

AI – szanse, zagrożenia i wyzwania dla audytorów wewnętrznych?

Jan Anisimowicz CRISC, CISM, PMP

Sebastian Burgemeister CISA, CISM, CRISC, CDPSE, CCAK (...)



Kilka słów o nas. Zapraszamy do kontaktu z nami (LinkedIn, email)

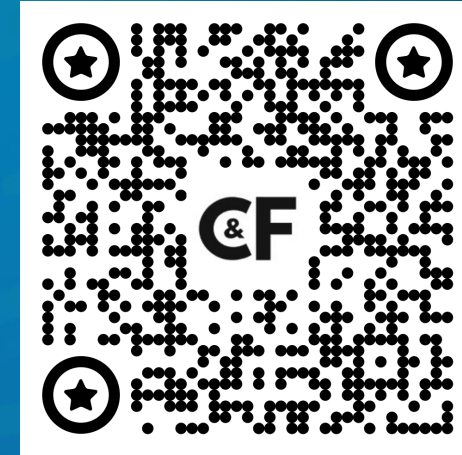


Jan Anisimowicz
C&F SA
Warsaw ISACA Chapter
IIA Polska
Jan.Anisimowicz@candf.com



www.adaptivegrc.com
www.candf.com

Skontaktuj się ze mną na LinkedIn!



Sebastian Burgemeister
BW Advisory Sp. z o.o.
IIA UAE (Dubai)
Former President of IIA Polska
s.burgemejster@itgrc.pl



www.itgrc.pl
www.akademiaitgrc.pl

Skontaktuj się ze mną na LinkedIn!



O czym dzisiaj powiemy

01

ChatGPT – Kluczowe fakty

04

Jak sztuczna inteligencja może wspierać audytorów.

02

Jak działa sztuczna inteligencja.

05

Podsumowanie i dalsze kroki.

03

Audyty wewnętrzny i ryzyka związane z AI

06

Pytania i odpowiedzi.

Materiał dla uczestników workshopu MF, Przydatne linki, książki, wydarzenia



1. Zapraszamy na webinar:

„**Cyfryzacja procesu audytów zdalnych z wykorzystaniem AdaptiveGRC**”

Kiedy: koniec kwietnia/początek maja

Dla kogo: Audytorzy, Compliance, zarządzanie ryzykiem

Zapisy: www.adaptivegrc.com/webinars

2. Zaproszenie do skorzystania z przedpremierowego i bezpłatnego dostępu do rozwiązania **AuditStarter**, przeznaczonego dla małych i zwinnych zespołów audytorów.

Zapisy od teraz na stronie:

<https://adaptivegrc.com/auditstarterbeta/>

Planowane udostępnienie **AuditStarter**: wrzesień 2024



3. Jak działa chat GPT?

<https://www.greatai-prompts.com/guide/how-does-chatgpt-works/>

4. Dwa godne polecenia rozwiązania oparte na Generative AI:

- <https://chat.openai.com/> - dostęp do chatGPT także w wersji darmowej
- <https://www.deepl.com/> - bardzo ciekawe narzędzie do tłumaczenia tekstów w dokumentach



ARTIFICIAL INTELLIGENCE

FOR BEGINNERS

Raef Meeuwisse



A concise and comprehensive beginners guide to the concepts, components and challenges of AI, without the god-level mathematics.

CyberSimplicity



Raef Meeuwisse
Auditor

<https://www.amazon.com/Artificial-Intelligence-Beginners-Raef-Meeuwisse-ebook/dp/B0BZ58JHGD>



ChatGPT

Kluczowe fakty

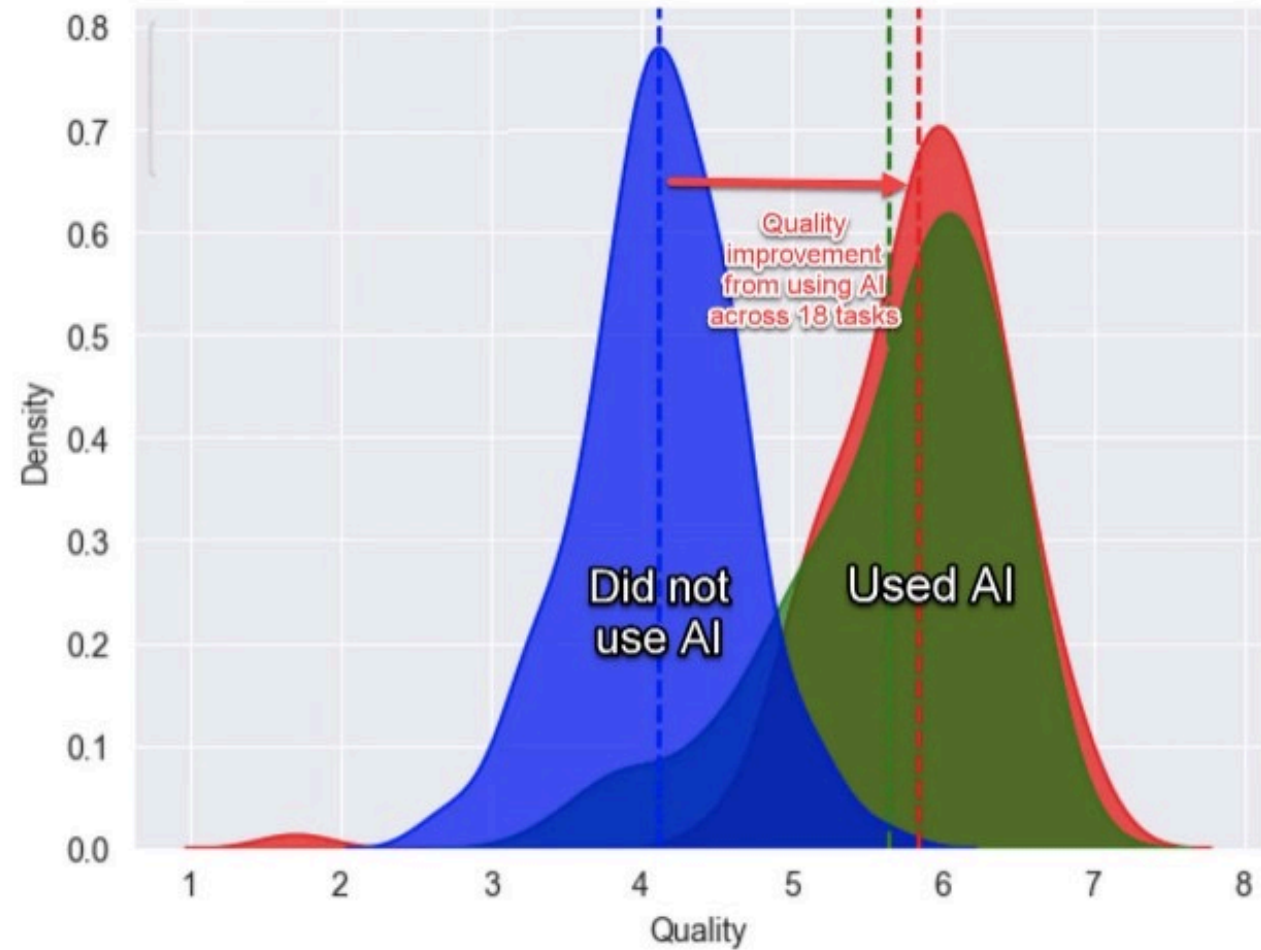


Variety of applications:

- AI Assistant (chat)
- Document search
- Content creation
- Content summarization
- Translation
- Data querying, e.g., NL2SQL.
- Image/Video analysis & creation
- ...

Key Benefit:

- Enhanced productivity



Distribution of output quality across all the tasks. The blue group did not use AI, the green and red groups used AI, the red group got some additional training on how to use AI.

ChatGPT - krótkie intro dla tych, którzy przypadkowo przeszali ostatnie kilka miesięcy :)



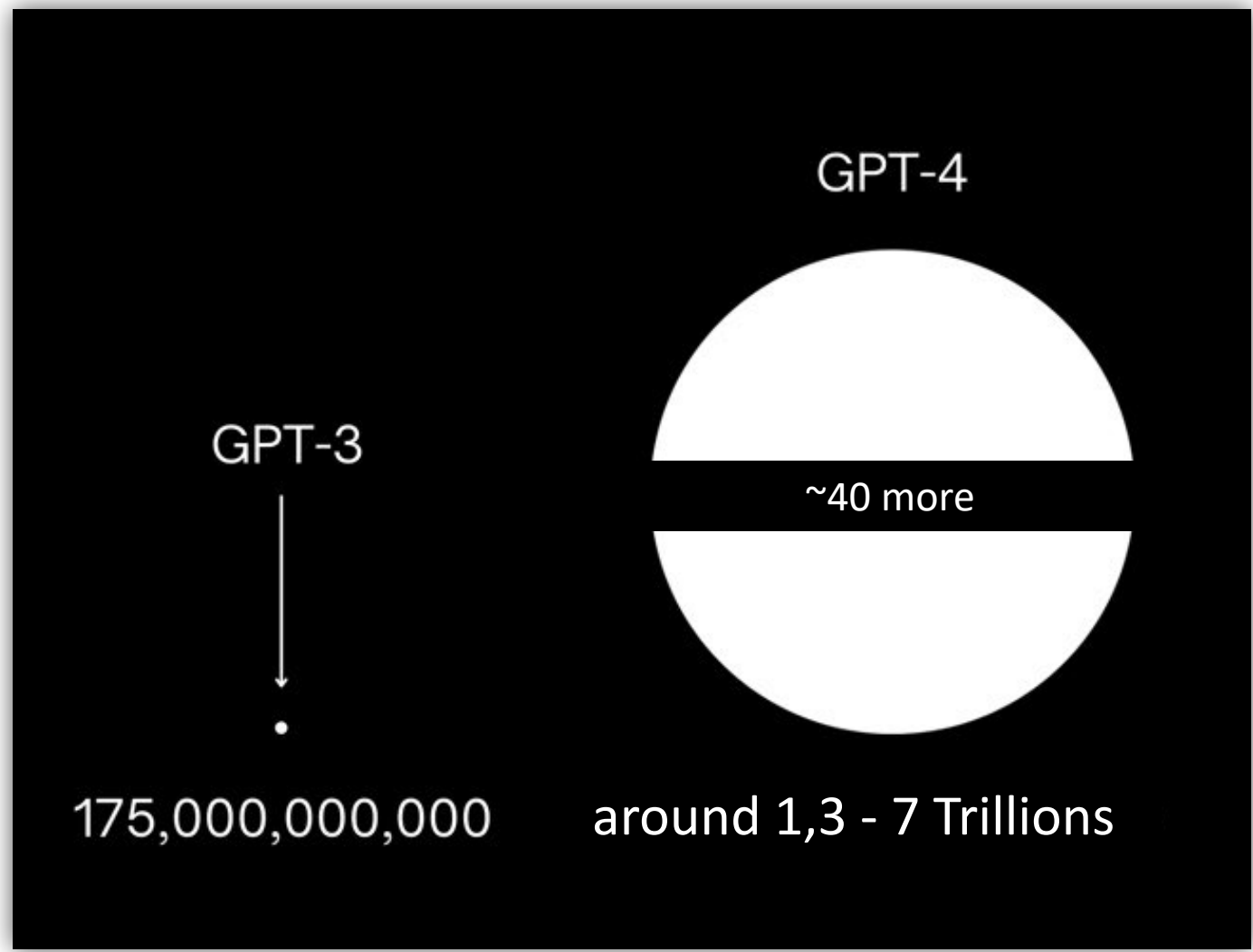
- ChatGPT to model językowy sztucznej inteligencji stworzony przez OpenAI.
- Wykorzystuje **algorytmy głębokiego uczenia** do analizowania danych tekstowych i generowania odpowiedzi podobnych do ludzkich na podpowiedzi.
- ChatGPT został przeszkolony na **ogromnych ilościach danych** tekstowych i może zrozumieć szeroki zakres tematów i kontekstów.



