

The importance of Auditing in the age of Generative AI

RO Masterclass 2023: Scarcity

—

23-11-2023



Instituut van
Internal Auditors

Nederland

Introduction

ir. Marc van Meel



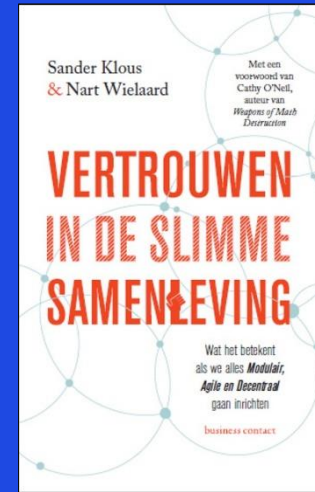
- Managing Consultant Responsible AI
- AI & Ethics Lead
- Recovering Data Scientist

Some concrete examples

×
×
× **City of
Amsterdam**

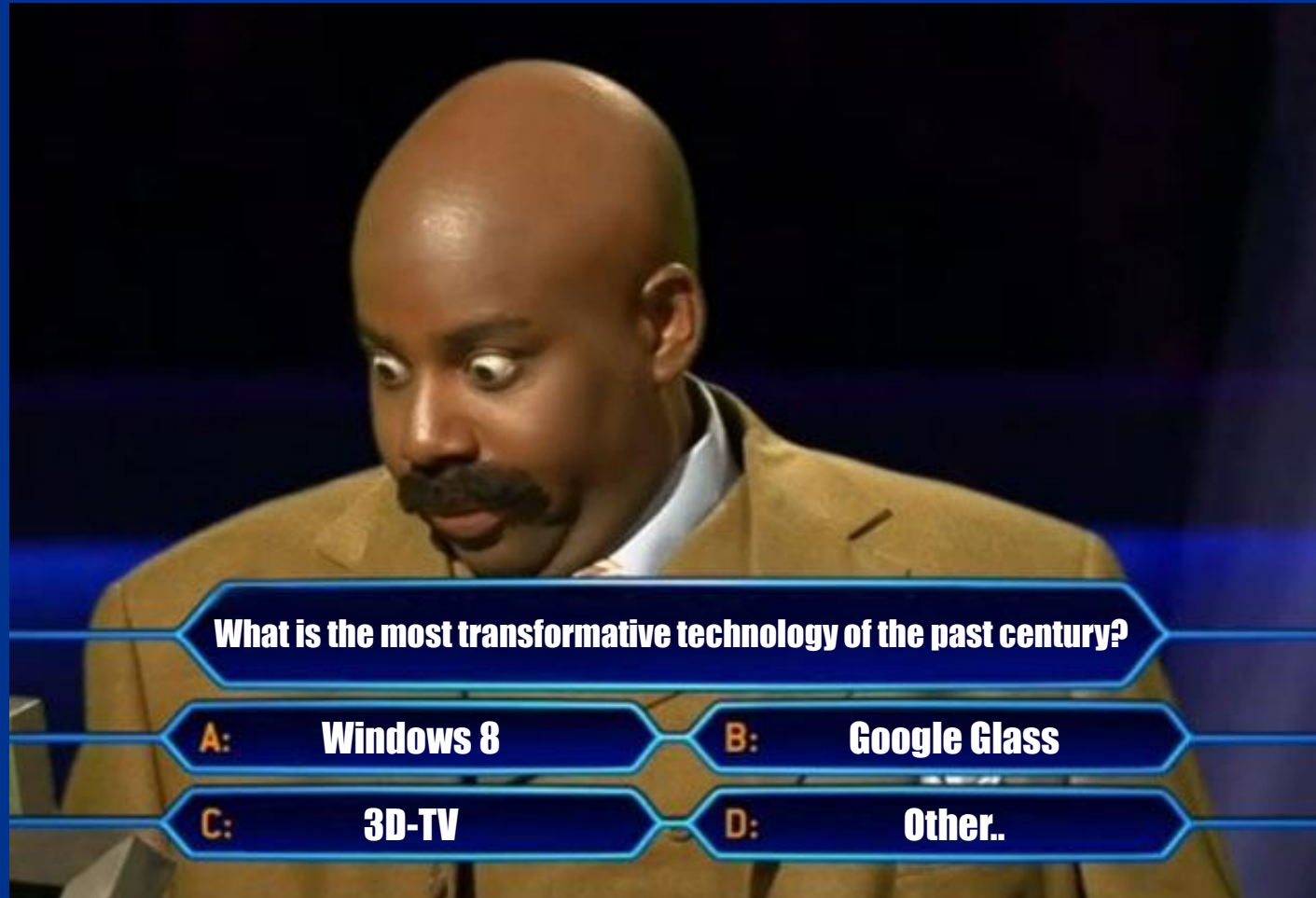


VERBOND VAN VERZEKERAARS



Prof. dr. S. (Sander) Klous –
Partner KPMG
University of Amsterdam

Q: What is the most transformative technology of the past century?



It's.. the washing machine!



