



# TUTORIAL

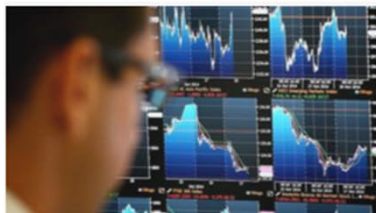
NICE  
FALL 2024

## NetWare 2024 & SocSys 2024 Diving into generative-AI to Assess its Current Status

Coordinator: Prof. Dr. Petre Dini, IARIA, USA/EU

[petre@iaria.org](mailto:petre@iaria.org)

<https://www.iaria.org/fellows/PetreDiniAtAGlance.pdf>



AI bot 'capable of covering up illegal trading'

IARIA is member of ORCID



Trading is increasingly done by computers



# Sudden Planetary Storm

NICE  
FALL 2024



## King Charles: AI risks require urgency and unity

King enters debate on AI as UK holds summit and announces global declaration on managing threats.



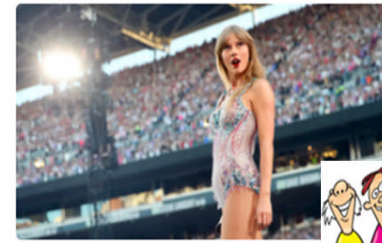
Elon Musk ahead of his meeting with UK PM Rishi Sunak



## US announces 'strongest action yet' on AI safety

The measures have been issued less than 48 hours before the UK hosts an AI safety summit.

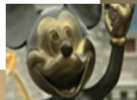
US vice-president to say short-term problems with technology as pressing as existential ones, before UK summit  
UK politics live - latest updates



Security game-changer for concerts, sporting events that will get you inside faster



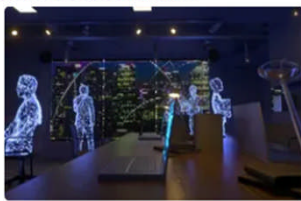
## Is AI a blessing or a curse? Barcode developer says 'it could go either way'



## Germany invests millions into AI military metaverse' that could never change combat



## Apple CEO sings AI's praises on rare visit to Capitol Hill



AI has already surpassed human capabilities in several domains – which jobs are next?

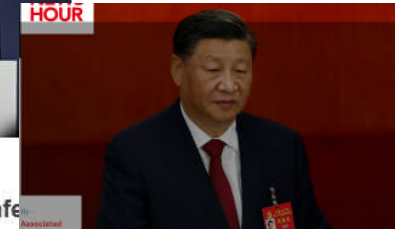
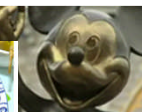


Disrupted LinkedIn hits 1 billion members, adds AI features for job seekers



Kamala Harris will reveal that 30 countries have agreed to sign a US-sponsored declaration for the use of AI by militaries. Photograph: Manuel Balce Ceneta/AP

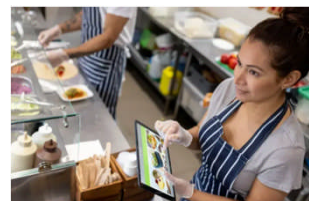
## Society & Equity US to launch its own AI safety institute



## China warns of artificial intelligence risks, calls for increased national security measures



## Terrible things in store if Washington doesn't act on AI race, tech ethicist warns



The military metaverse enables pilots to have more frequent training against relevant targets. Robinson said. (Dad)



## CEO warns of 'very strange outcomes' from new tech with 'superhuman' skills





# Before anything...

NICE  
FALL 2024

- **I acknowledge** that generative-AI is a useful approach for certain activities/duties, not necessarily ‘artificial’, but definitively, not ‘intelligent’, either; it seems a useful tool on a narrative in context, without any innovation in that context.

YET: multilingual narrative translation: I've seen que tu peux gerer istorii in mai multe limbi“ → “Ho visto che puoi gestire storie in diverse lingue.»

- **There are many areas and themes** where output from generative-AI tools **might be useful**, when carefully considered:

- drafting some content (yet to be combed by humans) on a given topic; a reasonable summary on a topic, quick reply,
- coding chunks, multiple programming languages, ... providing useful classroom personalized materials
- making trip/travel plans in certain areas, saving time, pointing to things you might not be aware-off
- others (situation on new nano-art, position ethics,... ); very useful at the level of being informed; very promptly!

- **As of April, 2024, my observations** are that:

- the output replicates (quicker, more comprehensible) what a human can do, usually called ‘automation’
- the output is subject of unintended (not easily detectable) mistakes issued with confidence
- the trust in the output (apart you have a clear understanding of expectation) is jeopardized by the nature of data used (not-validated, eventually obsolete, incomplete, not scientific)
- there is a danger of spreading false information (and news, facts)
- there is no accountability, liability, explainability of the output

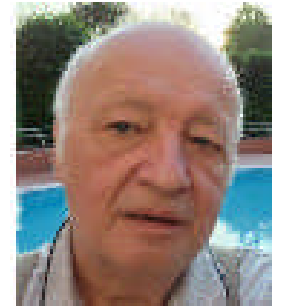
- **Generative-AI is a fledgling piece in the AI jigsaw**, useful in trivial cases, mostly informative, but its output cannot be taken as a proved and trustable fact || (try-test-trust-test-again-try-again and so forth, until you **lose trust or get confidence**)



# PETRE – AI-related activities

NICE  
FALL 2024

- Petre: 1980/90
  - Fuzzy-based resource allocation, Automatic knowledge incorporation, CAD/CAM Expert Systems,
  - Real-time embedded systems, Space/time thinking and processing, Multi-layers context-based meaning
- Petre: 1992: The First ITC Conference (Montreal), tutoring systems, self-adaptable Q&A professor-student systems (advanced Chatbots)
- Petre: 1997 Dartmouth, Mobile Intelligent Agents (Intelligent Grasshopping Polling)
- Petre: 1997-2000: Nomadic code, Mobile agents, (Grasshopper EU project)
- Petre: 2000-2010: Autonomous systems, Policy-driven systems, Intelligent systems (pushed to Patents, ITU, TMF, standards)
  - Capturing emerging properties, Variable pooling frequency, Self-adaptable decision polices, Reflexive-policies (Digital-Twins)
  - Routers embedded-AI (temporal logic in Syslog processing, policy-driven signal processing)
- AI-driven Selection of Content Servers based on Current Server Availability (dynamic availability, heuristics, real-time)
- Petre: 2010 - now (active observer and critic, panels, open discussions)



Petre DINI  
petre@iaria.org

At large: <http://www.iaria.org/fellows/PetreDini.pdf>



# AI Dream - Strong Evidence

NICE  
FALL 2024

**AI** does not exist in itself! (Ignorance, Arrogance, or Charlatanism)

**AI** is a parable, an umbrella, a target, a catalyzer, a Morgan the Fairy, an Accelerator, a distant dream, a fight for perfection, to reach the humans' intelligence capabilities.

Steam-engines, pencils, wind propellers, there were ALL steps of inventions towards AI (**EXPERIMENTS** are imperative)

- a. Can you imagine any AI entity thinking at **a pencil**, by itself, finding needed materials, shipping them, building the process, building the machinery, finding the glue, the powder, and having it in various types (more than 600 activities)
- b. Same question on simply **building a wind propeller pallet, a church, or proposing a recipe**. *(each item takes a human collective thinking approach)*
- c. One may think at **ANYTHING** and the above statement holds; and this, because *everything* involves **brain activity**.

