

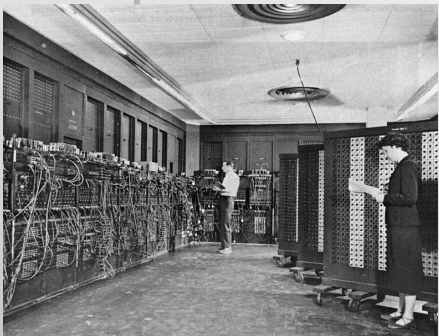
EURO Greece

A bit of history - how did we get here



Υπερυπολογιστές

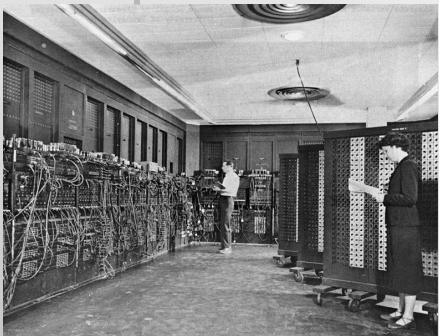
1946



ENIAC in BRL building
328. (U.S. Army photo)

Υπερυπολογιστές

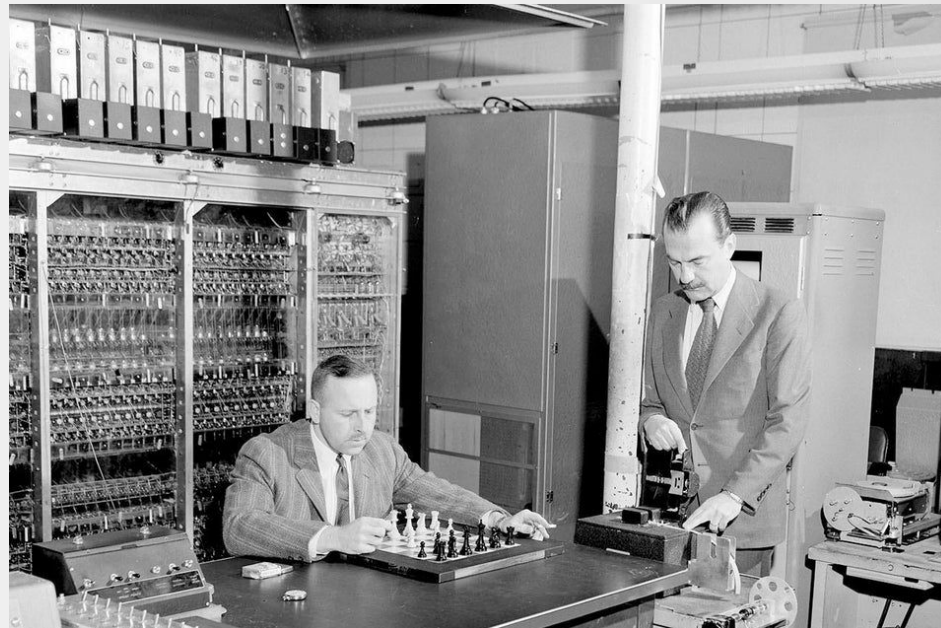
1946



ENIAC in BRL building 328. (U.S. Army photo)

Scientific Computing

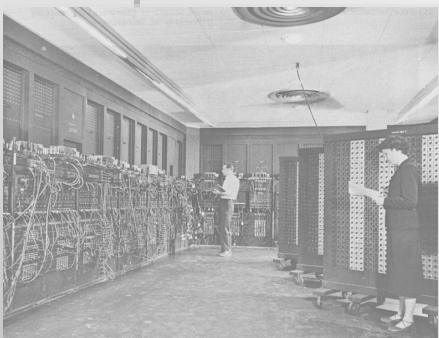
Χρήση υπολογιστών για τη μελέτη επιστημονικών προβλημάτων



Υπερυπολογιστές

1946

1985



ENIAC in BRL building 328. (U.S. Army photo)

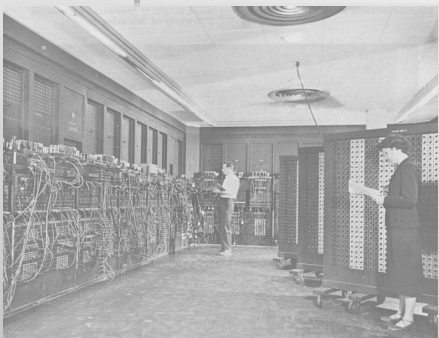


Cray-2 central unit

Υπερυπολογιστές

1946

1985



ENIAC in BRL building 328. (U.S. Army photo)



Cray-2 central unit

CRAY-2 (1985)

1.9 GFlop/s

1 core

x2



Smartwatch (2015)

3 GFlop/s

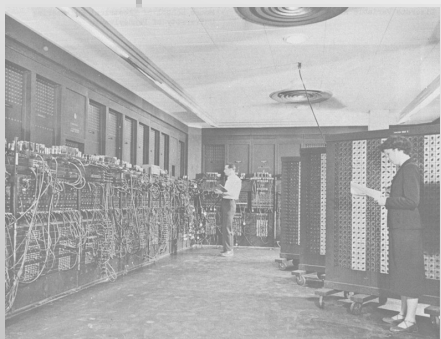
2 cores

Υπερυπολογιστές

1946

1985

1993



ENIAC in BRL building 328. (U.S. Army photo)



