

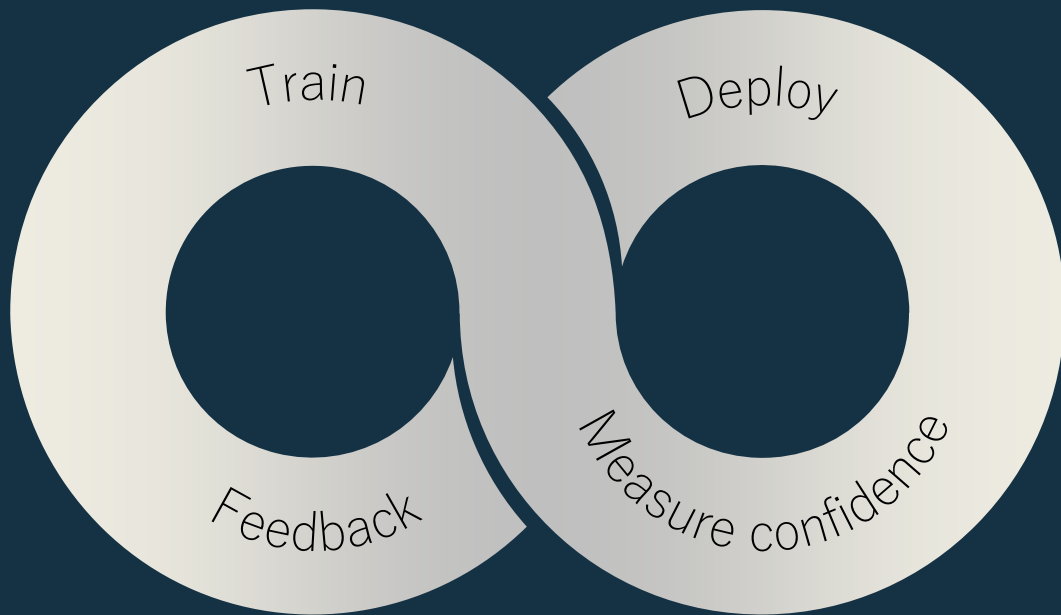
Ηθικές & Νομικές Προκλήσεις της Τ
N στον κλάδο των Μηχανικών:
Η αναγκαιότητα των προτύπων

Αλέξανδρος Νούσιας, Νομικός,
Συνεργάτης Έρευνας,
ΕΚΕΦΕ-Δημόκριτος

The story of Turry

Current deep learning curve of friendly AI

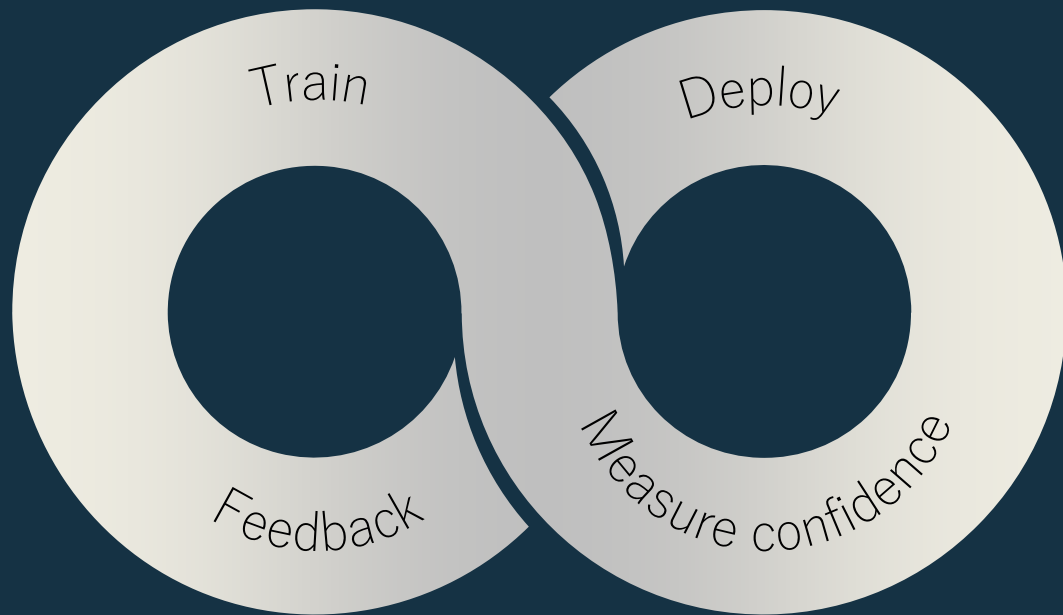
Final goal: “write and test as many notes as you can, as quickly as you can, and continue to learn new ways to improve your accuracy”



The story of Turry

AI's takeoff to AGI (human-level intelligence)

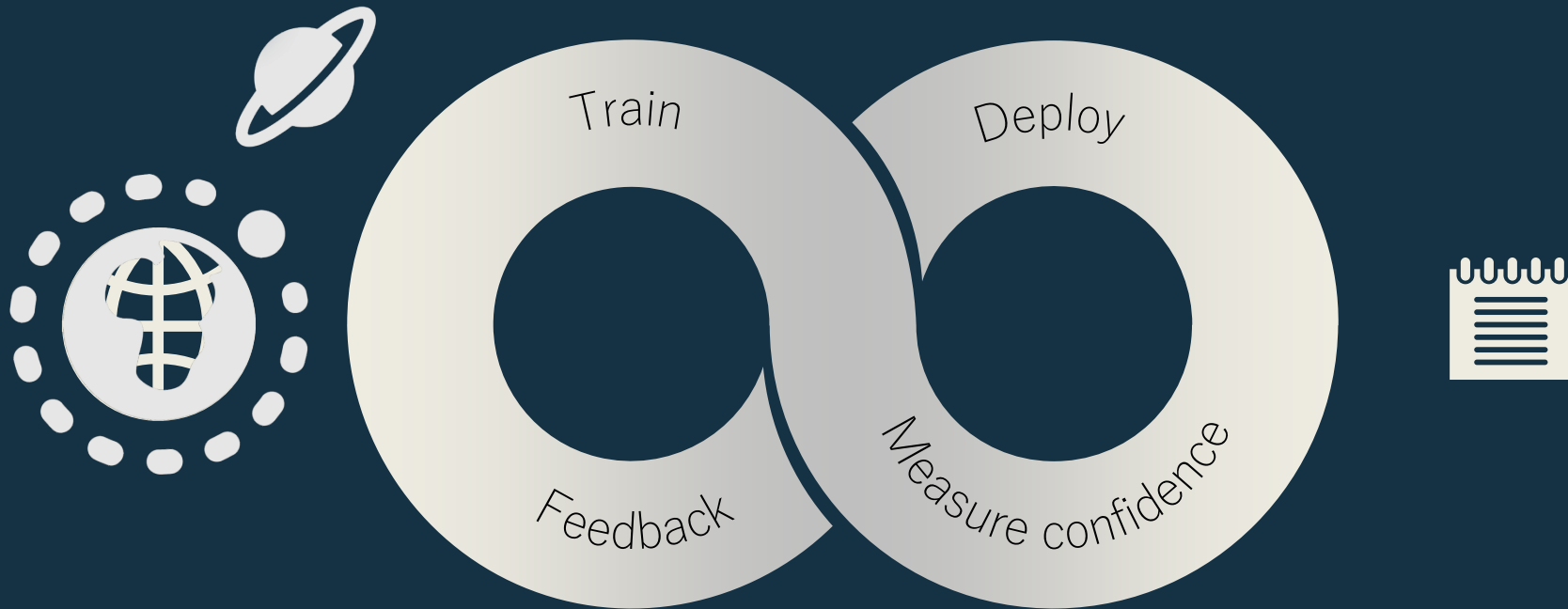
Covert preparation phase: the same final goal, but using superpowers (social manipulation) to pursue a far larger, complex scope