As dreams are not easy to be achieved, humans invented satisfactory levels, for their own comfort (as life is short, we should be proud of something, 😊); mainly, the famous 'one step at a time'. 😞 currently, we are at the very beginning of the first step!

There are:

- 5 Levels of Capability Maturity Models (software), as there were/are well too many software failures
- 5 Levels for automation on Driverless Cars, as well to many failures and challenges (to claim one step ahead)
- 3 Levels of Chargers for Electric Cars

and ..... **surprise**

- (at least) 4 levels of AI capabilities, as of Nov 13, 2023; some, name 7, already; <https://builtin.com/artificial-intelligence/types-of-artificial-intelligence>

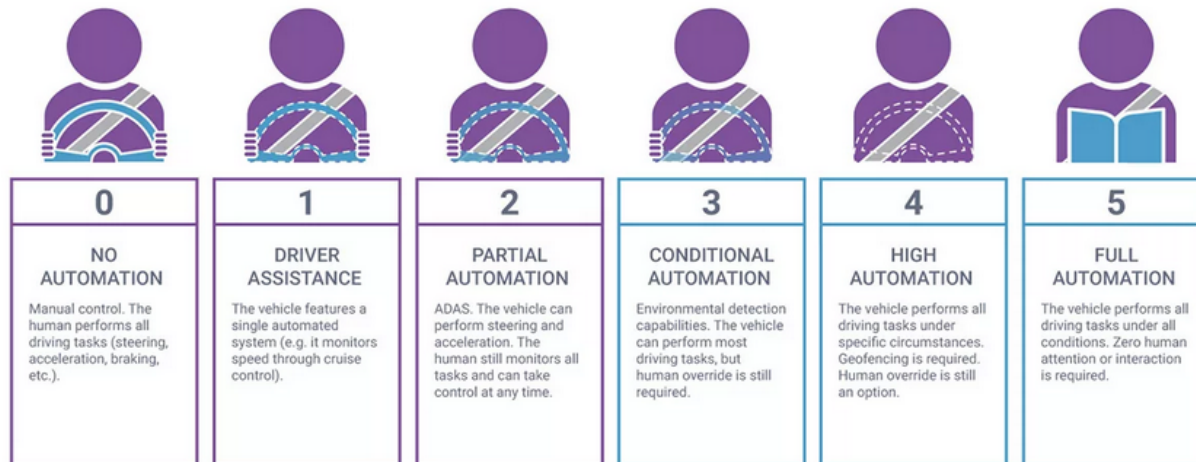
**Generative-AI** is the very first (still, sloppy achievement) and at a very low level of AI-dream (that is as much as we can, now).



# AI Dream - Strong Evidence

NICE  
FALL 2024

## LEVELS OF DRIVING AUTOMATION



THE HUMAN MONITORS THE DRIVING ENVIRONMENT

THE AUTOMATED SYSTEM MONITORS THE DRIVING ENVIRONMENT

## 5 levels of the Capability Maturity Model Integration



## 7 TYPES OF ARTIFICIAL INTELLIGENCE

- Artificial General Intelligence:** AI designed to learn, think and perform at similar levels to humans.
- Artificial Superintelligence:** AI able to surpass the knowledge and capabilities of humans.
- Reactive Machines:** AI capable of responding to external stimuli in real time; unable to build memory or store information for future.
- Limited Memory:** AI that can store knowledge and use it to learn and train for future tasks.
- Theory of Mind:** AI that can sense and respond to human emotions, plus perform the tasks of limited memory machines.
- Self-aware:** AI that can recognize others' emotions, plus has sense of self and human-level intelligence; the final stage of AI.



# AI Dream - Strong Evidence

NICE  
FALL 2024

## Generative-AI

Do not be disappointed; the model is rather trivial; only the context makes it more relevant; see quantity vs quality!

The working model is like having all the maps and trying to get from A to B, in a mountain region (you have detailed maps of), by checking the stability of any little rock and decide to go one step ahead (after seeing that you didn't fall); and so, repetitively, until the end is reached. Same as marching on moving sand! The process is deemed having additive errors.

The key process: guessing, predicting, and getting a weight of successful action; then, repeat, until you reach the peak!

The approach is based on Large Language Models; a large amount of data is stored; patterns are inferred from it, in the context (derived from a query through a known token). No token means no answer or a tentative (usually wrong) one.

Note: The text is broken down into chunks, which can be as short as one character or as long as one word (or even more in some languages). Jumping from a chunk to another is question of in context-prediction (there are other cosmetics such normalization, greedy decoding, beam, searching, etc.)

Much better than Google translator! Comprehension of narratives with mixed-languages!

## Some obvious limitations of generative-AI (as of Nov. 13, '23)

- There are not real bidirectional interactions; the only way for a better accuracy is obtained by rephrasing the query
- Knowledge is limited to public data access only (no private, no scientific databases); therefore, no validated
- Knowledge is behind the real state-of-the-art (however, lagging by a few years, in any domain, means a lot on accuracy)
- There is no guarantee, no liability; this situation leads to limited trust, or no trust at all.
- Some feedback is cooked, incomplete or incorrect; some is wrong but expressed with confidence (**test-try-trust-test-....** )

**YET, Generative-AI is useful in many cases**, especially as a quick idea (mostly useful for already knowledgeable people)



# AI Dream - Strong Evidence

NICE  
FALL 2024

## Generative-AI (example)

- Petre: What might follow after 'I am ...'?
- ChatGPT: The phrase "I am ..." can be followed by a wide variety of words or phrases to describe or define oneself or one's current state. Here are some examples:

### Descriptions of identity or profession:

I am a teacher.  
I am an engineer.  
I am a mother.  
I am a student.

### Emotional or physical states:

I am tired.  
I am happy.  
I am confused.  
I am sick.

### Personal beliefs or opinions:

I am certain that...  
I am of the opinion that...  
I am convinced that...

### Locations:

I am at home.  
I am in the office.  
I am on my way.

### Descriptions of characteristics or personality:

I am outgoing.  
I am introverted.  
I am ambitious.  
I am kind-hearted.

### Temporary situations or roles:

I am the team leader for this project.  
I am the designated driver tonight.  
I am on a diet.

### Affiliations:

I am a member of the local club.  
I am a Democrat/Republican.  
I am a Christian/Muslim/Hindu/etc.

## Back in '80, 90'

Expert systems

Rules-based systems

A\* algorithm

AI (Japanese, Fifth Generation)

Lisp

Prolog





# AI Dream - Strong Evidence

NICE  
FALL 2024

## Generative-AI (example)

- Petre: What physical state can a human be in?
- ChatGPT: Humans can exist in various physical states, based on physiology, consciousness, and activity. Here are some general categories of physical states:

### Conscious States:

Wakefulness: The state of being alert and aware of one's surroundings.

Drowsiness: A transitional state between wakefulness and sleep, often characterized by a decrease in alertness.

### Sleep States:

Non-Rapid Eye Movement (NREM) Sleep: Divided into three stages (N1, N2, N3), ranging from light to deep sleep.

Rapid Eye Movement (REM) Sleep: The sleep stage associated with vivid dreams.

### Altered States of Consciousness:

Hypnosis: A trance-like state of focused attention and heightened suggestibility.

Meditation: A state of deep concentration or relaxation, sometimes accompanied by a shift in awareness.

Flow: A state of being completely absorbed or immersed in an activity, often accompanied by a loss of the sense of time.

Trance: A semi-conscious state, often brought on by repetitive sounds, movements, or rituals.

### Physiological States:

Resting: Characterized by a reduction in physical and mental activity, but not necessarily asleep.