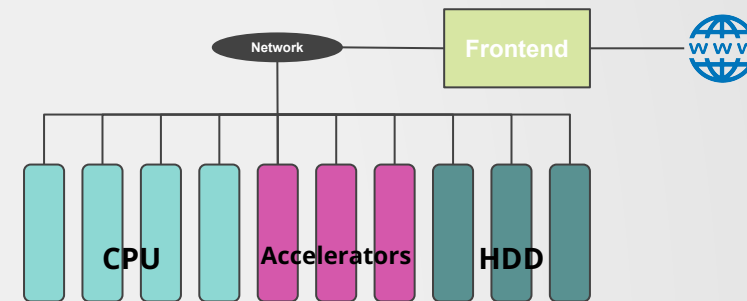
Beowulf cluster in NASA's Goddard Space Flight Center



Cray-2 central unit

Thomas Sterling: "Not only did nobody care, but there were even a number of people hostile to this project."

Because it was different. Because it was completely outside the scope of the supercomputing community at that time."



Υπερυπολογιστές

1946



ENIAC in BRL building 328. (U.S. Army photo)

1985



Cray-2 central unit

1993



Beowulf cluster in NASA's Goddard Space Flight Center

2000



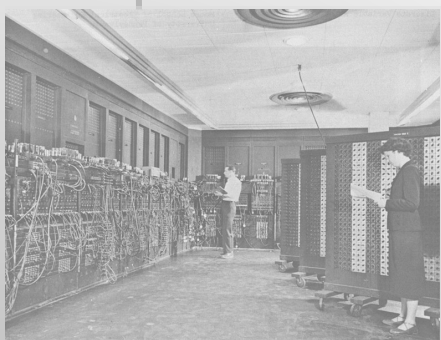
GPU

GPGPU - Use GPUs to accelerate a range of scientific applications

Late 1990 (2000), computer scientists, along with researchers in fields such as medical imaging and electromagnetics, started using GPUs to accelerate a range of scientific applications. This was the advent of the movement called GPGPU, or General Purpose GPU computing.

Υπερυπολογιστές

1946



ENIAC in BRL building 328. (U.S. Army photo)

1985



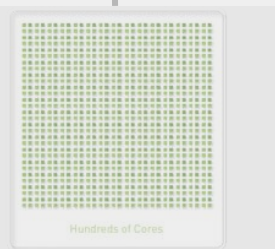
Cray-2 central unit

1993



Beowulf cluster in NASA's Goddard Space Flight Center

2000



GPU

GPGPU - Use GPUs to accelerate a range of scientific applications

2012



AlexaNet 2012



2009

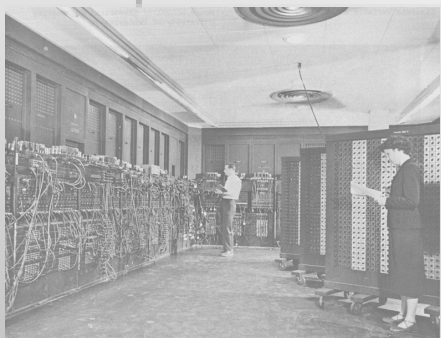


2007

Big Data - efficient data manager infrastructures

Υπερυπολογιστές

1946



ENIAC in BRL building 328. (U.S. Army photo)

1985



Cray-2 central unit

1993



Beowulf cluster in NASA's Goddard Space Flight Center

2000



GPU

GPGPU - Use GPUs to accelerate a range of scientific applications

2012



Big Data - efficient data manager infrastructures

2015

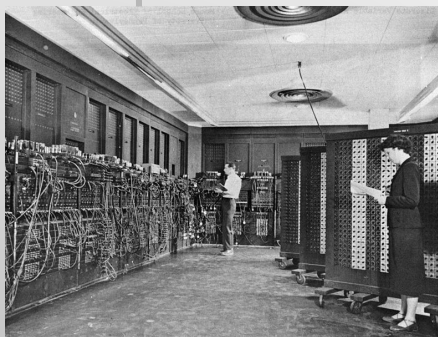


ARIS - Greek Supercomputer

2022

Υπερυπολογιστές

1946



ENIAC in BRL building 328. (U.S. Army photo)

