τύπους των εφαρμογών AI—από 62% πέρυσι.



Πηγή: Deloitte's State of AI in the Enterprise, 5th Edition report (October 2022)

Οικονομία – Ιδιωτικός Τομέας

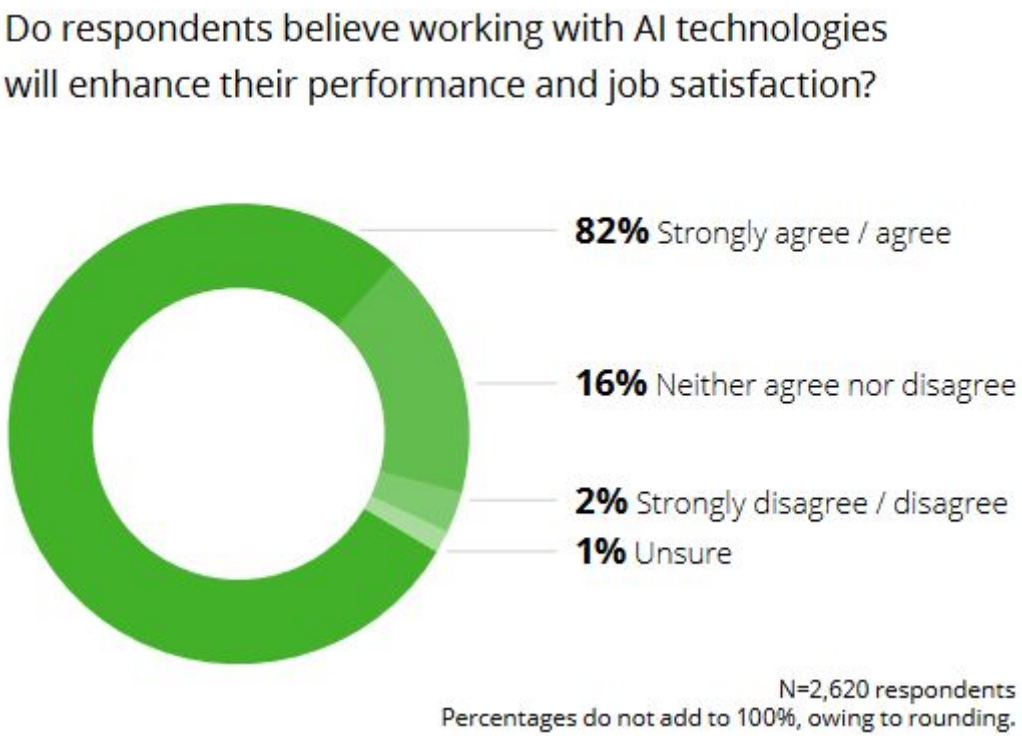
Κύρια εμπόδια για την εφαρμογή της Τεχνητής Νοημοσύνης



Πηγή: Deloitte's State of AI in the Enterprise, 5th Edition report (October 2022)

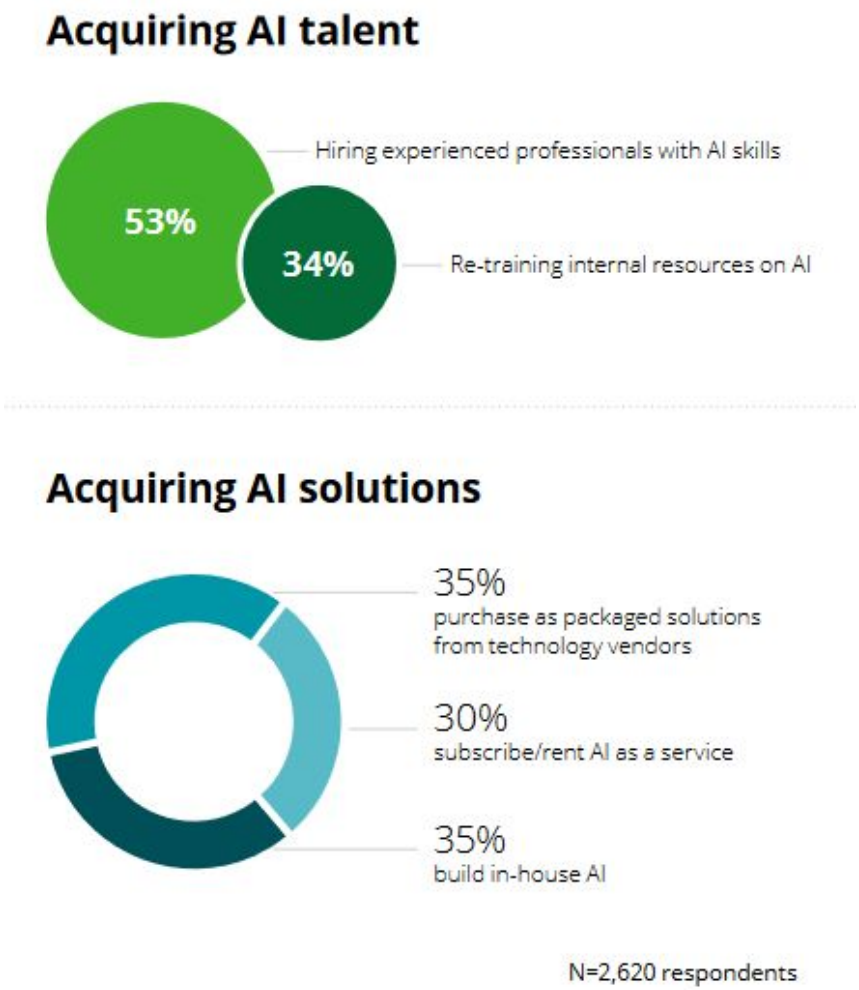
Οικονομία – Ιδιωτικός Τομέας

82% των ερωτηθέντων ανέφεραν ότι οι υπάλληλοί τους πιστεύουν ότι η εργασία με τεχνολογίες AI θα ενισχύσει την απόδοσή τους και την εργασιακή τους ικανοποίηση



Πηγή: Deloitte's State of AI in the Enterprise, 5th Edition report (October 2022)

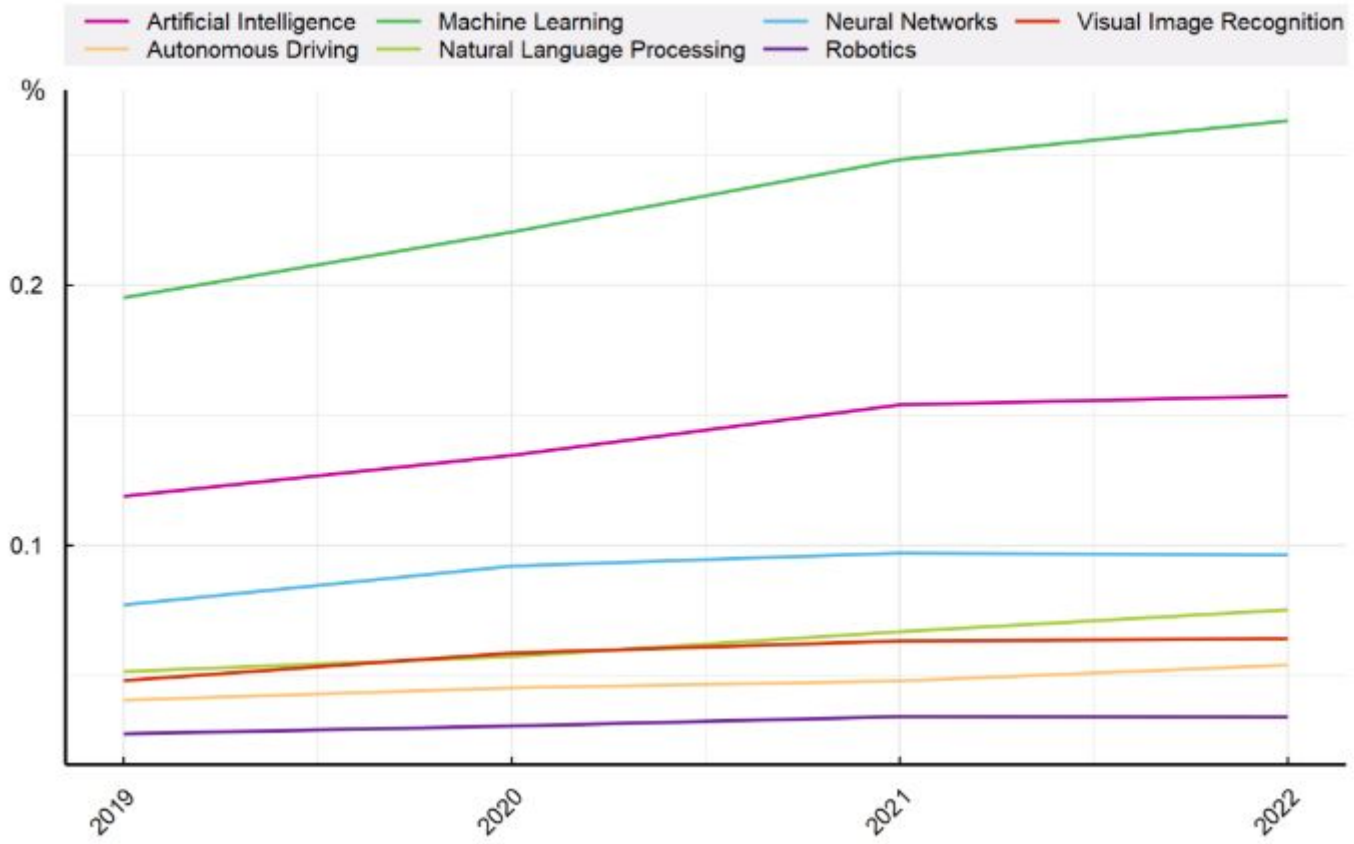
Οικονομία – Ιδιωτικός Τομέας



Πηγή: Deloitte's State of AI in the Enterprise, 5th Edition report (October 2022)