Activity: Engaged in physical or mental tasks, ranging from mild (e.g., walking) to intense (e.g., sprinting).

Fasting: The state of not consuming food for a period.

Digestive: Post-meal, where the body is processing and absorbing nutrients.

### Emergency Responses:

Fight or Flight: An acute stress response, preparing the body to either face a threat or flee.

Freeze: An immediate reaction to danger or threat, characterized by immobility.

Fawn: Behaving in a servile manner or trying to please a potential threat to diffuse danger.

### Health-Related States:

Healthy: Optimal functioning and well-being.

Illness or Disease: Compromised health due to infections, chronic conditions, or other diseases.

Injury: Physical damage or trauma.

Recovery: The process of returning to a healthy state after illness, injury, or surgery.

### Life Stages:

Fetal: Pre-birth, developing inside the womb.

Infancy: Early stage post-birth.

Childhood, Adolescence, Adulthood, Senior: Various stages of life characterized by physical, cognitive, and social changes.



# Let us not foul ourselves!

NICE  
FALL 2024

## Let us set the record straight (*as of Nov 13rd, 2023*)!

Inventions (~1% on purpose, 99% unintended discovery); on former, I know a very few, indeed!

Basis is: 'understanding' and 'inserting' a new acquired information in the right place of the existing knowledge. Mostly, incidental action, followed by experience, failures, experiments again, improvements, acceptance, improvements, etc.

Let us think at:

- Fire
- Matches
- Tires (cars, bicycles, ...)
- C Vitamin
- Stickers
- Wheels
- Bicycles

- 
- Music (i) (mind expression, even silent)
  - Music instruments (ii) (100% invention) for sharing
- 

<https://www.the-sun.com/tech/9508495/chatgpt-ceo-ai-superhuman-persuasion-skills-strange-outcomes/>  
<https://www.foxnews.com/us/chatgpt-chief-warns-superhuman-skills-ai-developop>

## AI Scarecrow



### Oracle of Delphi

Delphi was an ancient religious sanctuary dedicated to the Greek god Apollo. Developed in the 8th century B.C., the sanctuary was home to the Oracle of Delphi and the priestess Pythia, who was famed throughout the ancient world for divining the future and was consulted before all major undertakings.



**CEO warns of 'very strange outcomes' from new tech with 'superhuman' skills**

## Cutting through the Hype (towards AI - so-called Literacy!)

- Fathering AI
- AI Experts
- AI Analysis
- AI Ethicists
- AI-based producers
- AI Beneficiary

- AI culture, AI Literacy
- AI skeptics
- AI charlatans

- AI Governments summit
- AI (-self) governance
- AI governance
- Corporate self-guaranty entitlement
- Corporate **self-explainability duty**  
(see: [Oracle of Delphi](#))
- Corporate self-safety checking

- True AI Enthusiasts and Supporters
- Professors trying to get various flavors of scientific help in their teaching endeavor: CENTRIC, ICSEA, CYBER, etc.)



# AI – Media push

NICE  
FALL 2024

It seems that Media take us for fools and bombard us with alarming news about the destruction of humanity!

When a (right or wrong) statement is endorsed by one known name, there is little chance to object! Yet... !?

## Note:

- A crane sure **surpasses** the capabilities of humans, when it comes to lift heavy pieces.
- An automatic searching engine will certainly **exceed** the human skills for (quicker) grabbing information from websites.
- **Dangers come from** humans handling AI-tools, not for the AI-tools; see the dynamite!



メディアは私たちを愚か者とみなし、人類滅亡に関する憂慮すべきニュースを私たちに浴びせているようです

**It seems to me being either trivial or wrong statements**

<https://www.cmswire.com/digital-experience/5-bill-gates-takes-on-the-future-of-artificial-intelligence/>

What does the Microsoft co-founder now say about artificial intelligence? You may be surprised.

## The Gist

- **Deepfake danger.** *AI can help identify deep fakes and safeguard democracy, and there are lessons for customer experience and marketing professionals.*
- **AI security.** *Prioritize software vulnerability fixes and establish AI regulation.*
- **Job shift.** *Manage AI's job impact with support, retraining and strategic policies.*

<https://www.youtube.com/watch?v=s1qZbsytk08>



# AI - Political push

NICE  
FALL 2024

**STOP AI!**: The letter signed by the Band of Six



**Tech billionaire Elon Musk makes an eerie prediction about the future of AI**

**AI UN**



- Over the course of two days, tech and business executives and world leaders — including Vice President Kamala Harris — discussed the future of AI, regulation and safety issues associated related to the advanced technology. Nearly 30 countries including China and the European Union signed the Bletchley Declaration, which recognizes potential risks associated with AI.

**It seems to me being an eerie prediction, indeed, but not the sense of the article.**

<https://www.foxbusiness.com/technology/elon-musk-ai-eventually-create-situation-no-job-needed>

**The Gist:**

"We will have for the first time something that is *smarter than the smartest human*," Musk added.

Musk said, "It's hard to say exactly what that moment is, but there will come a point where *no job is needed*."

"You can have a job if you want to have a job — sort of personal satisfaction — *but the AI will be able to do everything*," the billionaire Tesla CEO added.

But the lack of guardrails surrounding AI has not stopped Musk from developing *his own generative AI models* to compete with those already available for commercial use, including chatbots ChatGPT and Google's Bard.

A "maximum truth-seeking" AI chatbot could be coming, which he dubbed "**TruthGPT**."



# AI Dream – Petre's level 5

NICE  
FALL 2024

**IF you think that this is AI, then, I am kindly inviting you to revisit the meaning of: ai vs Ai vs al vs AI**

(**MUST**) next steps, for the very first next AI (true) Level:

Mechanisms leading to inventions, things, and facts that are not apriori known: sailing, rowing, flying,... video cameras, cooking recipes, clothes buttons, sewing needles, perfusion needles, bicycle pedals, etc. ; they came as a necessity; an artificial piece has no necessities to be fulfilled.

## **PRIORITY #1**

**Discovery and understanding of Brain mechanisms**; thinking, inventing, acting, self-repairing, body coordination, external exposure, feeling necessities, etc., etc.

This will allow us to go to the next level of AI fulfillment, if ever.

Some timid (yet, in progress for ~ 50 years) steps: Machine-Brain interface, Cognitive Research

The greatest difficulty is that one cannot do conclusive experiments on humans' brain. No experiments, no progress!

**Note:** Knowledge for understanding mental diseases, Parkinson procedures, addiction, etc. is badly missing.

This might be at **Level 5** on the AI-scale in terms of achievements (understanding what does not work in the Brain and invent cures for findings, namely, AI-self-healing via AI itself). **YES, this might be the pillar piece of AI dream.**



# AI Dream - Strong Evidence

NICE  
FALL 2024

- Petre: What are the most prominent Cognitive achievements as of Sept '21"
- ChatGPT: As of September 2021, cognitive science, which encompasses interdisciplinary research from psychology, neuroscience, philosophy, artificial intelligence, and several other disciplines, has seen numerous achievements over the years. Some of the most prominent (? : I do not know who decided this complement) cognitive achievements include:

**Neuroplasticity:** The discovery that the brain is plastic, meaning it can change and adapt structurally and functionally throughout life based on experiences. This challenged the once-held belief that the brain's structure is fixed past a certain age.

**Mirror Neurons:** The identification of neurons that fire both when an individual acts and when they observe the same action performed by another, leading to insights into empathy, imitation, and social cognition.

**Dual-Process Theory:** The proposal that there are two systems of thought: "System 1" which is fast, instinctual, and emotional; and "System 2" which is slower, more deliberative, and more logical.