1985



Cray-2 central unit

1993



Beowulf cluster in NASA's Goddard Space Flight Center

2000



GPU

GPGPU - Use GPUs to accelerate a range of scientific applications

2012



Big Data - efficient data manager infrastructures

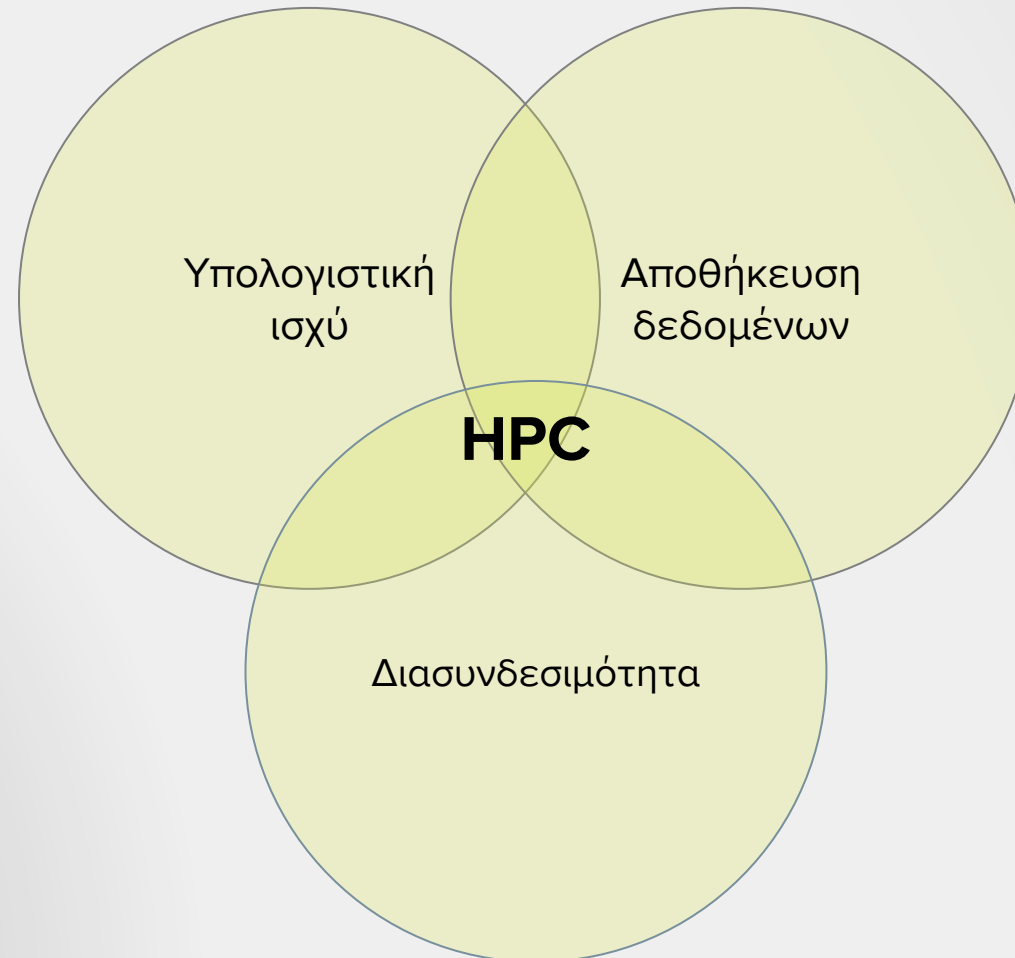
2015



ARIS - Greek Supercomputer

2022

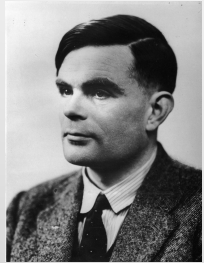
Υπερυπολογιστές



Τεχνητή Νοημοσύνη



Τεχνητή Νοημοσύνη



Alan Turing test -
*If machine tricks
human*

1950



**AI term is
introduced (John
McCarty)**

1956

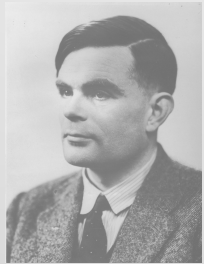


Deep Blue - Computer
defeats chess champion
Garry Kasparov

1997



Τεχνητή Νοημοσύνη



Alan Turing test -
*If machine tricks
human*

1950



**AI term is
introduced (John
McCarty)**

1956



Deep Blue - Computer
defeats chess champion
Garry Kasparov

1997

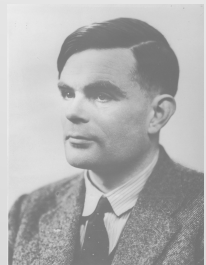


Apple Siri -
Intelligent virtual
assistant in iPhone

2011



Τεχνητή Νοημοσύνη



Alan Turing test -
If machine tricks human

1950



AI term is introduced (John McCarthy)

1956



Deep Blue - Computer defeats chess champion Garry Kasparov

1997



Apple Siri - Intelligent virtual assistant in iPhone

2011

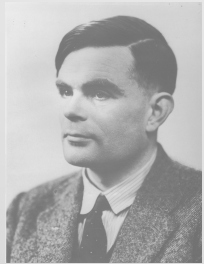


Big Data - Better data beats fancier algorithm

2012



Τεχνητή Νοημοσύνη



Alan Turing test -
If machine tricks human

1950



AI term is introduced (John McCarthy)

1956



Deep Blue - Computer defeats chess champion Garry Kasparov

1997



Apple Siri - Intelligent virtual assistant in iPhone

2011



Big Data - Better data beats fancier algorithm

2012

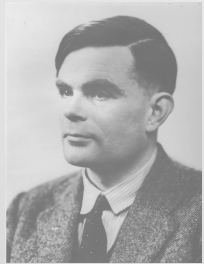


Tesla introduce self-driving car

2015



Τεχνητή Νοημοσύνη



Alan Turing test -
If machine tricks human

1950



AI term is introduced (John McCarthy)

1956



Deep Blue - Computer defeats chess champion Garry Kasparov

1997



Apple Siri - Intelligent virtual assistant in iPhone

2011



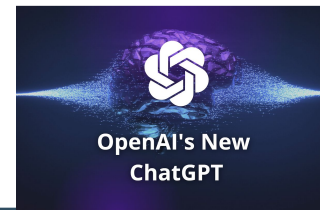
Big Data - Better data beats fancier algorithm