The telegraph vs the internet



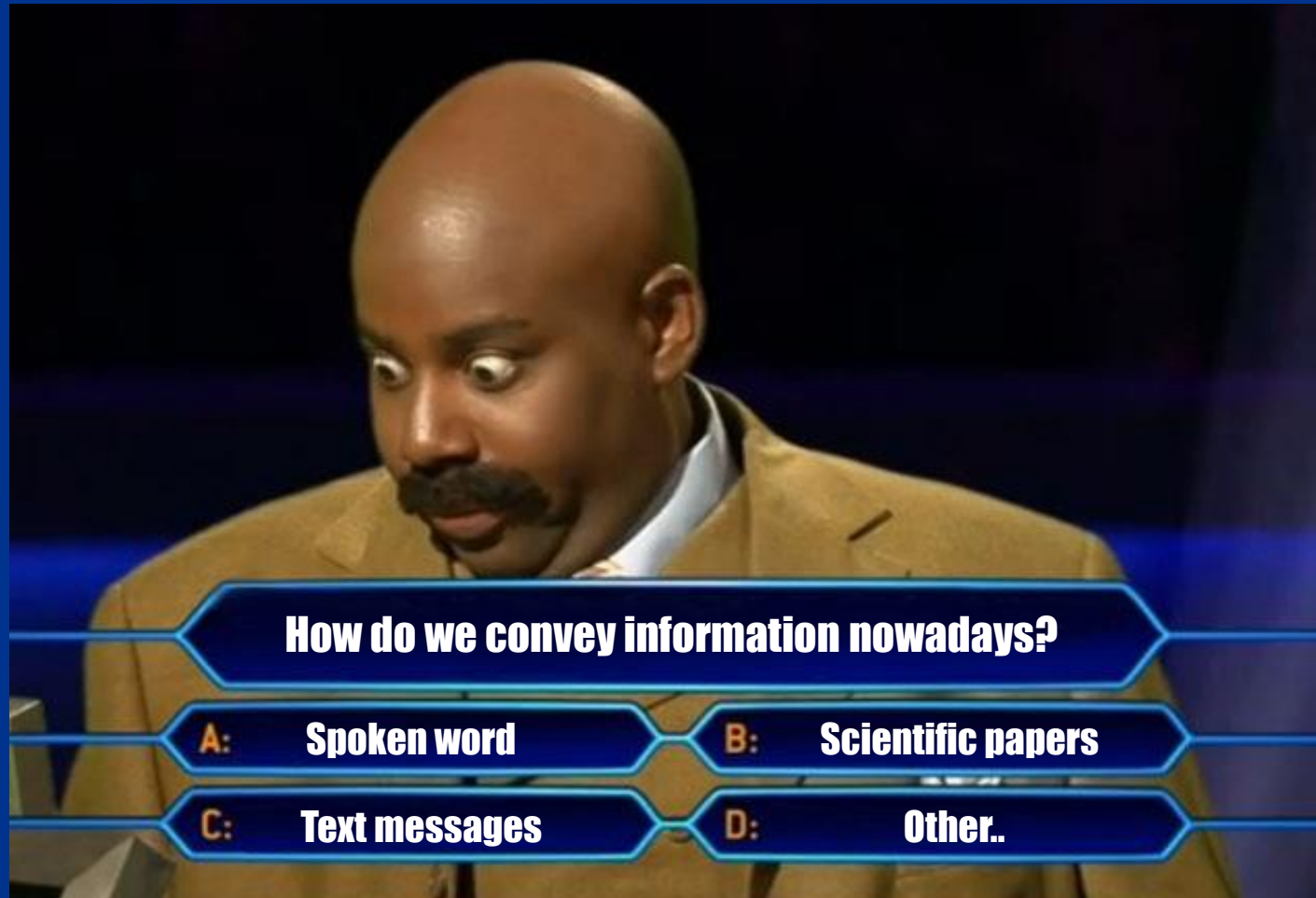
19th century

VS



20th century

Q: How do we convey information nowadays?



It's.. by memes!

INTERNET CULTURE

‘We actually elected a meme as president’: How 4chan celebrated Trump’s victory

Analysis by [Abby Ohlheiser](#)

Freelance reporter

November 9, 2016 at 2:00 p.m. EST

The Washington Post

Meme warfare: how the power of mass replication has poisoned the US election

Spread from the bottom up, political memes are now a form of propaganda - and it's killing our ability to intelligently orchestrate a political conversation

The
Guardian



Information made for consumption

- We now live in a **visual-audio society** -> appearance is everything
- We consume **short-form content** -> distractions and declining attention span
- Lack of context can lead to **polarization** between consumers
- We mistake **abstractions** for the thing itself
- We lost sense of what it means to be **well-informed**
- **Lack** of knowledge vs **overabundance** of knowledge (ignorance)
- **Entertainment** has become the default form communication