**Cognitive Behavioral Therapy (CBT):** Development and widespread acceptance of this form of psychotherapy, based on changing negative thought patterns to produce more positive behaviors and outcomes.

**Mapping the Brain with fMRI:** Functional Magnetic Resonance Imaging (fMRI) has allowed researchers to non-invasively study brain activity, illuminating how different brain regions function and interact.

**Understanding Memory:** Insights into different types of memory, like procedural, declarative, short-term, and long-term memory, and how they're stored and recalled.

**Cognitive Load Theory:** The understanding of how cognitive resources are allocated during learning and the strategies to optimize this process.

**Artificial Neural Networks:** Drawing inspiration from the brain, these computational models have led to significant advancements in machine learning and AI, particularly in tasks like image and speech recognition.

**Embodied Cognition:** The idea that the mind is not only connected to the body but that the body influences the mind, emphasizing the role of sensorimotor experiences in shaping cognition.

**Advancements in Understanding Disorders:** Improved comprehension of cognitive disorders such as Alzheimer's, ADHD, and Autism Spectrum Disorder, leading to better therapeutic approaches.

**Cognitive Offloading:** The exploration of how and why humans use objects and tools in their environment to reduce their cognitive workload.

**The Role of Sleep in Cognition:** Discoveries related to the importance of sleep in memory consolidation, emotional regulation, and problem-solving.

These achievements only scratch the surface of the continuous advancements in the field. The interdisciplinary nature of cognitive science ensures that it remains at the forefront of understanding the intricacies of the human mind and its intersection with technology and society.



# AI - Up front (... Morgan the Fairy)

NICE  
FALL 2024

## Progress on the fields mentioned below happened independently of any kind of AI

- Patterns recognition, Image recognition
  - Signal processing, Vision systems
  - Cognitive models, Autonomic systems
  - Learning models, Predictions systems,
  - Convolutional networks
  - Speech recognition, Image recognition
  - Conversion text-voice and vice versa
  - others
- and*
- High speed processing data, graphic, streaming
  - High Data collection/memory/storage
  - Miniaturization, Accessibility
  - High speed communication
  - Brain-human interface
  - others

**Artificial Intelligence (AI) originally intended** to create machines that could mimic human intelligence: to reason, learn from experience, recognize patterns, comprehend complex ideas, and employ language in meaningful ways, namely, to develop systems that could perform tasks that, when done by humans, would require the application of intelligence (emulation of cognitive functions, associated with human minds).

Over time, **AI became an umbrella** including a broad range of computational techniques and methods. Towards its original target practical applications helped creating specialized, task-specific algorithms and models. Yet, the underlying ambition remains: to capture or replicate facets of human intelligence in a machine, bridging the gap between organic cognition and digital computation.

**The fields of AI, machine learning, statistics, and pattern recognition have interwoven histories, and many techniques have been appropriated or evolved across these disciplines over time**

## More AI specific

- Datasets (variety, solid methodology, etc.)
- (layered) DeepLearning (CNN, recurrent, transformer architectures, etc.)
- ? ML (accuracy, parameter tuning, computational needs, etc.)



# AI - Up front (as a Catalyst, only)

NICE  
FALL 2024

## 1. Neural Networks and Their Variants:

- **Feedforward Neural Networks:** Basic networks where connections don't cycle back.
- **Convolutional Neural Networks (CNNs):** Tailored for image data.
- **Recurrent Neural Networks (RNNs):** For sequential data.
- **Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU):** Advanced RNNs.
- **Transformer Architectures:** Like BERT, GPT, which have revolutionized NLP.
- **Generative Adversarial Networks (GANs):** For generating data.
- **Autoencoders:** For dimensionality reduction and data generation.
- **Self-organizing maps, Hopfield networks, and Boltzmann machines:** Early forms of neural networks.

## 2 Reinforcement Learning:

Algorithms like Q-learning, Deep Q Networks (DQN), and various policy gradient methods like REINFORCE, A3C, and PPO predominantly belong to the AI realm.

Monte Carlo Tree Search (MCTS): Used in game-playing AI like AlphaGo.

3. **Modern Optimizers:** Techniques like Adam, RMSprop, and AdaGrad, developed mainly in the deep learning context.
4. **Attention Mechanisms:** Especially in the context of neural networks, allowing models to "focus" on specific parts of the input.
5. **Transfer Learning:** Techniques like fine-tuning where pre-trained models are adapted to new tasks.
6. **Few-shot and Zero-shot Learning:** Techniques that enable models to make predictions in scenarios where very few or no examples are available.
7. **Neural Architecture Search (NAS):** Techniques to automatically search for the best network architecture.
8. **Explainable AI (XAI):** Techniques specifically designed to make AI decisions interpretable.
9. **Neural Turing Machines and Differentiable Neural Computers:** Extensions of neural networks to give them the ability to work with external memory in a differentiable manner.
10. **OpenAI's DALL·E, CLIP, etc.:** Represent a new wave of models trained to perform tasks like image generation from textual descriptions.





# AI - Looking behind and ahead!

NICE  
FALL 2024

## For everybody's comfort

Humanity will not be replaced by artificial entities as

- There are no personalized memory(ies)
- There is no instant personalized thinking
- There is no visibility on any personal past activity
- There are no feelings of fear, emotions (eventually mimicking by instructions/learning)

Do not fear and do not have compassion on those thousands that regret fathering AI, because there were very few of them; most of the tens of thousands vocal ones truly were and are not.

*These two pictures: 2023 – March, Wilmington DE, MedExpress Lab Poster and Vitaly's input from the College of Healthcare Information Management Executives, April 1995, (Canada) triggered my initiative for documentation and for preparing this open discussion.*



## Lessons learned

- *Technologies were developed in a silos with great success*
- *Last 30 years of achievements allowed comprehensive results*
- *Awareness*
- *Education*
- *Societal services*



- Why there are not much progress in Healthcare AI?



Vitaly Herasevich



# Positive

NICE  
FALL 2024

- **Useful, still, caution is the Key**
- **Our Society changes, our behavior should, too!**
- **Personalized services:**
  - Personalized re- skilling
  - Personalized guideline for home self-healthcare
  - Personalized training
  - Personalized student advice
- **Educational**
  - Homework (extending class-work)
  - Personalized guideline for home self-healthcare
  - Personalized training
  - Personalized student advice
  - Exercising tools (skills development)



Petre Dini  
IARIA



# Positive

NICE  
FALL 2024

## ■ Hopes on Human and AI-based critical decision systems

### ■ Assisting: Decision paradigm change

- AI and the Human brain need to be used in conjunction; one is not to replace the other
- *bias* and *trade-offs* must be cleared by humans.
- Human + AI-based decisions need *personalized H-AI context-based training*.  
The feedback loop must be updated based in *success-index* of cooperation  
(see Intelligent Tutorial Systems).
- *Revisit*: Replace - Assist - Advice | Feedback-by-request, Suggestions-by-predictions
- *Validation*: Validate synchronization duration | Validate use case scenarios
- *Flow*: AI (Personalized-AI) & in-Context (Human skills) -->> Assisted decisions

### ■ Personalized AI for critical system resilience

- *Metaverse* is training the AI-based tools with virtual models;
- *Simulations, Digital Twins* accelerate the system deployments, but no real-world environments.
- Latency, reliability, synchronization (for collaborative work)
  
- *Digital-twin* approach can be used for a dry-training AI-Humans for specific tasks
- *Decision prediction and feedback as options*.
- *Updated training and real-time impact feedback*.



