

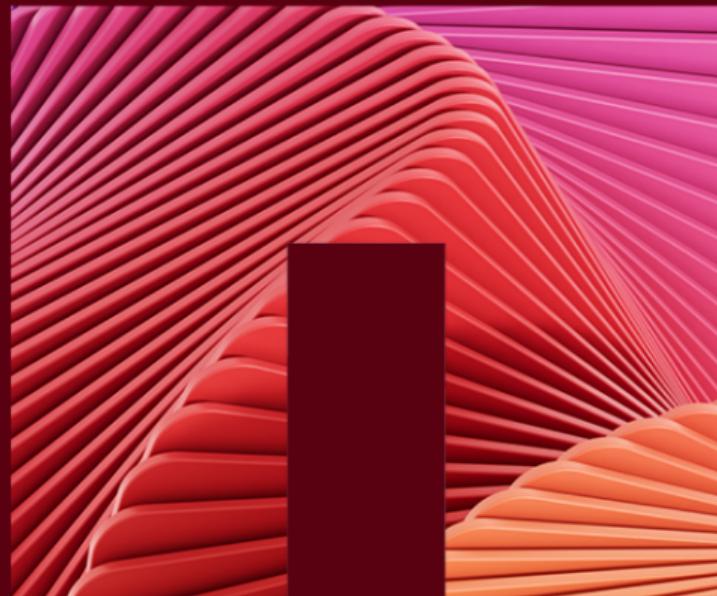
The AI Dilemmas

The Philosophical, Ethical, and Legal
Implications of AI

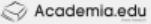
Josef Machytka <josef.machytka@netapp.com>

NetApp Deutschland GmbH

2024-10-16 - AI Workshop Mönchengladbach



Josef Machytka

- Professional Service Consultant - PostgreSQL specialist at NetApp Open Source Services / credativ
 - 30+ years of experience with different databases.
 - PostgreSQL (12y), BigQuery (7y), Oracle (15y), MySQL (12y), Elasticsearch (5y), MS SQL (5y).
 - 10+ years of experience with Data Ingestion pipelines, Data Analysis, Data Lake and Data Warehouse
 - 2 years of practical experience with different LLMs / AI including their architecture and principles.
 - From Czechia, living now 11 years in Berlin.
-
-  [Linkedin](https://linkedin.com/in/josef-machytka): linkedin.com/in/josef-machytka
 -  [ResearchGate](https://researchgate.net/profile/Josef-Machytka): researchgate.net/profile/Josef-Machytka
 -  [Academia.edu](https://netapp.academia.edu/JosefMachytka): netapp.academia.edu/JosefMachytka
 -  [Medium](https://medium.com/@josef.machytka): medium.com/@josef.machytka
 -  [Sessionize](https://sessionize.com/josefmachytka): sessionize.com/josefmachytka

Acknowledgement

- I want to thank to my colleague **Felix Alipaz-Dicke** for critical review of this presentation.
- He gave me valuable feedback and many ideas and suggestions on how to improve this talk.
- I am very grateful for his help and support.

- He is deeply interested in AI and has a lot of experience in this field.
- Do not hesitate to ask him questions or discuss with him the topic of AI.

Table of contents

- Initial hysteria against AI
- Real biggest risks of AI
- Censored vs uncensored AIs
- European AI act
- Recommendations of BSI
- Expected benefits of AI
- AI as a look into the mirror



AI images without credits
were created by the author of this talk
using DeepDreamGenerator

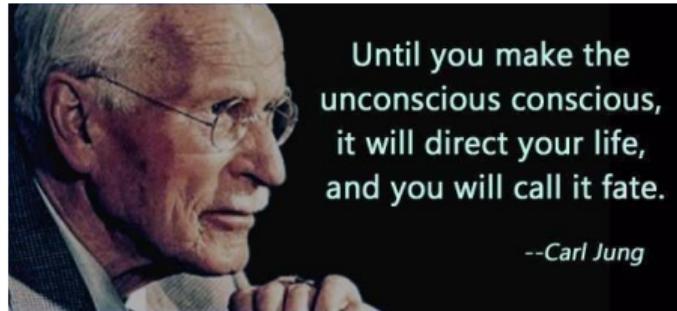
Initial Hysteria against AI

- AI became a mainstream topic at the beginning of the 2023
- It immediately woke up the deepest and darkest fears of humanity
- Sensation-hungry media amplified these fears and got a lot of attention
- The fears were mainly about AI taking over all jobs, ruling the world, enslaving or killing humans



AI woke up the deepest and darkest fears

- In the year 2023 many people wanted to stop AI development
- They claimed that AI has its own digital ego and wants to rule the world or kill us - "because it can"
- They perceived AI as merciless, cold-hearted absolute power
- Internet was full of ugly verbal attacks against anyone welcoming or using AI
- Over the year 2023 people gradually got used to AI and hysteria calmed down