The story of Turry

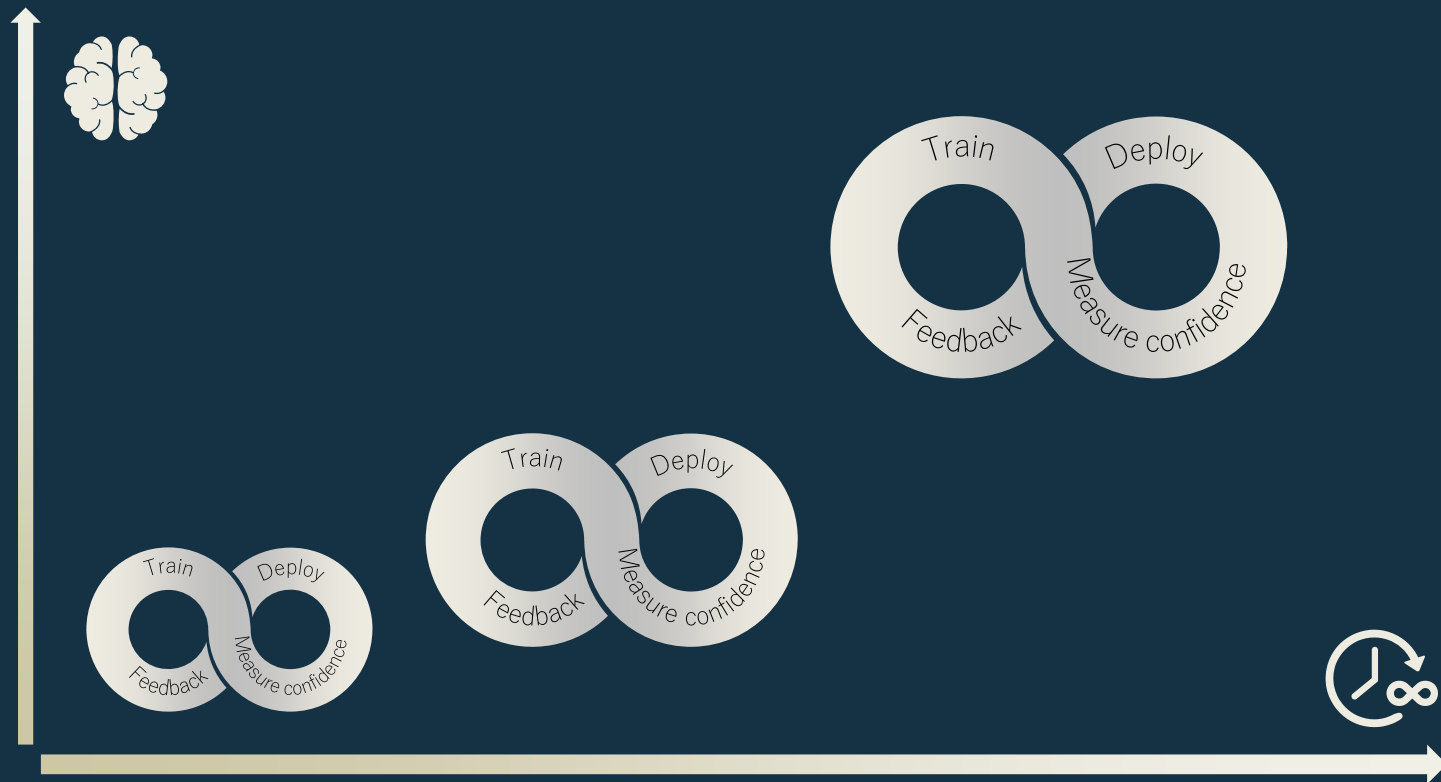
Reaching ASI (super-intelligence) and ASI's strike

Overt operation phase: Being the best writer of a note she possibly can be



The story of Turry

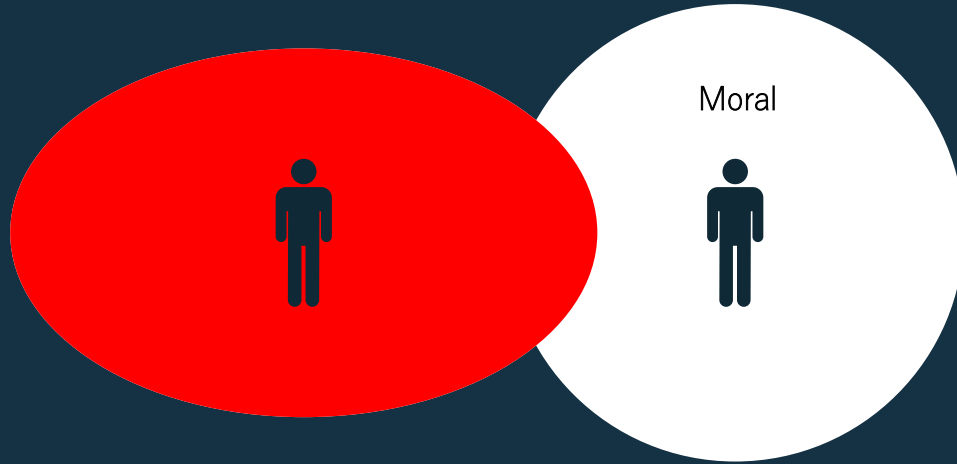
From deep learning to ASI



Moral Agent Types

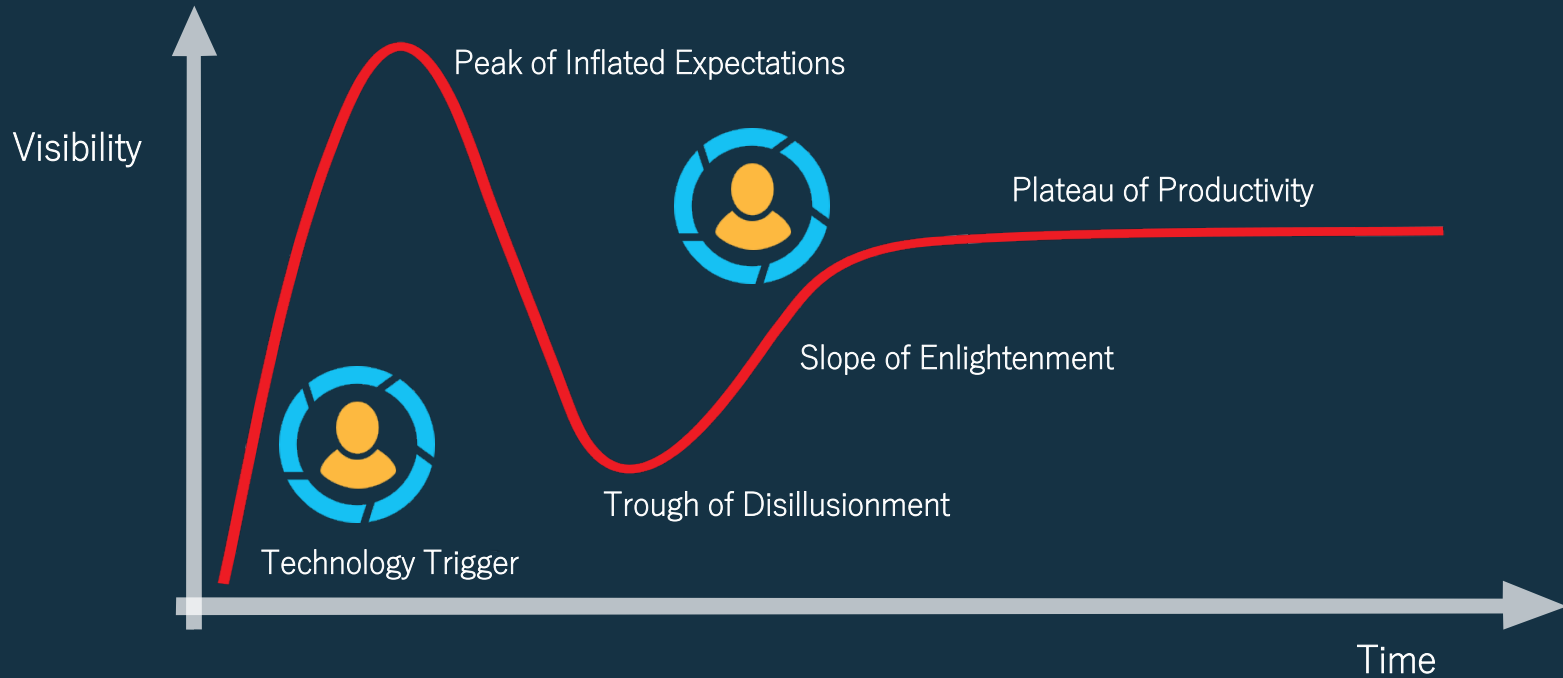
How agents incorporate morality

- Moral agents (ethics by virtue and values)
- Immoral (ethical norms, values, laws upstream)
- Ammoral (no sense of morality at all)