Petre Dini  
IARIA



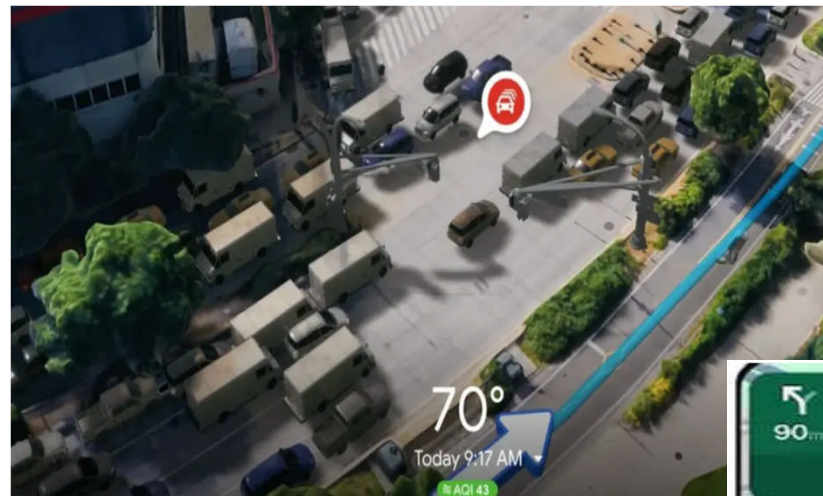
# Positive

NICE  
FALL 2024



Google Maps getting major upgrade thanks to new trend taking world by storm

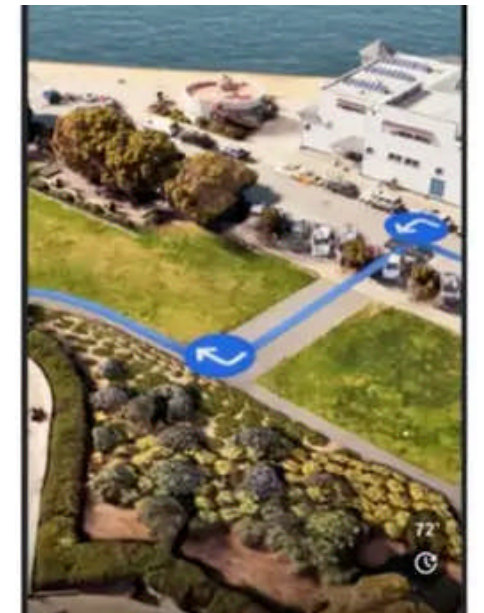
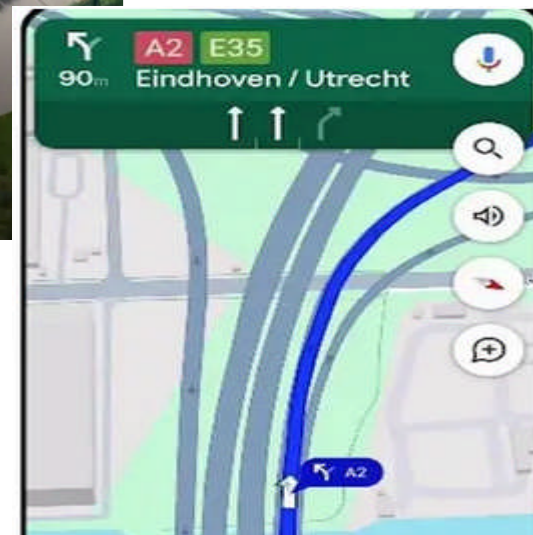
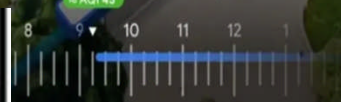
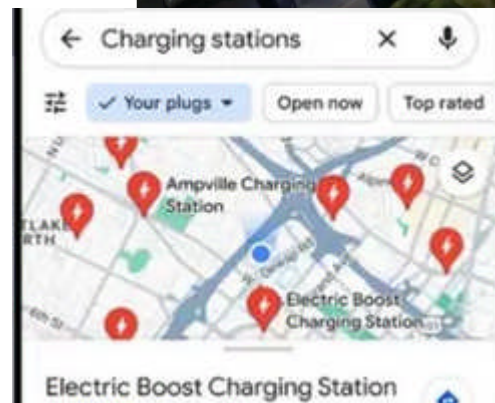
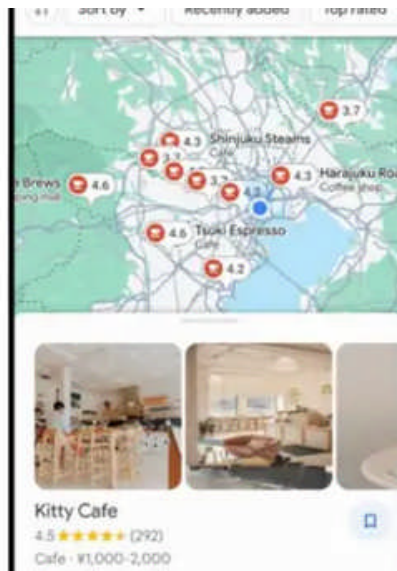
Google Maps gets a massive AI upgrade with 5 new features  
The latest updates to Google Maps makes it smarter and more helpful  
<https://www.foxnews.com/tech/google-maps-gets-massive-ai-upgrade-5-new-features>



3D with Immersive View



Petre Dini  
IARIA





+/-

NICE  
FALL 2024



Can you trust AI for financial advice? Or would you lose it all?  
Charles Schwab survey finds Gen Z wants to retire at 61, 75% of them would trust AI for financial advice



Petre Dini  
IARIA

Can you trust AI for financial advice?  
Or would you lose it all?



Biden's AI executive order could cement 'wokeness' into platforms: Kara Frederick

Heritage Foundation tech policy director Kara Frederick analyzes Biden's sweeping A.I. executive order and addresses the push to ban TikTok on 'Varney & Co.'





+/-

NICE  
FALL 2024



Visa launches consulting practice to advise clients on implementing AI  
Visa's new AI Advisory Practice aims to help payments industry clients implement AI for business needs

<https://www.foxbusiness.com/technology/visa-launches-consulting-practice-advise-clients-implementing-ai>



Petre Dini  
IARIA

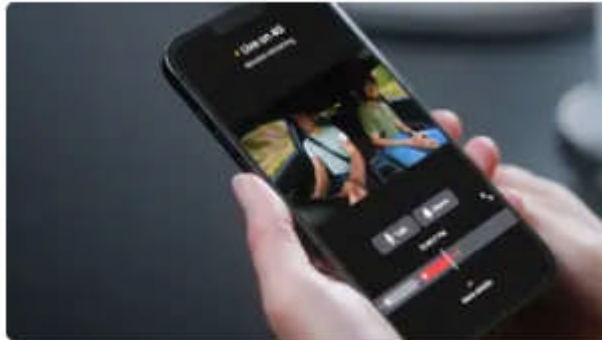


AI tools can be used to analyze large quantities of market data to detect trends and sentiment. (iStock / iStock)