Οικονομία – Αγορά εργασίας

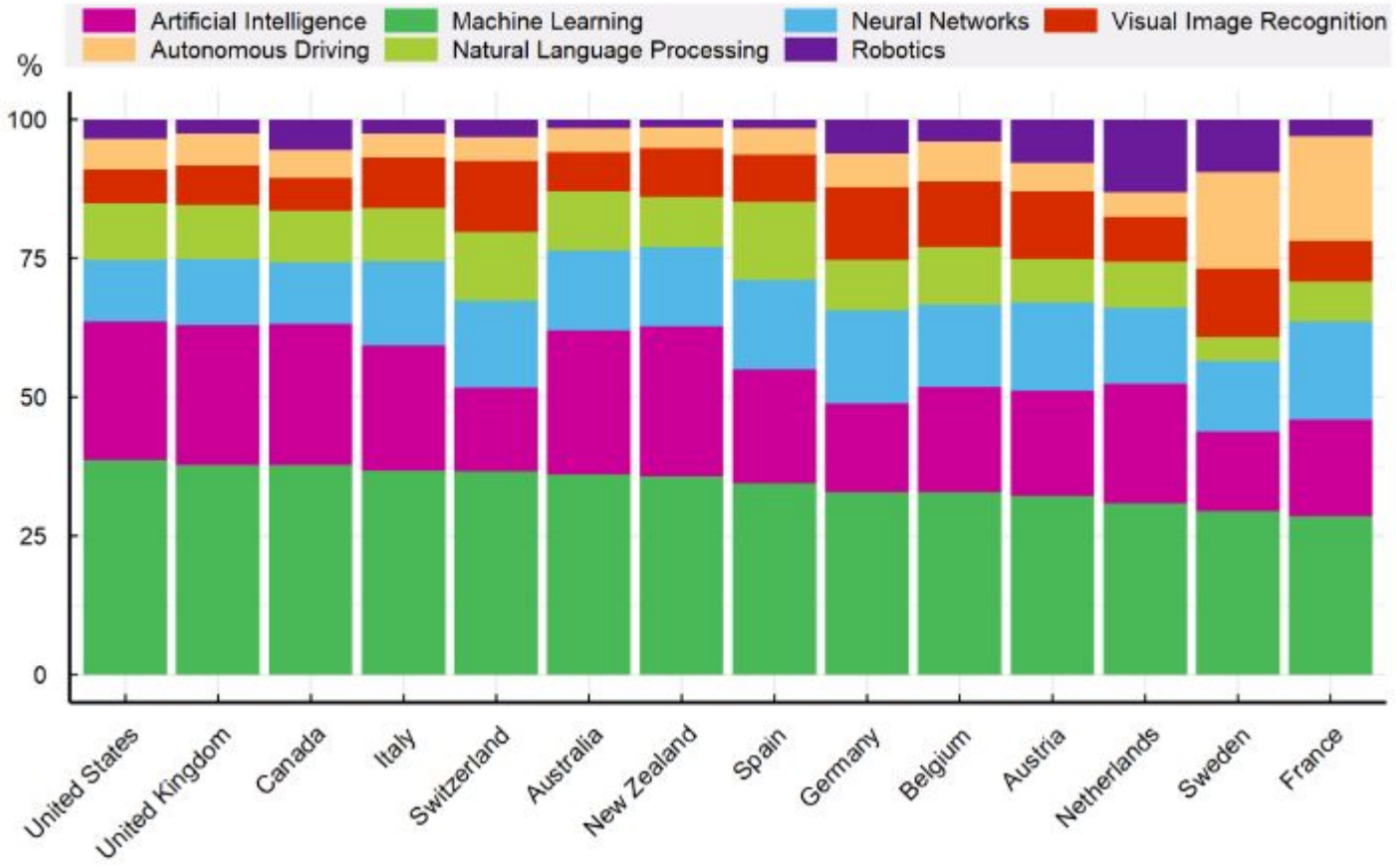
Ποσοστό κενών θέσεων στο διαδίκτυο που απαιτούν δεξιότητες τεχνητής νοημοσύνης, ανά ομάδα δεξιοτήτων και έτος



Πηγή: Emerging Trends in AI Skill Demand Across 14 OECD Countries (2023)

Οικονομία – Αγορά εργασίας

Το ποσοστό των κενών θέσεων στο διαδίκτυο που απαιτούν δεξιότητες τεχνητής νοημοσύνης κατά μέσο όρο την περίοδο 2019-22, ανά ομάδα δεξιοτήτων και χώρα



Πηγή: Emerging Trends in AI Skill Demand Across 14 OECD Countries (2023)

Οικονομία – Αγορά εργασίας

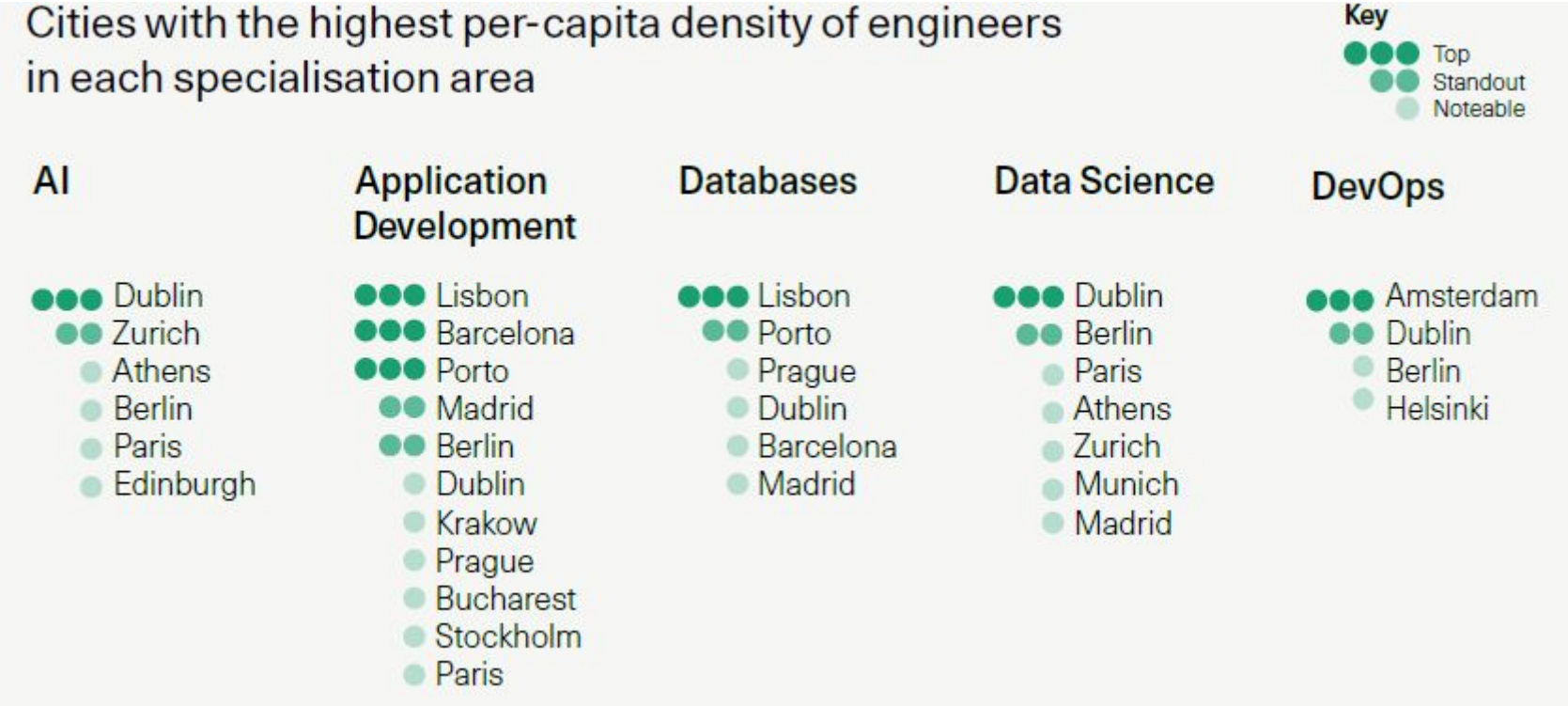
Ποσοστό διαδικτυακών κενών θέσεων τεχνητής νοημοσύνης από κορυφαίους και μη εργοδότες τεχνητής νοημοσύνης που απαιτούν τεχνικές δεξιότητες, Ηνωμένες Πολιτείες

Technical skill	Top AI	Other AI	Non-AI
Python (Programming Language)	43.6	46.3	1.3
Computer Science	40.0	39.1	3.0
Amazon Web Services	25.2	20.6	1.0
Data Science	24.8	27.6	0.0
SQL (Programming Language)	24.7	26.0	1.7
Data Analysis	24.5	22.9	2.7
Automation	21.0	20.1	2.3
Agile Methodology	21.0	20.7	2.6
Software Engineering	20.9	18.3	1.2
Java (Programming Language)	20.2	17.5	1.0
R (Programming Language)	18.6	18.4	0.0
Microsoft Azure	16.1	12.7	0.4
Algorithms	15.8	18.1	0.0
Software Development	15.2	16.1	1.2
Big Data	15.0	13.5	0.0
Scalability	14.0	13.2	0.2
Statistics	13.7	12.9	0.0
C++ (Programming Language)	11.6	13.4	0.0
Apache Hadoop	10.0	7.2	0.0
Business Intelligence	9.3	5.0	0.0

Πηγή: Emerging Trends in AI Skill Demand Across 14 OECD Countries (2023)

Οικονομία – Αγορά εργασίας

Πόλεις με την υψηλότερη κατά κεφαλήν πυκνότητα μηχανικών τεχνολογίας σε κάθε τομέα εξειδίκευσης