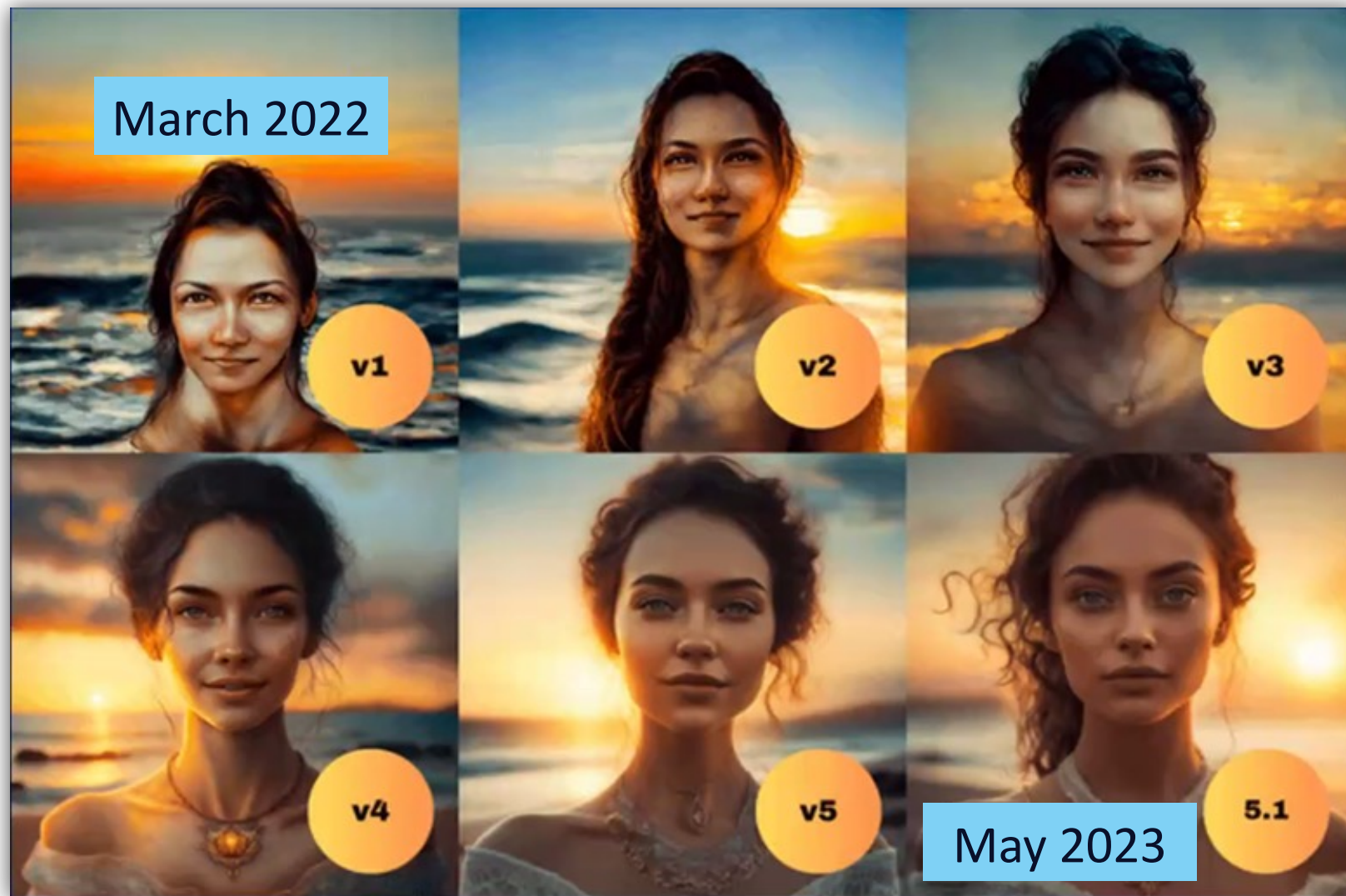
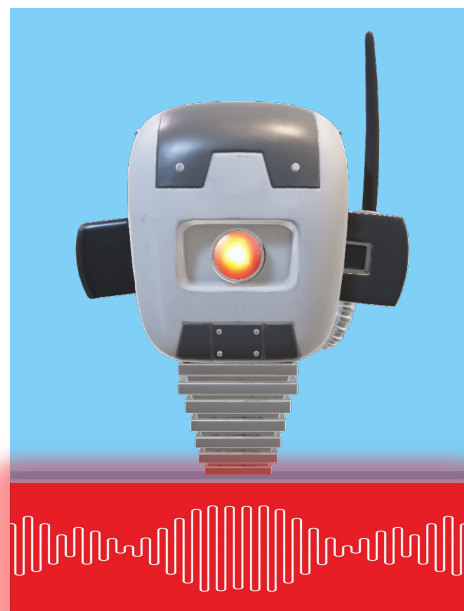
<https://www.greataiprompts.com/guide/how-does-chatgpt-works/>

ChatGPT – jak szybki jest rozwój

Porównanie ChatGPT w wersji 3 i 4 pokazujące, jak szybko rozwija się ten obszar



Obrazy generowane przez AI



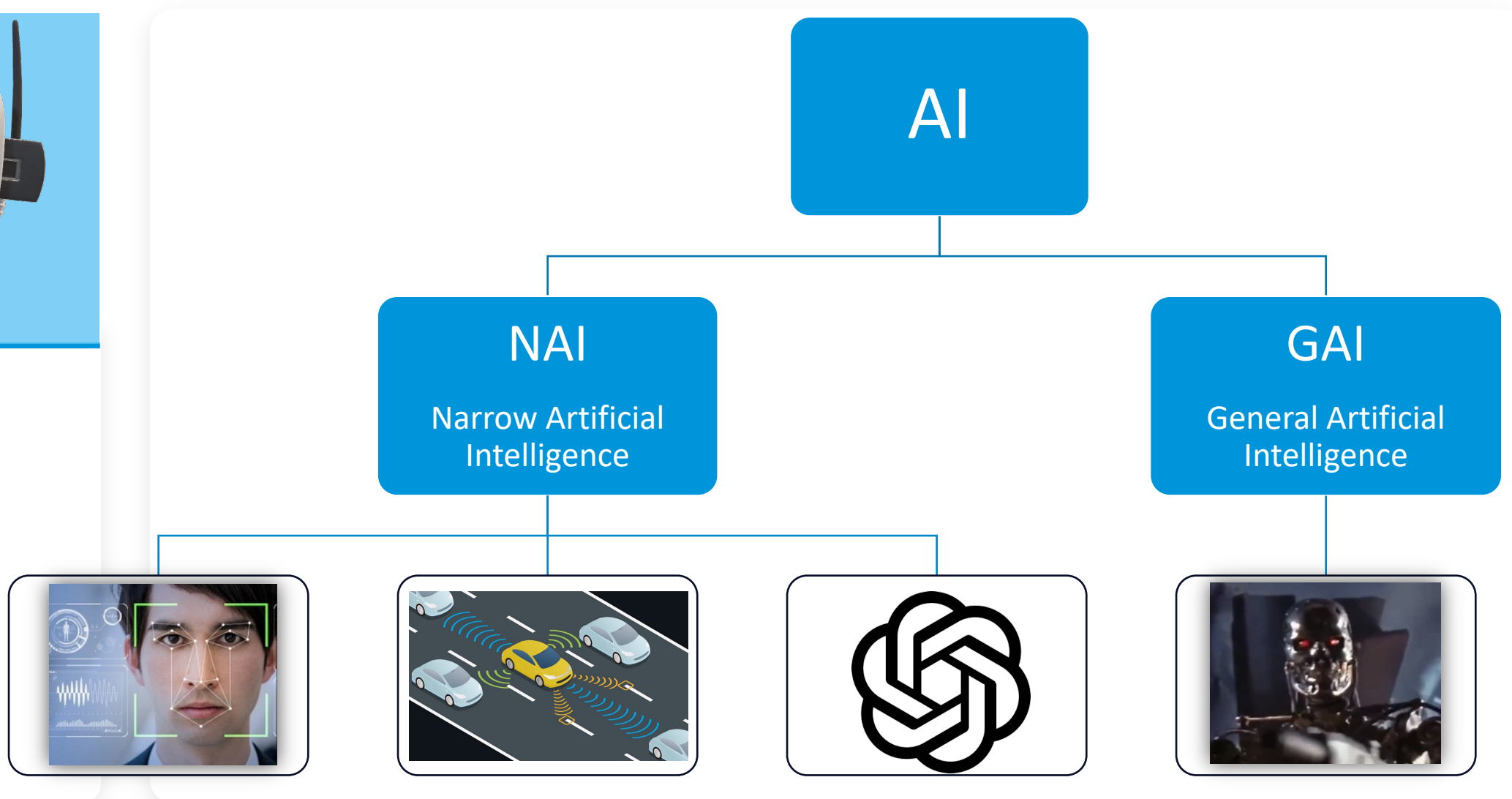


Sztuczna inteligencja

Jak to działa?

Hierarchia sztucznej inteligencji

ChatGPT as a part of NAI



ChatGPT – how to teach it

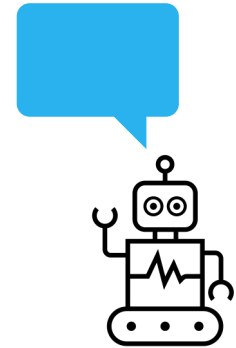
Approach



Data, tons of data.
And more..



Learning process



ChatGPT Ready

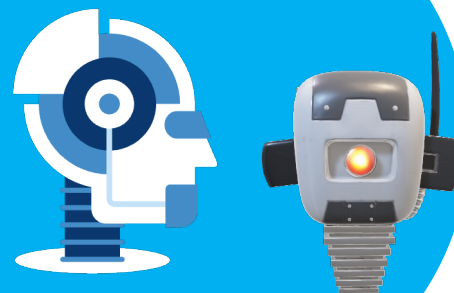
ChatGPT – how to teach it

Approach



Learning process

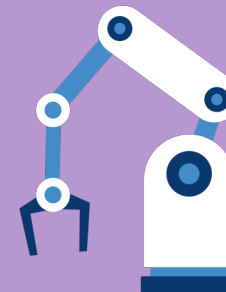
Artificial
Intelligence



“Sztuczna inteligencja to nauka i inżynieria polegająca na tym, że komputery zachowują się w sposób, który do niedawna uważaliśmy za wymagający ludzkiej inteligencji.”

Andrew Moore

Machine
Learning



“Uczenie maszynowe to nauka o algorytmach komputerowych, które ulepszają się automatycznie dzięki doświadczeniu.”

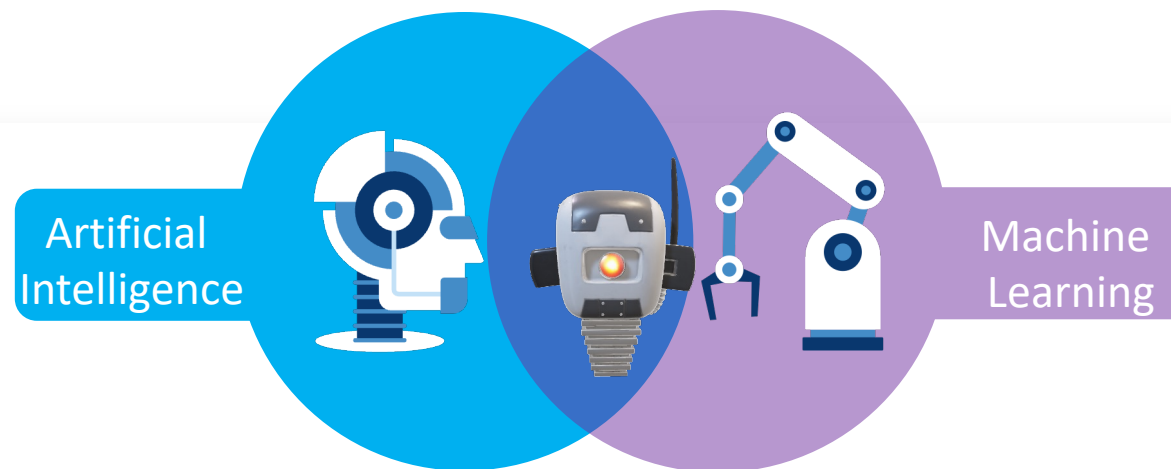
Tom M. Mitchell

ChatGPT – how to teach it

Approach



Learning process



Głęboka sieć neuronowa

Uczenie przez wzmacnianie

Transformatory (enkoder/dekoder)

Głęboka sieć neuronowa



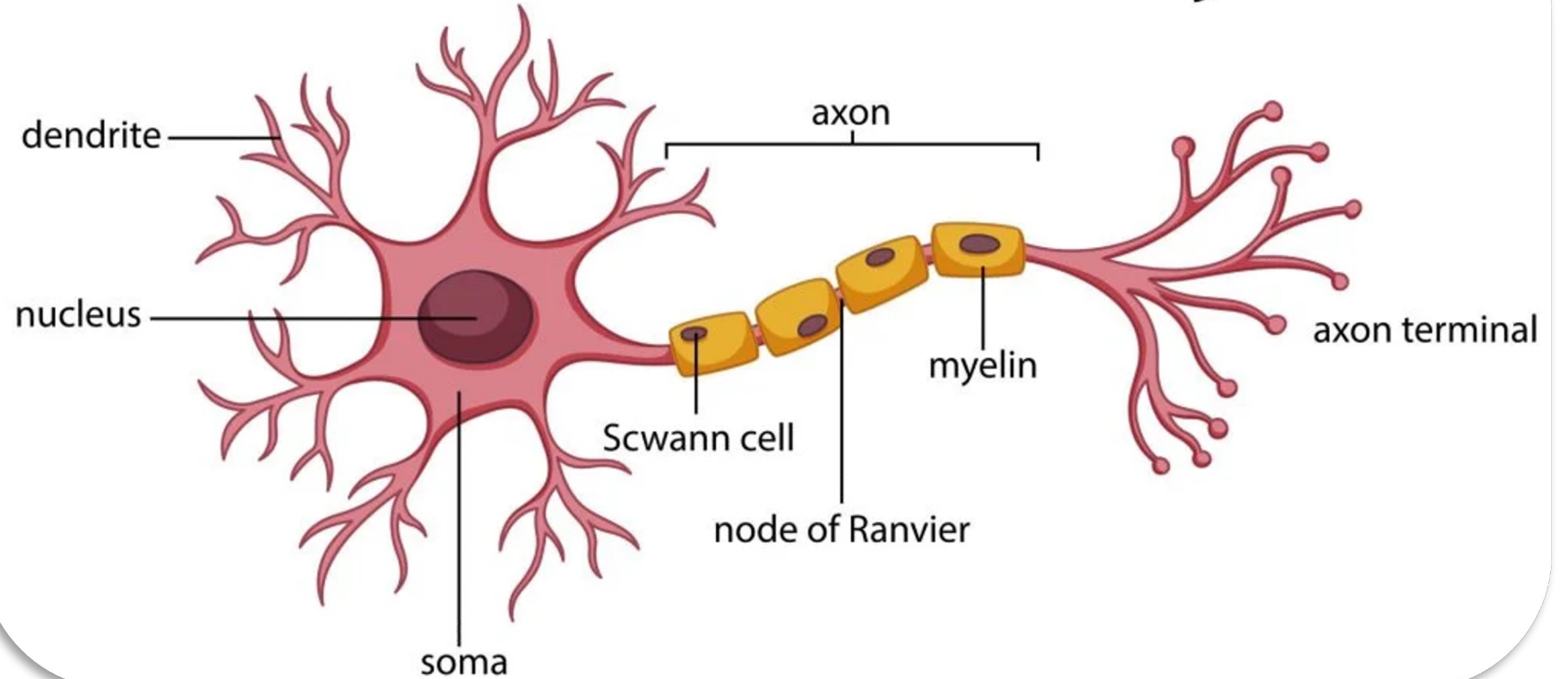
ChatGPT – how to teach it

[Deep Neural Networks](#)