NICE  
FALL 2024

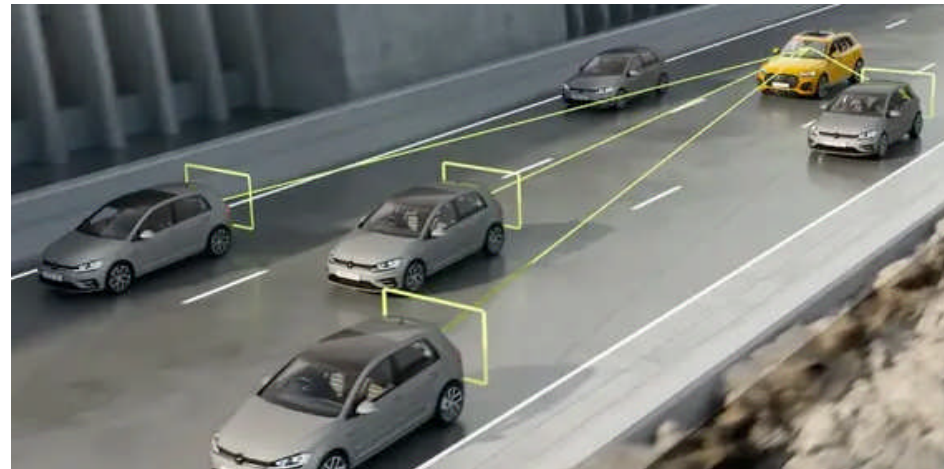
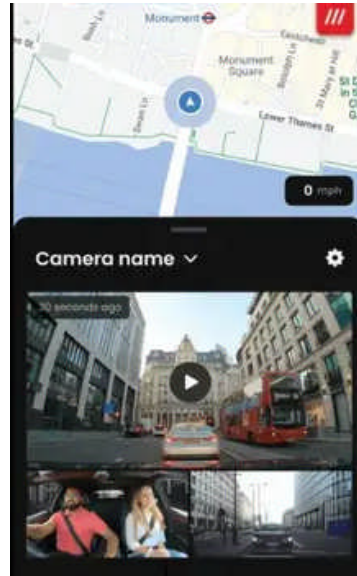
## Autos



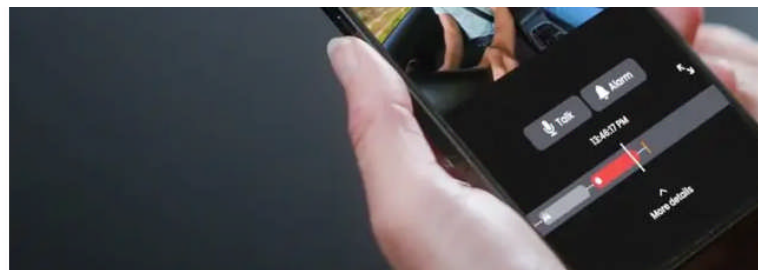
6 ways AI-powered dashcams can save your life and your money



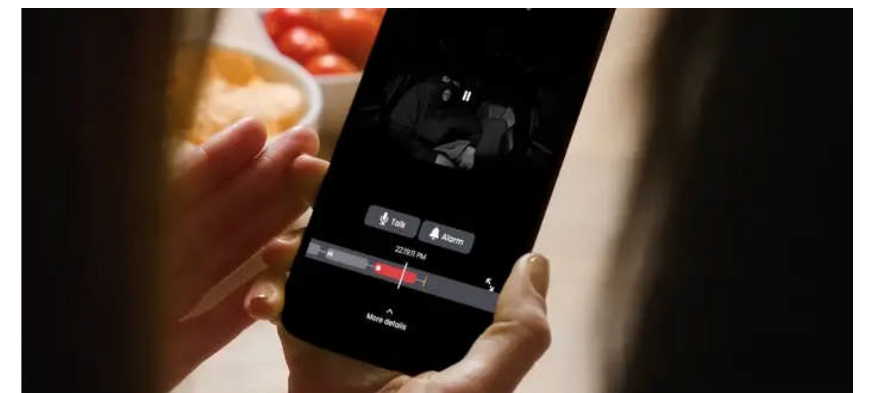
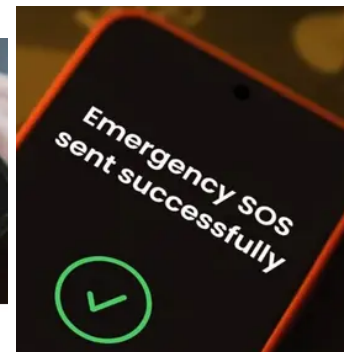
6 ways AI-powered dashcams can save your life and your money  
See what the world's first AI-powered dash camera can really do  
<https://www.foxnews.com/tech/6-ways-ai-powered-dashcams-can-save-life-money>

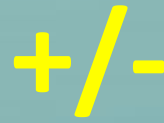


Petre Dini  
IARIA



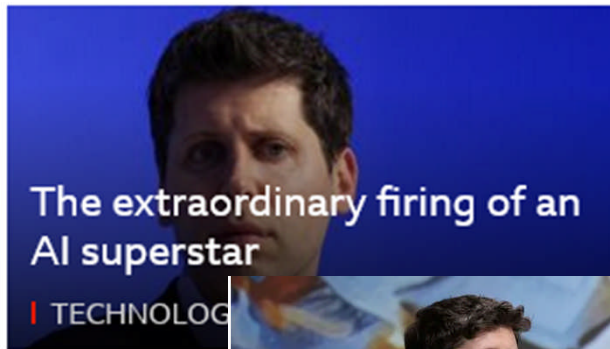
You can access your car's dashcam from your phone. (Nextbase)





# NICE FALL 2024

Sam Altman: The extraordinary firing of an AI superstar  
<https://www.bbc.com/news/technology-67461363>



OpenAI CEO Sam Altman gives a speech during a meeting in Paris in May. (Joel Saget/AFP via Getty Images/File / Getty Images)

Just two weeks ago he was in the UK at the world's first AI safety summit as one of only around 100 global delegates. He gave a speech last week about the future of his company and its tech.

I think it's safe to assume he genuinely had no idea what was coming.

Silicon Valley's big guns have so far rallied behind Mr Altman, including former Google CEO Eric Schmidt, who described him as a "hero of mine".

Microsoft boss Satya Nadella said he had "confidence" in the firm. Well, he needs to - Microsoft has invested billions in it, and the tech which underpins ChatGPT is now embedded in Microsoft's office apps.

One character who has been uncharacteristically quiet so far is Elon Musk. He and Mr Altman set up OpenAI together, along with others, but are said to have fallen out over a decision to move it away from being non-profit. There are rumours that it is this very issue which has once again divided opinion within the firm now.



Petre Dini  
IARIA



AI boss Sam Altman ousted after board loses confidence  
<https://www.bbc.com/news/business-67458603>

OpenAI investors reportedly trying to reinstate former CEO Sam Altman  
Altman said that he 'loved' his time at OpenAI

<https://www.foxbusiness.com/technology/openai-investors-reportedly-trying-reinstate-former-ceo-sam-altman>





**REALITY**

**NICE  
FALL 2024**



**Petre Dini  
IARIA**

**STAGE IS  
YOURS**



# REALITY vs. PERCEPTION

NICE  
FALL 2024



Petre Dini  
IARIA

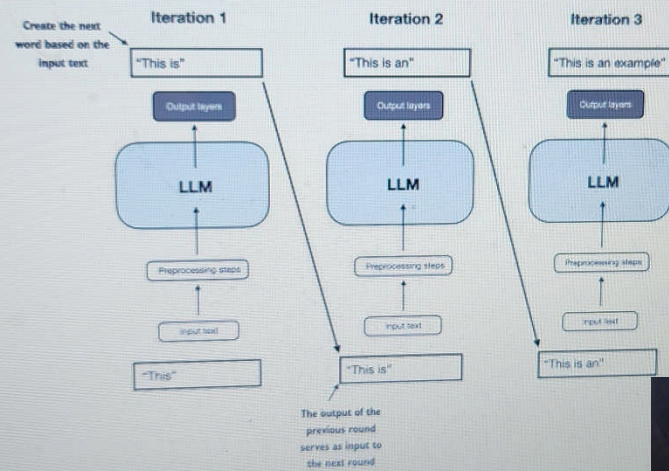
# LLMs



# LLMS/i

NICE  
FALL 2024

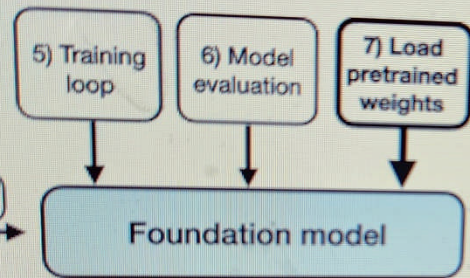
How do LLMs generate multi-word outputs?