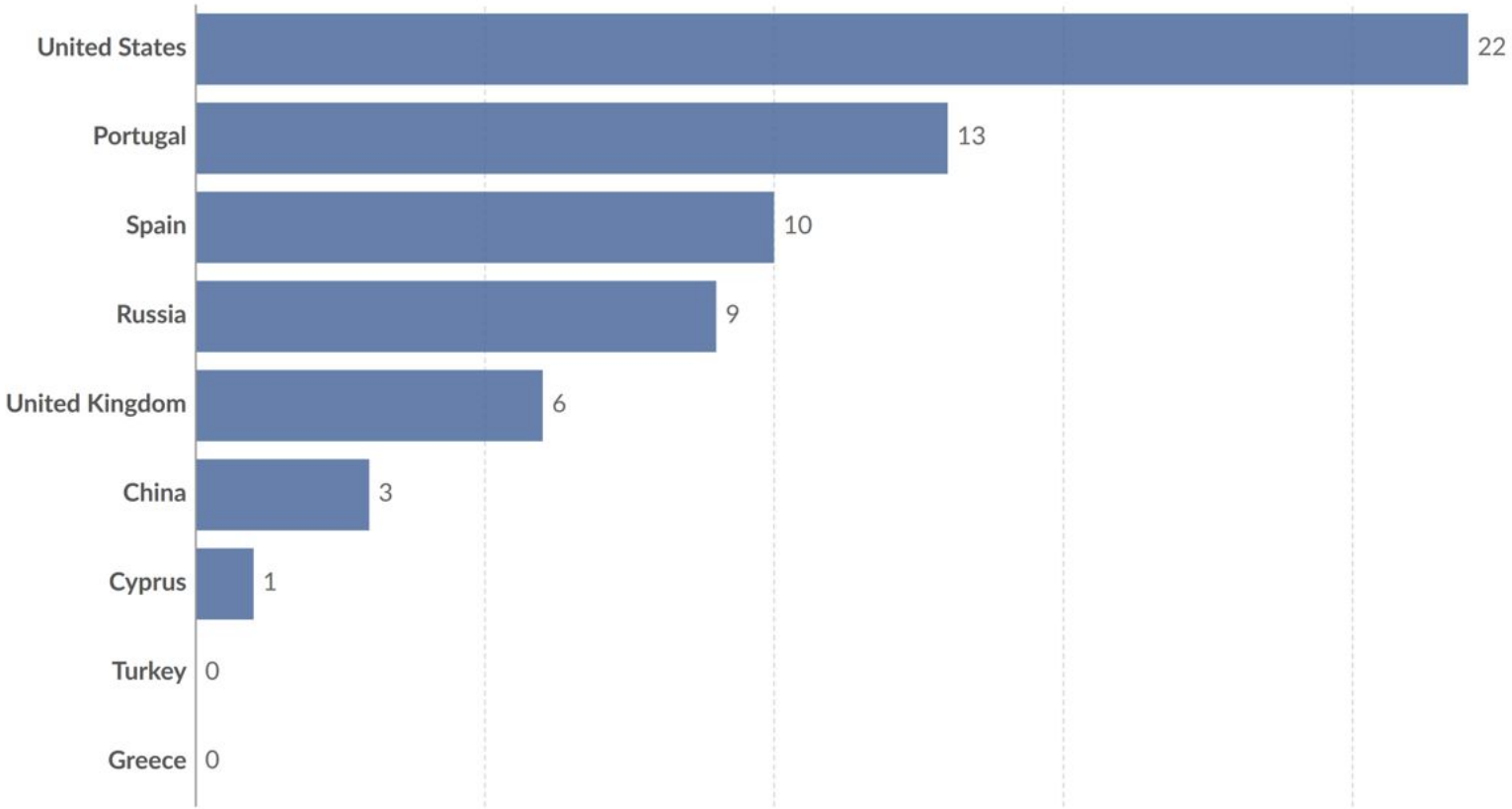


13% όλων των μηχανικών τεχνολογίας στην Αθήνα ειδικεύονται στην ΤΝ!

Πηγή: Atlas: Sequoia's interactive guide to Europe's technical talent (June 2023). https://atlas.sequoiacap.com/

Δημόσιος Τομέας - Νομοθεσία

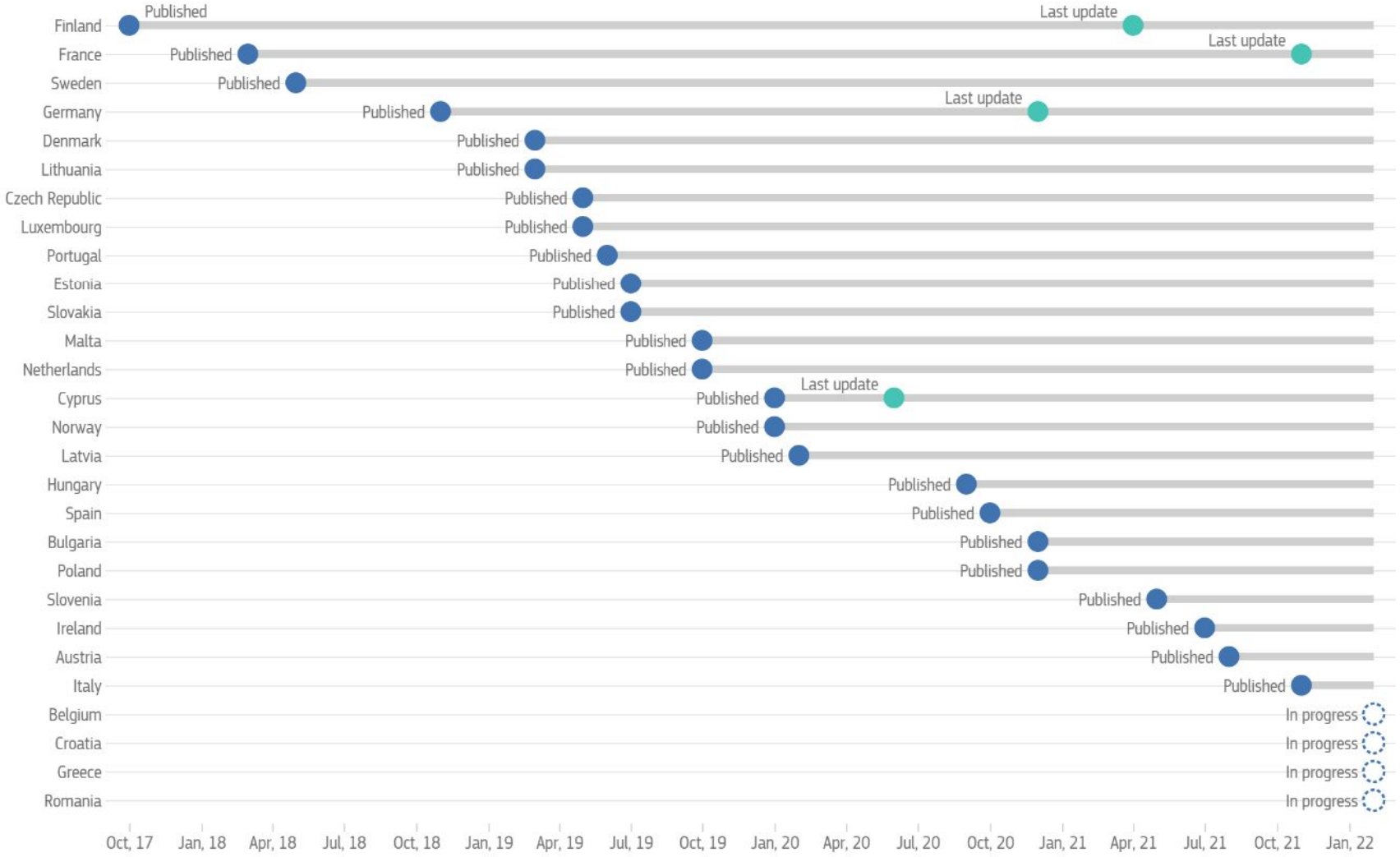
Νομοσχέδια* που σχετίζονται με την ΤΝ (έως το 2022)



* Στην Ελλάδα υπάρχει ένας νόμος για την ΤΝ (Νοέμβριος 2022)

Πηγή: https://ourworldindata.org/grapher/cumulative-number-artificial-intelligence-bills-passed

