ε.δε.μ²
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

infostrag

Η Επανάσταση της Τεχνητής Νοημοσύνης & Προκλήσεις της Ακαδημαϊκής Κοινότητας

Βασίλης Μάγκλαρης

Ομότιμος Καθηγητής Ε.Μ.Π.
Σχολή Ηλεκτρολόγων Μηχ. & Μηχ. Υπολογιστών
maglaris@netmode.ntua.gr www.netmode.ntua.gr

20ο Σεμινάριο της Ερμούπολης για την Κοινωνία της Πληροφορίας & την Οικονομία της Γνώσης

Συνεδριακή Αίθουσα Επιμελητηρίου Κυκλάδων
Παρασκευή 11 Ιουλίου 2025

Founding Fathers of Artificial Intelligence

Thomas Bayes (1701 -1761): Combinatorial Probabilities, Statistical Inference
https://en.wikipedia.org/wiki/Thomas_Bayes

Johann Carl Friedrich Gauss (1777 -1855): Statistical Inference, Distributions of Sample Data
https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss

Josiah Willard Gibbs (1839 -1903): Statistical Mechanics, Thermodynamics
https://en.wikipedia.org/wiki/Josiah_Willard_Gibbs

Ludwig Boltzmann (1844 -1906): Statistical Mechanics, Thermodynamics
https://en.wikipedia.org/wiki/Ludwig_Boltzmann

Andrey Markov (1856 -1922): Probability Theory, Stochastic Processes
https://en.wikipedia.org/wiki/Andrey_Markov

Alan Turing (1912 -1954): Computing Machinery, Codes, Artificial Intelligence, Logic
https://en.wikipedia.org/wiki/Alan_Turing

John von Neumann (1903 -1957): Statistical Modelling, Game Theory, Entropy
https://en.wikipedia.org/wiki/John_von_Neumann

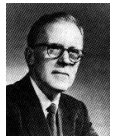
Andrey Kolmogorov (1903 -1987): Probability Theory
https://en.wikipedia.org/wiki/Andrey_Kolmogorov

Richard Bellman (1920 - 1984): Dynamic Programming
https://en.wikipedia.org/wiki/Richard_E._Bellman

Fathers of Arificial Intelligence - Machine Learning

Nicholas Metropolis - Μητρόπουλος (1915 - 1999): Monte Carlo Simulations, Simulated Annealing
https://en.wikipedia.org/wiki/Nicholas_Metropolis

Donald Hebb (1904 - 1985): Neurophysiology, Learning Rules
https://en.wikipedia.org/wiki/Donald_O._Hebb



Frank Rosenblatt (1928 - 1972): Psychology & Artificial Intelligence (AI), Neural Networks - Perceptron
https://en.wikipedia.org/wiki/Frank_Rosenblatt

David Rumelhart (1942 - 2011): Psychology & Artificial Intelligence (AI), Back Propagation Algorithm
https://en.wikipedia.org/wiki/David_Rumelhart

Vladimir Vapnik (1936): Statistical Learning, Support Vector Machines (SVM)
https://en.wikipedia.org/wiki/Vladimir_Vapnik

Teuvo Kohonen (1934 - 2021): Self-Organizing Maps (SOM)
https://en.wikipedia.org/wiki/Teuvo_Kohonen

John Hopfield (1933): Physics, Biology, Recurrent Neural Networks (RNN) (***Nobel Prize in Physics, 2024***)
https://en.wikipedia.org/wiki/John_Hopfield

Geoffrey Hinton (1947): Physics, Boltzmann Machines, Deep Belief Networks (***Nobel Prize in Physics, 2024***)
https://en.wikipedia.org/wiki/Geoffrey_Hinton

Demis Hassabis (1976): AI Research, Protein Structure Prediction (***Nobel Prize in Chemistry, 2024***)
https://en.wikipedia.org/wiki/Demis_Hassabis

Definitions: AI & ML

Artificial Intelligence – AI (Τεχνητή Νοημοσύνη):
Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind

IBM: https://www.ibm.com/topics/artificial-intelligence

Machine Learning – ML (Μηχανική Μάθηση):
Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy