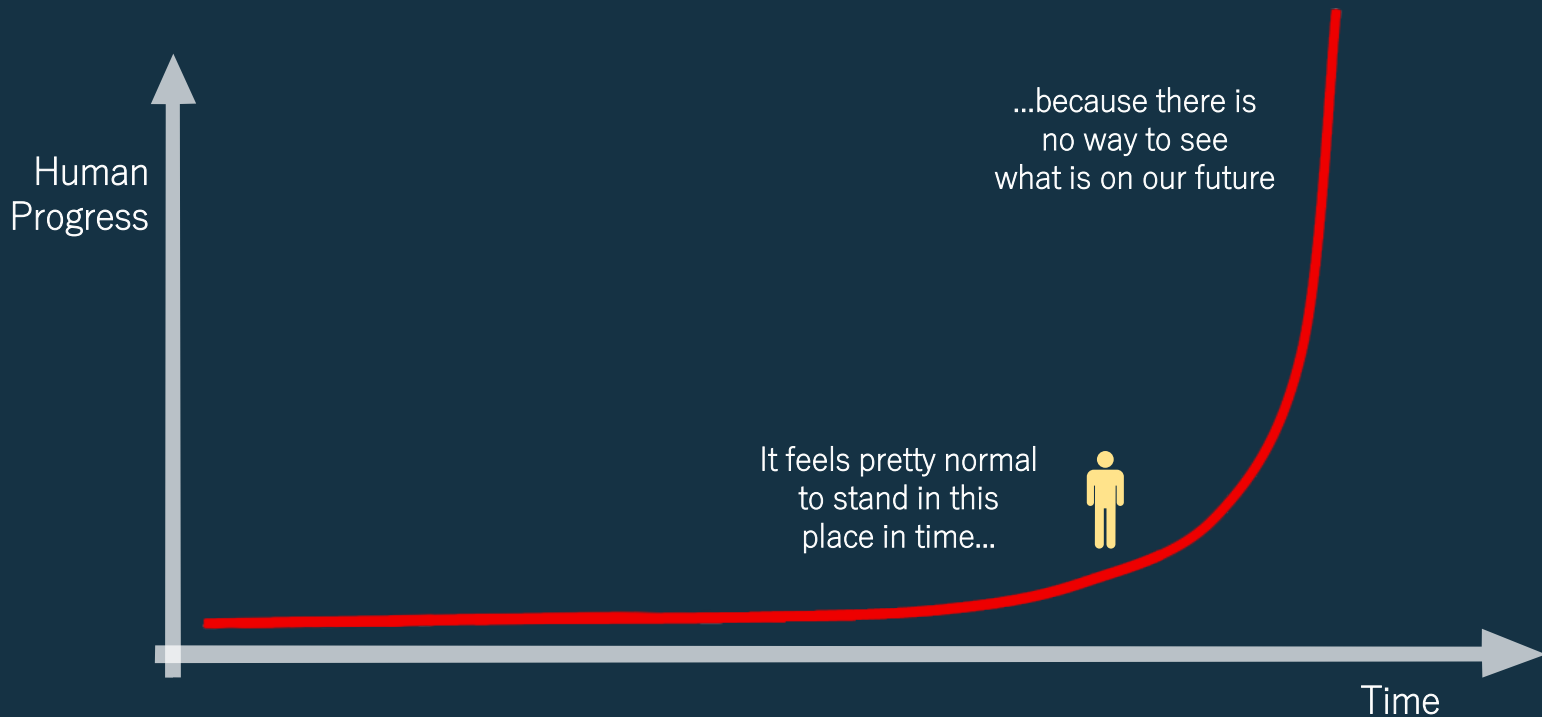
Amara's law

We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run



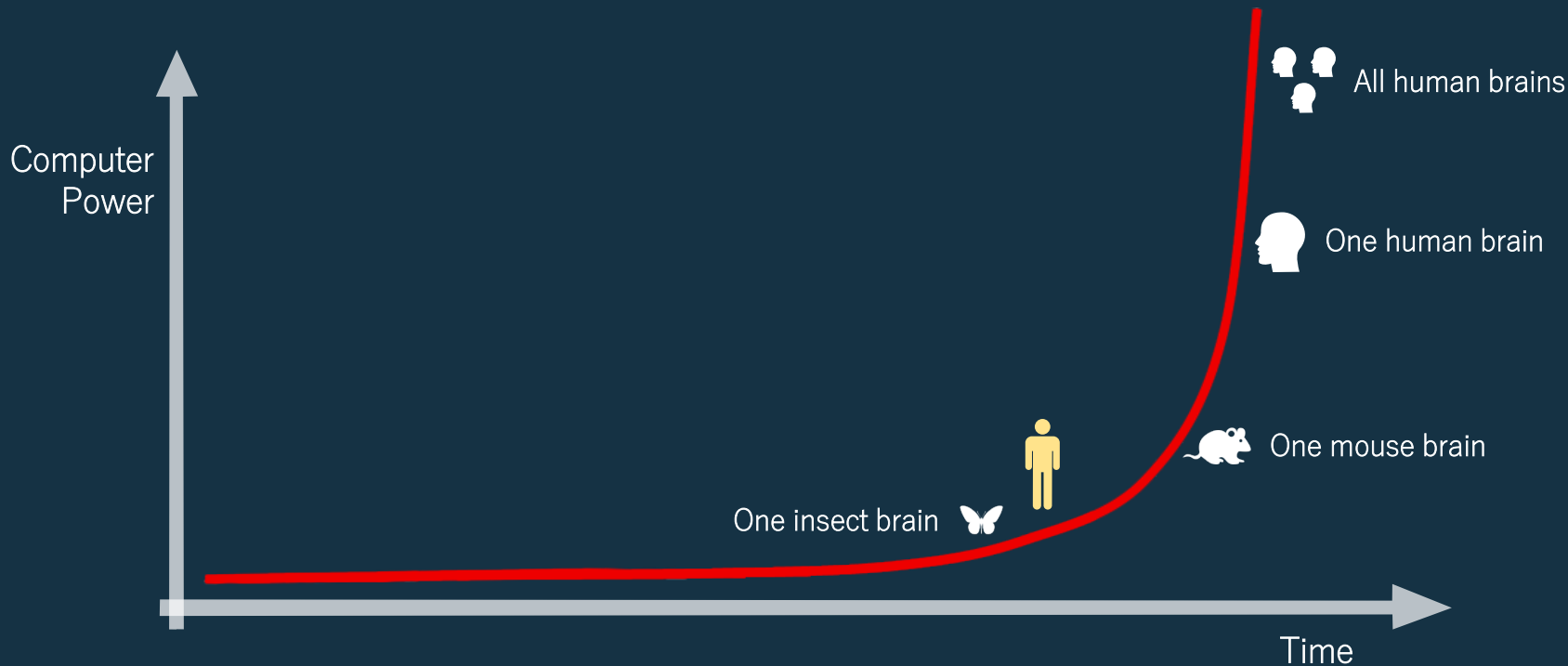
What does exponential growth feel like?

The AI Revolution and the Road to Superintelligence



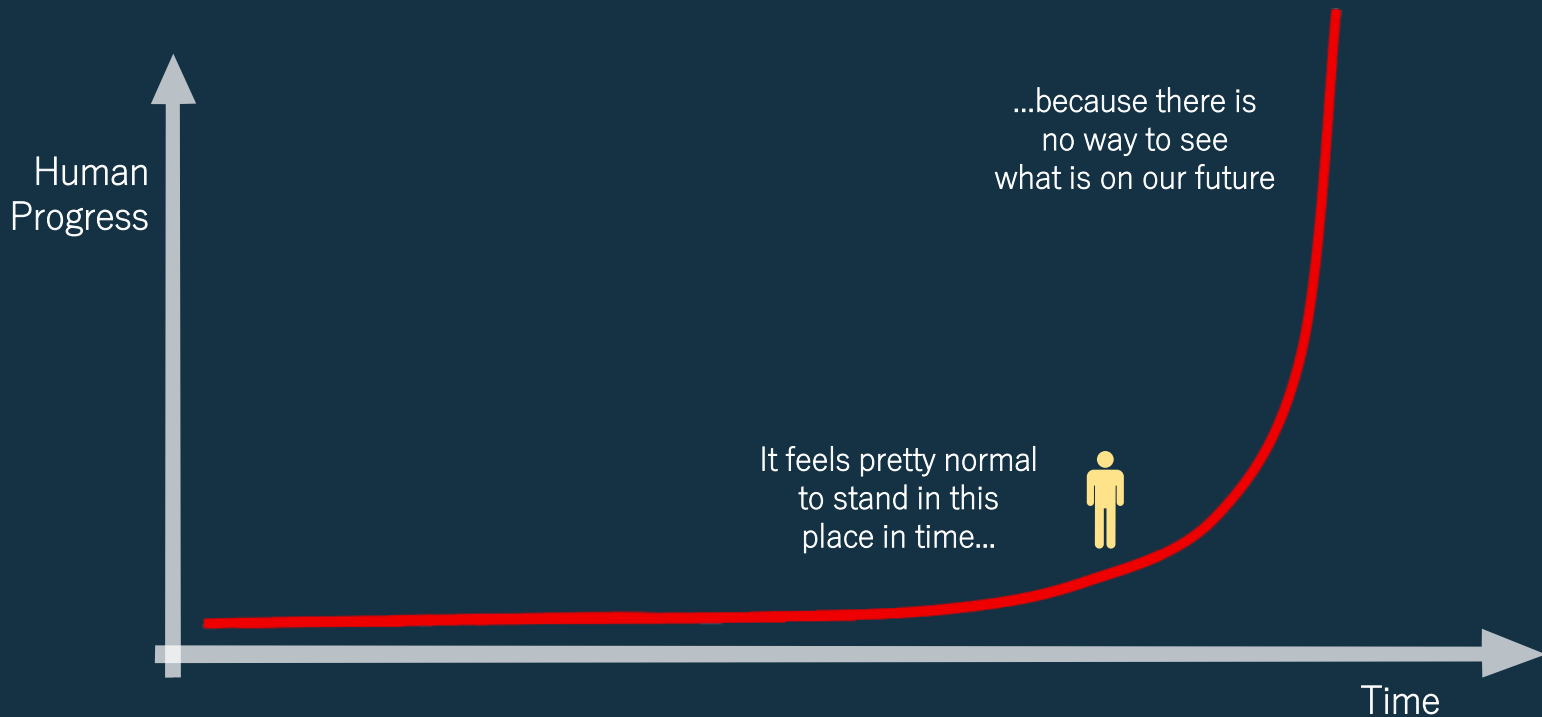
What does exponential growth feel like?