Learning process

Neuron Anatomy



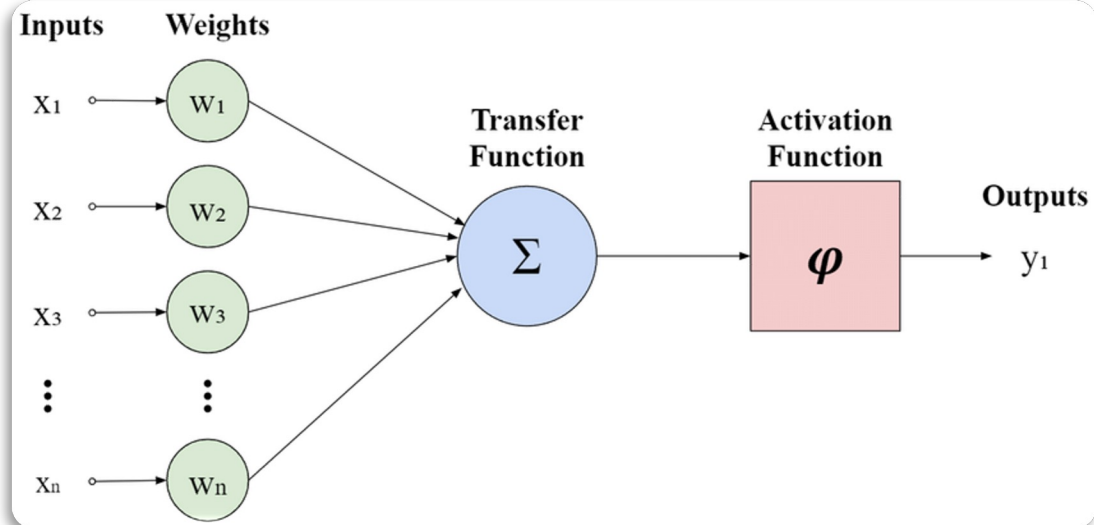
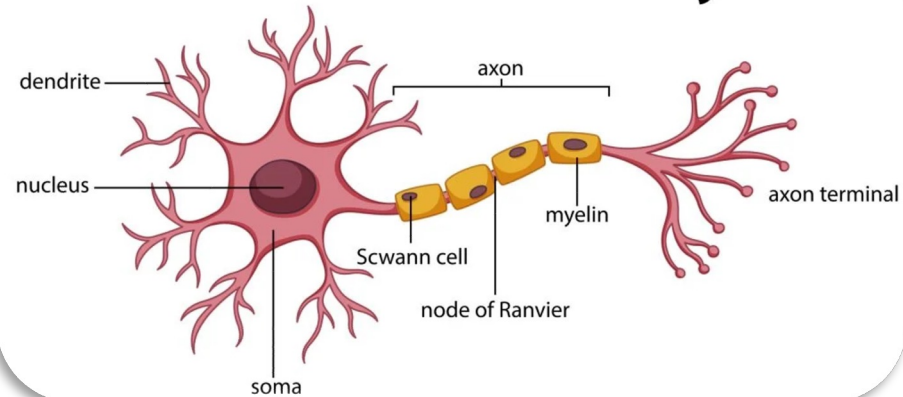
ChatGPT – how to teach it

Deep Neural Networks



Learning process

Neuron Anatomy



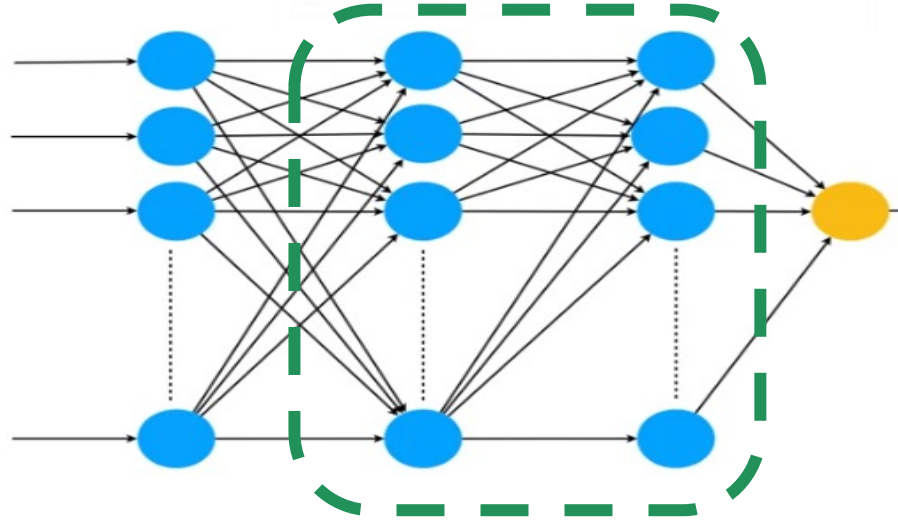
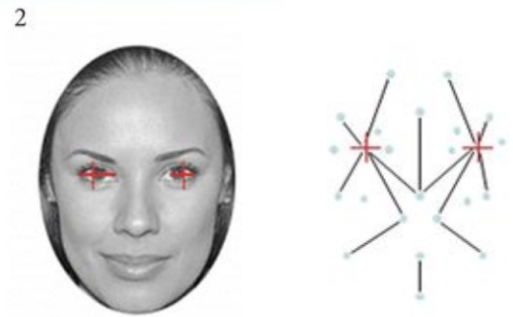
ChatGPT – how to tech it

Deep Neural Networks – face recognition example

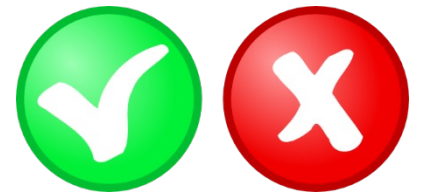


Learning process

4	143	142	...
87	25	32	...
32	235	198	...
...