IBM: https://www.ibm.com/topics/machine-learning

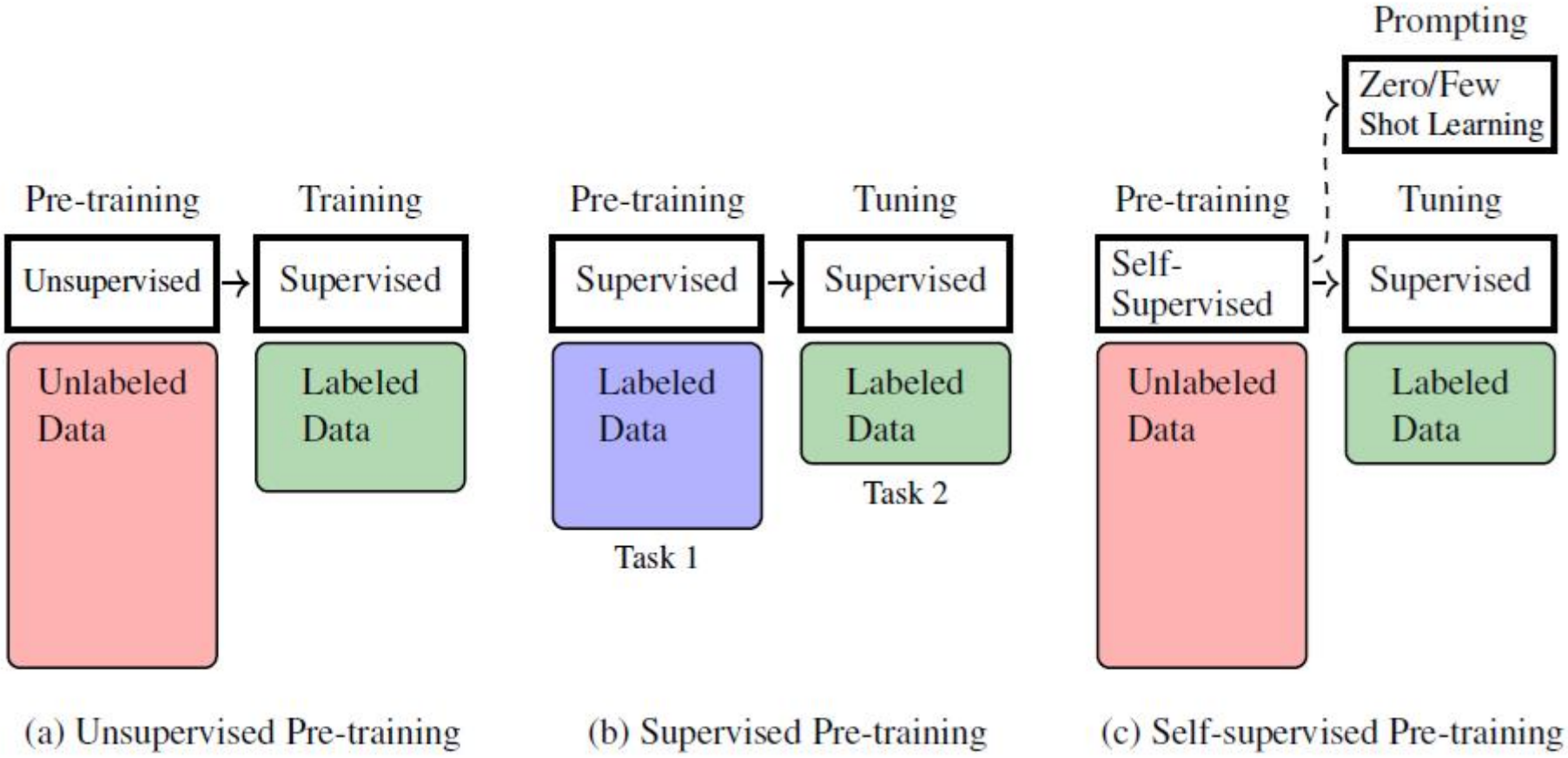
Large Language Models (LLM)