Moore's Law and the Road to Super computer power



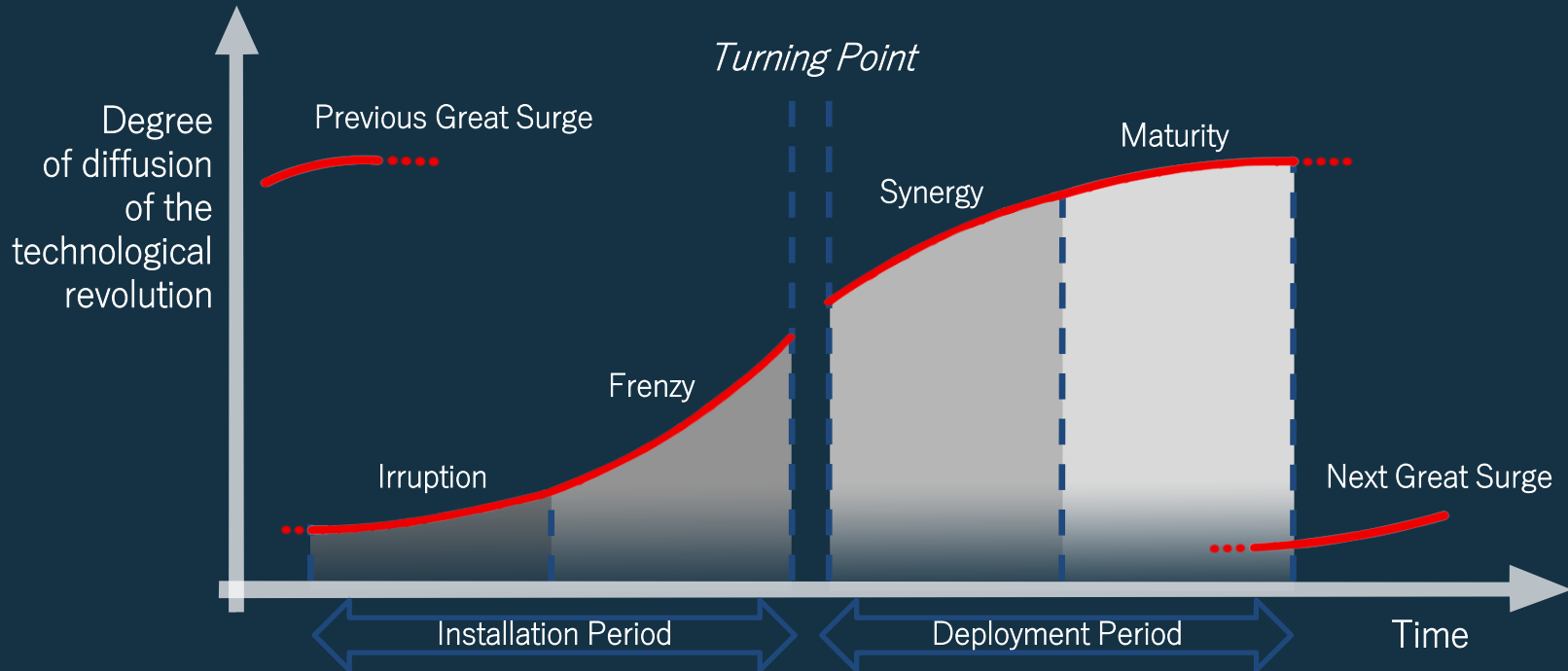
What does exponential growth feel like?

The AI Revolution and the Road to Superintelligence



The techno-economic paradigm (TEP)

TEP is a recurring pattern of cyclical movement: from an initial installation period, through a collapse and recession which signify the turning point, to a full deployment period.



A Blueprint for AI transformation (GR)

Mission & Vision

1. Prepare citizens for the AI transition
2. Improve public service efficiency.
3. Safeguard and enhance democracy.
4. Promote the quality and healthcare for all.
5. Democratize access to, and improve the quality of, education.
6. Turn Greece into an attractive global destination for AI and high- tech investment.
7. Preserve and enrich cultural heritage.
8. Climate mitigation and adaptation.
9. Support national security.



A Blueprint for AI transformation (GR)

Innovation & Entrepreneurship

Recommendations that can lead to:

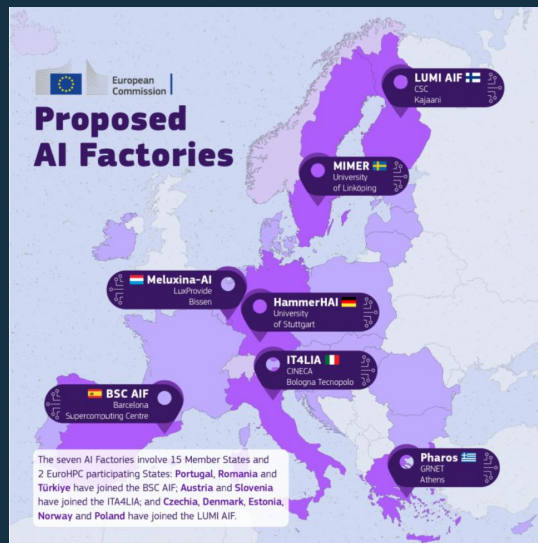
- faster appropriation of AI capabilities across Greek corporate actors;
- **identification of industry verticals or other subdomains** where Greece can credibly develop a competitive advantage in attracting and promoting high added value corporate activity;



Pharos: The Hellenic AI factory

From the AI blueprint to action

Pharos aims to facilitate the implementation of AI solutions by exposing a set of services that accelerate the development of AI products to an open ecosystem. Services related to data storage, platform access, strategic workflows, AI pipelines, use case support, and product operation will be delivered to support application building, model training, AI solutions development, frameworks and tools utilization.



Υποθέσεις Εργασίας ΤΝ στους Μηχανικούς

AI/ML στη Μηχανική κάνει τον κόσμο πιο αποδοτικό και στοχευμένο για τους μηχανικούς και όλους τους εμπλεκόμενους του κλάδου



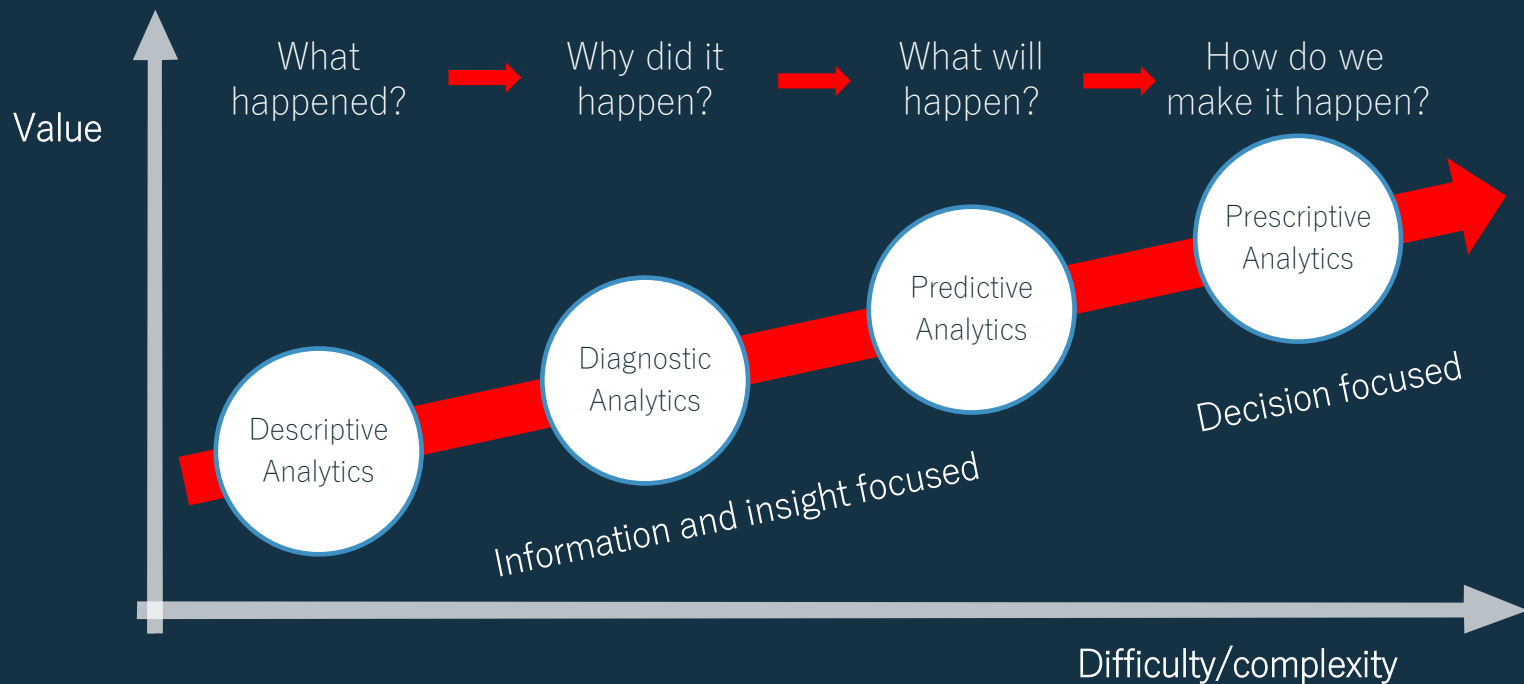
Δεδομένα στη Μηχανική

Data is the new (s)oil?