2012



Tesla introduce self-driving car

2015

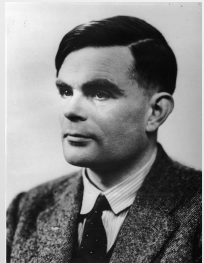


GPT-3 / ChatGPT - Available to public testing

2022



Τεχνητή Νοημοσύνη



Alan Turing test -
If machine tricks human

1950



AI term is introduced (John McCarthy)

1956



Deep Blue - Computer defeats chess champion Garry Kasparov

1997



Apple Siri - Intelligent virtual assistant in iPhone

2011



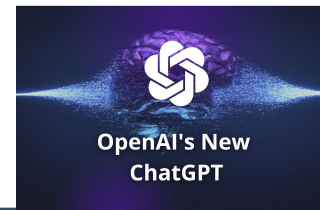
Big Data - Better data beats fancier algorithm

2012



Tesla introduce self-driving car

2015

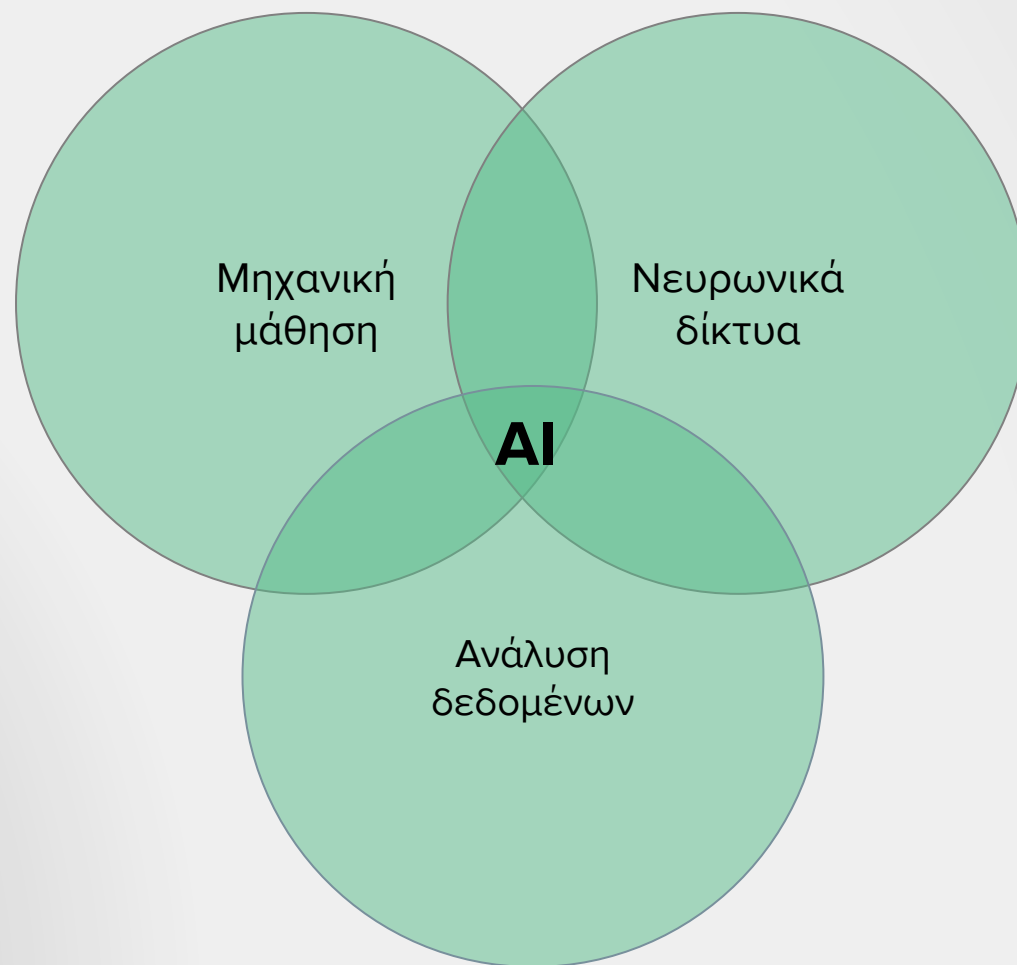


GPT-3 / ChatGPT - Available to public testing

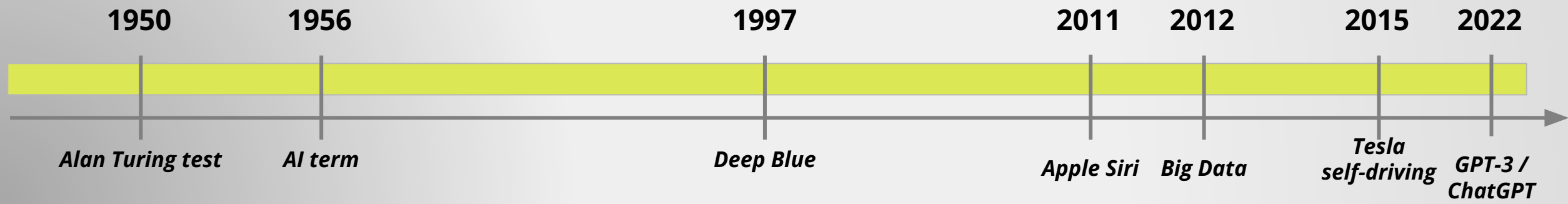
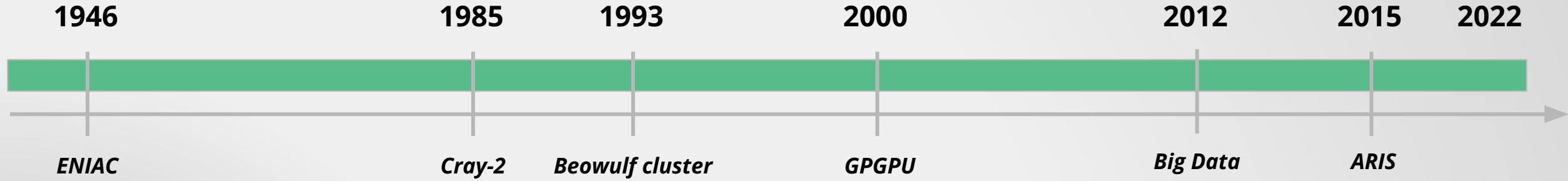
2022