https://en.wikipedia.org/wiki/Large_language_model

- **Current hype** involving combination of ***Natural Language Processing*** (**NLP**) and ***Artificial Intelligence*** (**AI**) - ***Machine Learning*** (**ML**) fields
- Builds on years R&D in **NLP** (e.g. BERT Models, ***Search Engines***, ***Automatic Translation***, ***Chatboxes***…) and **ML** (***Deep Learning***, ***Generative Models***, ***Autoencoders***, ***Transformers***…)
- Tuning of ***billions*** of parameters (***synaptic*** weights, NLP ***tokens***, ***embedding*** vectors…)
- Use of ***Unsupervised***, ***Supervised***, ***Self-supervised*** pre-training and ***Reinforcement Learning*** algorithms
- Deployment of extensive ***data-centers***, with very high ***energy*** requirements
- Need extensive resources, usually offered to end-users by ***Computing Clouds*** as ***S***oftware-***a***s-***a***-***S***ervice (**AaaS**), with downloading options
- Very lengthy pre-training for corpus (***foundation***) model, possible customization for specific use-cases
- Raised legal ***regulatory*** matters (property rights, confidentiality, openness), ***ethical*** & ***geopolitical*** concerns, far-reaching effects of ***labor realignment***, challenges that humanity faced in previous industrial and technological revolutions (reminiscent of violent ***Luddism*** reactions in the early 19th century against proliferation of looming machines etc.)

Large Language Models (LLM)

https://en.wikipedia.org/wiki/Large_language_model



For a comprehensive review see the **2025** book "***Foundations of LLMs***" by ***Tong Xiao*** & ***Jingbo Zhu***, NLP Labs, Northeastern University – China, https://arxiv.org/pdf/2501.09223)

Large Language Models (LLM)

https://arxiv.org/pdf/2501.09223

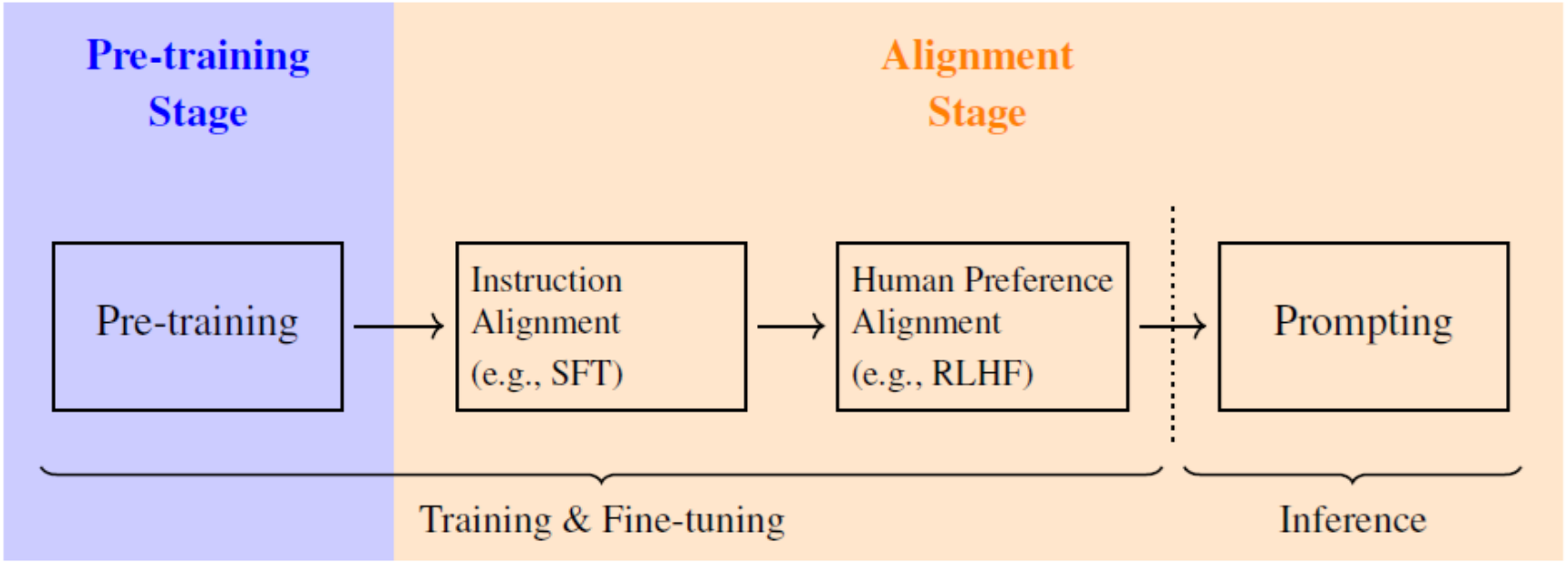
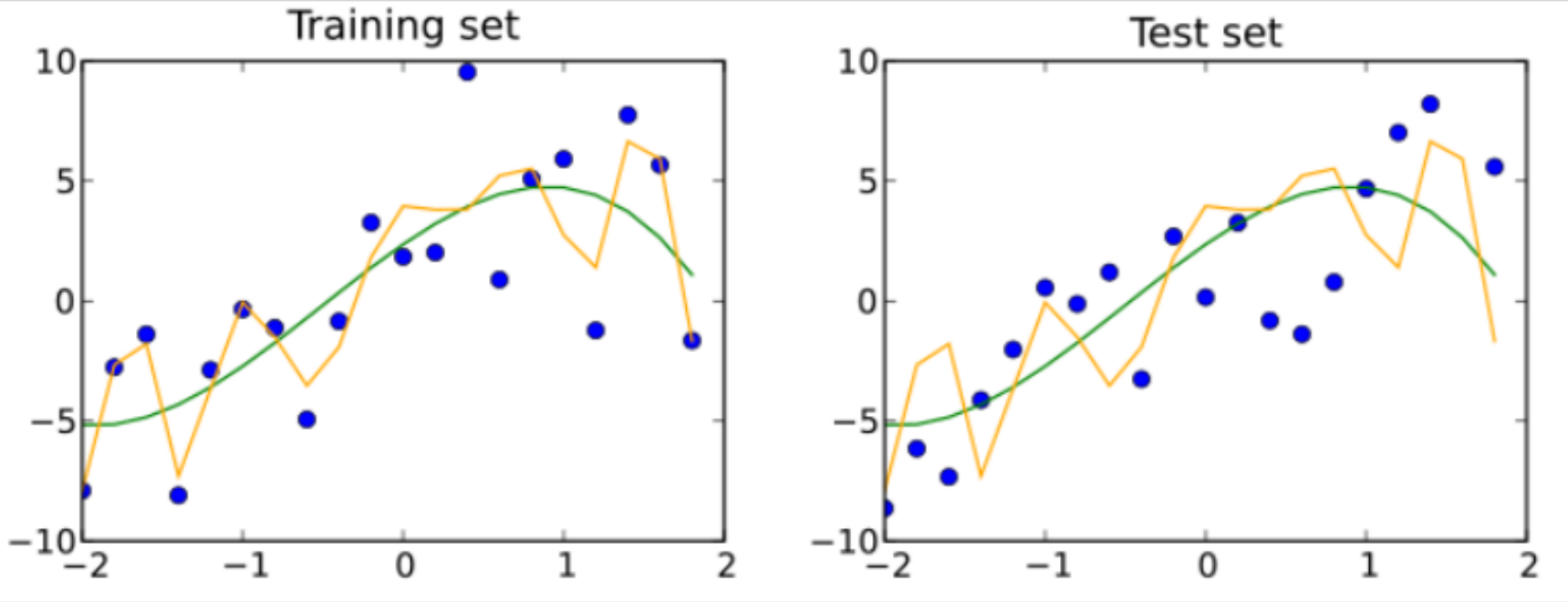