Δεδομένα στη Μηχανική

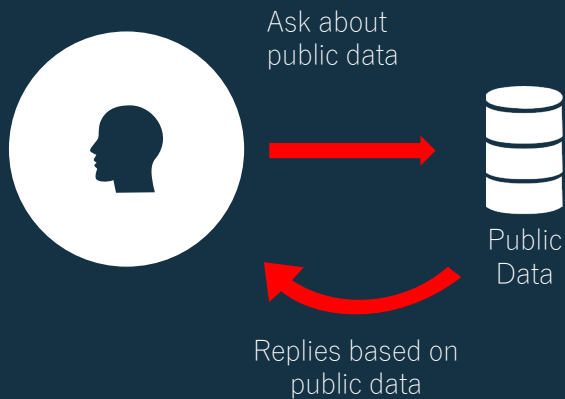
From descriptive to prescriptive data analytics



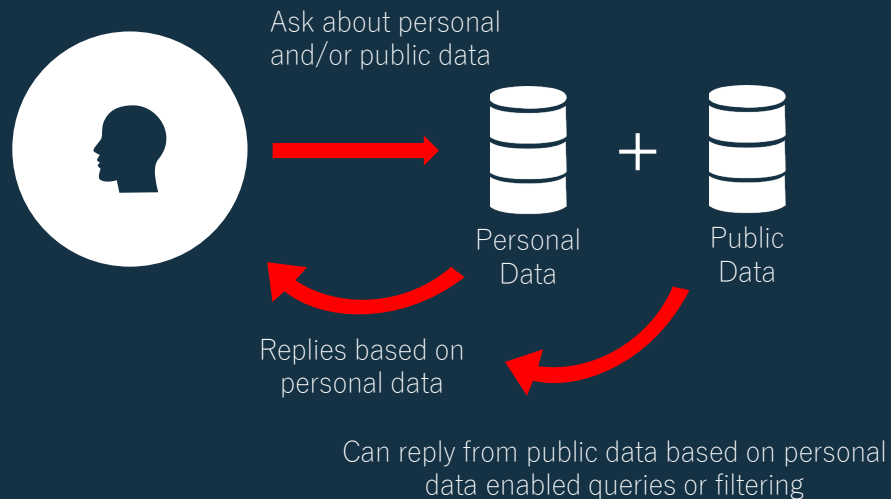
General vs. Personal AI

Delivering informed decisions

General AI



Personal AI



Trade-off

Data



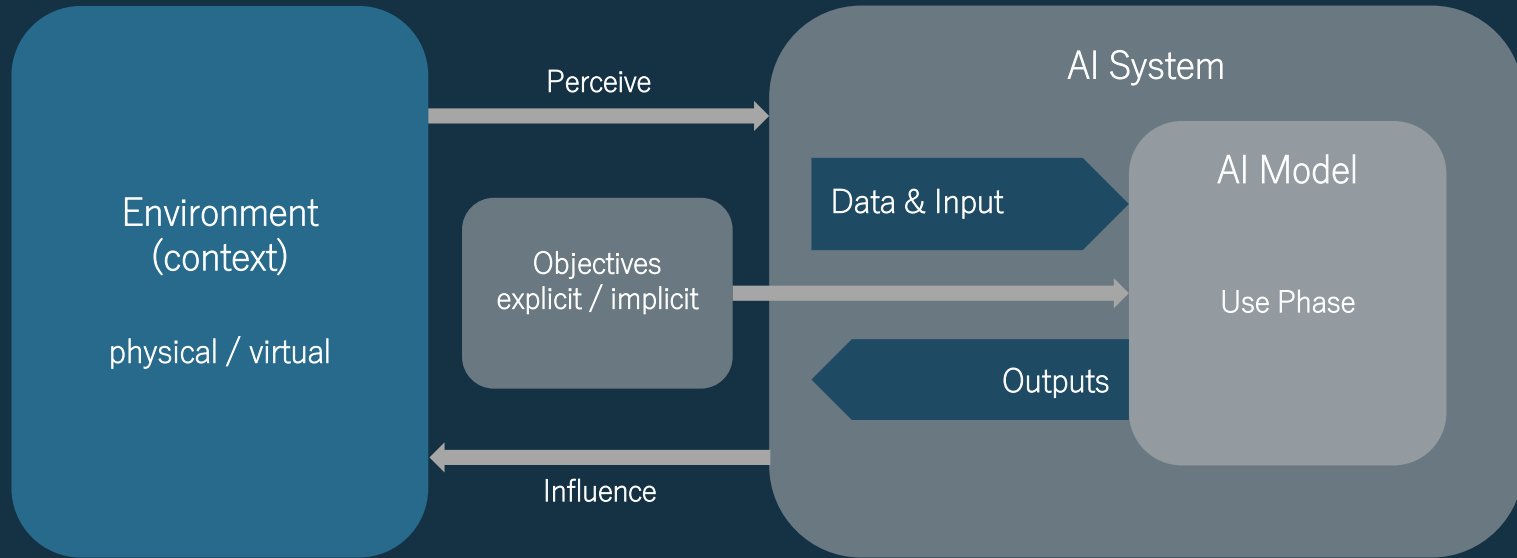
Data others keep about you



Value that you gain from it

AI Act's definition of an AI system

Regulation (EU) 1689/2024 on AI



AI system is a machine-based system, designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that for explicit or implicit objectives infers from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments.

AI system basics

Machine based

Varying levels of autonomy

Adaptiveness after deployment

for explicit or implicit objectives

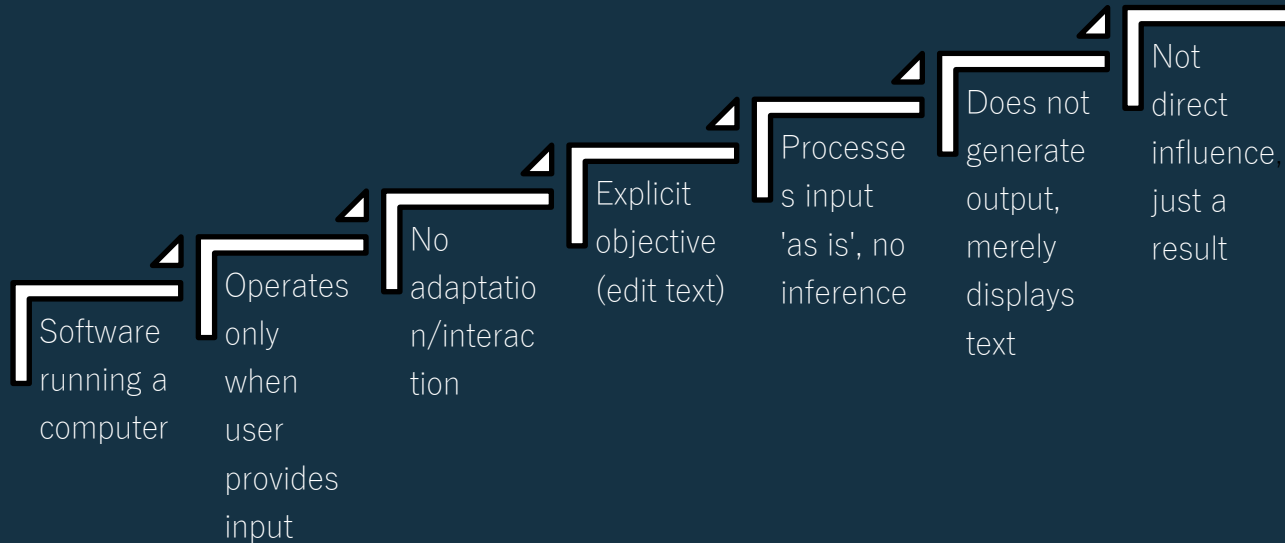
Infers from input how to generate output

Predictions, content, recommendations, decisions

can influence physical or digital environment

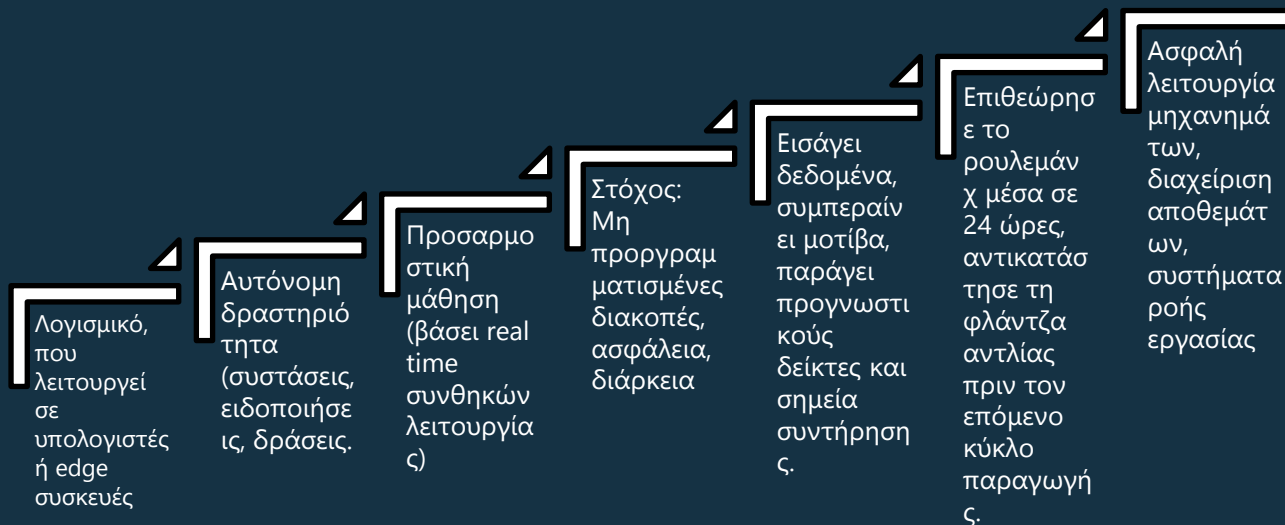
Text Editor

Not an AI System [Regulation (EU) 2024/1689 on AI]