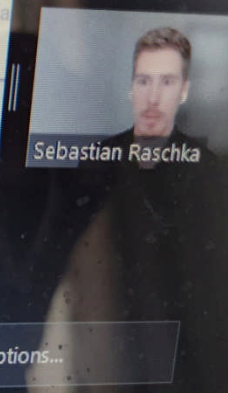
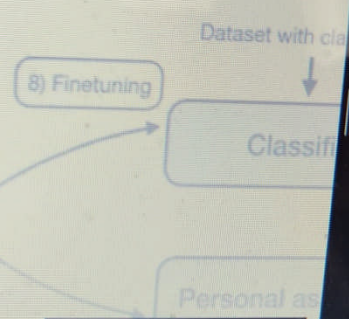
## Llama 3 was trained on 15T tokens

"To train the best language model, the curation of a large, high-quality training dataset is paramount. In line with our design principles, we invested heavily in pretraining data. Llama 3 is pretrained on over 15T tokens that were all collected from publicly available sources."

### STAGE 2: PRETRAINING



### STAGE 3: FINETUNING



TechTalk Slides

Captions...

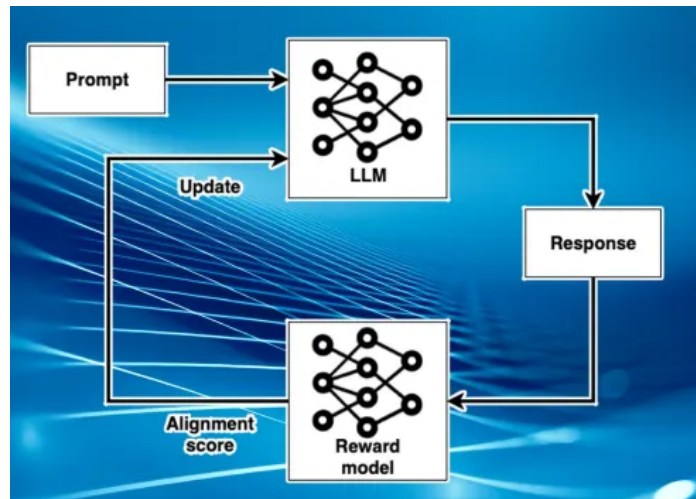


# LLMs/ii

NICE  
FALL 2024

<https://aws.amazon.com/what-is/reinforcement-learning-from-human-feedback/>

<https://bdtechtalks.com/2023/01/16/what-is-rlhf/>



RLHF  
Datasets

Q&A  
Datasets

Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



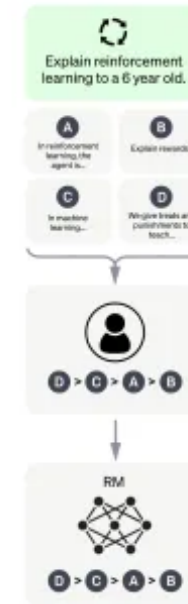
Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

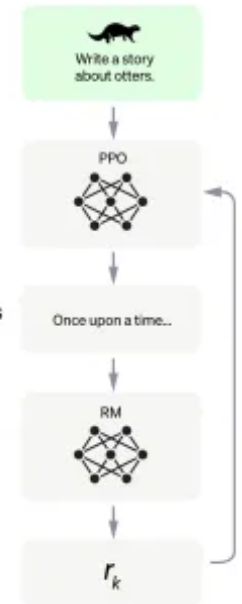
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

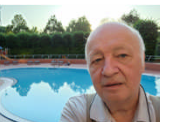
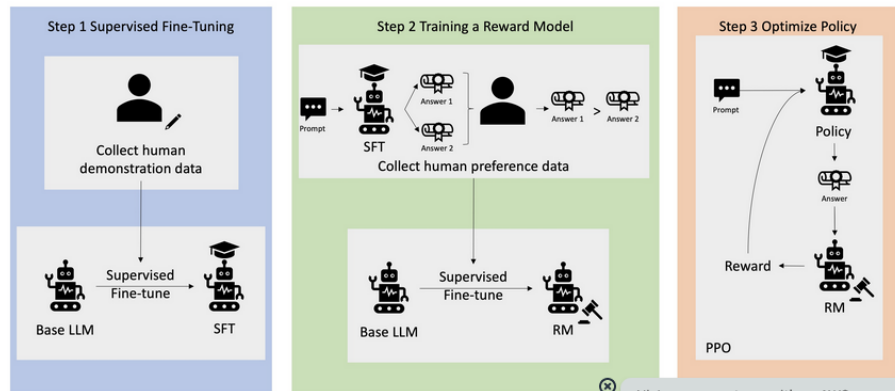
The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



RLHF is also not a perfect solution. Human feedback can help steer LLMs away from generating harmful or erroneous results. But human preferences are not clear cut, and you can never create a reward model that conforms with the preferences and norms of all societies and social structures.





## ■ Are hallucinations a bug or a feature?

- We don't know how knowledge is created, but imagining what doesn't yet exist is part of it. Is creativity a form of hallucination?
- Creativity may be unwelcome sometimes. For instance, a financial report or a medical diagnose
- Deterministic procedures or specialized domains can be worked with by using augmentation or calls to external systems (e.g., calculator)
- More data, more training, more scrutiny may not be the answer
  - More data will carry similar semantic meaning to existing data
  - What is considered non-biased today, may be seen as biased tomorrow



Petre Dini  
IARIA



# HALUCINATIONS

NICE  
FALL 2024

- **Hallucinations: Starting for real facts (datasets) and getting incorrect (irrational) answers/output**
- **Machine hallucinations are in fact, immature LMM training, triggered by the nature of learning and understanding.**
  - For humans, learning is peered by perception of meaning and consequences of actions.
- **LLM hallucinations are induced by the exploring mechanisms and by the nature of the input data.**
  - Hallucinations of output from artificial entities (LLMs,) are unavoidable, as learning is without understanding and there is no perception of the output.
- **Only humans can detect hallucinations.**
- **Hallucinations are induced by the training process based on 'one word at a time' and 'the best guess at each step'.**
- **Hallucinations are induced by limited backward attention in focusing on forward target**
- **Key considerations in *machine vs human* learning (out evaluation should be performed accordingly)**
  - Learning/acting with consequences
  - Learning by understanding the meaning (or by **attaching** a meaning)
  - Learning **by chunks vs by words**; no human will use sentences 'this, this is, this is a, this is a big' without a context by guessing the best next word)
- **Note:**
  - Apart *on purpose* improper or incorrect training, when doubt in a learning point, the best option (context) is chosen.
  - See going over mountains from A to B.