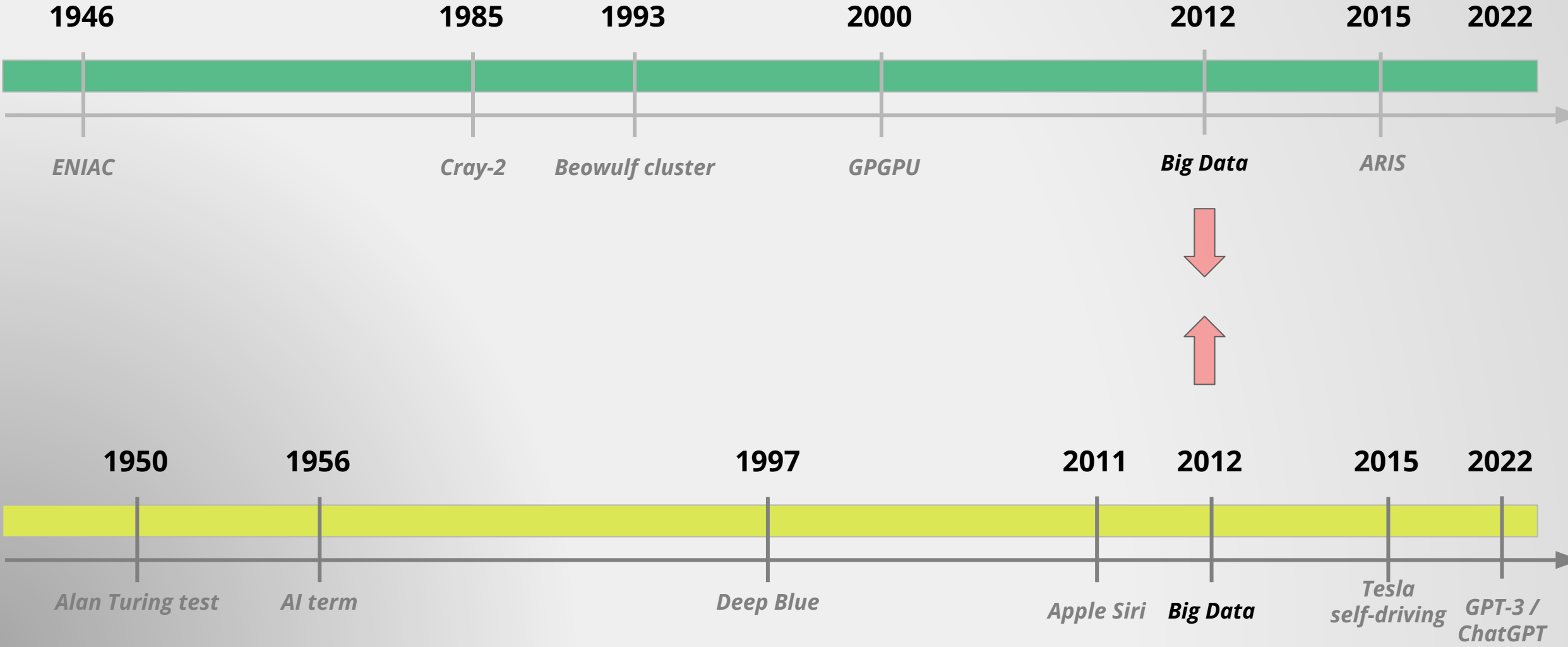
Τεχνητή Νοημοσύνη



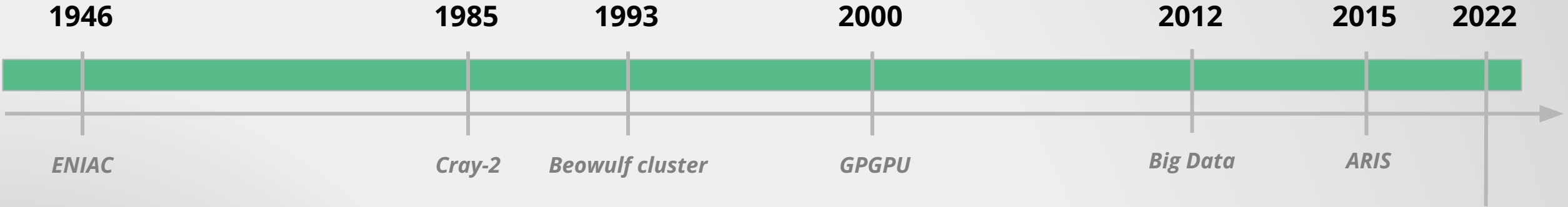
Τεχνητή Νοημοσύνη & Υπερυπολογιστές



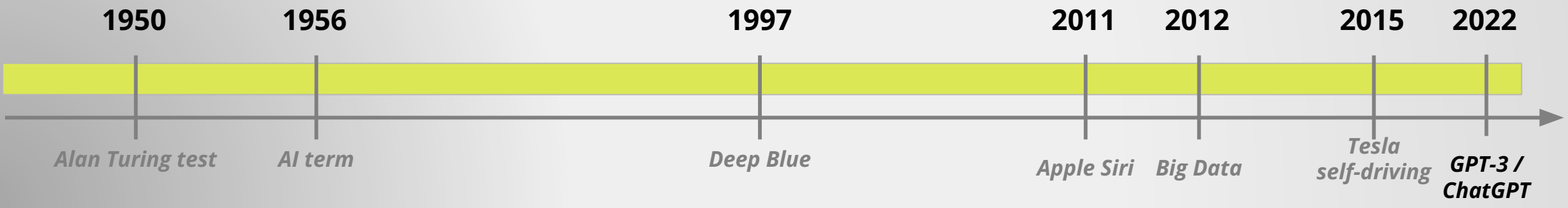
Τεχνητή Νοημοσύνη & Υπερυπολογιστές



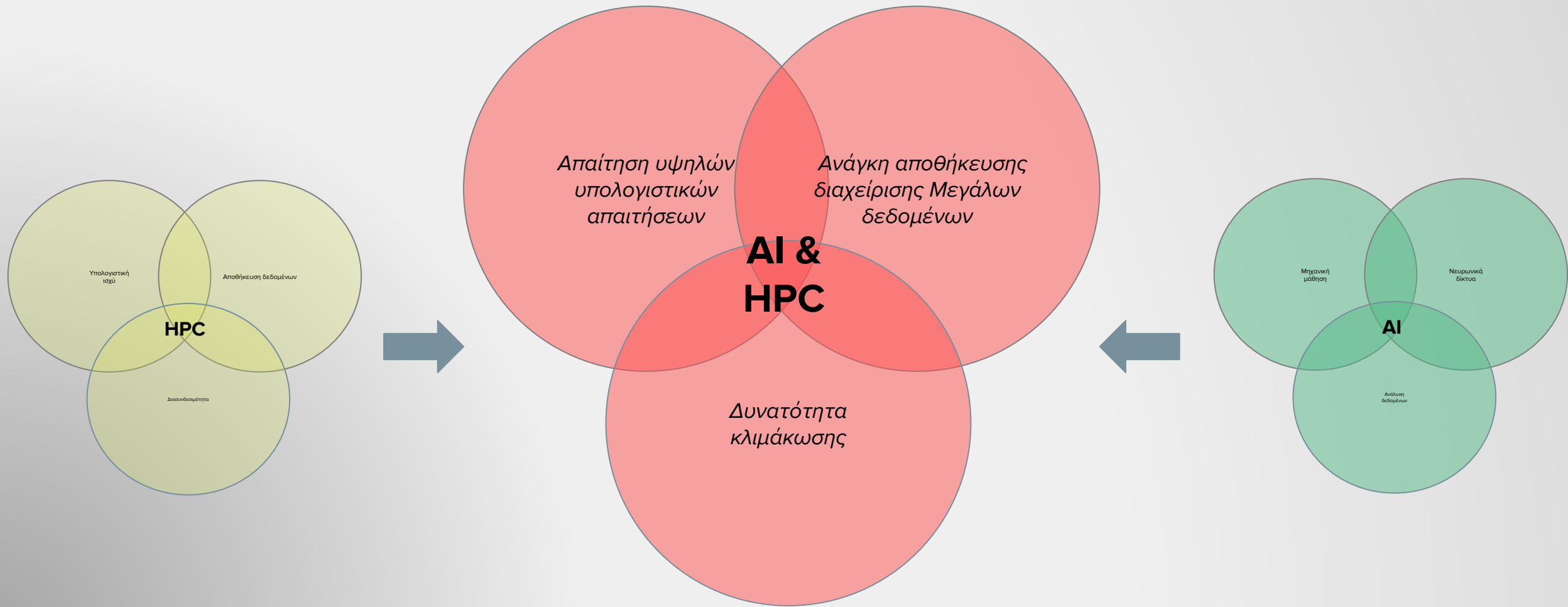
Τεχνητή Νοημοσύνη & Υπερυπολογιστές



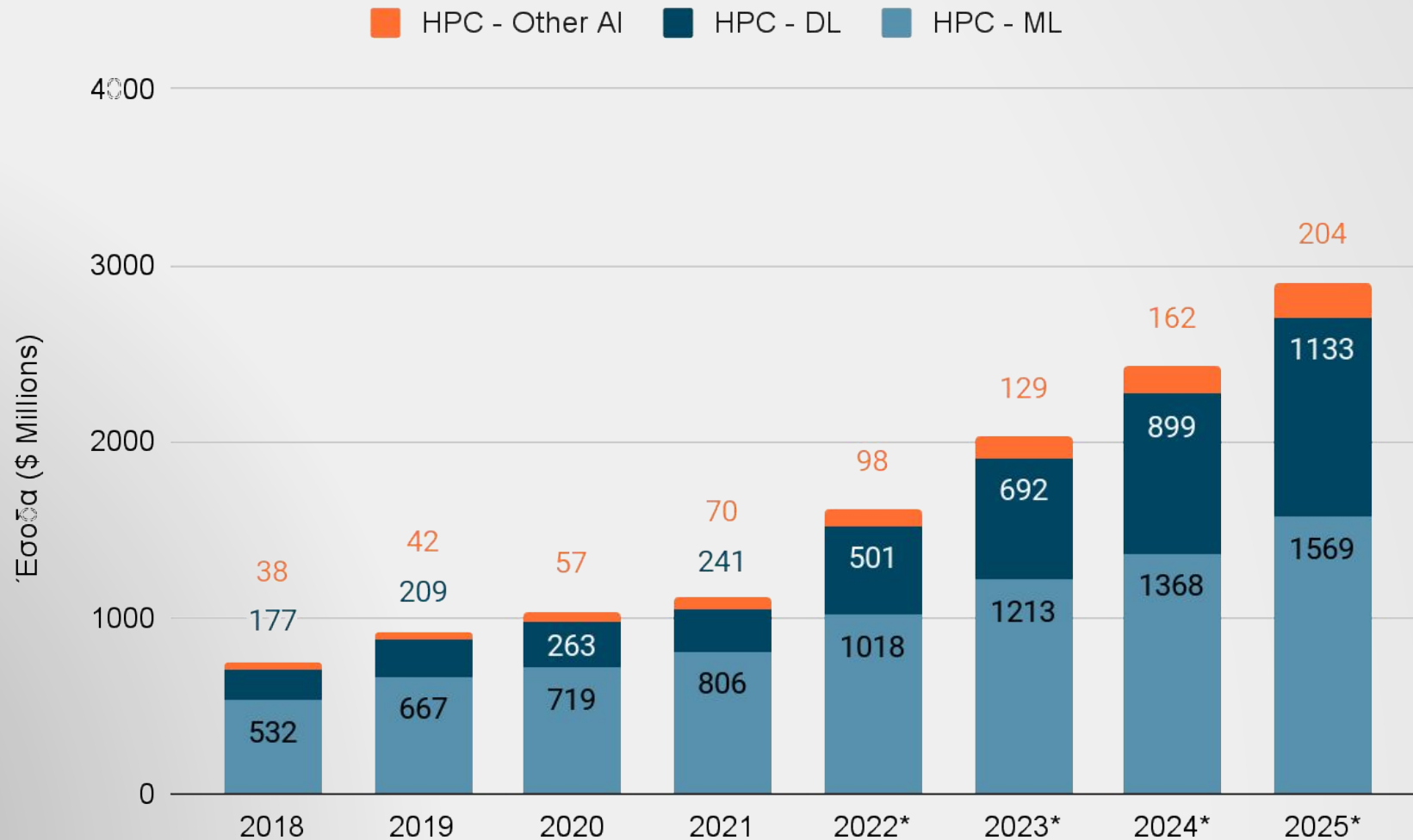
HPC essential for impactful AI applications (GenAI)



Τεχνητή Νοημοσύνη & Υπερυπολογιστές



Τεχνητή Νοημοσύνη & Υπερυπολογιστές