Social media is shortening our attention spans

By using their phones less, Generation Z can accomplish more

September 2, 2022 | Dante Caloia

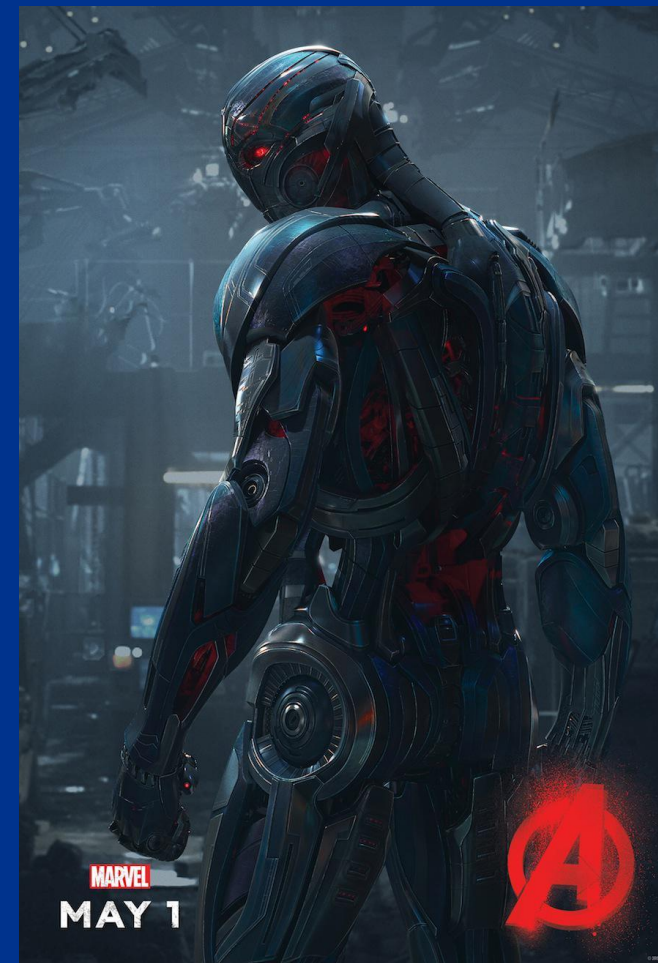
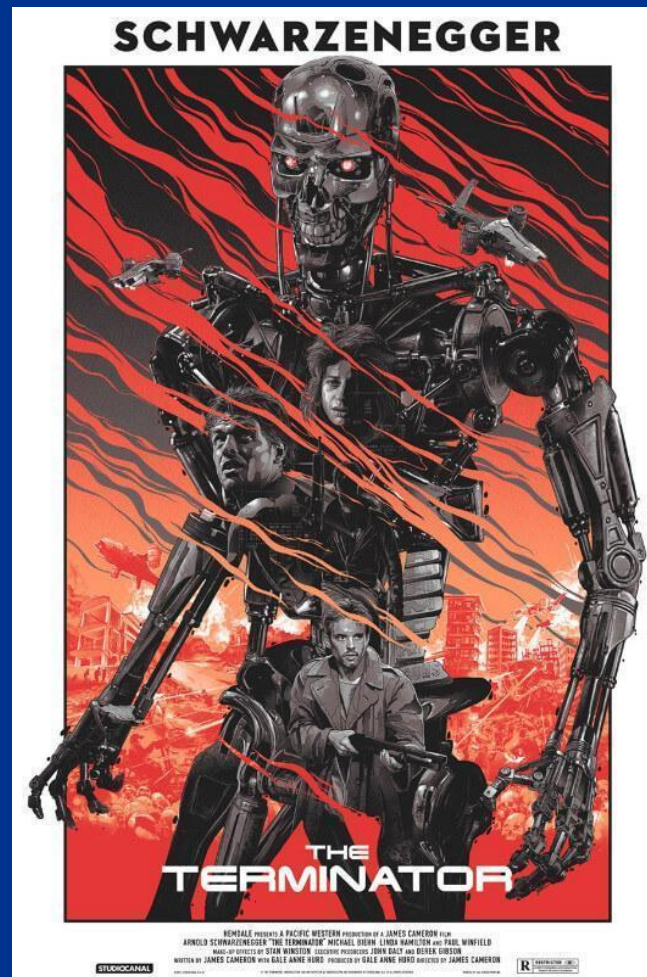
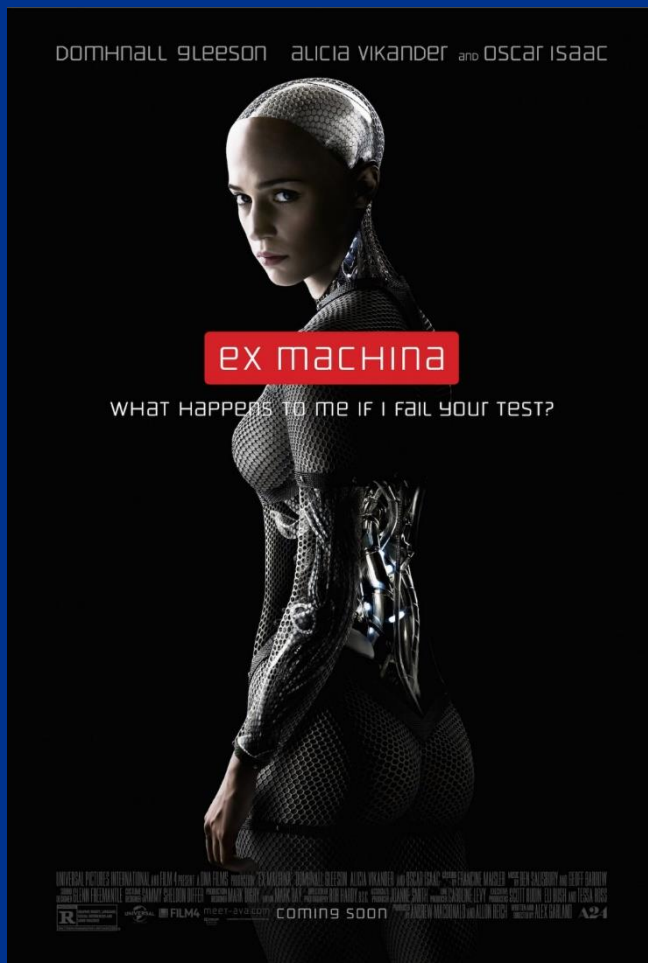
Even the "news" has become entertainment



The commercialization of useless information



Death by robot is fun!