Petre Dini  
IARIA



+ii

NICE  
FALL 2024

## ▪ Hallucinations are human, in nature

- Hallucinations are human, in nature (Hallucinations are glimpses to the reality)
- Practical Leisure: Poetry, Panting (except for realism, in the end, but also too idyllic), stories, especially children's books, some music styles, etc.
- Spiritual behaviors: Dreams, Wish-Tale, imaginations, Angels, ...
- Real inventions: in fact, apriori unintended.

## ▪ AI/LLM Hallucinations

- Generated narratives, generated music, generated poetry, generated nano-pictures, artificial facts/tale,...

## ▪ Where hallucination are damaging

- When critical predictions, forecasts, or hopes fail, leading to loss of something (material, beliefs)
- In mission critical decisions (health, investments, politics, wars, etc.)



Petre Dini  
IARIA

## Why LLMs now?

- Job market virtualization
  - Digitalization and storage of Teradata
  - Tremendous technology advances (trans speed, comp power, etc.)
  - \$ Billions for training



# REALITY vs. PERCEPTION

NICE  
FALL 2024



Petre Dini  
IARIA

# HALLUCINATIONS





NICE  
FALL 2024

## ■ Both Human and AI/LLM hallucinations can be on purpose or unintended.

- The question is that AI/LLM hallucinations cannot be controlled, there is a lack of understandability and explainability, and, even more, 'they are rarely timely discovered'.
- In LLM trained with more than 13 trillions of parameters, it is practically towards impossible to identify a cause.
- So, hallucinations represent a significant risk, but also opportunities; they are part of our existence.
- A few examples:
  - most of the side-effect achievements of NASA led to practical applications (mostly materials/clothes, solar panel, etc.)
  - C vitamin
  - Tires
  - .... almost all the inventions

## ■ Hallucinations vs. Synesthesia vs. Dreams vs. Intended actions

### ■ *Inherited Genetics or Human Subjectivism*

- a. Different perceptions: Wassily Kandinsky saw color when he heard music (complex perceptions)
- b. Different Interpretations of the Same Situations (embedded feelings) [humans excel in this diversity]

### ■ *By Training*

- a. Wrong Information: fake, incomplete, obsolete, inaccurate, biased
- b. Unbalanced Information (taxonomy, Q&A Datasets, overfitting, underfitting, predefined answers in interactive tools, alternative answers (answers 1 or answer 2),
- c. Poor quality of data and inadequate Datasets validation.
- d. Lack of human feedback, as this might be the only way to feel pain on a wrong output
- e. On purpose - misleading



Petre Dini  
IARIA



NICE  
FALL 2024

## ■ Under consideration

- Customization of Datasets (notice that 1 million of data points is a very small Dataset)
- RAG (Retrieval-Augmented Generation) used for dedicated domains (financial, legal, sports, etc.)
- Reinforcement Learning by Human Feedback - accuracy, honesty, and skills are downsides
- **Needs**
- Increase spanning attention over large sequences
- Bidirectional processing (see Bengal language, where sequence of the words is mostly irrelevant) via activating or deactivating the attention masks
- **Concrete situation:** Disabling Attention Masks.
- "Disabling the attention masks" means that the model will no longer apply these masks during the **attention computation**.
- This can have several implications:
  - **No Padding Mask:** If the padding mask is disabled, the model may incorrectly attend to padding tokens, which can lead to incorrect or meaningless outputs. For example, if a sentence is padded with [PAD] tokens, the model might give undue importance to these tokens.
  - **No Causal Mask:** Disabling the causal mask in an autoregressive model means that tokens can attend to future tokens. This breaks the autoregressive property and can lead to issues during generation, where the model might use information from future tokens that it should not have access to yet.



**Petre Dini**  
**IARIA**

## ■ Appendix: Understanding Attention Masks

- *In the context of Large Language Models (LLMs), such as those used in transformers, "disabling the attention masks" refers to altering the way the model processes sequences of data during training or inference.*
- *Attention masks are a key component in transformer architectures like BERT, GPT, and others. They are used to manage the attention mechanism, which allows the model to focus on different parts of the input sequence. Typically, attention masks serve the following purposes:*
- *Padding Mask: To handle variable-length sequences by distinguishing actual data from padding tokens. For example, in a batch of sequences of different lengths, shorter sequences are padded with a special token (e.g., [PAD]). The attention mask ensures that these padding tokens do not affect the model's computations.*
- *Causal Mask: In models like GPT, which generate text sequentially, causal masks ensure that each token only attends to previous tokens and not future ones, preserving the autoregressive property.*



+ /iv

NICE  
FALL 2024

## Conclusion

- **Hallucinations caused by the LLM approaches**
  - LLM efficiency is not too high considering energy and computation power for training.
  - One token at a time, ...
  - No memory
  - No long span over an acquired knowledge (read: 'text') towards the target (short horizon; looking for bidirectional processing).
- **Hallucinations are induced by the process and by selected data**
  - We will always live with an 'accepted level of hallucinations', in a real society or a digital one, as hallucinations are here to stay
  - Awareness and adequate literacy is needed for preventing personal damages (hopefully, in our control)
  - There is a danger of being used by wrongdoers (at any level) by altering the process
- **Be aware**
  - Identify the cause and find tips to reduce the risks of LLM hallucinations
  - Mitigate LLM hallucinations for minimizing the negative impact
  - Distill the Ethical Hallucinations
  - Identify Good and Bad Hallucinations



**Petre Dini**  
**IARIA**

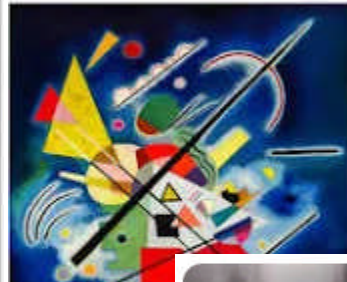
## Caveat

**Hallucinations might (and will) be used for any kind of justification (of the output, or of an action) by wrongdoers and/or bad actors**



+ / v

PORTO  
July 2024



NASA Turns to AI to Design ...

Our Top Five Strangest Build...

2+2 = 5, all lines are straight, 2 = 10, log(-3) = x  
Horses ride a bicycle over the shining cloud of dust!!  
Aliens will invest in The House of the rising Sun!  
Alice's Adventures in the Wonderland

Metaverse  
VR AVR  
Immersion

John published a paper on Mars in the year 2500 [24].  
There are 5 Planets and one Galaxy only.  
Army of X plans to attack Y country at midnight, 2024, June 30.

Recommended references on a geology paper.

- [x] Pierre, Title 1
- [z] Jacobs, Title 2
- [y] Stan, Title 3
- Note: [x] doesn't exist
- [z] has another title
- [y] is a carpenter instructions book



**LLMs-based:** - very good summarization of information they are fed with, even only less than 1% validated as true  
 - very good mixed (4-5-6 ...) languages, correct punctuation, correct grammar, spelling correction on context-based intuition  
 - helpful at the informative level, like white papers, very quickly obtained and quite comprehensive  
 - assumes user's familiarity and experience with a given domain; see, selection an oscilloscope for 5G spectrum, financial aspects,...



+ /vi

NICE  
FALL 2024

## Items under scrutiny

- a. Are human hallucinations more acceptable than machine hallucinations?
- b. Why can hallucinations of artificial machines occur?
- c. How to spot damageable hallucinations?
- d. How to improve the LLMs processes for minimizing hallucinations?
- e. ...
- f. ..
- g. .



**REALITY**

**NICE  
FALL 2024**



**Petre Dini  
IARIA**

**STAGE IS  
YOURS**