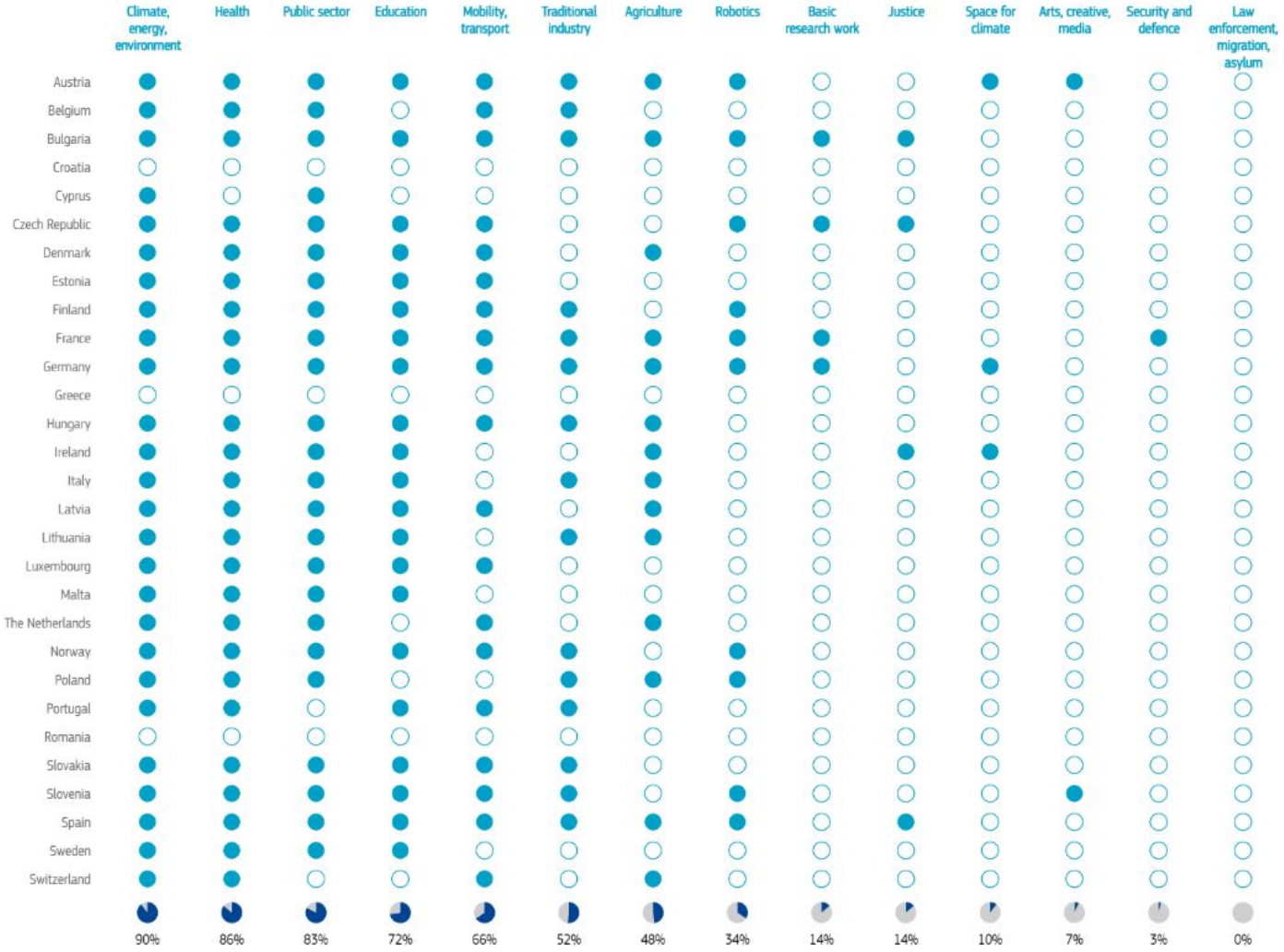
Δημόσιος Τομέας – Εθνικές Στρατηγικές



Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

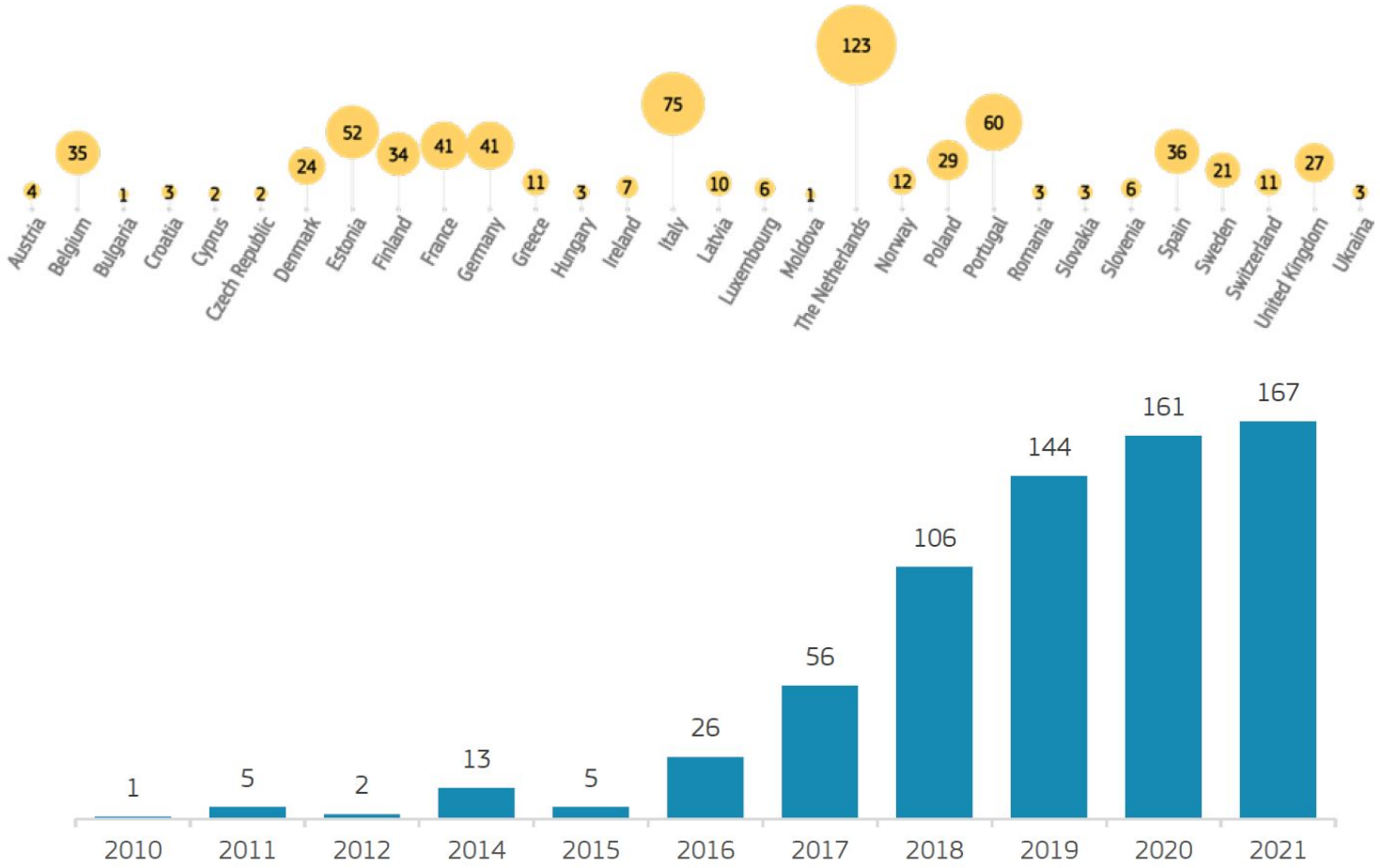
Δημόσιος Τομέας – Εθνικές Στρατηγικές

Σύγκριση τομέων προτεραιότητας με βάση την εθνική στρατηγική ΤΝ κάθε χώρας



Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

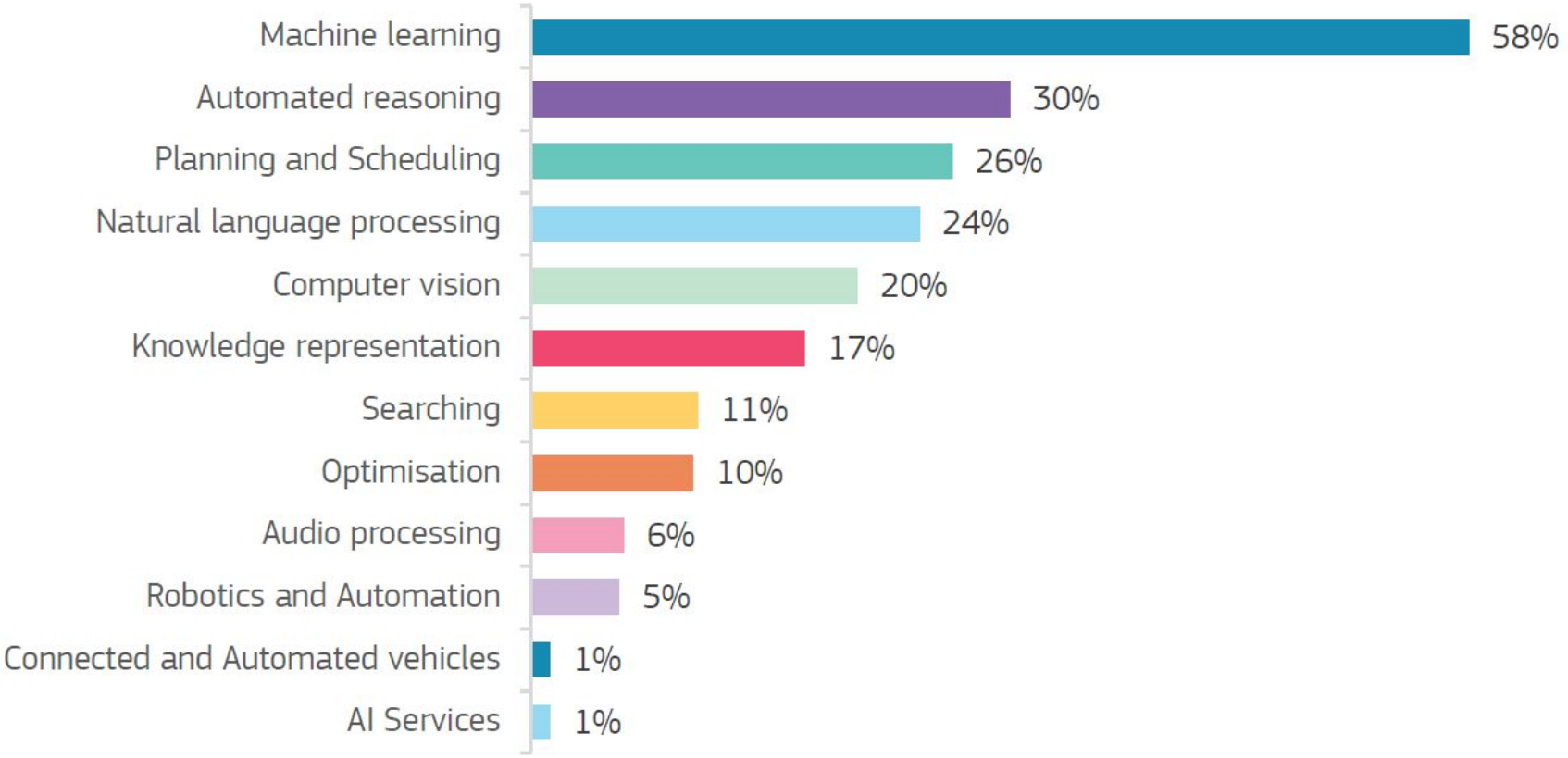
Δημόσιος Τομέας – Έργα* ΤΝ στην ΕΕ



* περιλαμβάνονται έργα μη λειτουργικά και υπό σχεδιασμό

Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

Δημόσιος Τομέας – Έργα* στην ΕΕ ανά τεχνολογία ΤΝ



* περιλαμβάνονται έργα μη λειτουργικά και υπό σχεδιασμό

Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

Δημόσιος Τομέας – Ενδεικτικά Έργα ΤΝ

#	Case	Country	Short description	Status	AI Classification
1	Intelligent Control Platform	Denmark	A digital platform that provides an automated assessment of how a selected company/businesses is more likely to commit fraud compared with others.	Implemented	Machine learning
2	Eva, targeted COVID-19 border checking	Greece	Between August and November 2020, in the midst COVID-19 crisis, the Greek Government trialled an AI system in border control points that helps the selection of travellers to test upon arrival. The purpose was to effectively allocate scarce PCR tests during the summer tourism season.	No longer in use	Machine learning
3	Reducing night noise through nudging	Belgium	To solving an issue of too much noise in crowded streets sound meters were installed and an application for citizens reporting developed. This will allow proper corrective actions, also through nudging.	In development	Audio processing
4	Unlocking digitised documents and correcting OCR	Luxemburg	The Luxemburg National Library developed an AI system that operates on top of the results of the different OCR (Optical Character Recognition) used over the years for digitising historical newspapers and books. The system aims at improving the quality of the result, identifying and correcting mistakes.	Implemented	Computer vision
5	Object Detection Kit	The Netherlands	The AI solution automatically identifies rubbish on the street and shares this with the garbage management services of the city to act and solve the issue. This is done by analysing imagery collected from the pictures taken by smartphones installed into vehicles driving around in the city.	No longer in use	Computer vision
6	OTT – decision-support tool for consultants	Estonia	An AI system used in the Estonian Unemployment Insurance Fund which aims to assist its consultants with providing insights predicting the chances of an unemployed person getting a new job.	Implemented	Machine Learning