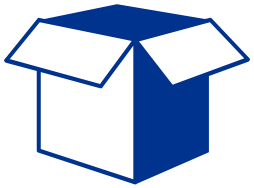


Let's talk AI - common characteristics of AI Systems

Q: Where is AI being applied within your organization?



Making predictions based on large quantities of (historical) **data**



Contains a (self)learning aspect, with limited transparency between input and output relationships (“**black box**”)



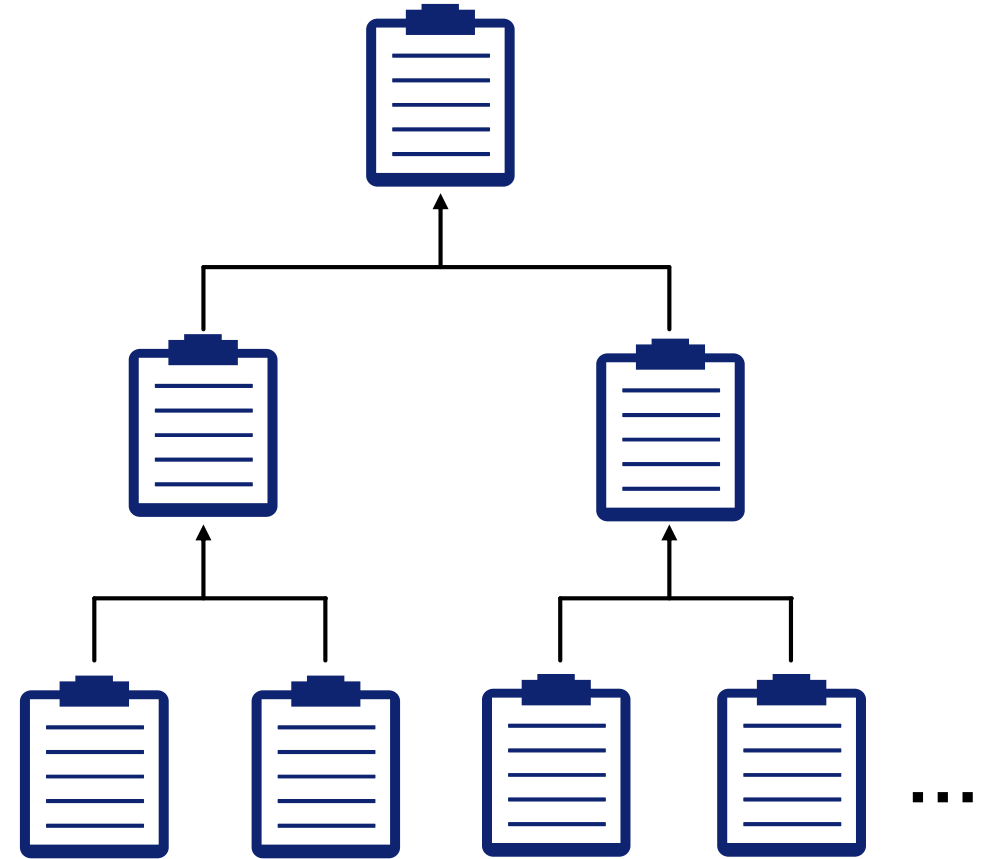
Often applied in combination with some degree of **automated decision-making**

Traditional IT vs AI



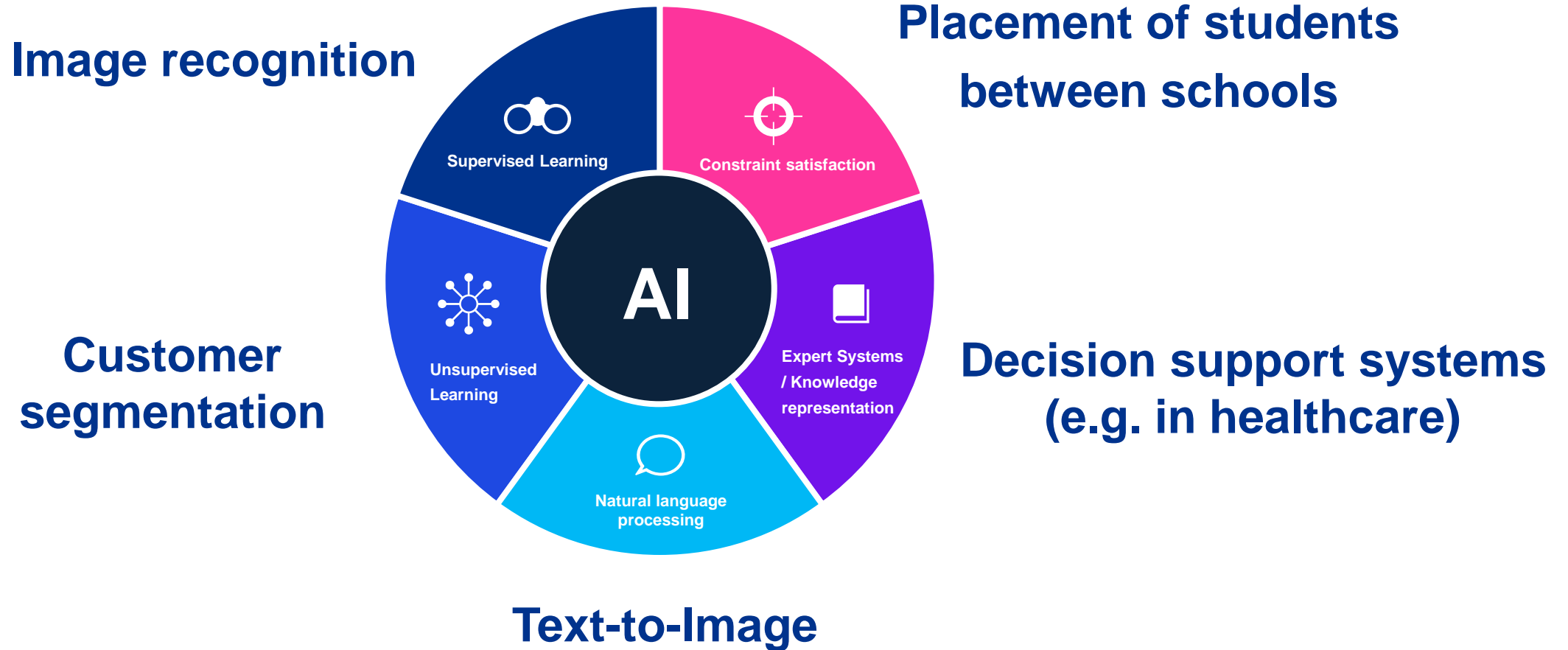
IT (explicit)