Fig. 4.1: Schematic illustration of the pre-train-then-align method for developing LLMs. In the pre-training stage, we train an LLM on vast amounts of data using next token prediction. Then, in the alignment stage, we align the LLM to user instructions, intents, and preferences. This includes instruction alignment, human preference alignment, and prompting.

Datasets - Illustration of Overfitting

https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets



A training set (left) and a test set (right) from the same statistical population are shown as blue points. Two predictive models are fit to the training data. Both fitted models are plotted with both the training and test sets. In the training set, the MSE of the fit shown in orange is 4 whereas the MSE for the fit shown in green is 9. In the test set, the MSE for the fit shown in orange is 15 and the MSE for the fit shown in green is 13. The orange curve severely overfits the training data, since its MSE increases by almost a factor of four when comparing the test set to the training set. The green curve overfits the training data much less, as its MSE increases by less than a factor of 2.

Discriminative Machine Learning Models

Definition:

Classification or *Regression* (estimation) of *data elements* via *conditional probability* estimates of plausible outputs (*label*) given *input sample* elements, based on what the model learns by iteratively feeding sample elements of a *training dataset* and checking *generalization* by applying elements of a *testing dataset*

Applications:

- Classification of sample elements based on their characteristics (features)
- Pattern recognition based on principal sample element features
- Medical imaging, diagnostics semi-automation tools
- Prediction (regression) of output based on pre-stored pairs of input-output elements

Generative AI Models

Definition:

Generation of sample elements conforming to joint input-output statistics estimated by iterative input of *training sample* elements from which the system infers *joint probabilities* of the *output* with input *features* (virtual reality output, risk of hallucinations)

Applications:

- Bayes classifiers, a very popular and simple generative classification method
- Current hype, with massive training datasets and lengthy training times (months!), expensive environmentally hazardous datacenters requirements, cloud-hosted multiprocessor GPUs (Graphics Processing Units) and highly specialized staffing
- Generation of text elements and chatboxes based on *Large Language Models* - *LLM*, text translation, voice recognition, production of simulated (virtual reality) images, idealized background screens, animated cartoons…
- Extensive training of *Search-Engines* (*Google*, *MS Bing*…), *OpenAI* - *MS ChatGPT* (*Chat Generative Pre-trained Transformer*), *DeepSeek* chatbox…
- Offered As-a-Service (*AaS*) to customers, stirring fierce competition for supremacy amongst the US, China, Europe (?)

Causes-and-Effects of the Artificial Intelligence Revolution

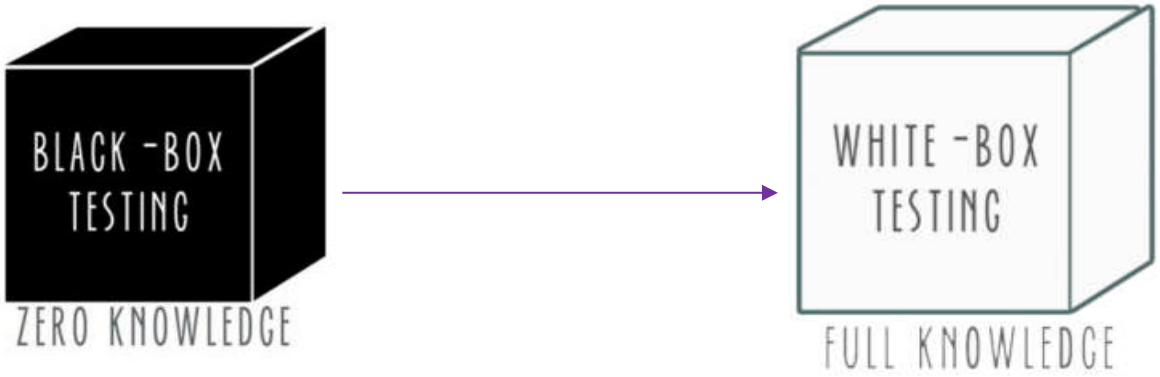
- The cataclysmic developments of distributed (*cloud*) computing and storage infrastructures enables extremely complex algorithms of statistical inference and stochastic optimization, based on large historical datasets

- Processing of multi-dimensional huge data (*big data*) with a massive number of characteristics (*features*) triggers novel data mining algorithms to *estimate*, *predict*, *classify* and *generate* new sample elements, statistically close to pre-stored historical data

- The ever-deepening understanding of learning methods in biological systems, leads to emulation of *Human Intelligence* via *Artificial Intelligence* algorithms that characterize or generate new instances, always with some *error probability* (expected in statistical inference decisions) and hopefully minute danger to lead to hazardous *hallucinations*

- The advances in *Natural Language Processing* (*NLP*) and *Text Processing* fields, coupled with technology breakthroughs and the ubiquitous *Internet* availability, lead to generative massive models often referred to as *Large Language Models* (*LLMs*), with human-friendly attributes and tremendous commercial potential and geopolitical might

eXplainable Artificial Intelligence (XAI)



- Just ***accuracy*** may not justify ***Artificial Intelligence*** – ***Machine Learning*** choices
- Mistrust of users – analysts – operators – regulators in ***uninterpretable*** decision systems of ***black-box*** methods
- Necessity of ***eXplainable Artificial Intelligence*** (***XAI***) methods in selecting models and parameter tuning algorithms
- Justification of ***ML*** recommendations required to users – clients on their personal profiles (***local interpretations***)
- Required ***feature engineering*** and model justification choices. Current wide-spread adoption by major commercial ***LLMs*** of ***reasoning*** of parameter setting, leading to development of user-friendly GUI's (leveraging on GenAI and GPUs)