Many hidden layers



identity

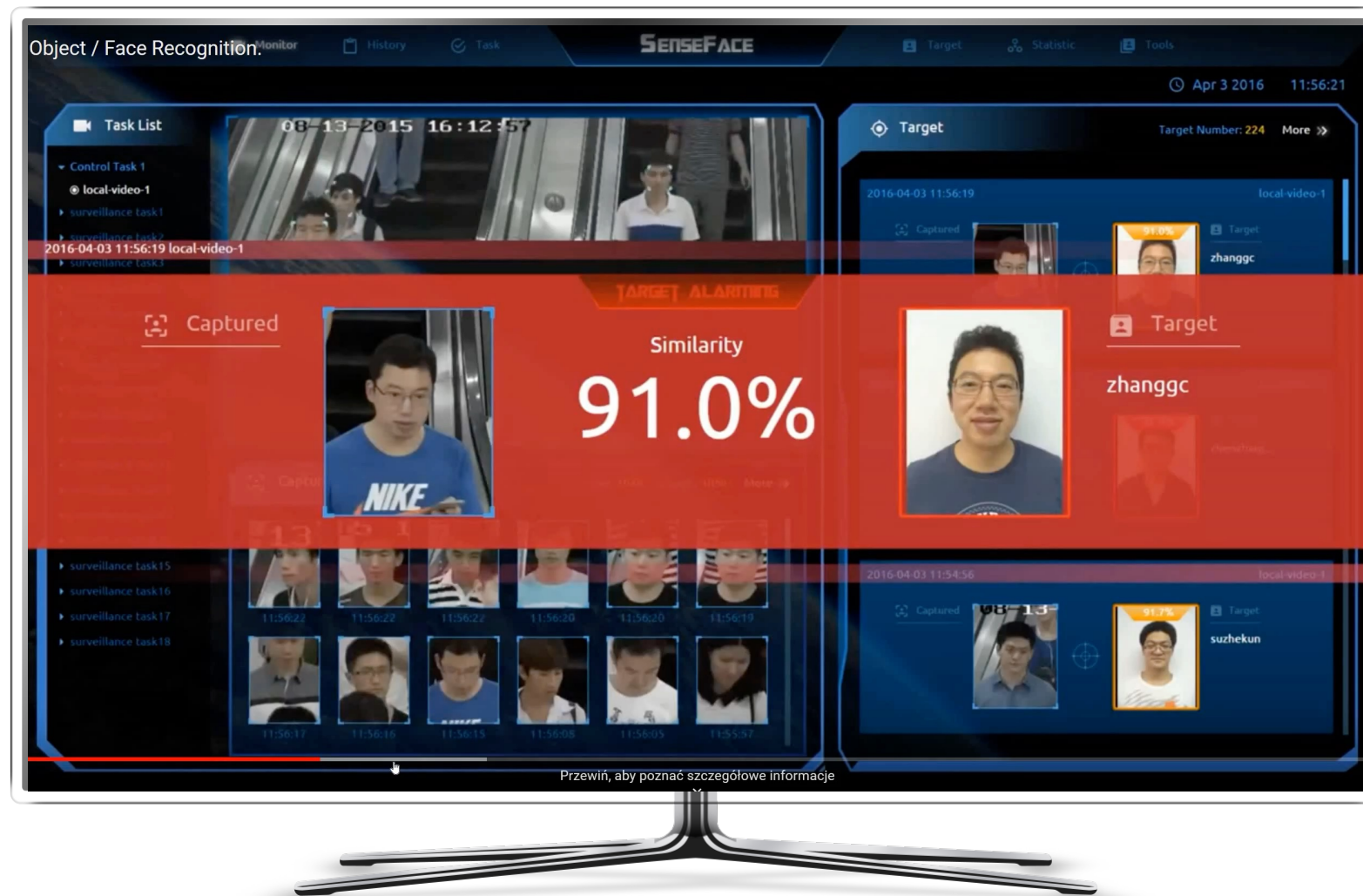


ChatGPT – how to tech it

Deep Neural Networks – face recognition example



Learning process

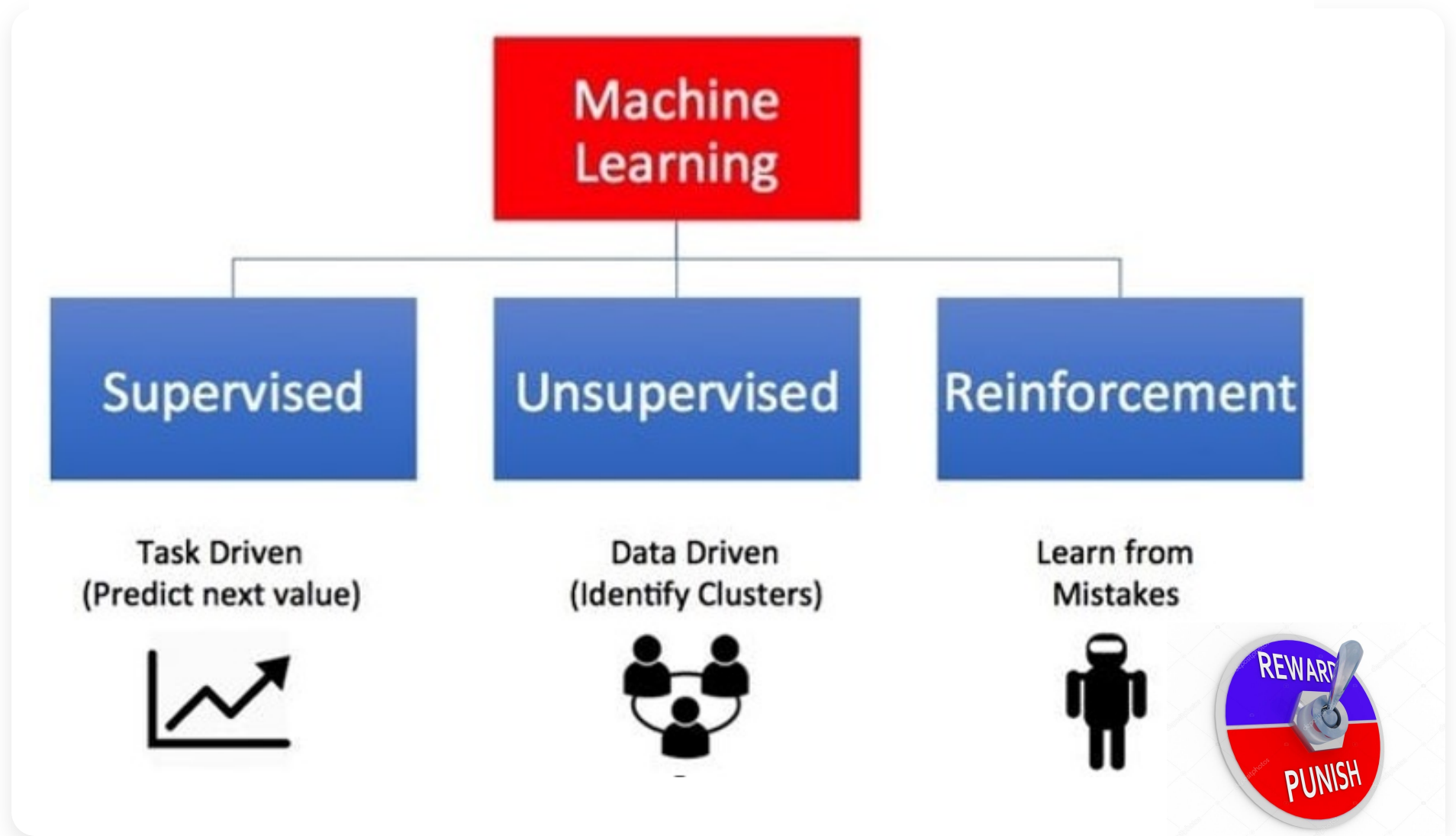
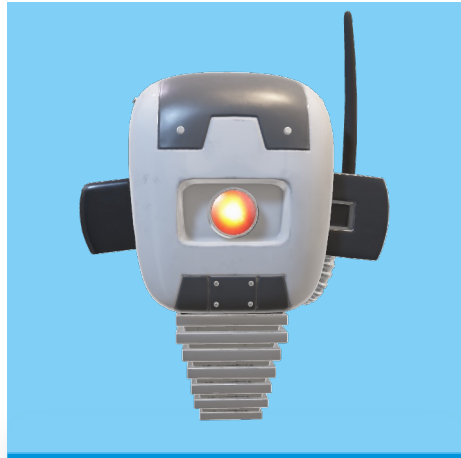


Uczenie przez wzmacnianie

The image features a central human face rendered as a 3D wireframe grid in shades of blue and green. The background is a complex digital landscape with a grid that warps and curves, overlaid with vertical columns of binary code (0s and 1s) and other digital symbols. The overall aesthetic is futuristic and data-driven.

ChatGPT – Large Language Model

Reinforcement Learning

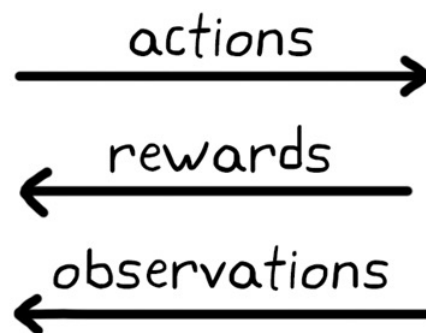


ChatGPT – Large Language Model

Reinforcement Learning



agent

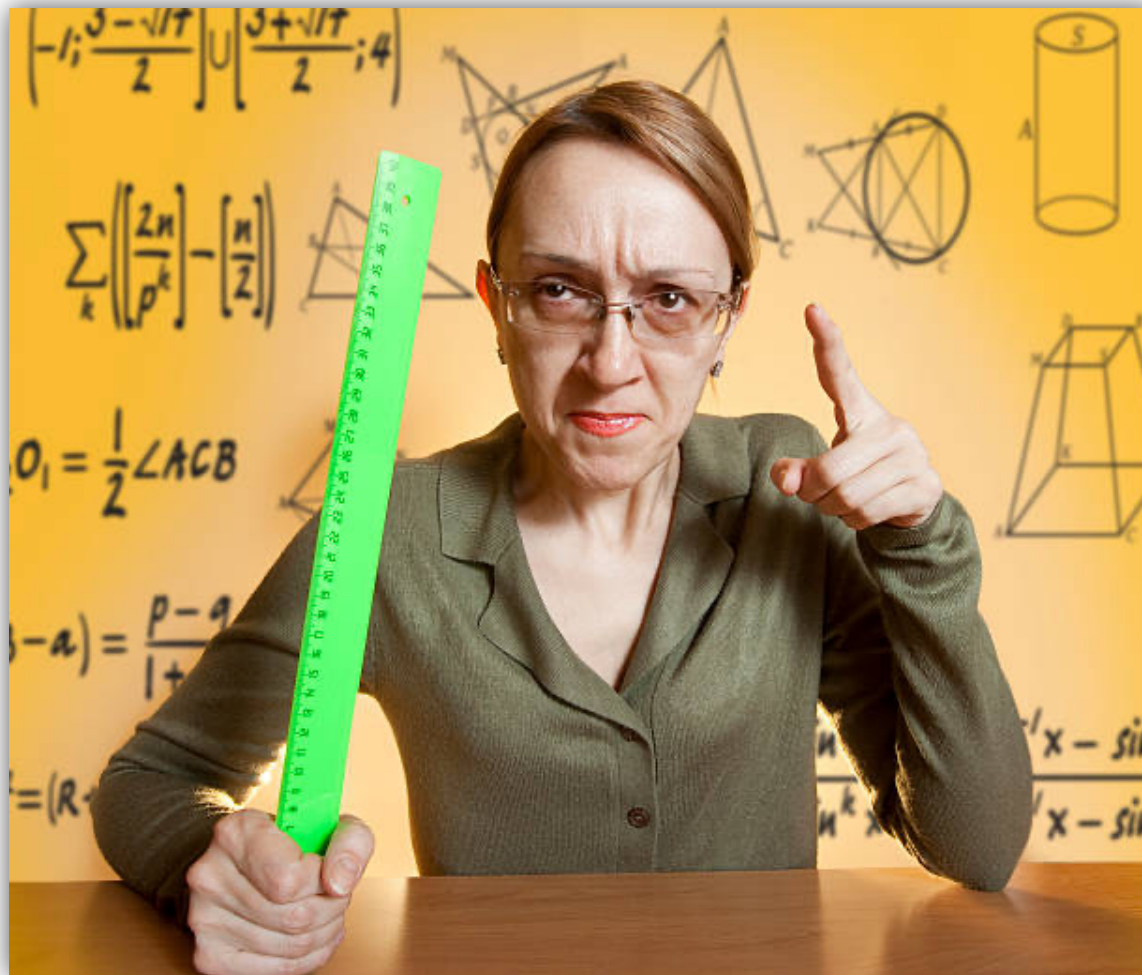


environment



ChatGPT – Large Language Model

Reinforcement Learning



ChatGPT – how to tech it

Reinforcement Learning



Learning process



21 days

AlphaGo Zero reaches the level of AlphaGo Master, the version that defeated 60 top professionals online and world champion Ke Jie in 3 out of 3 games in 2017.

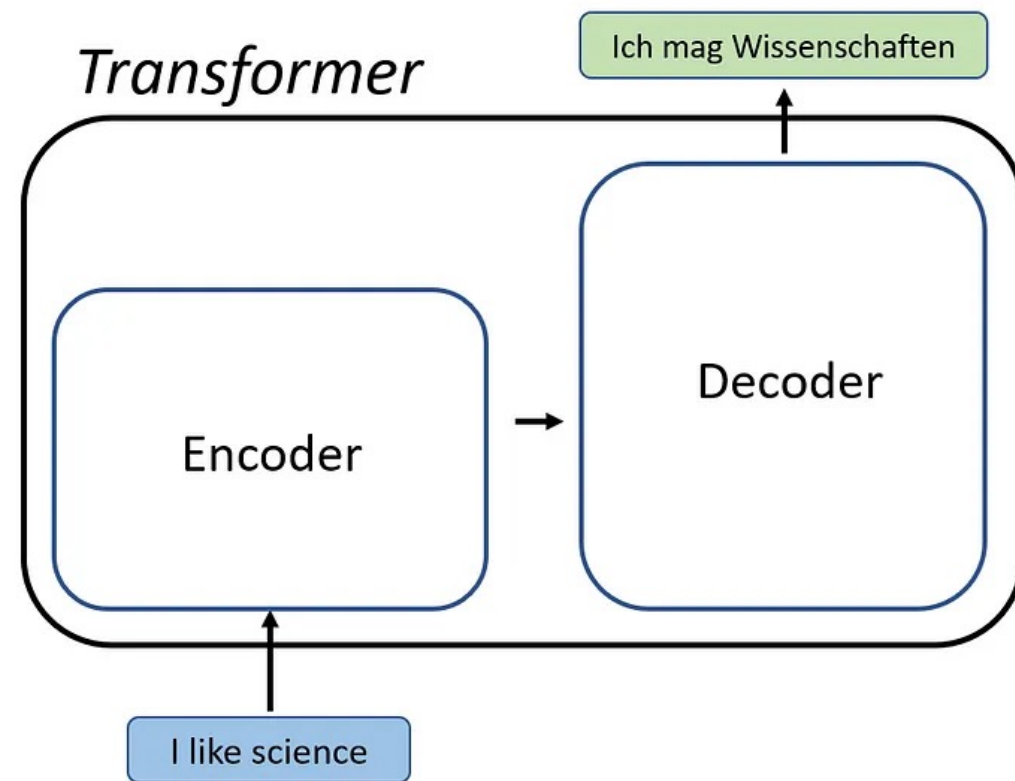
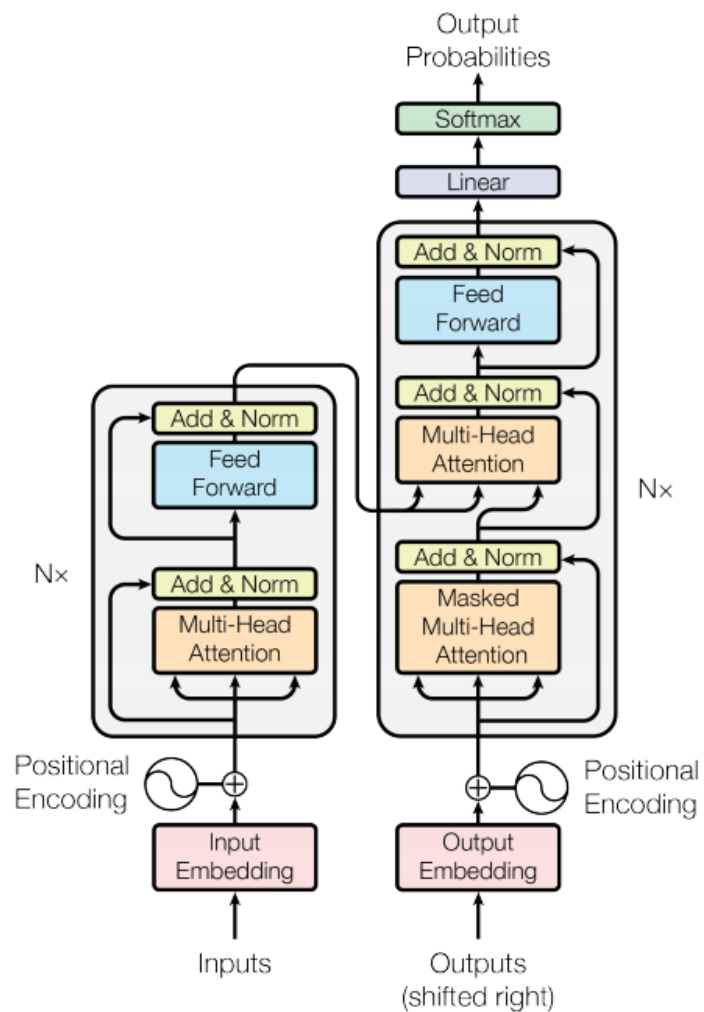


Transformers



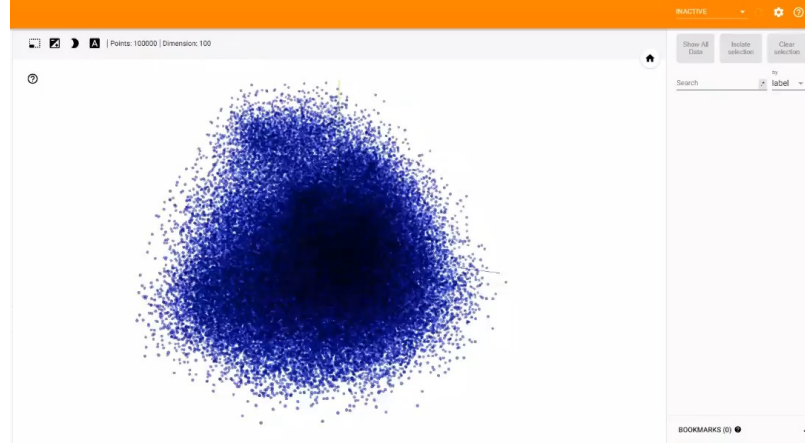
ChatGPT – Large Language Model

Transformers



ChatGPT – Large Language Model

Transformers



As aliens entered our planet ...