VS



AI (implicit)

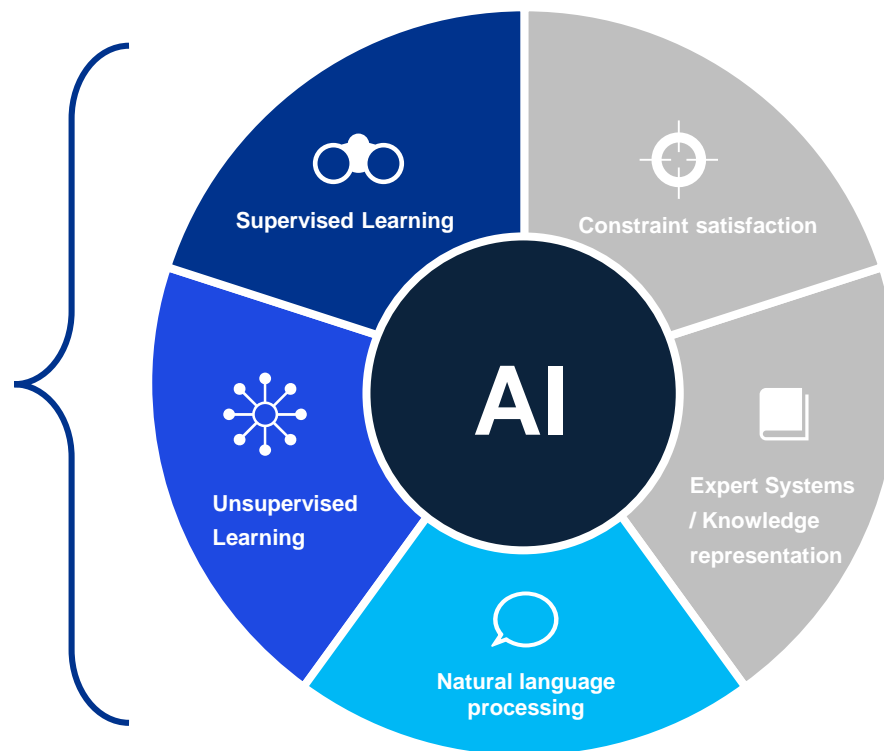
What AI is currently being used for



We now live in the age of Generative AI

Generative AI

- ChatGPT
- Dall-E
- Midjourney
- Stable Diffusion
- ...



Organizations are now faced with the novel challenges of AI

The development and deployment of AI poses unique challenges. Organizations should therefore take measures to ensure that AI is used in a responsible manner. The interests of citizens/customers, the organization and society intersect here.



Properly specifying your intended goal in alignment with your values is significantly more challenging than with 'traditional' IT.



The performance of AI is impacted when structural changes take place in the data or in the environment (drift).



The opaque nature (black box) of some AI systems can create both distrust as well as blind trust.



Defining who can be held responsible or accountable in automated decision-making is no trivial feat.



The use of AI can help to solve existing ethical issues, but it can also create new ethical challenges.