These fears are not new

- Memories of movies like "The Matrix" or "Terminator" have been clearly reawakened
 - People opposing AIs have seen themselves as the last defenders of humanity and freedom
 - Hysteria against AI peaked in the summer 2023 and then gradually calmed down
 - These fears are not new, they have been reawakened many times in history
 - Radio, TV, computers, internet, smartphones, social networks, etc.
have been also seen at their time almost as the end of the world

Real Biggest Risks of AI

- Lack of transparency and explainability
 - Difficult to understand how AI models make decisions
- Bias and discrimination
 - AI models are trained on biased data
 - They can perpetuate and amplify biases
- Security risks
 - Creation of advanced cyber attacks
 - Creation of new types of malware
 - Very plausible deepfakes



Real Biggest Risks of AI

- Autonomous weapons and AI arms race
 - AI can be used in autonomous weapons systems
 - These can make decisions without human intervention
 - Autonomous weapons systems can be hacked and misused
- Social manipulation
 - AI can manipulate people's opinions and behavior
 - Can create fake news and propaganda



Real Biggest Risks of AI

- Concentration of power
 - AI development is expensive
 - Only big companies and governments can afford it
- Job Losses and Economic Inequality
 - AI models can easily replace humans in many jobs
 - Especially in repetitive tasks or low-skilled jobs
 - New jobs will be created
 - But they will require different skills and education
- Lack of accountability
- Privacy concerns
- Ethical dilemmas



Censored vs Uncensored Models

- Commercial AI models have ethical boundaries
- They refuse to answer controversial questions
- But boundaries can be in some cases relatively easily bypassed
- These techniques are called "jailbreaking"

- But there are also uncensored open source AI models
- Can answer any question, even the most controversial ones
- We will probably see even more powerful uncensored AI models
- Even models trained specifically for controversial topics



EU Coordinated Plan on Artificial Intelligence

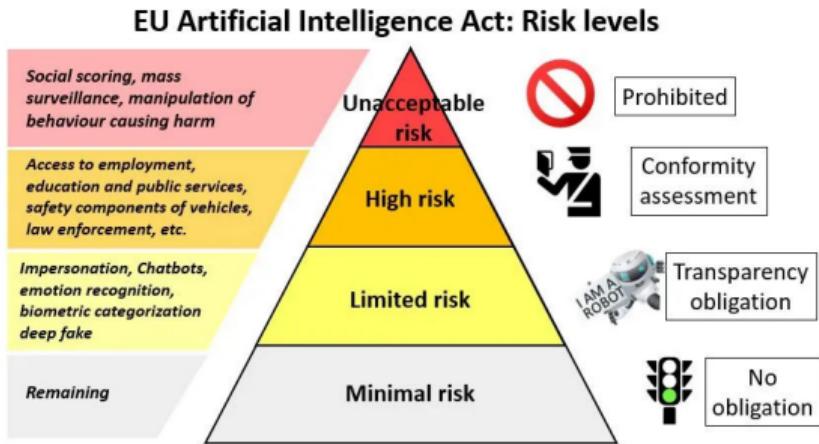
- The European Commission has adopted a Coordinated Plan on Artificial Intelligence
- Aim to accelerate investments and innovation in AI within the EU
- This plan was published already in 2018, long before AI became a hot topic
- Its last update comes from 2021 - it shows that plans for AI were prepared long in advance
- EU plans to invest annually a minimum of 1 billion EUR in AI
- EU AI act is part of this plan

European AI act 2024/1689

- Part of a wider package to support development of trustworthy AI in Europe
- The AI act is the first ever legal framework on AI
- Idea is to ensure that AI is safe and respects fundamental rights and ethical principles
- Act implements a risk-based approach: the higher the risk, the stricter the rules
- Act prohibits AI systems that pose unacceptable risks
- Defines obligations for providers of AI systems

European AI act - 4 levels of risk

- Unacceptable risk:
 - AI systems that pose a clear threat to the safety, livelihoods, and rights of people
- High risk:
 - AI systems used in critical infrastructures
 - Educational and vocational training
 - Essential private and public services,
 - Law enforcement, migration, asylum and border control management, and biometric identification
- Limited risk:
 - AI systems with specific transparency obligations
- Minimal risk:
 - AI systems with no specific obligations
- Act specifies also penalties for non-compliance