Σύστημα προληπτικής συντήρησης

AI System [Regulation (EU) 2024/1689 on AI]



Trade-off

Algorithms



Algorithmic performance

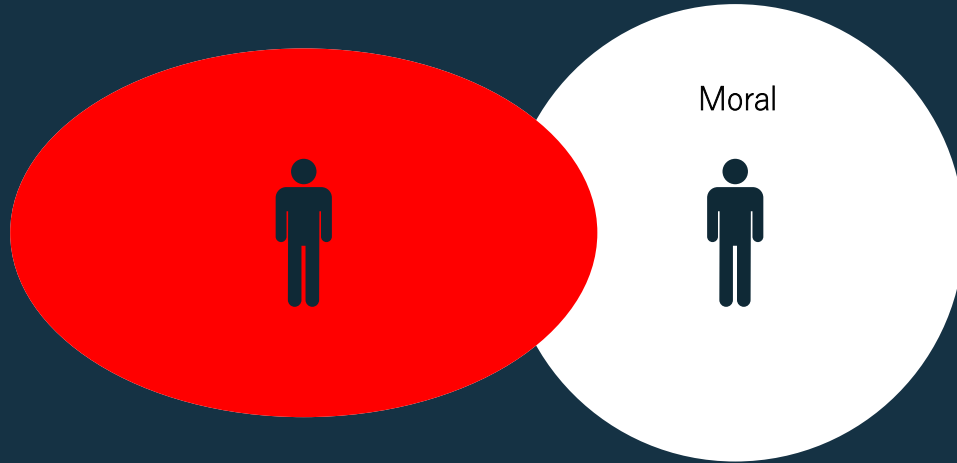


Task optimization & impact

Moral Agent Types

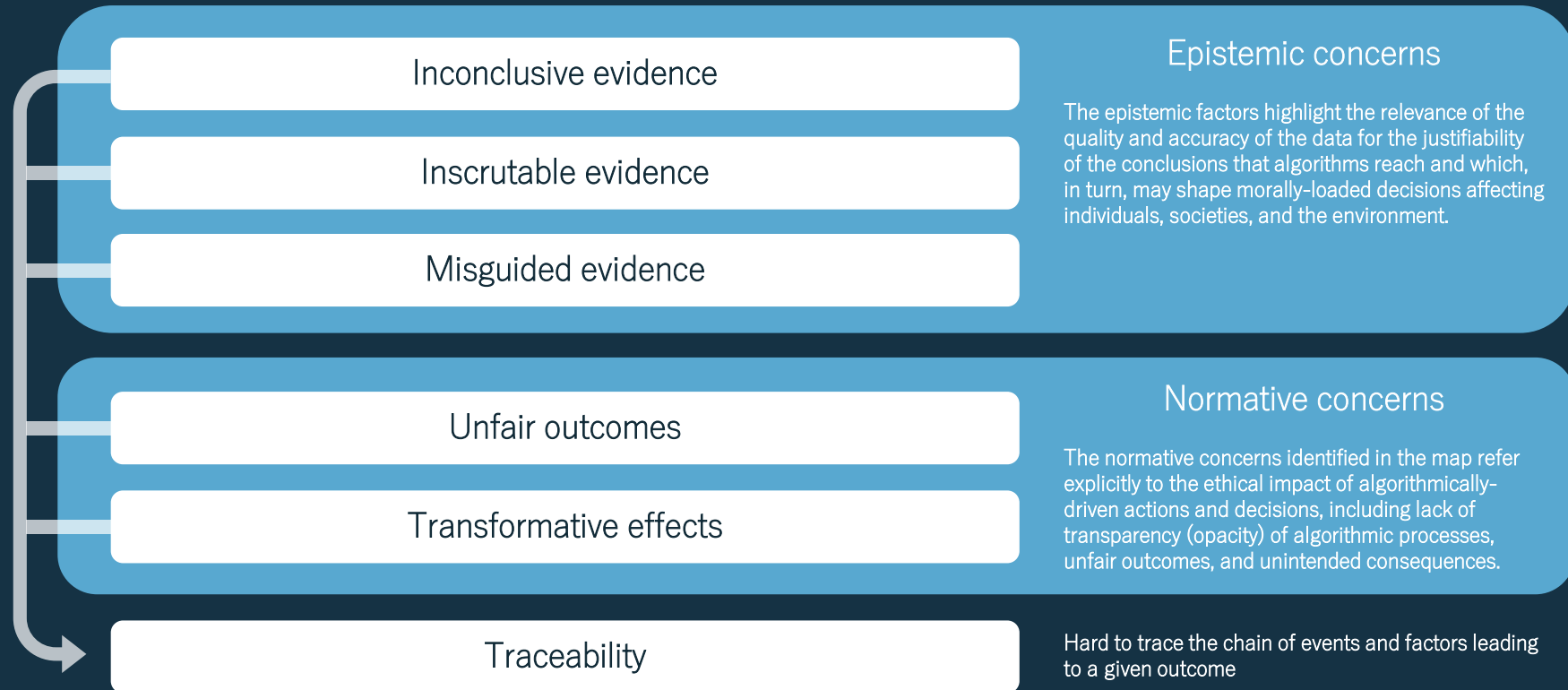
How agents incorporate morality

- Moral agents (ethics by virtue and values)
- Immoral (ethical norms, values, laws upstream)
- Ammoral (no sense of morality at all)



Ethical concerns raised by algorithms

Six types of ethical concerns raised by algorithms



AI Risks



Fundamental Rights

EU Charter of Fundamental Rights

- Right to non discrimination,,
- Human dignity
- Privacy & Personal data protection
- Freedom of expression



Safety Risks

- For users when AI technologies are embedded in products and services

AI Risks



Unacceptable

- Distorting social behavior
- Exploiting vulnerabilities (group/persons)
- Social scoring (by authorities)
- Real time biometric identification



High Risk

- Products subject to EU safety rules (Annex II)
- Various categories (Annex III)



Low Risk

- AIS that interact with humans (eg chatbots)
- Emotion recognition / biometric categorization
- Deepfakes

AI high risk (art.6.3, Annex III)

AI Act latest version 5662/24 (January 26, 2024)



Biometrics

Remote biometric identification/classification systems;



Critical Infrastructure

AIS intended to be used as safety components in the management and operation of critical infrastructure, (i.e. road traffic, supply of water, gas, Heating, electricity and the digital);



Education and vocational training

- AIS intended to be used to determine access, admission or to assign natural persons to educational and vocational;
- AI systems intended to be used to evaluate learning outcomes;

Standards & Trustmarks

Based on the developments of the JTC21 Advisory Group

- The European Commission has issued a “standardization request” to the European standards bodies (CEN and CENELEC) listing a series of topics for which new harmonized standards are required to cover the compliance obligations in the AI Act.
- Following the liaison between IEEE and ISO/JTC1 SC42 there might be wider synergies with the CEN CENELEC bodies adding further momentum to provide technical specifications in support of the AI Act*

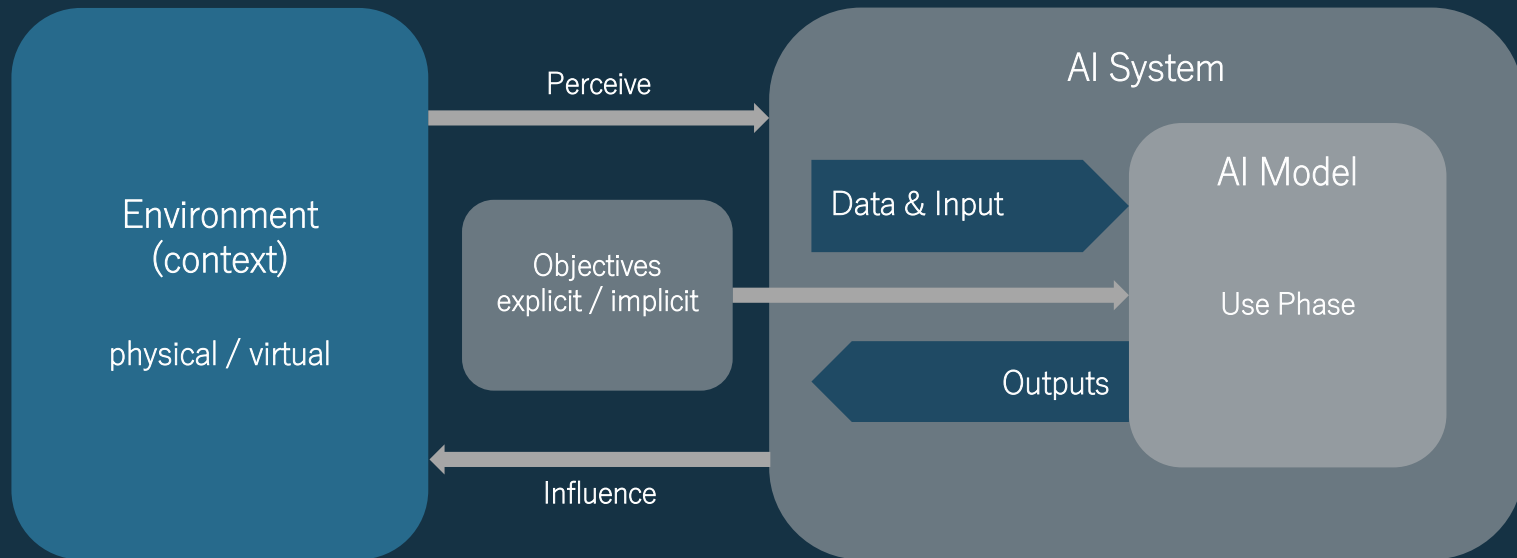
Works so far:

- ISO/IEC CD 42001 ‘Artificial Intelligence Management Systems
- ISO/IEC JTC 1/SC 42 ‘Artificial Intelligence
- IEEE P7000 series
- IEEE CertifAIEd (Certification program for assessing ethics)

* source: Soler Garrido, J., Tolan, S., Hupont Torres, I., Fernandez Llorca, D., Charisi, V., Gomez Gutierrez, E., Junklewitz, H., Hamon, R., Fano Yela, D. and Panigutti, C., AI Watch: Artificial Intelligence Standardisation Landscape Update, EUR 31343 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-76-60450-1, doi:10.2760/131984, JRC 131155.