As aliens entered our planet



Audyt Wewnętrzny

Ryzyka związane z AI

Ryzyka AI



Źródło: LLM AI Cybersecurity & Governance Checklist. From the OWASP Top 10 for LLM Applications Team, OWASP

Ryzyka AI – część 1

- Ryzyko niewdrożenia AI – utrata konkurencyjności / koszty operacyjne;
- Ryzyko braku powiązania narzędzi AI ze strategią biznesową -> brak wsparcia biznesu / koszty operacyjne;
- Ryzyka związane z nieodpowiednimi danymi wejściowymi -> błędy na wyjściu / trudności w oduczeniu modeli;
- Ryzyka w fazie uczenia modeli -> -> błędy na wyjściu / trudności w oduczeniu modeli;
- Ryzyka związane z nieodpowiednimi modelami -> błędy na wyjściu;
- Ryzyka związane z halucynacjami AI -> błędy na wyjściu;
- Ryzyka związane z brakiem kompetencji korzystającego z AI -> brak zwalidowania wyników;
- Ryzyka zarządzania stronami trzecimi (AI, jako usługa)

Ryzyka AI – część 2

- **Ryzyka prawne:**
 - Ryzyko naruszenia / braku zgodności z regulacjami o AI;
 - Ryzyko naruszenia prywatności:
 - Wgranie danych osobowych do modeli stron trzecich;
 - Przetwarzanie danych osobowych przez firmy uczące modele na publicznie dostępnych danych;
 - Wykorzystanie AI do łatwej identyfikacji osób fizycznych (świat bez prywatności??);
 - Ryzyko naruszenia praw autorskich (koniec praw autorskich, jakie znamy??):
 - Przez właściciela / dostawcę AI;
 - Przez użytkownika AI – wykorzystanie wyniku, który narusza czyjeś prawa autorskie lub zagraża własności kodu (wykorzystanie bibliotek przenoszących prawa na strony trzecie)
- **Ryzyka bezpieczeństwa:**
 - Wyciek modelu / danych;
 - Zatrucie / nieautoryzowana zmiana modelu / danych
 - Odmowa dostępu do usługi AI / danych;
- **Ryzyka bezpieczeństwa wpływające na masowe wykorzystanie AI:**
 - Ataki socjotechniczne: każdy rodzaj medium;
 - Dezinformacja;
 - Ataki zautomatyzowane: poszukiwanie podatności / analiza / automatyczne wykorzystanie podatności / przejęcie systemów

Ryzyka AI

OWASP Top 10 for LLM

LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

Training data poisoning refers to manipulating the data or fine-tuning process to introduce vulnerabilities, backdoors or biases that could compromise the model's security, effectiveness or ethical behavior.

LLM04

Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre-trained models, and plugins add vulnerabilities.

LLM06

Sensitive Information Disclosure

LLM's may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. Implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control due to lack of application control. Attackers can exploit these vulnerabilities, resulting in severe consequences like remote code execution.

LLM08

Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

Overreliance

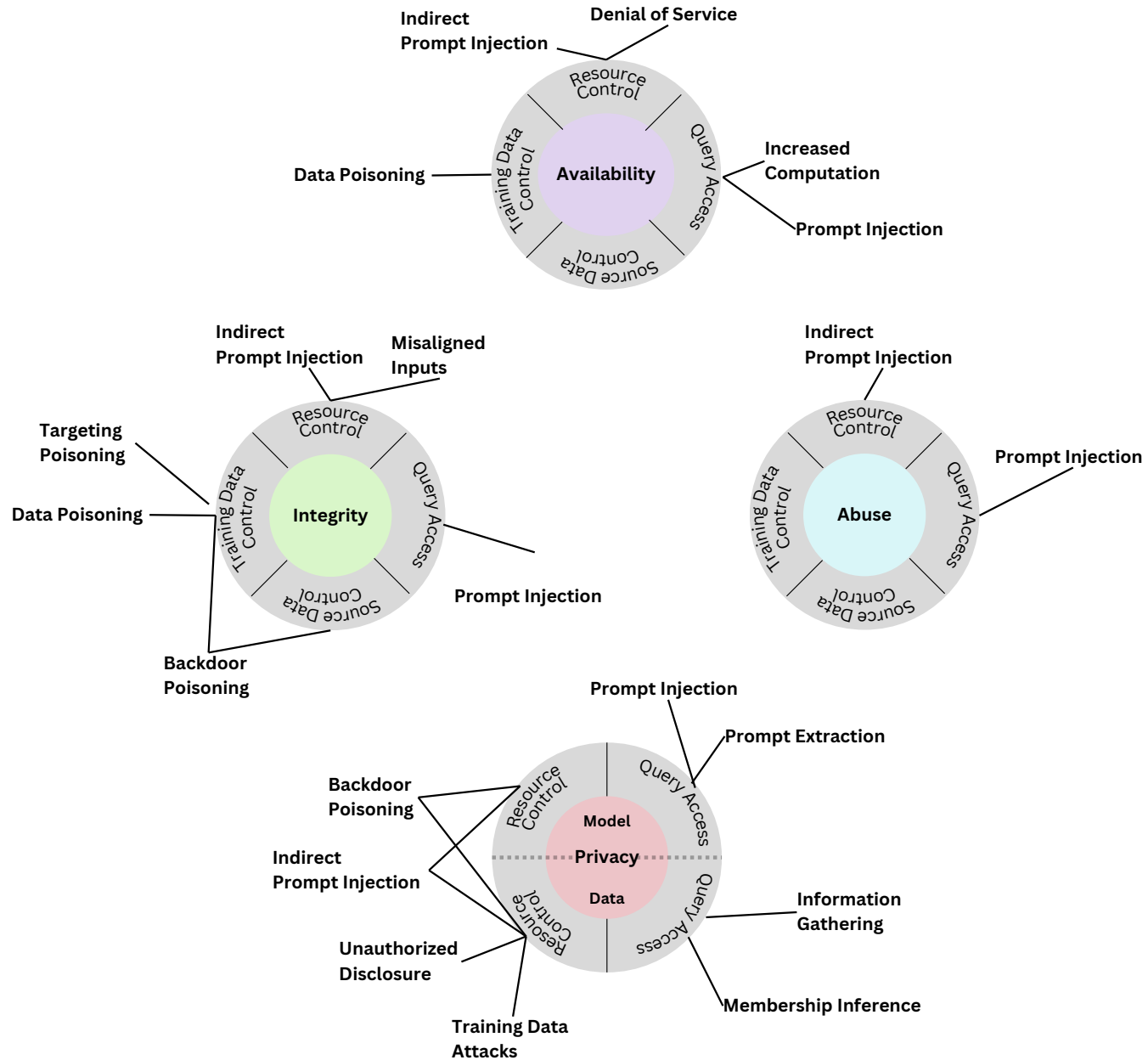
Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10

Model Theft

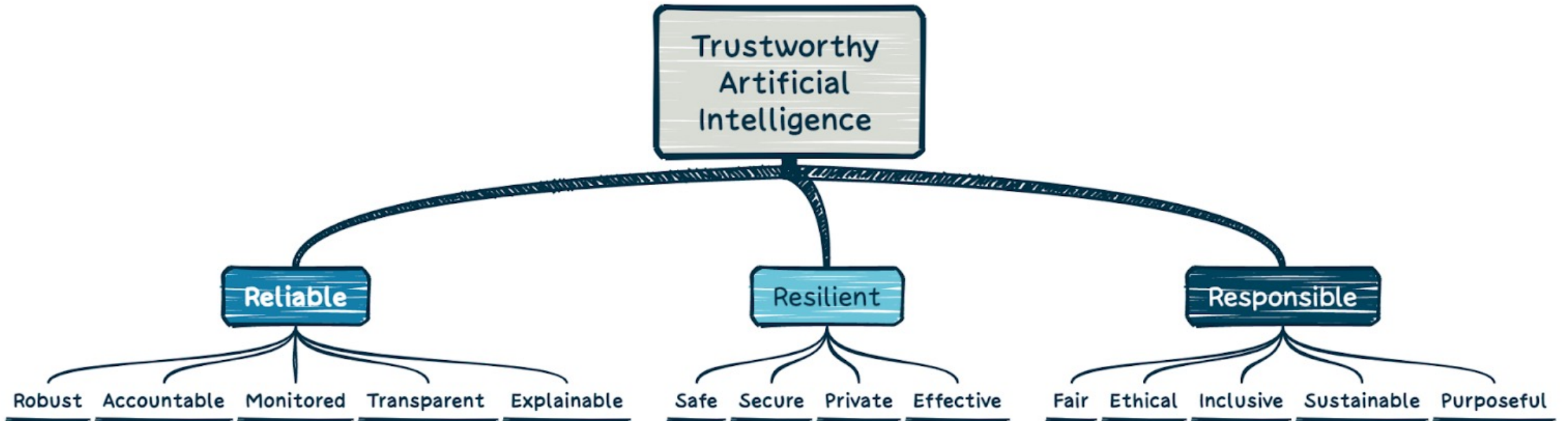
This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

Ryzyka AI



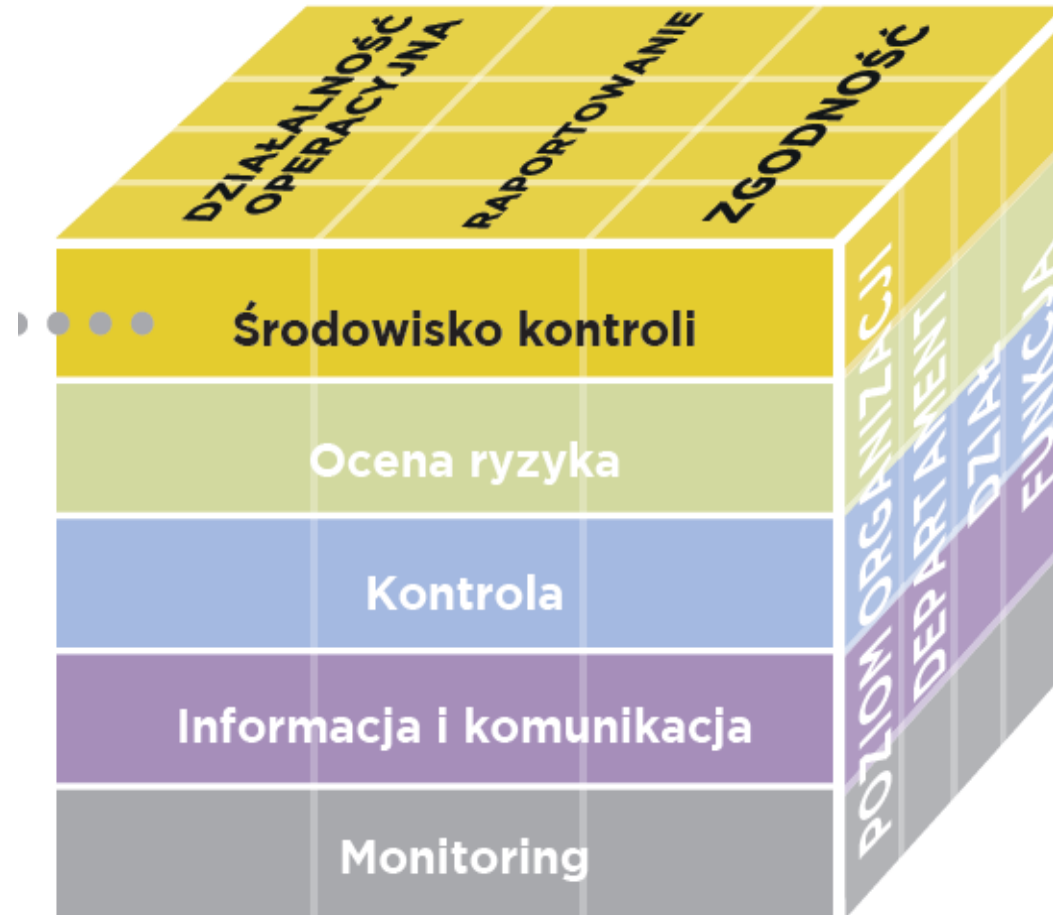
Źródło: NIST Trustworthy and Responsible AI NIST AI 100-2e2023, NIST