Risks of the Artificial Intelligence Revolution

- Artificial Intelligence exhibits risks associated with all (r)evolutions (e.g. widening of global inequalities, re-alignment of work-force, new employment rules) and new challenges (e.g. how to protect *Individual Privacy Rights* – *IPR* & enforce *Intellectual Property*)
- Use of (Generative) AI to spread fake news, promote plagiarism, infringe Intellectual Property (e.g. unauthorized use of Wikipedia texts by *OpenAI* for *ChatGPT* training)
- Humanity will cast *regulations* (e.g. 2016/6/91 EU *General Data Protection Regulation* - *GDPR*) to harness use of big data and smart algorithms (*perhaps a wishful thinking*…)

Risks of the Artificial Intelligence Revolution

- Artificial Intelligence exhibits risks associated with all (r)evolutions (e.g. widening of global inequalities, re-alignment of work-force, new employment rules) and new challenges (e.g. how to protect *Individual Privacy Rights* – *IPR* & enforce *Intellectual Property*)
- Use of (Generative) AI to spread fake news, promote plagiarism, infringe Intellectual Property (e.g. unauthorized use of Wikipedia texts by *OpenAI* for *ChatGPT* training)
- Humanity will cast *regulations* (e.g. 2016/6/91 EU *General Data Protection Regulation* - *GDPR*) to harness use of big data and smart algorithms (*perhaps a wishful thinking*…)

Geoffrey Hinton: British – Canadian, Born in London UK 1947

- 1977: Ph.D., University of Edinburgh, Scotland, UK
- Academic Career UK, USA, Canada
- Pioneer in Neural Networks research (Boltzmann Machines, Deep Belief Networks, Generative AI…)
- 2013-2023: Scientific Advisor of Google & Professor, University of Toronto
- May 2023: Resigned from Google to freely speak of uncontrollable AI risks
- **Sept. 2024**: ***Nobel Prize in Physics***

NY Times, May 2023 on *G. Hinton*: *"The Godfather of A.I." Leaves Google and Warns of Danger Ahead: Generative A.I. can already be a tool for misinformation. Soon, it could be a risk to jobs. Somewhere down the line, tech's biggest worriers say, it could be a risk to humanity*
https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html

Responsibilities of the Academic Community

- Train a generation of skilled *AI* - *ML* professionals, with understanding of fundamentals of computing & statistical learning, beyond glorified *Python Programmers*
- Promote inter-disciplinary *AI* potential, while alerting on associated deficiencies and risks
- Emphasize on IPR protection, condemn illegal & immoral applications, avoid war games
- Continue raising environmental concerns for uncontrollable deployment of huge, power hungry *data centers*
- Participate in structuring the regulatory frameworks to protect user rights, while enforcing open-access to data from (*anonymized*) case-studies
- The *ML – AI* methods are building on statistical models with error probabilities & hallucinations:
  - The output of *Artificial Intelligence* requires thorough controls and verification by trained *Human Intelligence*
  - Automation will complement but not substitute (at least in the foreseeable future) the expertise and unlimited potential of human brain
- Let's avoid the *Luddism* exaggerations of the Industrial Revolution of the late 1800's !!!

Responsibilities of the Academic Community

- Train a generation of skilled *AI* - *ML* professionals, with understanding of fundamentals of computing & statistical learning, beyond glorified *Python Programmers*
- Promote inter-disciplinary *AI* potential, while alerting on associated deficiencies and risks
- Emphasize on IPR protection, condemn illegal & immoral applications, avoid war games
- Continue raising environmental concerns for uncontrollable deployment of huge, power hungry *data centers*
- Participate in structuring the regulatory frameworks to protect user rights, while enforcing open-access to data from (*anonymized*) case-studies
- The *ML – AI* methods are building on statistical models with error probabilities & hallucinations:
  - The output of *Artificial Intelligence* requires thorough controls and verification by trained *Human Intelligence*
  - Automation will complement but not substitute (at least in the foreseeable future) the expertise and unlimited potential of human brain
- Let's avoid the *Luddism* exaggerations of the Industrial Revolution of the late 1800's !!!

Διαχρονικά τα Πανεπιστήμια ήταν χώροι επιστημονικής προόδου & ανοικτής καινοτομίας, χωρίς περιορισμούς εμπορικών συμφερόντων, καθώς και φορείς κριτικής μετάδοσης γνώσης στις επερχόμενες γενιές

Ας το κρατήσουμε και στην εποχή της Τεχνητής Νοημοσύνης