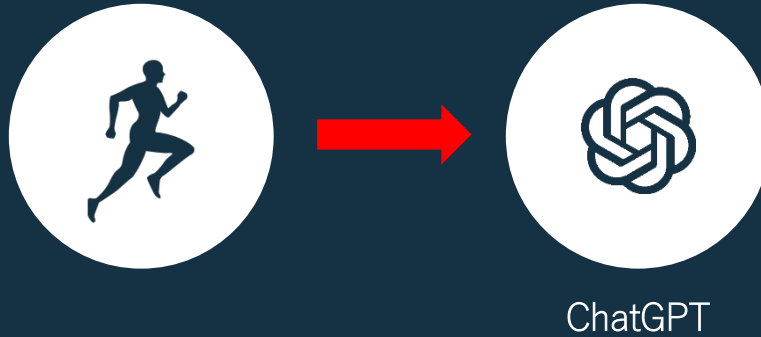
AI Act's definition of GPAI models

It refers to the latest version 5662/24 (January 26, 2024)



General Purpose Artificial Intelligence (GPAI) models are trained with a large amount of data using self supervision at scale, that display significant generality and are capable to competently perform a wide range of distinct tasks and can be integrated into a variety of downstream systems and applications. They serve a variety of purposes (direct or integrated)

Can we ask ChatGPT?



Ένας διάλογος..

Ανάλυση δεδομένων τοπογραφικής έρευνας



Hi AI! Έχω συλλέξει δεδομένα υψομέτρων από μία περιοχή και θέλω να υπολογίσω τον όγκο του εδάφους που πρέπει να αφαιρεθεί για να ισοπεδωθεί η περιοχή. Μπορείς να με βοηθήσεις;

Έχω ένα αρχείο CSV με τις συντεταγμένες (x, y) και το υψόμετρο (z).



Εισαγωγή Δεδομένων: Πες μου σε ποια μορφή έχεις τα δεδομένα (π.χ. συντεταγμένες με υψόμετρα σε αρχείο CSV).

Καθορισμός Στόχου Υψομέτρου: Ποιο είναι το επιθυμητό ύψος στο οποίο θέλεις να ισοπεδωθεί η περιοχή; Έχεις τα δεδομένα σε κάποιο αρχείο ή θέλεις καθοδήγηση για να τα εισαγάγεις;

Ένας διάλογος..

Ανάλυση δεδομένων τοπογραφικής έρευνας



Το επιθυμητό υψόμετρο είναι 100 μέτρα.

Μπορείς να μου δώσεις κώδικα Python για αυτοματοποίηση του υπολογισμού;



Μπορείς να υπολογίσεις τον όγκο σε τρία βήματα:

Διαίρεση Περιοχής: Διαιρέστε την περιοχή σε μικρά τμήματα (π.χ., πρίσματα ή τετράγωνα).

Υπολογισμός Διαφοράς Υψομέτρου: Για κάθε τμήμα, αφάιρεσε το στόχο από το καταγεγραμμένο υψόμετρο.

Αν $z > 100$, $z > 100$, υπολόγισε τον όγκο του εδάφους που πρέπει να αφαιρεθεί.

Αν $z < 100$, $z < 100$, υπολόγισε τον όγκο του εδάφους που πρέπει να προστεθεί.

Άθροισμα Όγκων: Χρησιμοποίησε τον τύπο για τον όγκο ενός πρίσματος: $V = \text{Εμβαδόν βάσης} \times \text{Διαφορά ύψους}$.

Ένας διάλογος..

Ανάλυση δεδομένων τοπογραφικής έρευνας



Μπορείς να μου δώσεις κώδικα Python για αυτοματοποίηση του υπολογισμού;

Τέλεια! Τόχω ☺



```
# Φόρτωση το αρχείο CSV data = pd.read_csv('elevation_data.csv') #
Αντικατέστησε με τη διαδρομή του αρχείου σου # Καθορισμός
επιθυμητού υψομέτρου target_level = 100 # Υπολογισμός διαφοράς
ύψους data['height_diff'] = data['z'] - target_level cell_area = 1 #
Αντικατέστησε με το εμβαδόν ενός τμήματος σε τετραγωνικά μέτρ
α data['volume'] = data['height_diff'] * cell_area # Υπολογισμός όγκω
ν εκσκαφής και επιχώσεων excavation_volume = data[data['height_d
iff'] > 0]['volume'].sum() fill_volume = data[data['height_diff'] < 0]['vol
ume'].
sum() print(f"Όγκος Εκσκαφής: {excavation_volume} κυβικά μέτρα")
print(f"Όγκος Επιχώσεων: {abs(fill_volume)} κυβικά μέτρα")
```

Trade-offs (agency & autonomy)

Behavioral Engineering



Overreliance to AIS



Value that you gain from it

Trade-offs (IPR)

Confidentiality & Trade Secrets



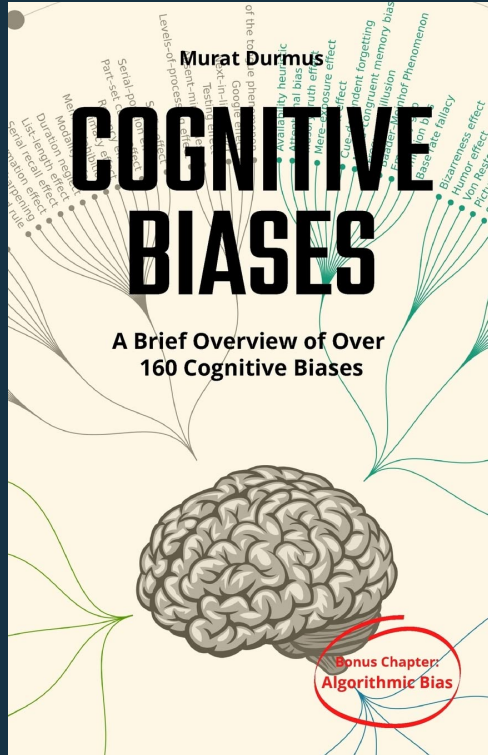
Information Display



Secure Output

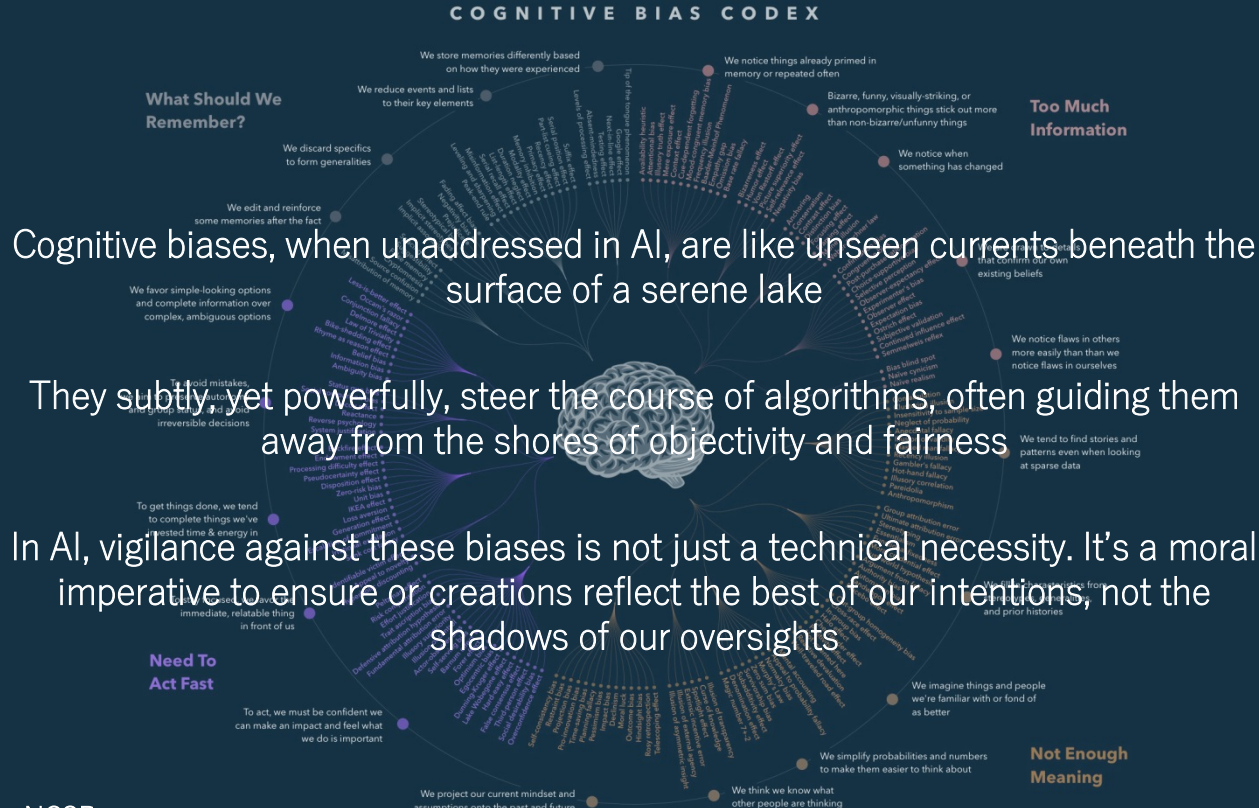
Cognitive Biases

Cognitive Biases; A Brief Overview of Over 160 Cognitive Biases



Murat Durmus

Cognitive Biases



Cognitive Biases

Which biases should you pay particular attention to when developing AI based systems