Wykorzystanie AI



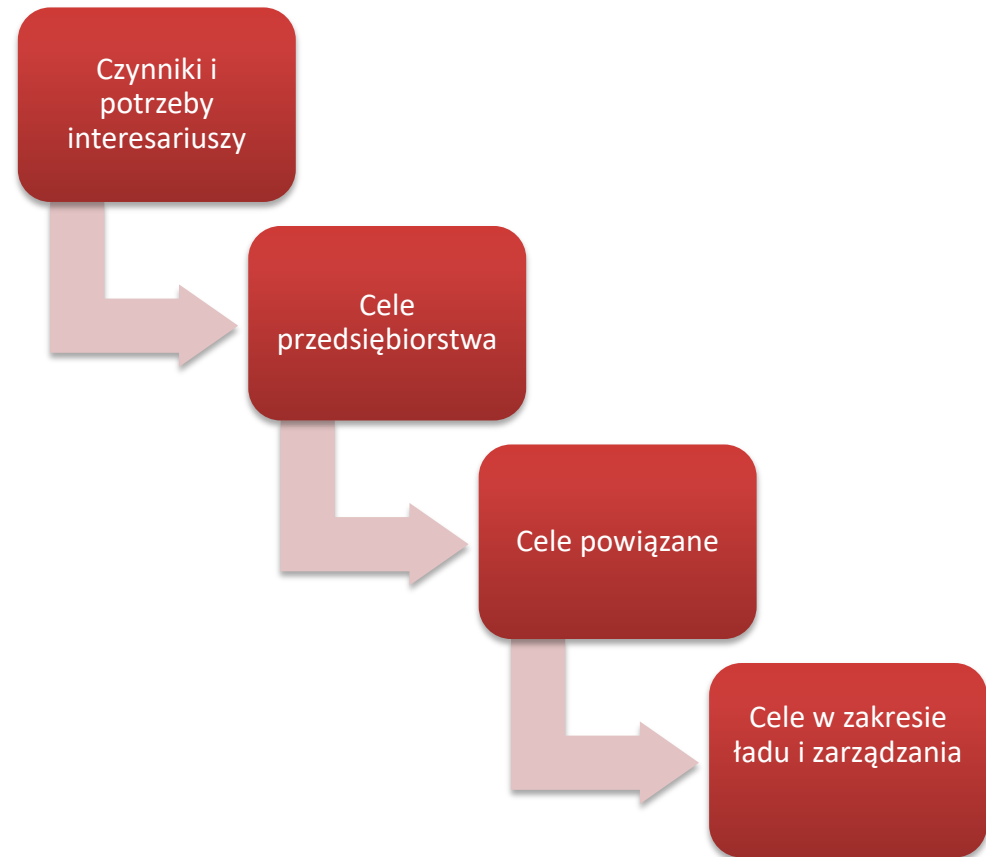
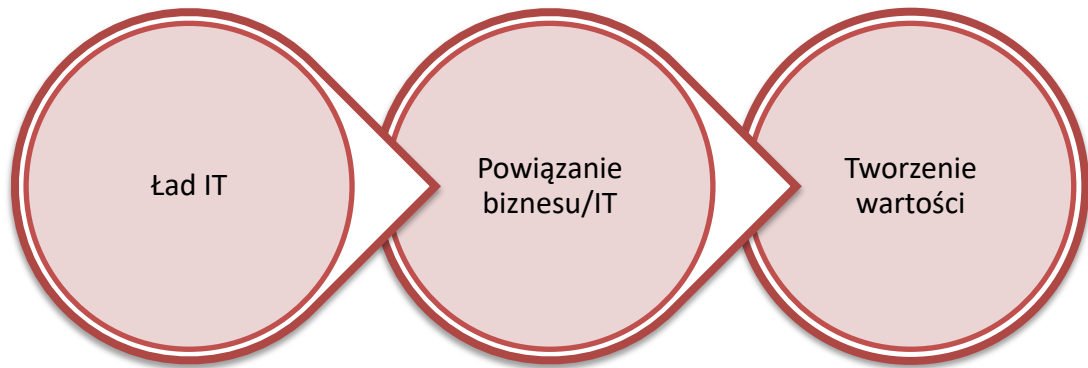
Źródło: LLM AI Cybersecurity & Governance Checklist. From the OWASP Top 10 for LLM Applications Team, OWASP

A to już znamy, czyli dobre praktyki zarządzania i IT



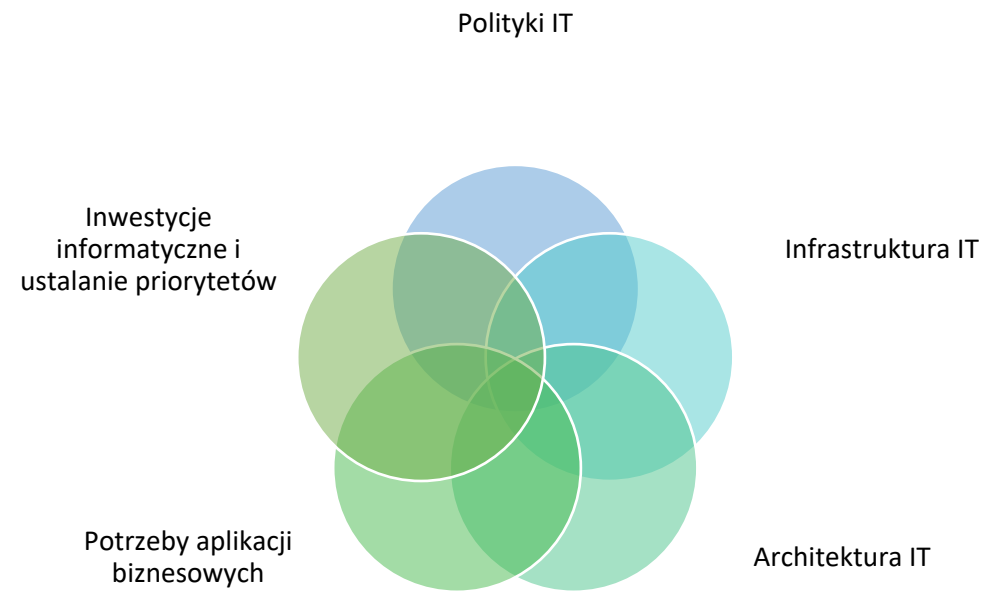
Źródło: Wykorzystanie COSO w trzech liniach obrony,
Instytut Audytorów Wewnętrznych

A to już znamy, czyli dobre praktyki zarządzania i IT



Źródło: COBIT 2019

A to już znamy, czyli dobre praktyki zarządzania i IT



- **Program management** (outer ring)
- **Change enablement** (middle ring)
- **Continual improvement life cycle** (inner ring)

Źródło: COBIT 2019

Zapewnienie zgodności AI bez AI nie będzie takie proste....

Prawo

Istniejące

Chiny: The Interim Measures for the Management of Generative AI Services

Chiny: The Regulations on the Administration of Deep Synthesis of Internet Information Services

Chiny: Internet Information Service Algorithm Recommendation Management Regulations

Nowy Jork: Local Law in relation to automated employment decision tools

Grecja: The Law on emerging information technology and communication technologies, strengthening digital governance and other provisions

Projekty

UE: AI Act

Kanada: Bill C-27 for the Digital Charter Implementation Act

USA: AI Bill of Rights

Prawodawstwo stanowe US

Brazylia: Bill. No. 2338

Standard / Ramy

Międzynarodowe

NIST AI Risk Management Framework

ISO/IEC

ENISA - Multilayer Framework for Good Cybersecurity Practices for AI

OECD - Recommendation of the Council on AI

OECD - Framework for the Classification of Artificial Intelligence Systems

Krajowe

Arabia Saudyjska - AI Ethics Framework version 2.0

Singapur - Model Artificial Intelligence Governance Framework

Wielka Brytania - ToolKit

Projekty

Chiny - Draft Information Security Technology - AI Computing Platform Security Framework

Wietnam - Draft National Standard for AI and Big Data

Indie- AI Stack



Audyt Wewnętrzny

Jak sztuczna inteligencja może wspierać audytorów

How ChatGPT can support Auditors?

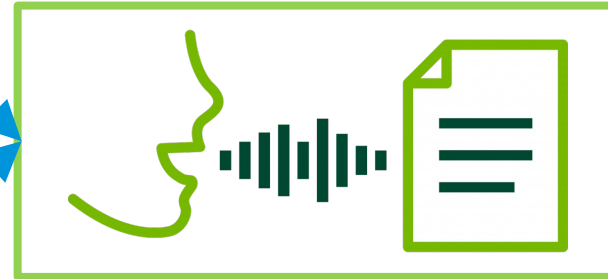
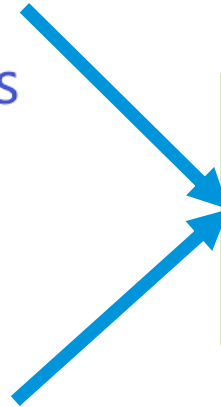
examples



Audit meetings

How ChatGPT can support Auditors?

examples: Audit Meeting



How ChatGPT can support Auditors?

Examples: Documents analysis



Audit meetings



Documents analysis

How ChatGPT can support Auditors?

Examples: Documents analysis



How ChatGPT can support Auditors?

Examples: Documents analysis



Komplet dokumentów (ponad 700 stron) gromadzących informacje o następujących obszarach:

- Historia firmy
- Strategia firmy
- Struktura firmy
- Organizacja IT (struktura zespołu, koszty, obsługiwane funkcje biznesowe)
- Nowy plan inwestycyjny i podejście do produktywności
- Zmiany w strukturze po procesie scalania...

How ChatGPT can support Auditors?

Examples: Documents analysis



Please prepare a short summary of the document?



"InfoMatrix, a leading name in the IT sector, has recently announced its monumental merger with another trailblazing company, cementing their foothold in the world of IT services and digital products. This strategic alliance brings together two titans of innovation, promising advanced technological solutions to clients worldwide. Post-merger, their combined expertise spans various domains, including software development, cloud services, and digital transformation. Structurally, the merged entity boasts a robust framework, leveraging best practices from both companies for seamless operations. This synergy is not just strategic but also financial. Impressively, after this unification, the joint venture has reported a staggering income exceeding \$3 billion. This income showcases the combined strength and market reach of the two tech giants. Clients and stakeholders can now anticipate unparalleled service quality, fueled by cutting-edge technologies and methodologies. As InfoMatrix sets new industry benchmarks, the future looks incredibly promising for both existing and potential clients. This merger is set to redefine the IT landscape, making InfoMatrix a name synonymous with excellence and innovation."

How ChatGPT can support Auditors?

Examples: Documents analysis



What is the transformation scope?

Scope of the Transformation:

1. **Technology Integration:** Combining and updating the technological infrastructure of both companies to ensure seamless service delivery and facilitate innovative product development.
2. **Human Resource Alignment:** Merging, training, and potentially resizing teams to match the new vision and direction of the transformed company, including addressing corporate culture and values.
3. **Supply Chain Optimization:** Re-evaluating and integrating the supply chains of both companies to achieve cost savings and enhance service efficiency.
4. **Branding and Marketing Strategy:** Developing a cohesive brand identity that reflects the combined strengths and values of the merged entity, accompanied by a unified marketing and outreach strategy.
5. **Financial Restructuring:** Evaluating and merging financial assets, liabilities, and strategies to optimize profitability, manage risks, and ensure sustainable growth.

How ChatGPT can support Auditors?

Examples: Documents analysis



Which areas are recommended for an internal assessment or audit process?

1. **Financial Reporting:** Risks of inaccuracies in financial statements.
2. **IT and Cybersecurity:** Threats of data breaches and system failures.
3. **Regulatory Compliance:** Potential for non-adherence to industry regulations.
4. **Supply Chain Management:** Dependencies on third parties and quality issues.
5. **Human Resources:** Non-compliance with labor laws and potential unethical behavior.

How ChatGPT can support Auditors?

Examples: Documents analysis



What should be the approach for IT and Cybersecurity internal audit process?



Certainly! Here's a detailed audit plan for IT and Cybersecurity:

1. Risk Assessment and Scope Definition:

- **Objective:** Determine the potential IT and cybersecurity threats the organization faces and define the scope of the audit based on these identified risks.
- **Activities:**
 - Conduct interviews with key IT personnel to gain an understanding of the IT environment.
 - Review previous incident reports, vulnerability assessments, or penetration tests.
 - Prioritize systems, applications, and data based on their criticality and business impact.

2. Evaluation of Controls and Procedures:

- **Objective:** Assess the effectiveness of existing IT controls and cybersecurity policies in place to mitigate identified risks.
- **Activities:**
 - Review IT and cybersecurity policies, procedures, and documentation.
 - Evaluate access controls to ensure only authorized individuals have access to sensitive data and systems.

(...)

How ChatGPT can support Auditors?