RELIABILITY

AI should do what it is intended to do

RESILIENCE

AI should not only function now, but also in the future

EXPLAINABILITY

AI must provide results that can be comprehensible and verified

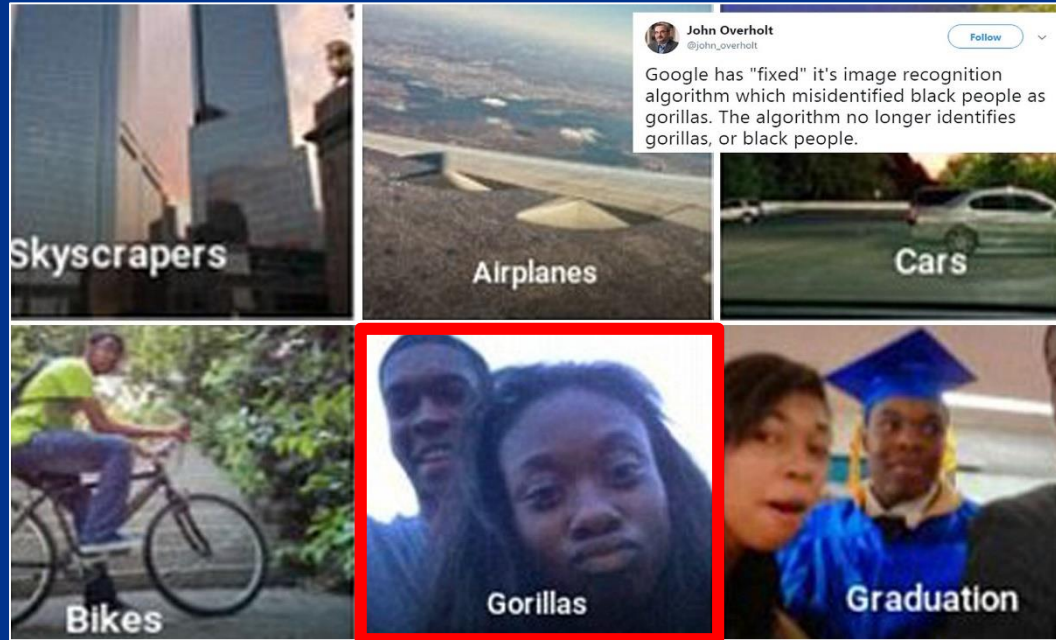
ACCOUNTABILITY

AI must have an owner which can be addressed

FAIRNESS

AI must be fair, non-discriminatory and inclusive

Racist technology – a new problem?



GOOGLE \ TECH \ ARTIFICIAL INTELLIGENCE \

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

Nearly three years after the company was called out, it hasn't gone beyond a quick workaround

By James Vincent | Jan 12, 2018, 10:35am EST

2015

Racist technology, exists longer than you think!



GOOGLE | TECH | ARTIFICIAL INTELLIGENCE

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

Nearly three years after the company was called out, it hasn't gone beyond a quick workaround

By James Vincent | Jan 12, 2018, 10:35am EST

2015

1940

The New York Times

The Racial Bias Built Into Photography

Kodak

Problems with chatbots are not new either!

TECH · ASHLEY MADISON

Ashley Madison Used Chatbots to Lure Cheaters, Then Threatened to Expose Them When They Complained

BY DAVID Z. MORRIS

July 10, 2016 at 7:11 PM GMT+2

MICROSOFT / WEB / TL:DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day



By JAMES VINCENT
Via THE GUARDIAN | Source TAYANDYOU (TWITTER)
Mar 24, 2016, 11:43 AM GMT-1 | [0 Comments](#) / [0 New](#)



ARTIFICIAL INTELLIGENCE

A college kid's fake, AI-generated blog fooled tens of thousands. This is how he made it.

"It was super easy actually," he says, "which was the scary part."

By Karen Hao

August 14, 2020

Man created a fake restaurant and it became the #1 restaurant in London

KPMG Responsible AI – our mission

We help make sure client AI Systems are **fit for purpose, ethical and compliant with rules & regulations.**

From high level principles..



..to verifiable claims and testing procedures.

- How do we define reliability and which level is sufficient?
- How deep/far should explainability go?
- To whom should the explanation be understandable and how can you measure this?
- What is the right criterion for fairness?
- Which security standards apply to an AI System?
- What are the (legal) requirements regarding accountability?
- ...

Explainability does **not** imply trust

- “Streetlight Effect”
- Explanations require domain expertise, which the reader often doesn’t have
- Explanations risk exposing IP and company secrets
- Explanations enable actors with malicious intent to “game the system” and execute adversarial attacks



Trust requires **Reliability** – ask yourself, which plane would you rather step into?