7	Automation of subtitling videos and audios	Finland	The AI system is based on understanding speech and transforming it into text. It is used to provide subtitles on videos and is part of a wider initiative within the administration to use Speech-to-Text technologies in various use cases.	Implemented	Audio processing
8	Estimation of income for those paying by modules	Spain	An AI system which estimates the income of Small and Medium Enterprises (SMEs) as well as of self-employed individuals who have decided to pay their taxes in phases rather than defining an exact income.	Implemented	Automated reasoning

Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

Δημόσιος Τομέας – Ενδεικτικά Έργα ΤΝ

Reducing night noise through nudging (City of Leuven in Belgium)

- A trial project used sound monitors to map night-time noise levels, along with behavioural ‘nudges’ to reduce noise.
- The nudges included adapting public lighting in response to noise levels, and using signage and light projections.
- The projections, for example, cut sound peaks by 30 percent.
- The effect of the projection tapered off after around 1am.

Πηγή: https://cities-today.com/leuven-tackles-night-noise-with-smart-nudging/

Δημόσιος Τομέας – Ενδεικτικά Έργα ΤΝ

Μέθοδος ανάπτυξης των έργων ΤΝ

#	Case	Development method
1	Intelligent Control Platform	Developed in-house with a data science unit
2	Eva, targeted COVID-19 Border Checking	Developed in collaboration with academia
3	Reducing night noise through nudging	Developed by an external company
4	Unlocking digitised documents and correcting OCR	Developed in-house (on top of OCR developed by private vendors)
5	Object Detection Kit	Developed in-house
6	OTT – decision-support tool for consultants	Developed in collaboration with academia and private vendors
7	Automation of subtitling videos and audios	Developed in collaboration with a private vendor
8	Estimation of income for those paying by modules	Developed mostly in-house

Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

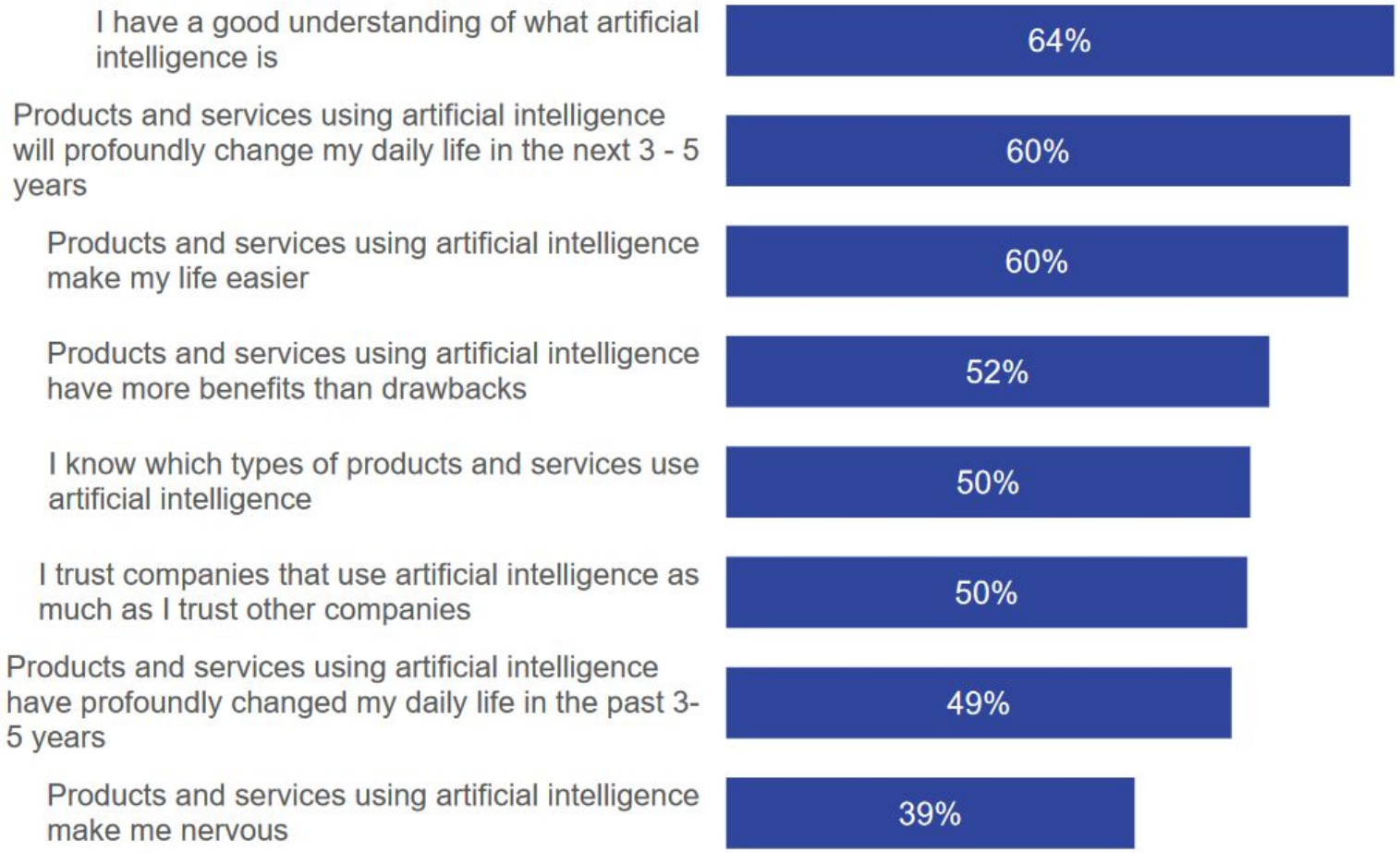
Δημόσιος Τομέας: Το in-house AI or not?

- Ανάπτυξη “in-house” ή από εξωτερική εταιρία;
  - Αρκετοί εξέφρασαν επιφυλάξεις σχετικά με επάρκεια σε δεξιότητες ΤΝ στους δημοσίους υπαλλήλους.
  - Από την άλλη, συστήματα που έγιναν από εταιρίες, λόγω έλλειψης ευελιξίας και ελέγχου δεν ανταποκρίθηκαν στις προσδοκίες. Π.χ. στις περ. 4, 7 το Δημόσιο αναγκάστηκε να φτιάξει νέα συστήματα «on-top».
  - Εσωτερική ειδίκευση στην ΤΝ επίσης διασφαλίζει ότι η ΤΝ χρησιμοποιείται με ηθικό τρόπο, μετριάζοντας τυχόν πιθανούς κινδύνους τόσο κατά τη διάρκεια της ανάπτυξης όσο και όταν χρησιμοποιείται ενεργά σε παραγωγική διαδικασία.