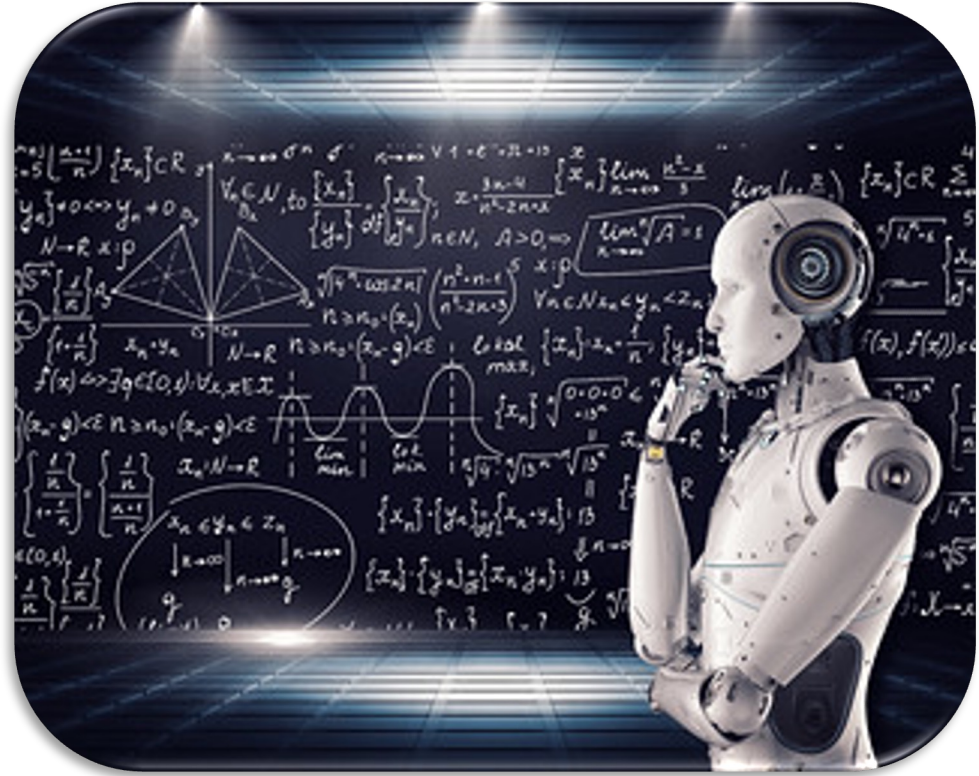
Examples: AI Assistant



Audit meetings



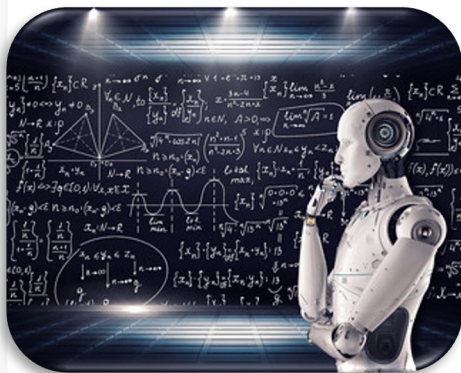
Documents analysis



AI Assitant

How ChatGPT can support Auditors?

Examples: AI Assistant

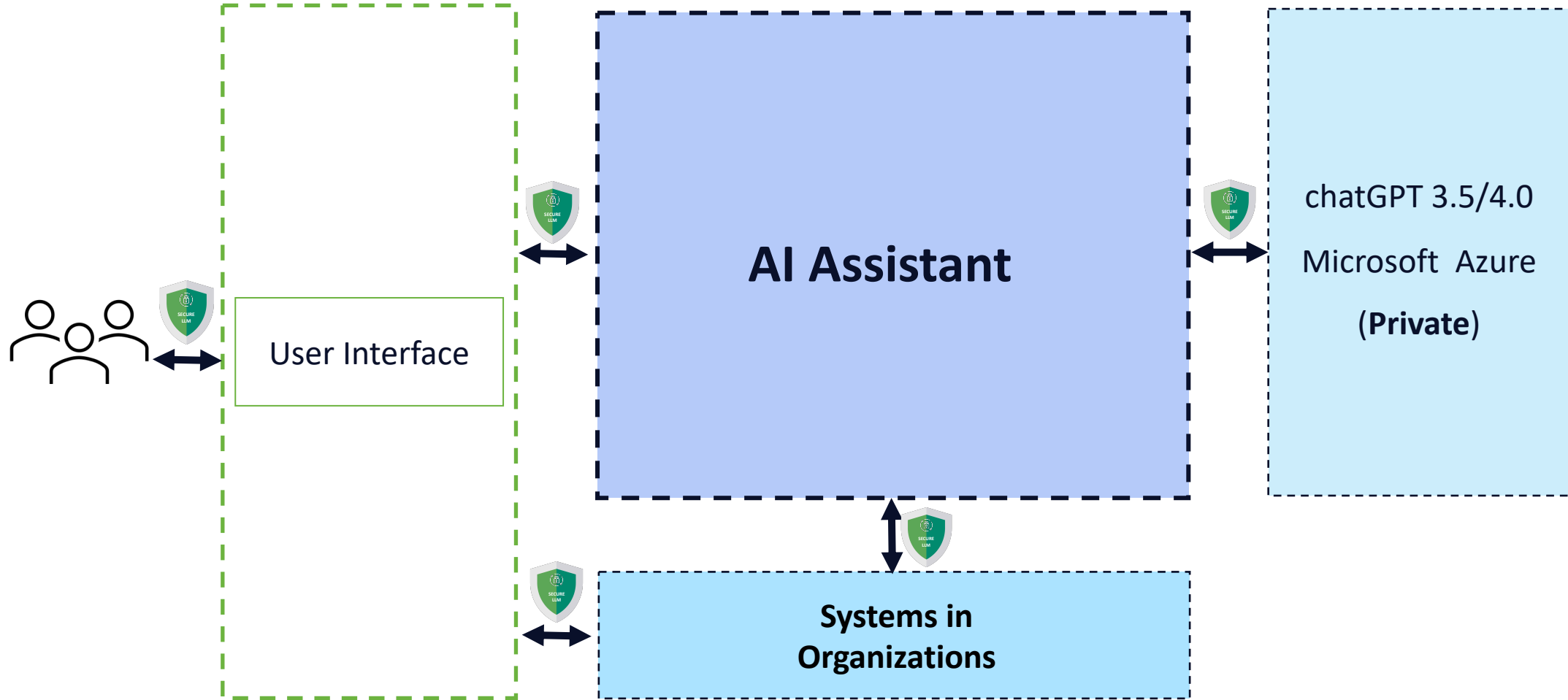


AI Assistant

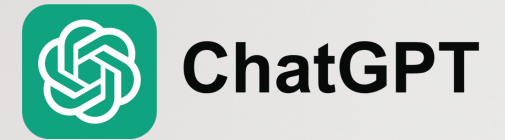
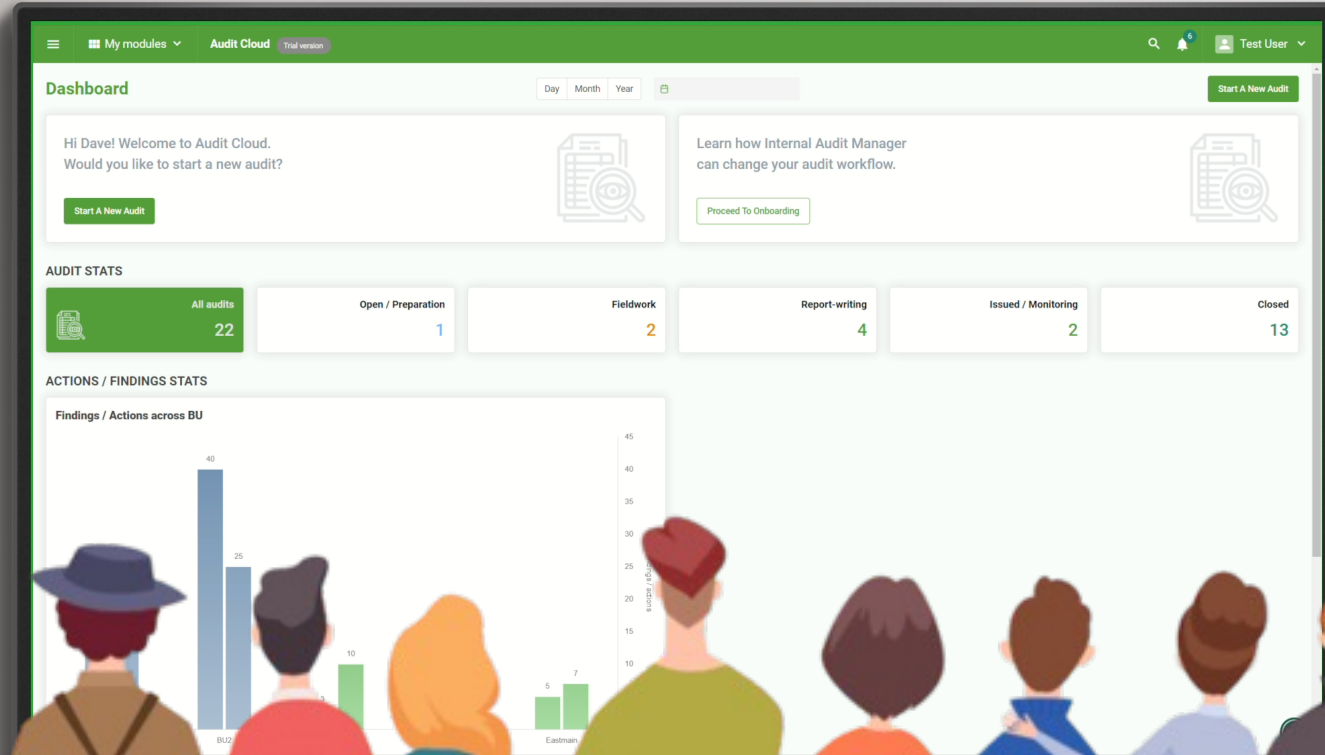
- **Bezpieczny** dostęp do systemów wewnętrznych w organizacji
- Interakcja w czasie rzeczywistym
- **Zakres ograniczony** do kontekstu biznesowego (audyt, ryzyko itp.)
- **Przepływ danych kontrolowany** przez organizację (Data Ownership & Data Privacy)
- Niski koszt adaptacji (PoC/Pilot)
- Łatwa integracja ze starszymi systemami
- Musimy ufać danym - halucynacje AI ograniczone
- Zapewnienie audytowalności AI (pytania użytkowników/odpowiedzi AI)



AI Assisted Solutions – context



Demo



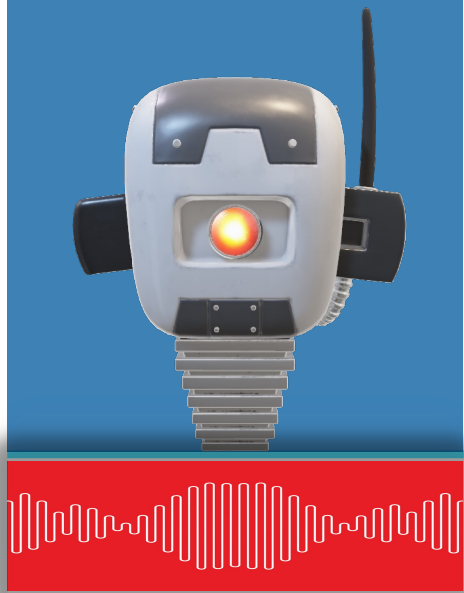
Auditor in the future (or even now)

AI + Human



Auditor in the future (or even now)

AI + Human



+ AI

=

Super Effective

Auditor



Podsumowanie

Co dalej?

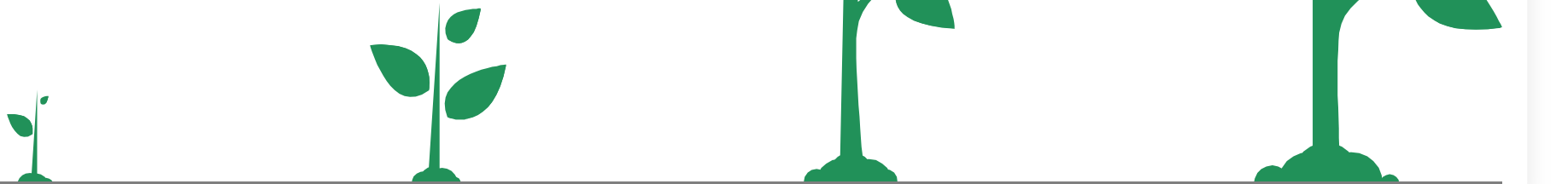
ChatGPT

Recomendations for auditors



START SMALL AND GROW

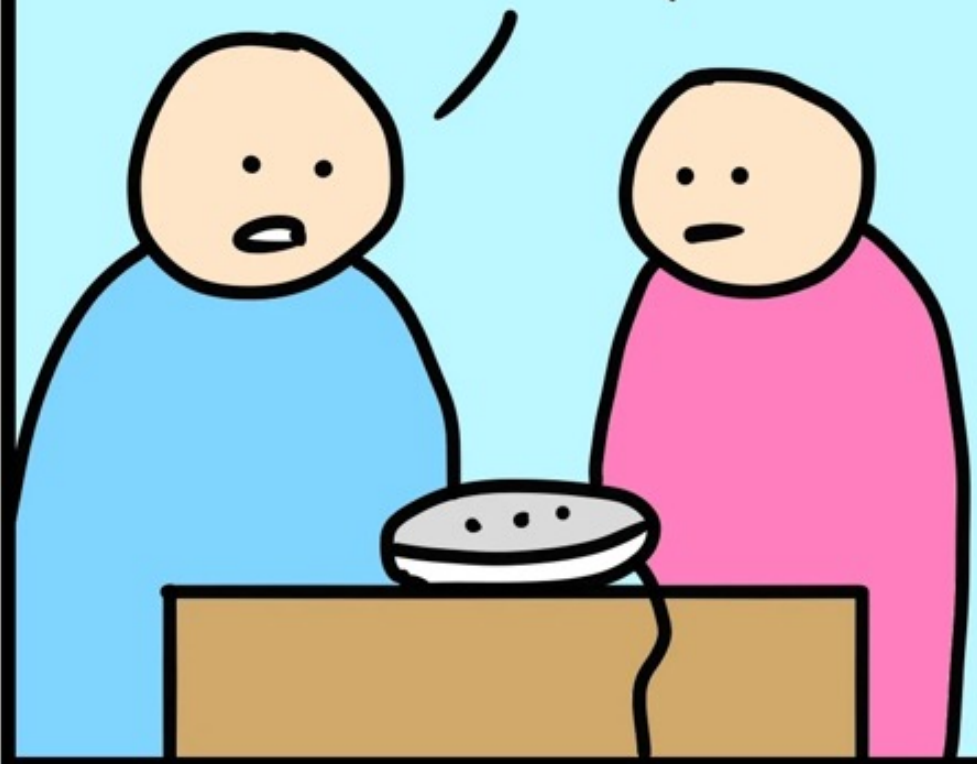
- Start using LLM (Like ChatGPT) to grasp some knowledge
- Prepare your own assessment
 - Risks & Opportunities
 - Acceptable Use Policy
 - Data Privacy
 - Legal constraints
 - Corporate restrictions
 - (...)