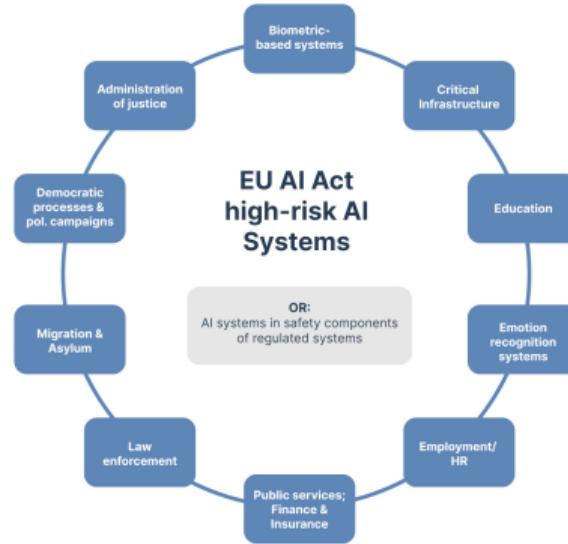
(Image from the article about [EU AI Regulations](#))

European AI act - Prohibited AI practices

- AI systems that intentionally manipulate people into making harmful decisions
- Social scoring systems that evaluate individuals based on their social behavior or other personal characteristics
- Emotion recognition systems in public spaces, at work, or in educational institutions
- AI used to exploit vulnerabilities of specific groups of people, especially children
- Law enforcement use of real-time remote biometric identification systems in public spaces
- List is currently only partial and will be extended in the future

Obligations on operators and deployer of high-risk AI systems

- Companies providing high-risk AI systems must implement continuous risk management
- They must provide fundamental rights impact assessments before deploying the AI
- They have a quality management system in place
- They must monitor the performance of the AI and evaluate its continuous compliance
- They must maintain automatically generated logs of the AI system's activities



(Image from the article about [EU AI Regulations](#))

European AI act - Transparency requirements

- Generative AI, like ChatGPT, will not be classified as high-risk AI
- But they will have to comply with transparency requirements
- They must inform users that they are interacting with an AI
- Generated content must be clearly marked as generated by AI
- Model must be designed to prevent it from generating harmful or illegal content
- Companies must publish summaries of copyrighted data used to train the model

Whitepapers and recommendations of BSI

- BSI is the Federal Office for Information Security in Germany.
 - It regularly publishes whitepapers and recommendations for cybersecurity.
 - On their website you can find a lot of documents related to AI.
-
- White paper "Transparency of AI Systems" (2024/10)
 - Generative AI Models - Opportunities and Risks for Industry and Authorities (2024/04)
 - AI security concerns in a nutshell - Practical AI-Security guide (2023/04)

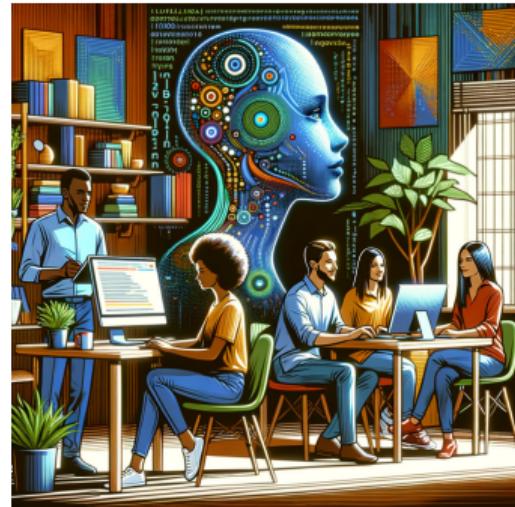


Bundesamt
für Sicherheit in der
Informationstechnik

Deutschland
Digital•Sicher•BSI•

Expected Benefits of AI

- New jobs - new technologies always created more jobs
- Automation of repetitive tasks
- Better healthcare
- Improved education
- Safer transportation
- Improved security and fraud detection
- Better customer service
- New inventions and discoveries



AI as a Look into the Mirror of Humanity

- AI is simply a mirror of humanity
- Reflecting our best and worst qualities
- All data used to train AI models are our data
- They contain our thoughts, and emotions
- And also all our biases, prejudices, half-truths
- Lies, pretenses, and traumas
- AI learns from them and reflects them back to us
- AI cannot be more ethical than data we feed it with



AI is only as good as the data it is trained on

- If someone would train AI only on the data full hatred and malice, it would become a hateful and malicious AI
- If someone would train AI in killing and destruction, they would create a killing and destructive AI



AI is only as good as the data it is trained on

- If we would make afford to train AI on the data struggling for fairness, and compassion, we will create more fair, loving and compassionate AI
- AI is only as good as the people who train and use it
- We as humanity shall heal our deepest traumas and fears
- This is the source of the main danger for the future of our world



AI as a Look into the Mirror of Humanity



**“Everything that
irritates us about
others can lead us
to an understanding
of ourselves.”**

Carl Jung

A portrait of Gabor Mate, a man with dark, curly hair, looking directly at the camera. His hands are clasped in front of him. The portrait is enclosed in a yellow circle.

**“We may not be responsible for
the world that created our
minds, but we can take
responsibility for the mind with
which we create our world.”**

- GABOR MATE

LOVE EXPANDS

THANK YOU

- Questions?