Τεχνητή Νοημοσύνη & Υπερυπολογιστές



Υψηλό κόστος εγκατάστασης: Οι HPC clusters με GPUs και προηγμένο εξοπλισμό δικτύωσης είναι δαπανηρά στην απόκτηση και τη συντήρηση.



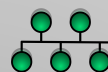
Πλοκοπτητα διαχείρισης: Η διαχείριση ενός συστήματος HPC είναι πολύπλοκη ακόμα και χωρίς να ασχολείται με φόρτους εργασίας AI. Οι οργανισμοί συχνά απαιτούν ειδικούς για τη διαμόρφωση, τη συντήρηση και τη βελτιστοποίηση των clusters HPC για φόρτους εργασίας AI.



Βελτιστοποίηση κλιμάκωσης: Τα clusters HPC είναι εξαιρετικά κλιμακώσιμα, αλλά η βελτιστοποίηση της διαδικασίας κλιμάκωσης για την εξυπηρέτηση φόρτων εργασίας AI είναι δύσκολη χωρίς να υποστούν πτώσεις στην απόδοση.



Κατανάλωση ενέργειας: Τα συστήματα HPC μεγάλης κλίμακας καταναλώνουν σημαντικές ποσότητες ενέργειας ανεξάρτητα από το τι τρέχετε σε αυτά. Ο ενεργειακά αποδοτικός εξοπλισμός και η ψύξη είναι απαραίτητα για την ελαχιστοποίηση του περιβαλλοντικού αντίκτυπου και τη μείωση των λειτουργικών δαπανών.