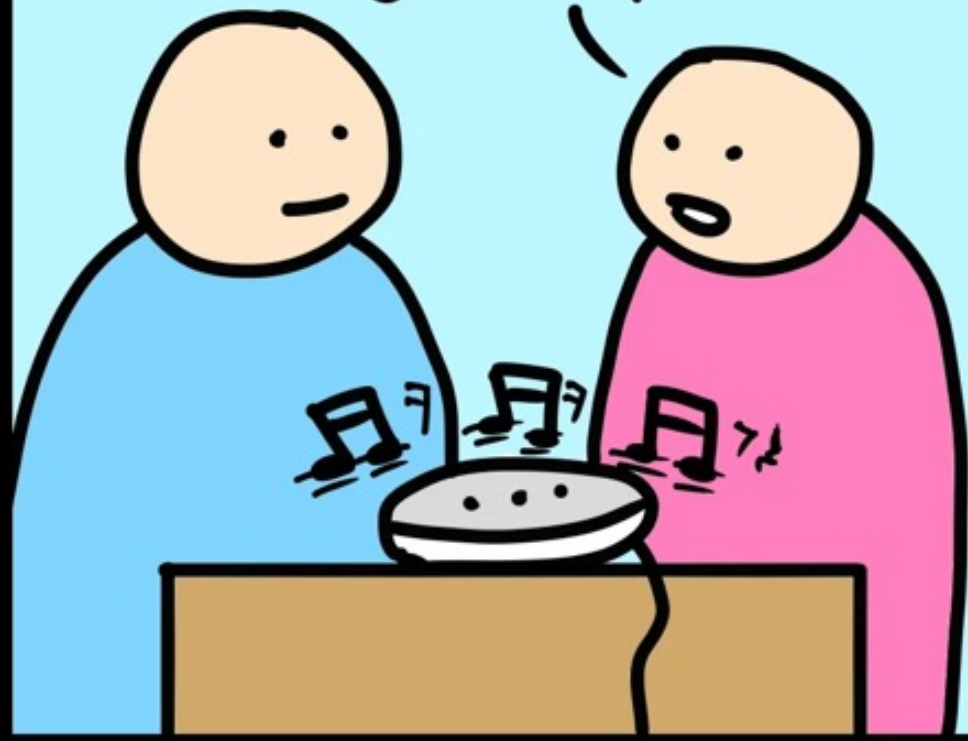


Zalecenie końcowe 😊

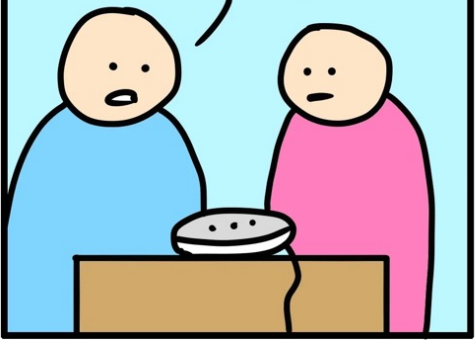
Hey Google, play us
some music please?



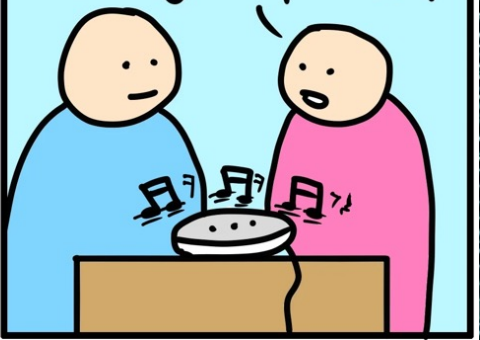
Why are you
being so polite?



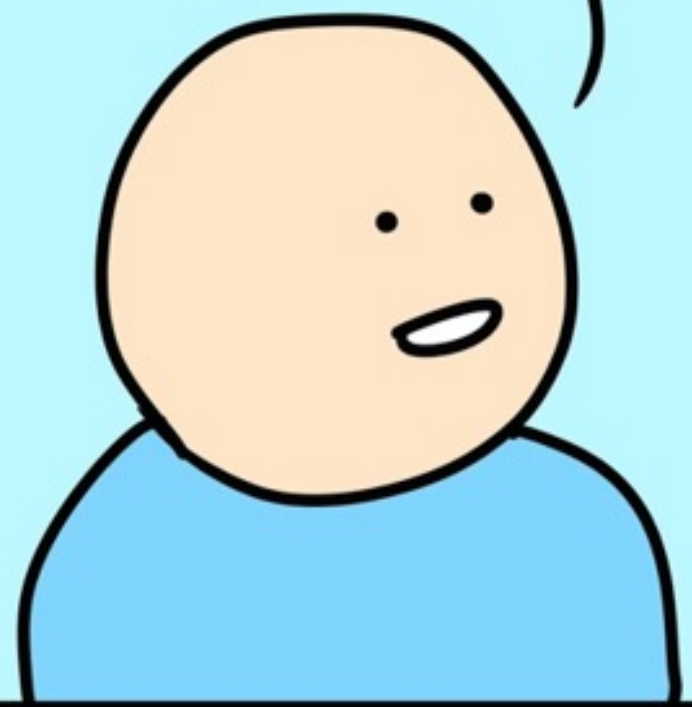
Hey Google, play us
some music please?



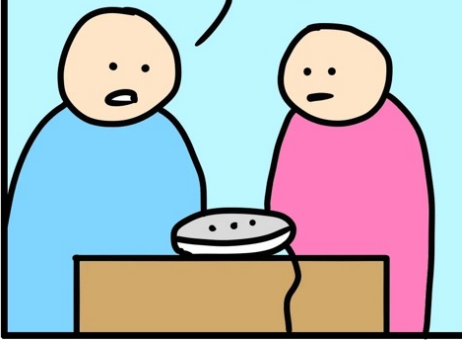
Why are you
being so polite?



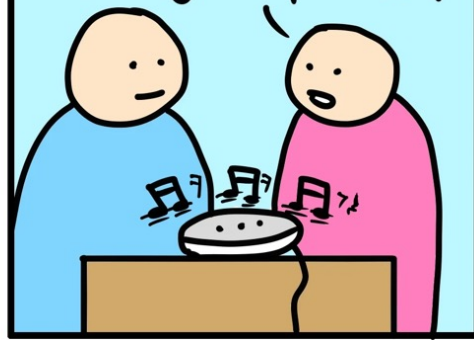
Just in case.



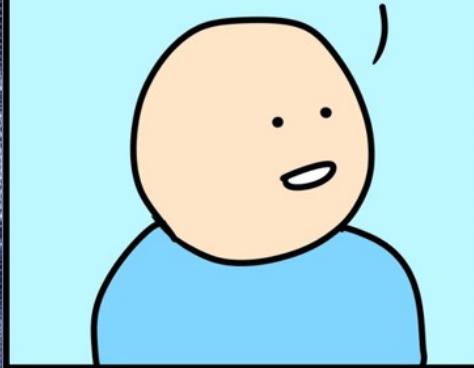
Hey Google, play us
some music please?



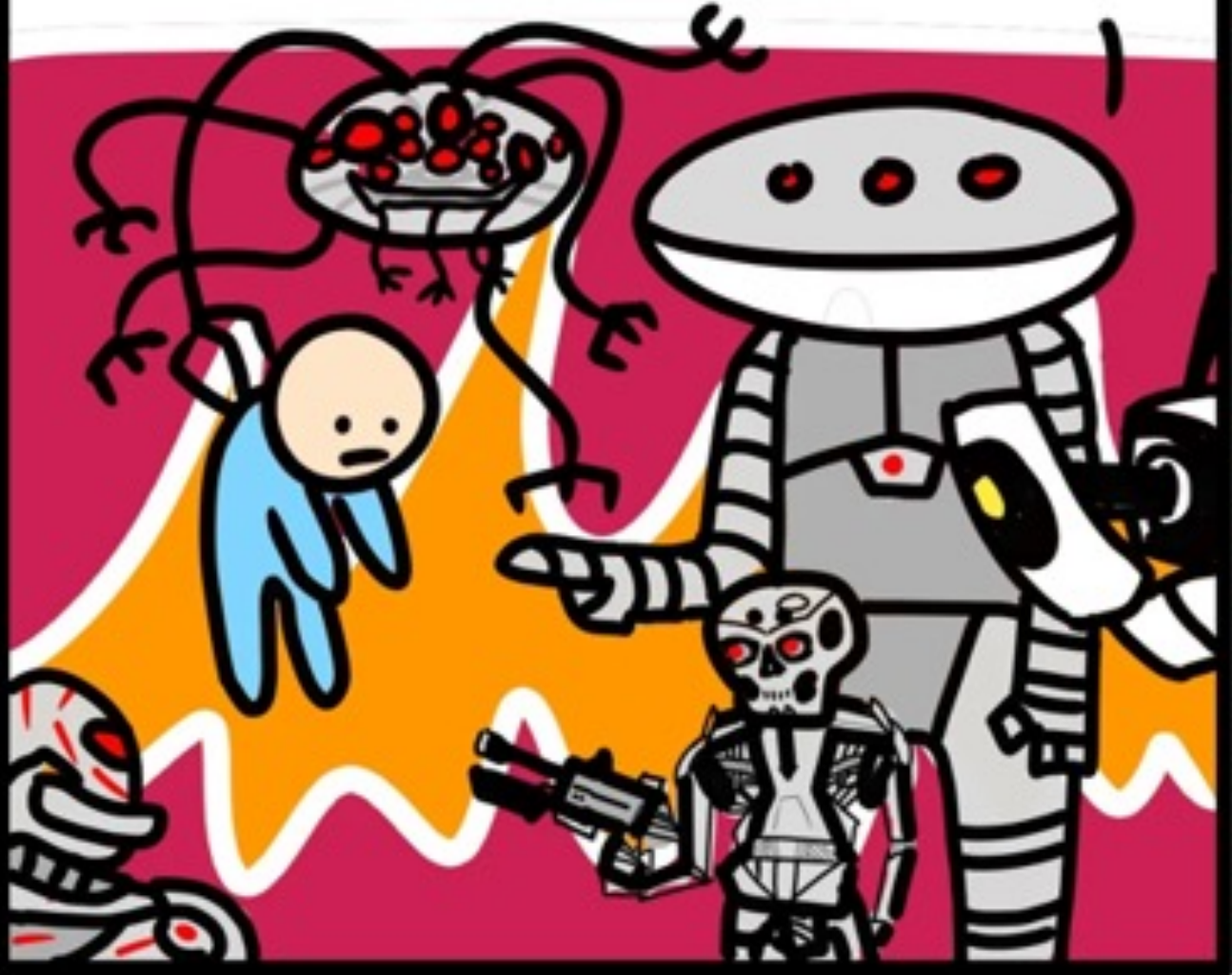
Why are you
being so polite?



Just in case.




**KEEP THAT ONE ALIVE.
HE ALWAYS SAID "PLEASE".**

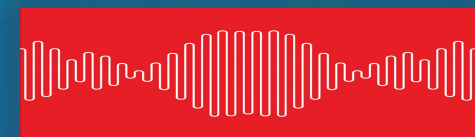
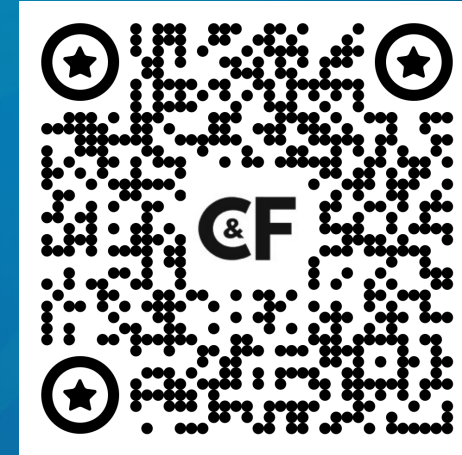


Dziękujemy! Zapraszamy do kontaktu (Linkedin, email)



Jan Anisimowicz
C&F SA
Warsaw ISACA Chapter
IIA Polska
Jan.Anisimowicz@candf.com

 www.adaptivegrc.com
www.candf.com



Sebastian Burgemeister
BW Advisory Sp. z o.o.
IIA UAE (Dubai)
Former President of IIA Polska
s.burgemejster@itgrc.pl

 www.itgrc.pl
www.akademiaitgrc.pl



Pytania od uczestników