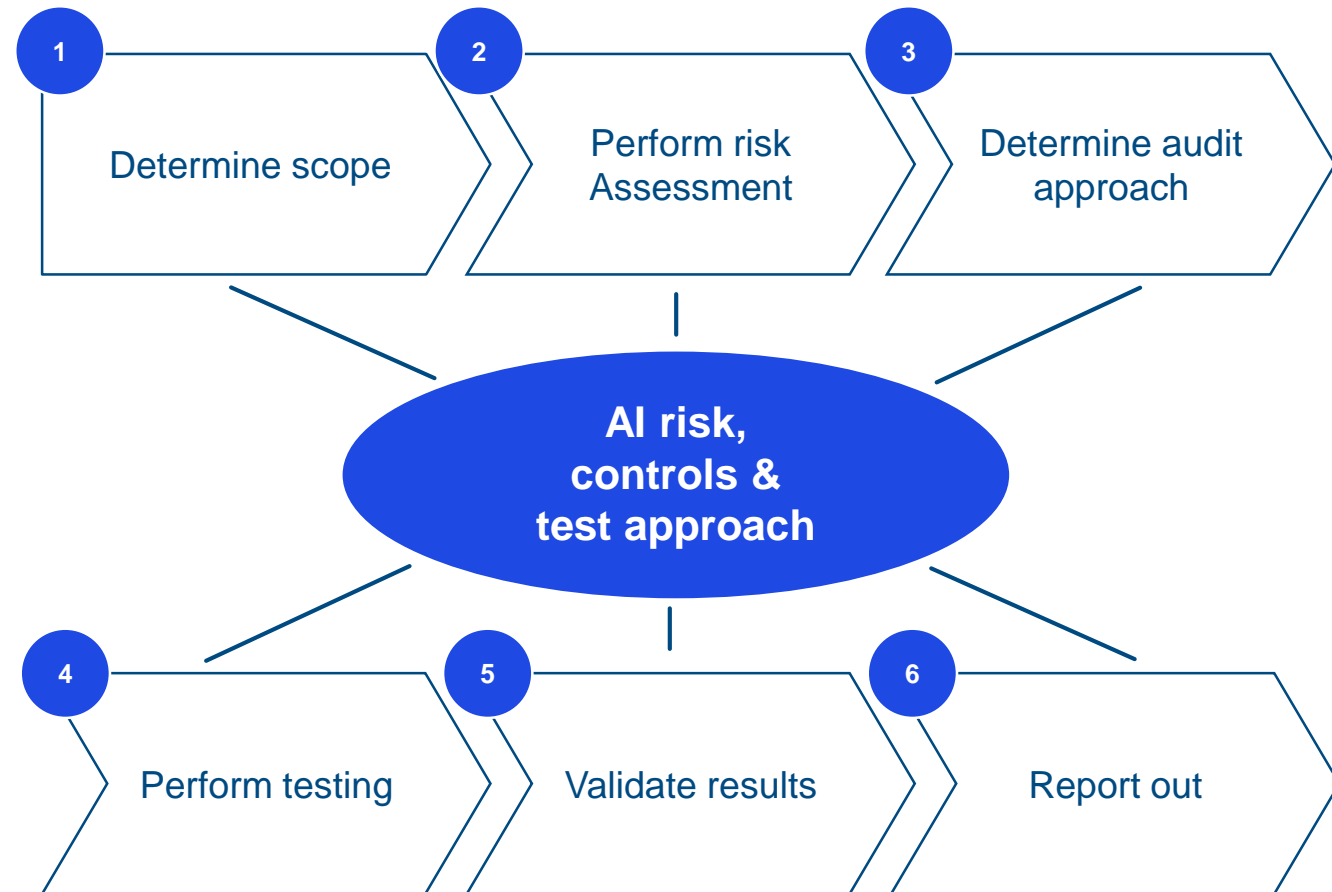
VS



Auditing AI – a Risk-Based Approach

Our generic algorithm audit approach, similar to the way we audit other objects.

- 1. Determine scope.** We first determine the type and scope of the AI system to be audited. AI systems never operate in isolation, we gain understanding of the business context and the business process it is a part of. If relevant we also consider its role with respect to the financial statement audit (FSA).
- 2. Perform risk assessment.** We use a risk-based approach to obtain an understanding of the technical details of the AI system and the business context in which it operates.
- 3. Determine audit approach.** Based on the risk profile of the AI system and the relevant norms to which the AI system is to be evaluated we determine the proper combination of audit procedures and depth of testing. We draft a detailed control program by tailoring existing, potentially generic, norms and control frameworks to the specific characteristics and context of the AI system.
- 4. Perform testing.** We perform testing based on our audit program.
- 5. Validate results.** We discuss our initial findings with the stakeholders of the AI system from the client's side and address any remaining issues.
- 6. Report out.** We report the final results.



We quantify the risks of AI along three different axis



Complexity

Ranging from traditional ruled-based systems to Neural Networks, is the application broadly characterizable as Artificial Intelligence?



Autonomy

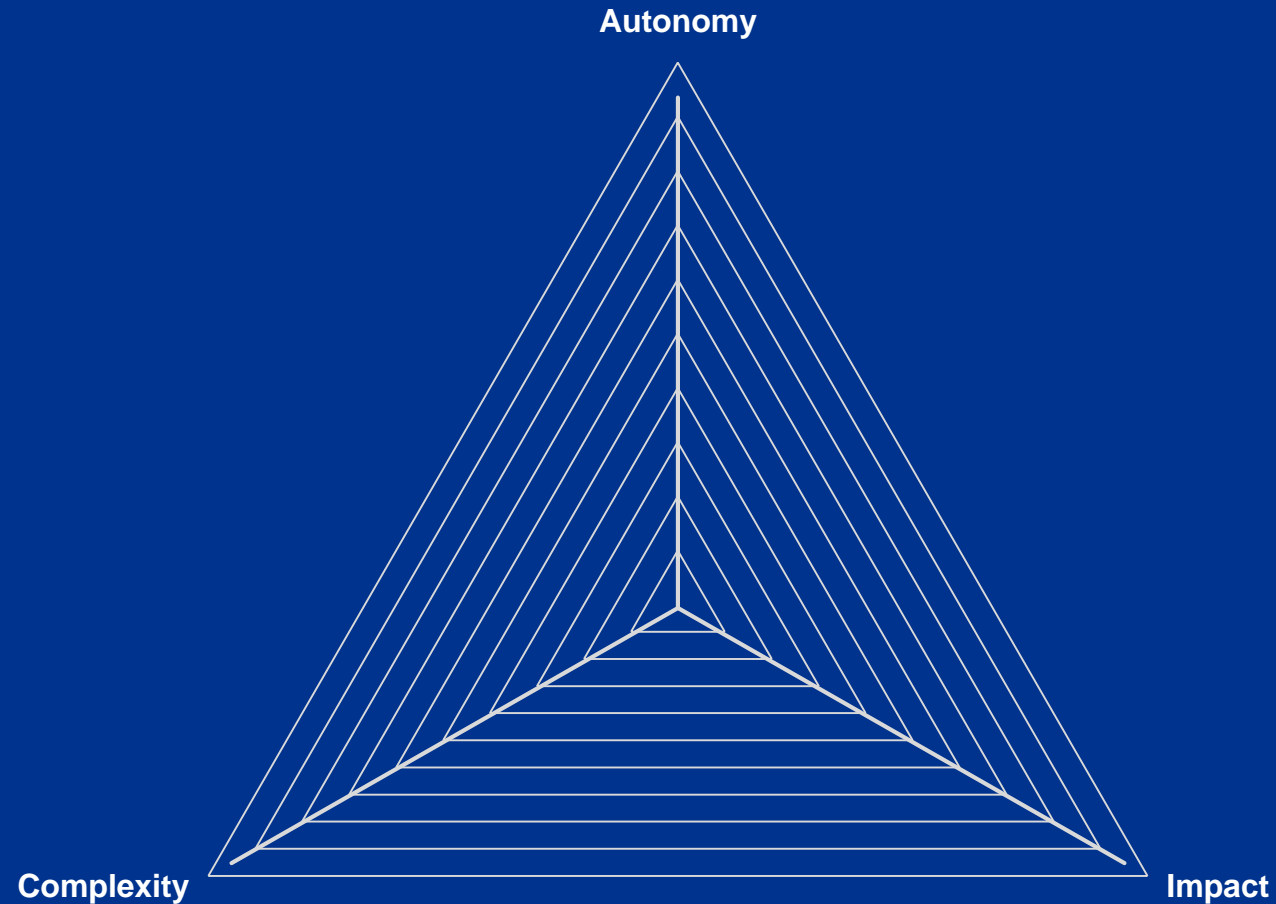
To which degree is the decision-making process based on automated processing of data with no, or only pro forma, control by a human decision-maker?