Πηγή: AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition

Κοινή γνώμη

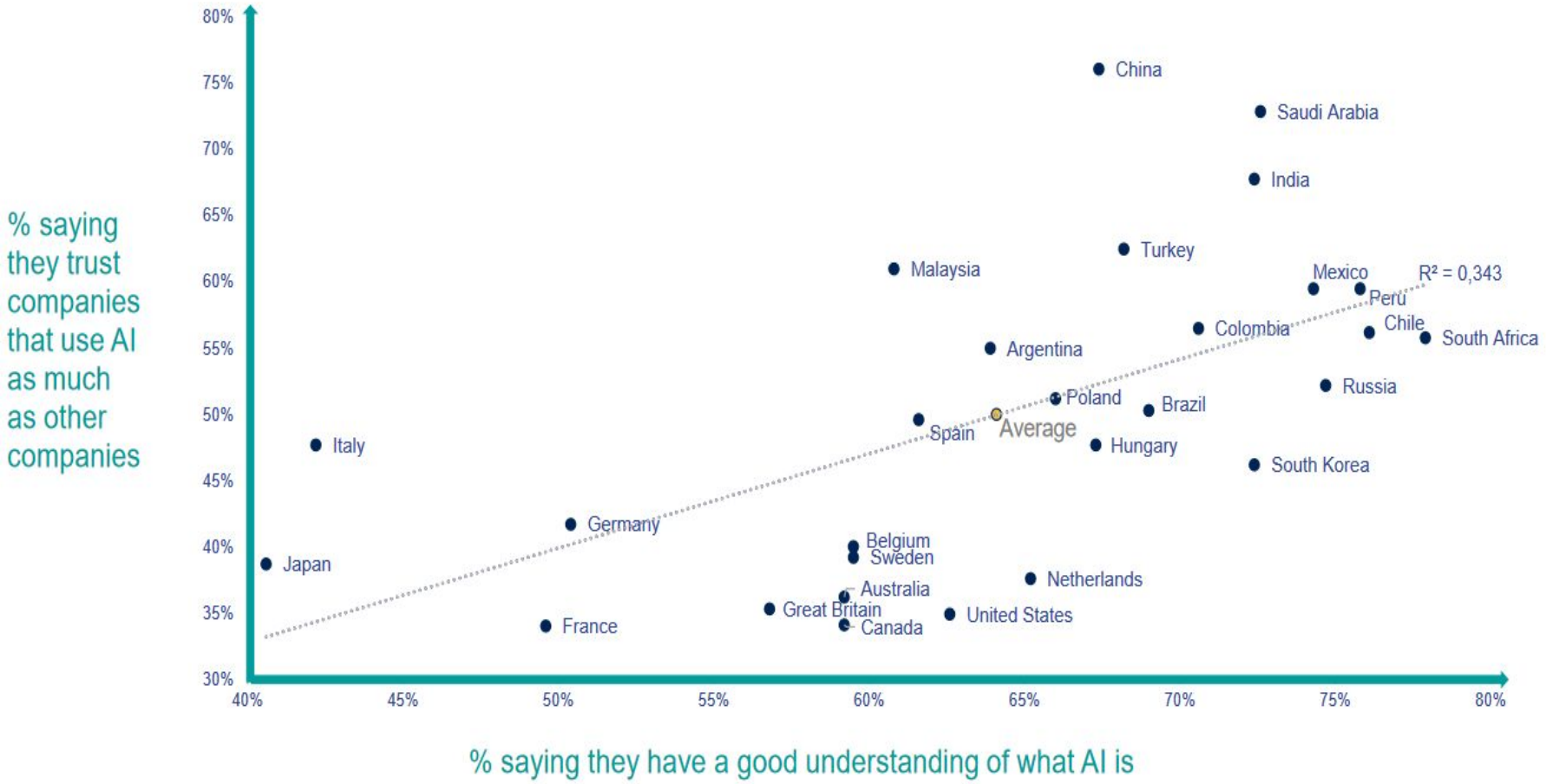
Απόψεις για την ΤΝ (Μέσος όρος όλων των χωρών)



Πηγή: © Ipsos - Global Opinions and Expectations about AI - January 2022
Base: 19,504 online adults aged 16-74 across 28 countries, Nov.–Dec. 2021

Κοινή γνώμη

Η εμπιστοσύνη στην ΤΝ συσχετίζεται με την κατανόηση. Και τα δύο είναι υψηλότερα στις υπό ανάπτυξη χώρες σε σχέση με τις χώρες υψηλού εισοδήματος



Πηγή: © Ipsos - Global Opinions and Expectations about AI - January 2022
Base: 19,504 online adults aged 16-74 across 28 countries, Nov.–Dec. 2021

Εξελίξεις και Τάσεις

Ερωτήσεις;

Περίγραμμα

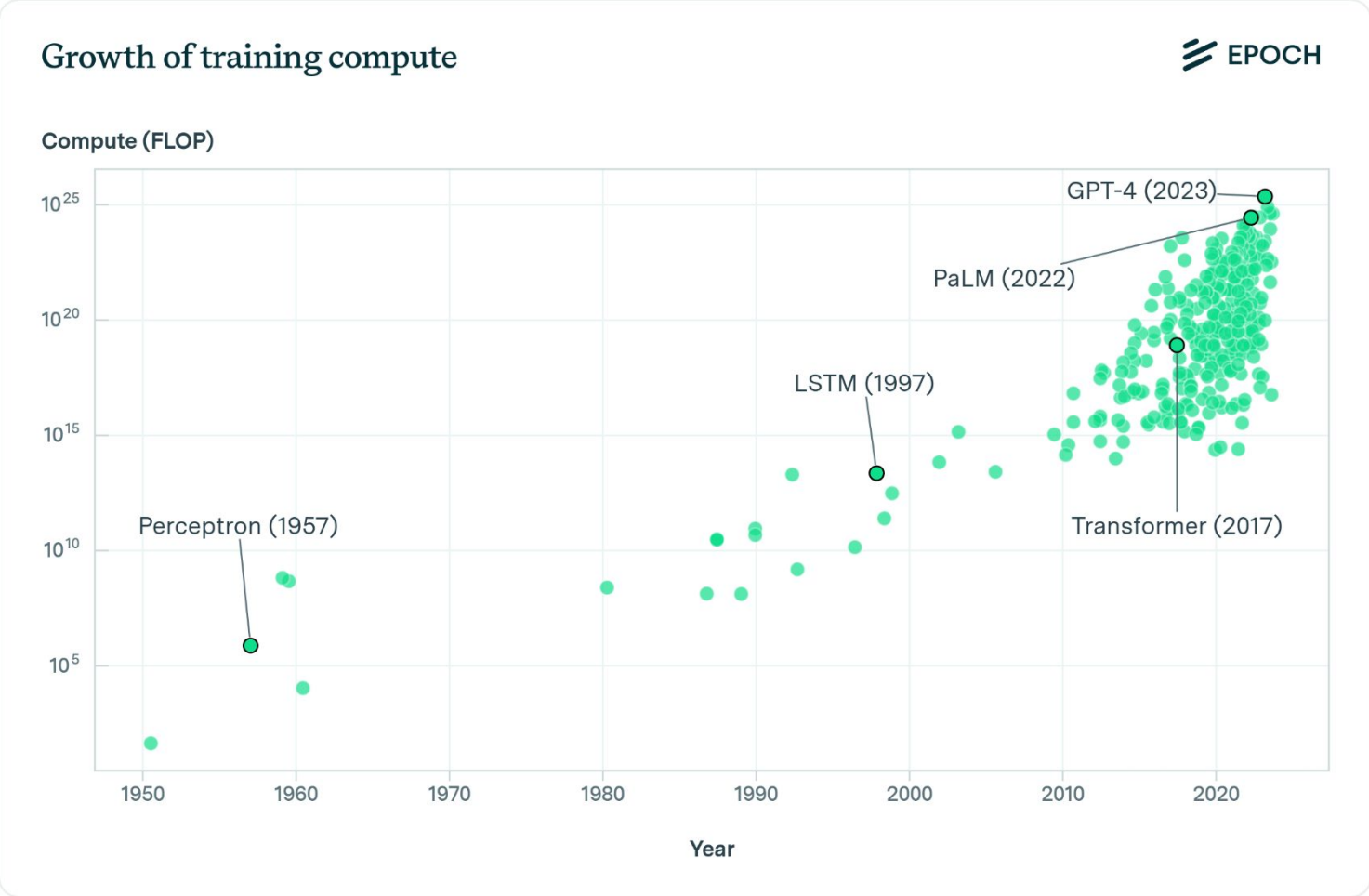
1. Σύντομη ιστορική αναδρομή και ορισμοί
2. Εξελίξεις και Τάσεις
3. Προκλήσεις και Κίνδυνοι
4. Συμπεράσματα

Προκλήσεις και Κίνδυνοι

- Υπολογιστική Ισχύ
- Μέγεθος Δεδομένων
- Κόστος εκπαίδευσης μοντέλων μηχανικής μάθησης
- DeepFakes
- Ποιότητα δεδομένων
- Προκαταλήψεις (Biases)
- Κίνδυνοι Ασφαλείας
- Επεξηγησιμότητα και διαφάνεια

Προκλήσεις: Υπολογιστική Ισχύ

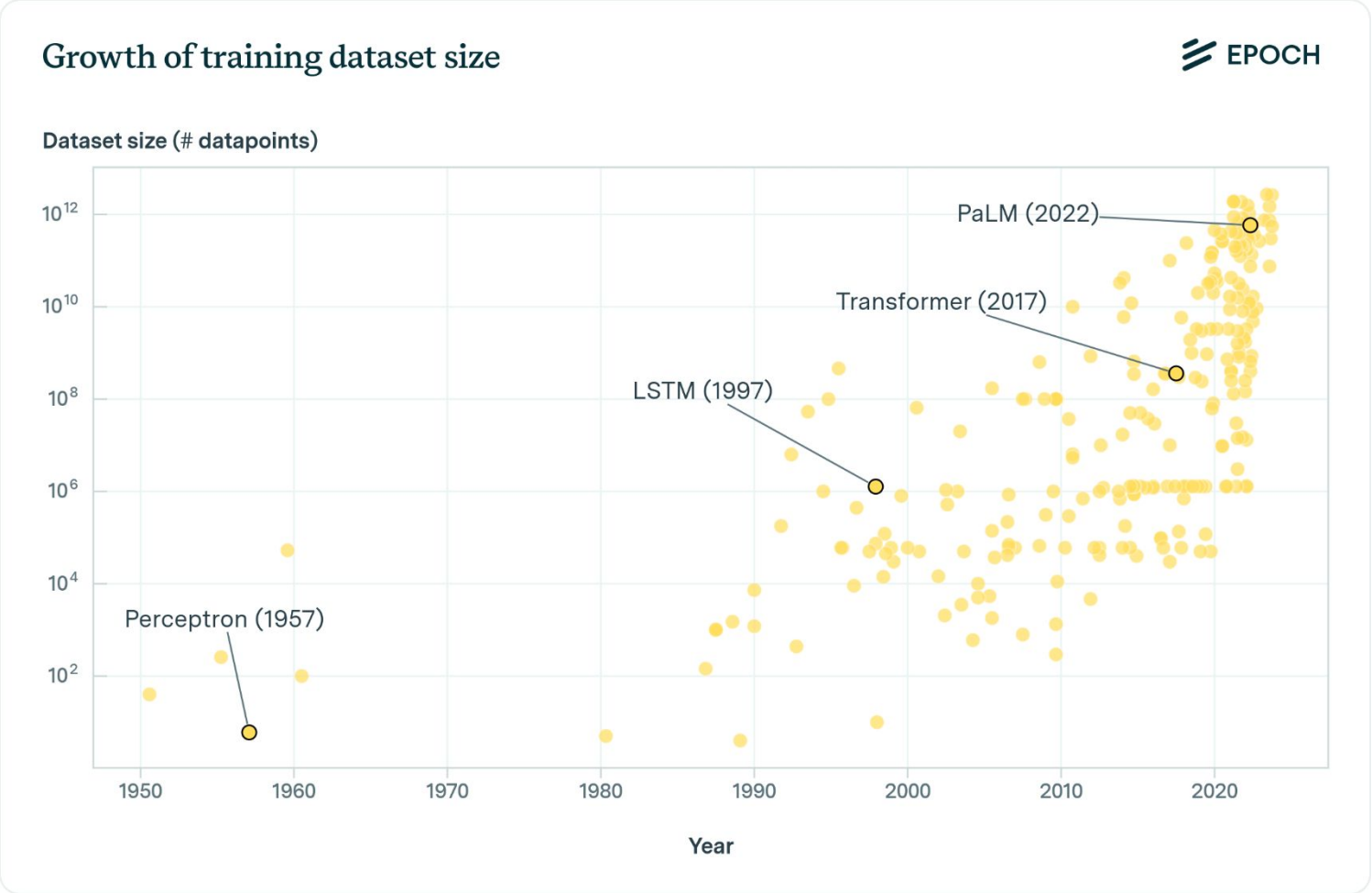
Υπολογιστική ισχύ για την εκπαίδευση αξιόλογων συστημάτων Μηχανικής Μάθησης