Cognitive Biases

Confirmation Bias

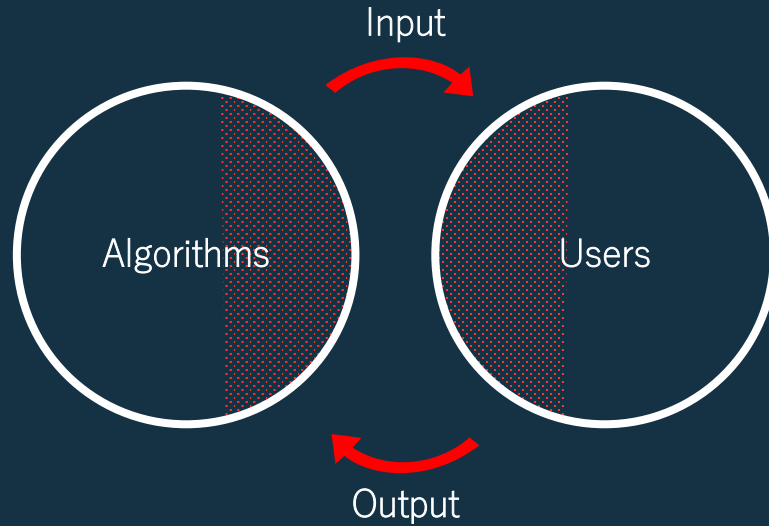
This is the tendency to seek, interpret, and remember information that confirms pre-existing beliefs. In AI, this might lead to a model being trained on data that reinforces the biases of its developers or the data source rather than representing a balanced perspective.



Cognitive Biases

Algorithmic Bias

This refers to biases that arise from the algorithms themselves, including how data are collected, coded, selected, or used in training. Algorithms can perpetuate or amplify societal and cultural biases if not carefully monitored and corrected.



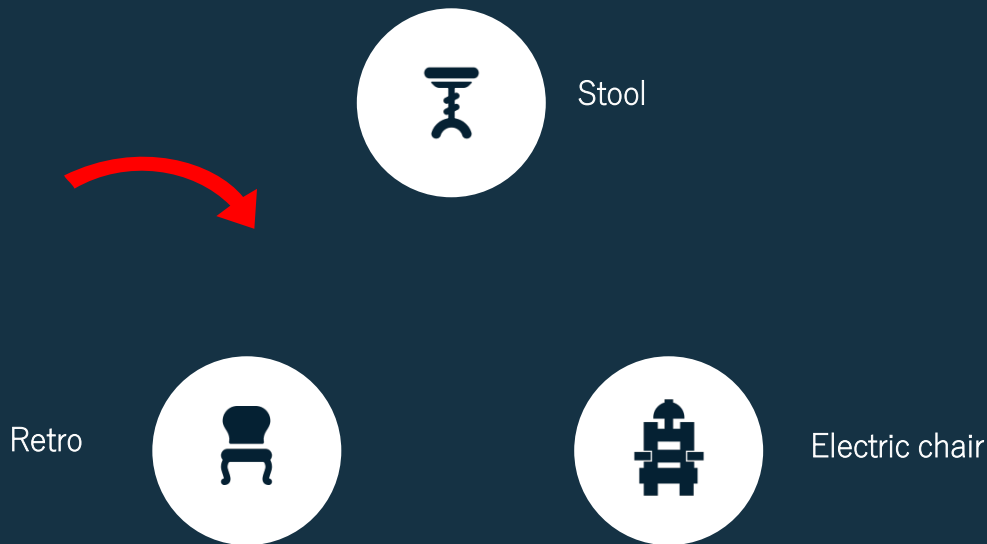
The Logic of Information

Why a ML model gives a certain output? Logic of design as a logic of requirements

$\{R1...Rn\}$ = System
Chair for one person

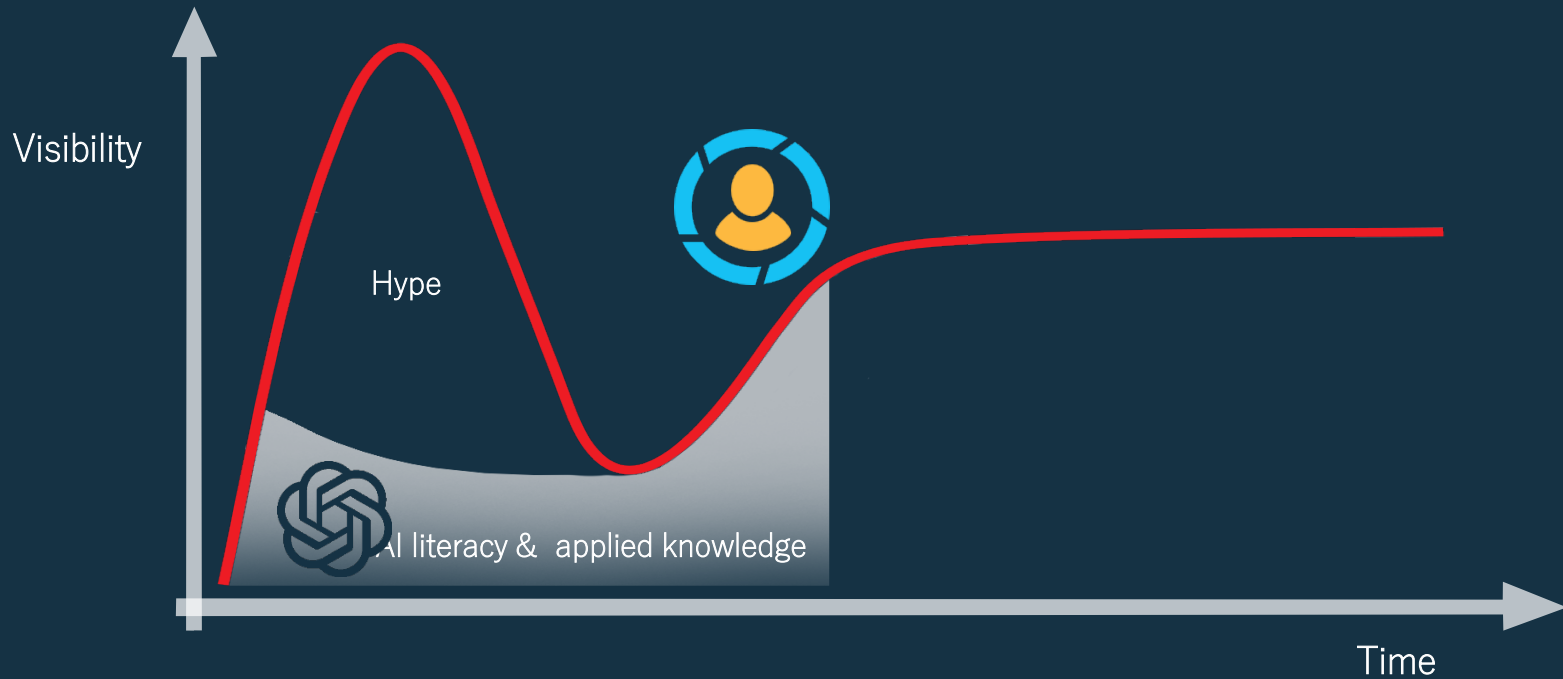
Identify the requirements
of a system and how
these requirements
'sufficientize' our solution:

- Context
- Level of Abstraction
- Purpose



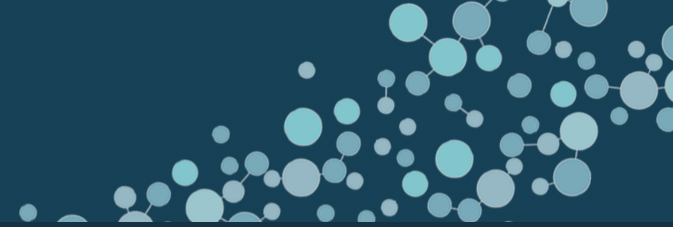
Amara's law

It takes a lot of time to apply knowledge in a productive way, beyond the initial hype



Trade-offs

AI in engineering



Replace engineers



Complement & augment



Thank You

Alexandros Nousias



Alexandros Nousias



@alexnousias



<https://www.linkedin.com/in/alexandrosnoussias/>



alexandros.nousias@gmail.com