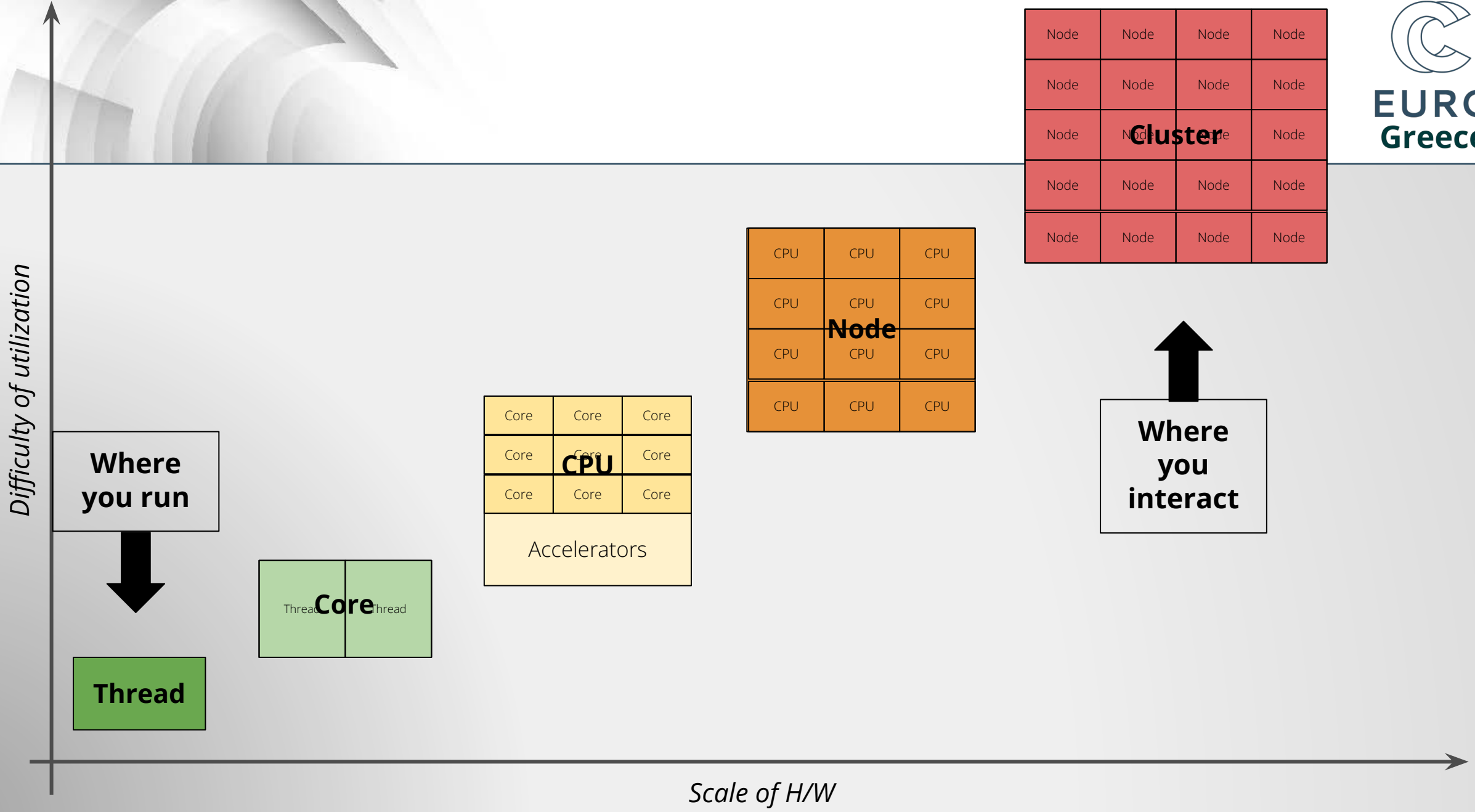
Προβλήματα μετακίνησης δεδομένων: Η μετακίνηση μεγάλων όγκων δεδομένων μεταξύ του χώρου αποθήκευσης και των κόμβων σε ένα cluster δημιουργεί συχνά καθυστερήσεις. Οι φόρτοι εργασίας AI που απαιτούν συχνή πρόσβαση σε δεδομένα πρέπει να εκτελούνται σε σύστημα HPC με συνδέσεις υψηλής ταχύτητας και αποτελεσματική στρατηγική διαχείρισης δεδομένων.



Διαθεσιμότητα προσωπικού: Επί του παρόντος, υπάρχει έλλειψη εξειδικευμένων επαγγελματιών που γνωρίζουν πώς να αξιοποιούν αποτελεσματικά τις δυνατότητες τόσο της HPC όσο και της AI.



Ενσωματώσεις λογισμικού: Η ενσωμάτωση frameworks λογισμικού AI με περιβάλλοντα HPC απαιτεί προσεκτική εξέταση των λογισμικών και των εξαρτήσεων. Διαφορετικά, ο χρήστης ενδέχεται να αντιμετωπίσει προβλήματα συμβατότητας και απόδοσης.



Thanks!



Co-funded by
the European Union



EuroHPC
Joint Undertaking

Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia under grant agreement No 101101903.

Δρ. Σ. Καρόζης

Συνεργαζόμενος Ερευνητής, ΕΚΕΦΕ “Δημόκριτος”

e-mail: skarozis@ipta.demokritos.gr