Πηγή: Epoch

Προκλήσεις: Μέγεθος δεδομένων

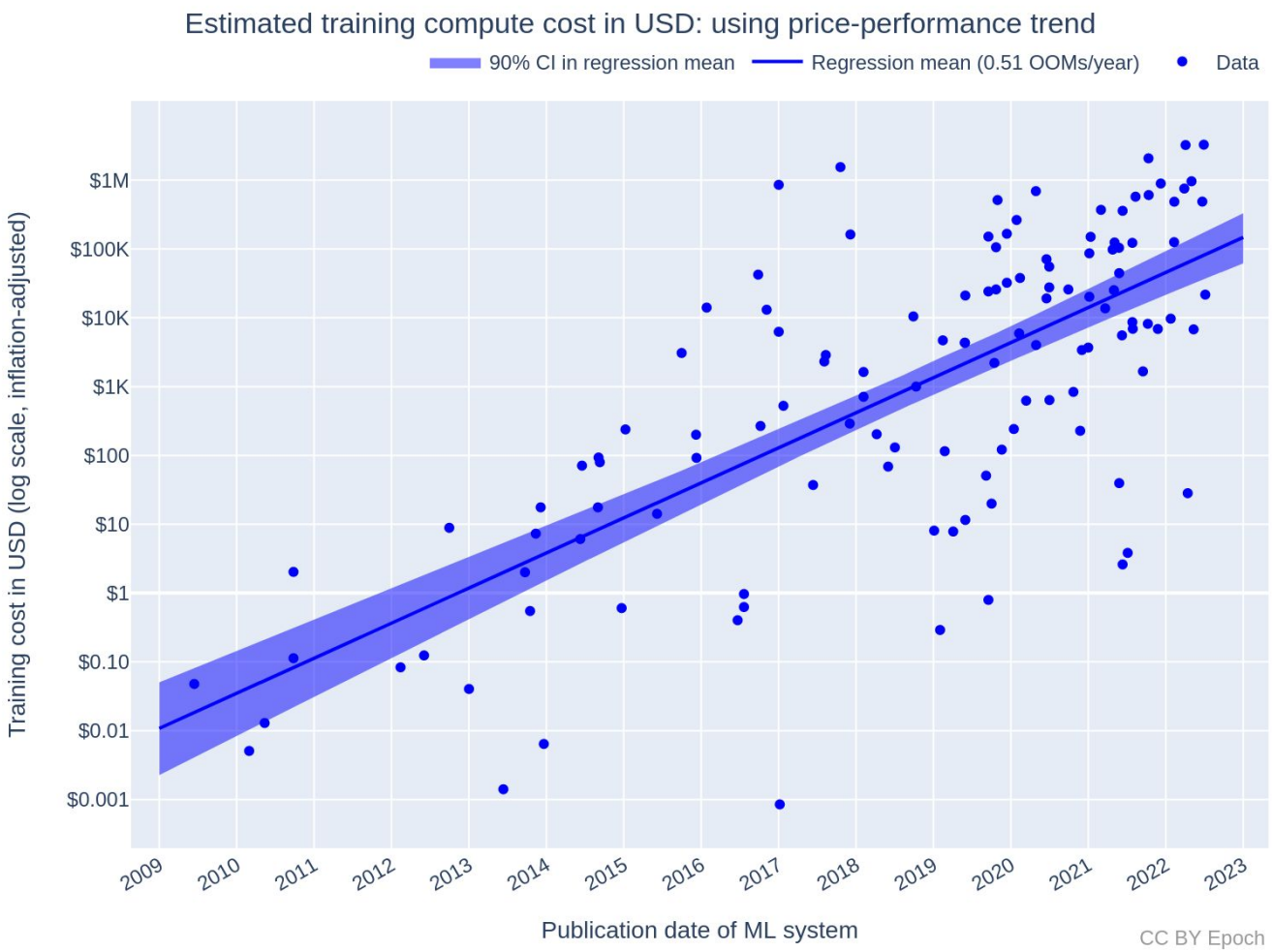
Μέγεθος δεδομένων για την εκπαίδευση αξιόλογων συστημάτων Μηχανικής Μάθησης



Πηγή: Epoch

Προκλήσεις: Κόστος εκπαίδευσης

Εκτίμηση κόστους εκπαίδευσης σημαντικών συστημάτων Μηχανικής Μάθησης



Πηγή: Epoch

Προκλήσεις: Deep Fakes

- Τα Deepfakes είναι βίντεο ή εικόνες που συχνά εμφανίζουν άτομα που έχουν αλλοιωθεί ψηφιακά, είτε πρόκειται για τη φωνή, το πρόσωπο ή το σώμα τους, έτσι ώστε να φαίνεται ότι «λένε» κάτι άλλο ή είναι εντελώς άλλος.

Προκλήσεις: Deep Fakes

- Κίνδυνοι από deep fakes
  - Κυκλοφορία ψευδών ειδήσεων / προπαγάνδα
  - Παραπλάνηση και να δημιουργία σύγχυσης για σημαντικά θέματα
  - Παρενόχληση, εκφοβισμός, υπονόμευση.
  - Ανήθικες ενέργειες όπως η εκδικητική πορνογραφία, με την οποία οι γυναίκες βλάπτονται δυσανάλογα.

Προκλήσεις: Παράδειγμα Deep Fake

16 Μαρτίου 2022: Ένα deepfake του Ουκρανού προέδρου που καλούσε τους στρατιώτες του να παραδώσουν τα όπλα τους φέρεται να «ανέβηκε» σε μια παραβιασμένη ουκρανική ειδησεογραφική ιστοσελίδα.

Προκλήσεις: Ποιότητα Δεδομένων Εκπαίδευσης

- Τα συστήματα ΤΝ βασίζονται σημαντικά στα δεδομένα για την εκπαίδευσή τους ώστε να κάνουν σωστές προβλέψεις.
- Κίνδυνοι:
  - Ανακριβή δεδομένα μπορούν να οδηγήσουν σε προκαταλήψεις και ανακρίβειες με αποτέλεσμα την κακή ποιότητα των προβλέψεων.
  - Έλλειψη δεδομένων μπορεί να οδηγήσει στην παράβλεψη σημαντικών μοτίβων και συσχετίσεων με αποτέλεσμα ελλιπείς και προκαταλειμμένες προβλέψεις

Προκλήσεις: Ποιότητα Δεδομένων Εκπαίδευσης

- Παράδειγμα: e-rater σύστημα αυτόματης διόρθωσης εκθέσεων στις ΗΠΑ.
  - Το 2018 βρέθηκε μετά από έρευνα ότι υποβαθμολογεί Αφροαμερικανούς.
  - Ένας από τους λόγους ήταν ανεπαρκή δεδομένα εκπαίδευσης από Αφροαμερικανούς.

Πηγές:

1. https://www.vice.com/en/article/pa7dj9/flawed-algorithms-are-grading-millions-of-students-essays
2. https://onlinelibrary.wiley.com/doi/full/10.1002/ets2.12192

Προκλήσεις: Προκαταλήψεις

- Προκαταλήψεις στα δεδομένα
  - **Historical bias:** Υπαρκτές προκαταλήψεις που έχουν περάσει στα δεδομένα εκπαίδευσης
  - **Representation bias:** Όταν τα δεδομένα δεν είναι αντιπροσωπευτικά για κάποιες ομάδες.
    - Παράδειγμα: Δεδομένα που έχουν συλλεγεί από κινητά τηλέφωνα ενδεχομένως να μην αντιπροσωπεύουν σωστά τον φτωχό πληθυσμό
- Προκαταλήψεις στα μοντέλα
  - **Evaluation bias:** Όταν ο τρόπος ελέγχου του παραγόμενου μοντέλου δεν είναι αντιπροσωπευτικός
  - **Aggregation bias:** όταν κατά την κατασκευή του μοντέλου διαφορετικοί πληθυσμοί συνδυάζονται ακατάλληλα.

Προκλήσεις: Προκαταλήψεις

- Παράδειγμα: Συστήματα δημιουργίας εικόνων από κείμενο
  - https://huggingface.co/spaces/society-ethics/DiffusionBiasExplorer

Προκλήσεις: Κίνδυνοι Ασφαλείας

- Τα συστήματα ΤΝ μπορεί να είναι ευάλωτα σε επιθέσεις, όπου κακόβουλοι παράγοντες μπορούν να παραπλανήσουν το σύστημα.

- **Παράδειγμα: Galactica (Meta, Νοέμβριος 2022)**
  - Ένα σύστημα σαν το chatgpt εκπαιδευμένο σε επιστημονικά άρθρα όπου μπορούσε κανείς να ρωτήσει ό,τι ήθελε σχετικά με την επιστήμη.
  - Πολύ γρήγορα διάφοροι χρήστες βρήκαν τρόπους να ξεγελάσουν το σύστημα ώστε να παράγει ρατσιστικό και προσβλητικό περιεχόμενο.
  - Μετά από μόλις 3 ημέρες η Meta αναγκάστηκε να τραβήξει την πρίζα.

Πηγή:
https://arstechnica.com/information-technology/2022/11/after-controversy-meta-scientific-papers/

Προκλήσεις: Επεξηγησιμότητα και Διαφάνεια

- Πολλά μοντέλα ΤΝ θεωρούνται "μαύρα κουτιά", καθώς είναι δύσκολο να κατανοήσει κανείς πώς καταλήγουν σε συγκεκριμένες αποφάσεις.
- Πώς μπορούμε να εμπιστευτούμε αποφάσεις που θα παίρνει η ΤΝ όταν δεν μπορεί να μας εξηγήσει το τρόπο που έφτασε σε αυτήν την απόφαση;
- Μπορούμε να εμπιστευτούμε ένα σύστημα ΤΝ όταν δεν ξέρουμε πάνω σε τι δεδομένα έχει εκπαιδευτεί;

Προκλήσεις: Επεξηγησιμότητα και Διαφάνεια

Οδηγία ΟΑΣΑ:

«Οι φορείς τεχνητής νοημοσύνης θα πρέπει να δεσμευτούν για διαφάνεια και υπεύθυνη αποκάλυψη σχετικά με τα συστήματα τεχνητής νοημοσύνης. Για το σκοπό αυτό, θα πρέπει να παρέχουν ουσιαστικές πληροφορίες, κατάλληλες για το πλαίσιο και συνεπείς με τις βέλτιστες πρακτικές:

- να προωθήσει μια γενική κατανόηση των συστημάτων AI,
- να ενημερώσουν τους ενδιαφερόμενους για τις αλληλεπιδράσεις τους με συστήματα τεχνητής νοημοσύνης, συμπεριλαμβανομένου του χώρου εργασίας,
- να επιτρέψει σε όσους επηρεάζονται από ένα σύστημα τεχνητής νοημοσύνης να κατανοήσουν το αποτέλεσμα και,
- να επιτρέψει σε όσους επηρεάζονται αρνητικά από ένα σύστημα τεχνητής νοημοσύνης να αμφισβητήσουν το αποτέλεσμά του με βάση απλές και κατανοητές πληροφορίες σχετικά με τους παράγοντες και τη λογική που χρησίμευσε ως βάση για την πρόβλεψη, τη σύσταση ή την απόφαση.»

Πηγή: https://oecd.ai/en/dashboards/ai-principles/P7

Προκλήσεις: Η περίπτωση της Γαλατά Σαράι

- 12/7/2021: Οι παίκτες της Γαλατά Σαράι αρνούνται τον έλεγχο rapid test στο αεροδρόμιο βάση αλγορίθμου.
- Την ίδια μέρα: Μετά από μεγάλη φασαρία επιστρέφουν στην Τουρκία. Το φιλικό με τον Ολυμπιακό ακυρώνεται.
- 12-14/7/2021: Ακολουθούν σφοδρές αντιδράσεις από την Τουρκική Κυβέρνηση
- Εάν υπήρχε τρόπος εξήγησης της απόφασης μήπως δεν θα υπήρχαν αντιδράσεις;

- 16/7/2021: Η Γαλατά Σαράι ανακοινώνει 2 κρούσματα κορωνοϊού

Προκλήσεις: Επεξηγησιμότητα και Διαφάνεια

Παράδειγμα καλής πρακτικής:

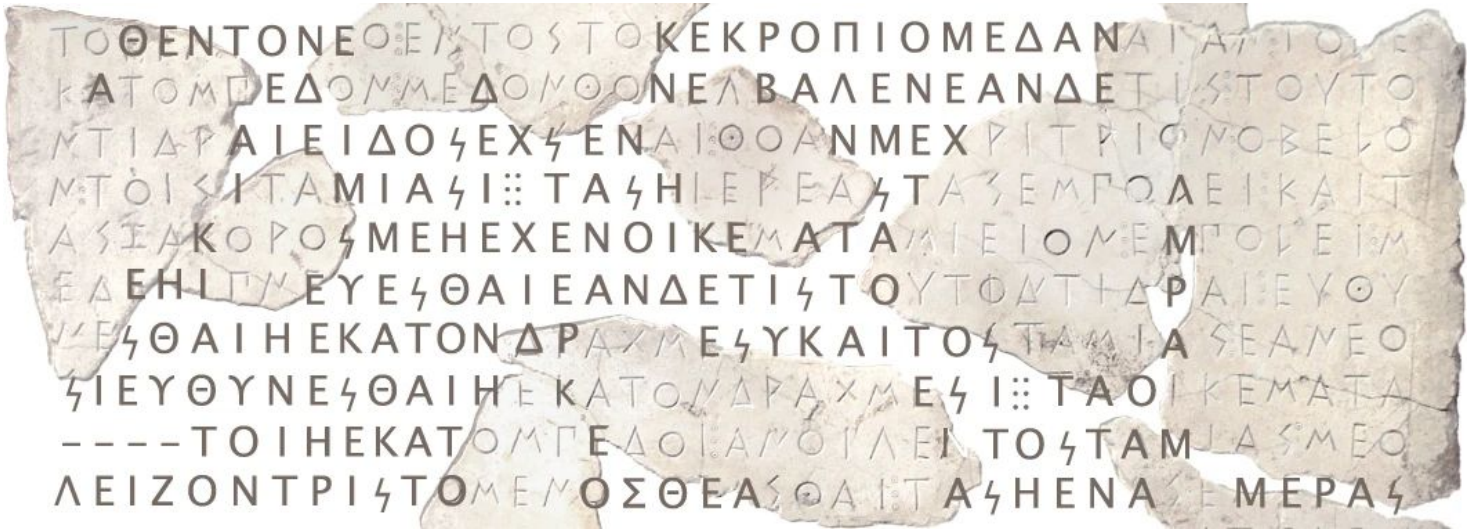
Ithaca: Restoring and attributing ancient texts using deep neural networks

- Σύστημα μηχανικής μάθησης για την κειμενική αποκατάσταση, τη γεωγραφική και χρονολογική απόδοση αρχαιοελληνικών επιγραφών.
- Έχει σχεδιαστεί για να βοηθά και να επεκτείνει τη ροή εργασίας του ιστορικού
- Η αρχιτεκτονική του επικεντρώνεται στη συνεργασία, την υποστήριξη αποφάσεων και την ερμηνευτικότητα.
- Απόδοση ομάδας ελέγχου αυξήθηκε από 25%

Προκλήσεις: Επεξηγησιμότητα και Διαφάνεια

Είσοδος:

Έξοδος:



Προκλήσεις: Επεξηγησιμότητα και Διαφάνεια

- Ζωντανή επίδειξη:
  - https://ithaca.deepmind.com

Προκλήσεις και Κίνδυνοι

Ερωτήσεις;

Περίγραμμα

1. Σύντομη ιστορική αναδρομή και ορισμοί
2. Εξελίξεις και Τάσεις
3. Προκλήσεις και Κίνδυνοι
4. Συμπεράσματα

Συμπεράσματα: Εξελίξεις

- Η Έρευνα στο χώρο της ΤΝ συνεχίζει να αυξάνεται με ραγδαίους ρυθμούς
- Κατακόρυφη αύξηση της απόδοσης των σημαντικών μοντέλων μηχανικής μάθησης τα τελευταία χρόνια
- Αύξηση όμως και στο κόστος εκπαίδευσης, λόγω της αύξησης της υπολογιστικής ισχύος και του όγκου δεδομένων που απαιτούνται
- Ιδιωτικός και Δημόσιος τομέας αναζητούν συνεχώς τρόπους να εντάξουν συστήματα ΤΝ
- Μηχανική Μάθηση και Python είναι οι πρώτες δεξιότητες που ζητούνται στην αγορά εργασίας της

Συμπεράσματα: ΤΝ στην ΕΕ

- Οι περισσότερες χώρες έχουν δημοσιεύσει εθνικές στρατηγικές
- Κλίμα, Υγεία, Δημόσια Διοίκηση και Εκπαίδευση οι κύριοι τομείς προς εφαρμογή
- Αρκετά έργα είναι σε λειτουργία, πολλά σε φάση ανάπτυξης
- Συνδυασμός «εσωτερικής-εξωτερικής» ανάπτυξης: μονόδρομος

Συμπεράσματα: Κίνδυνοι

- Ποιότητα δεδομένων
- Προκαταλήψεις (Biases)
- Κίνδυνοι Ασφαλείας
- DeepFakes
- Επεξηγησιμότητα και διαφάνεια

Συμπεράσματα: Η θέση της Ελλάδας

Προσωπική εκτίμηση της θέσης της Ελλάδας σε σχέση με τον υπόλοιπο κόσμο στην ΤΝ

Πεδίο	Αξιολόγηση θέσης
Έρευνα στην ΤΝ	≚
Εργατικό Δυναμικό ΤΝ	±
Νομοθεσία ΤΝ	∓
Εθνική Στρατηγική ΤΝ	∓
Έργα ΤΝ στο Δημόσιο	=

Σύμβολο	Ερμηνεία (Ελλάδα vs κόσμος)
≚	σε λίγο καλύτερη θέση
±	σε πλεονεκτική θέση
=	περίπου στην ίδια θέση
∓	σε λίγο χειρότερη θέση
∓	σε μειονεκτική θέση

Ευχαριστώ για την προσοχή σας!