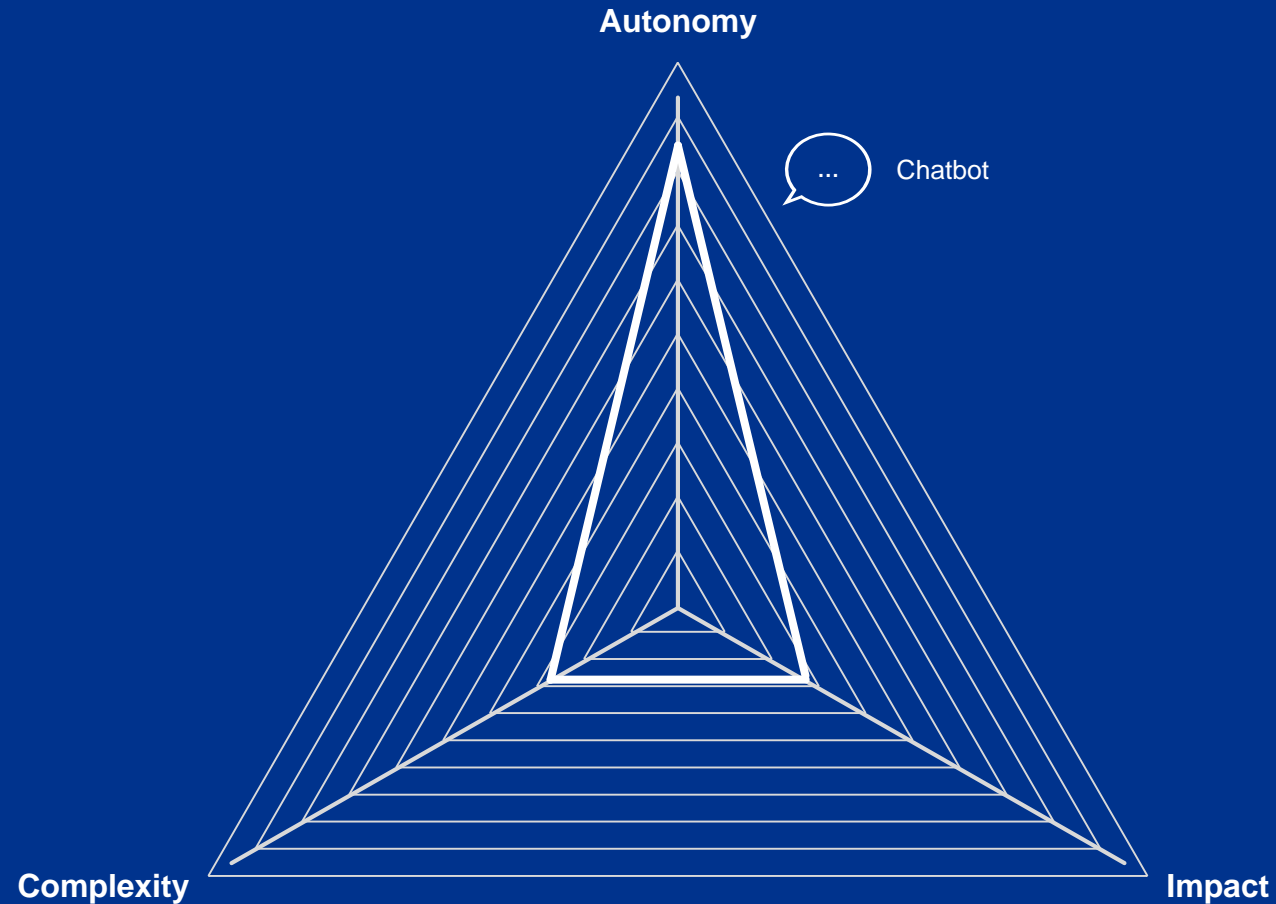
Impact

Does the interaction with the application affect rights, duties, powers, and/or liabilities of people, groups of people, or organizations?

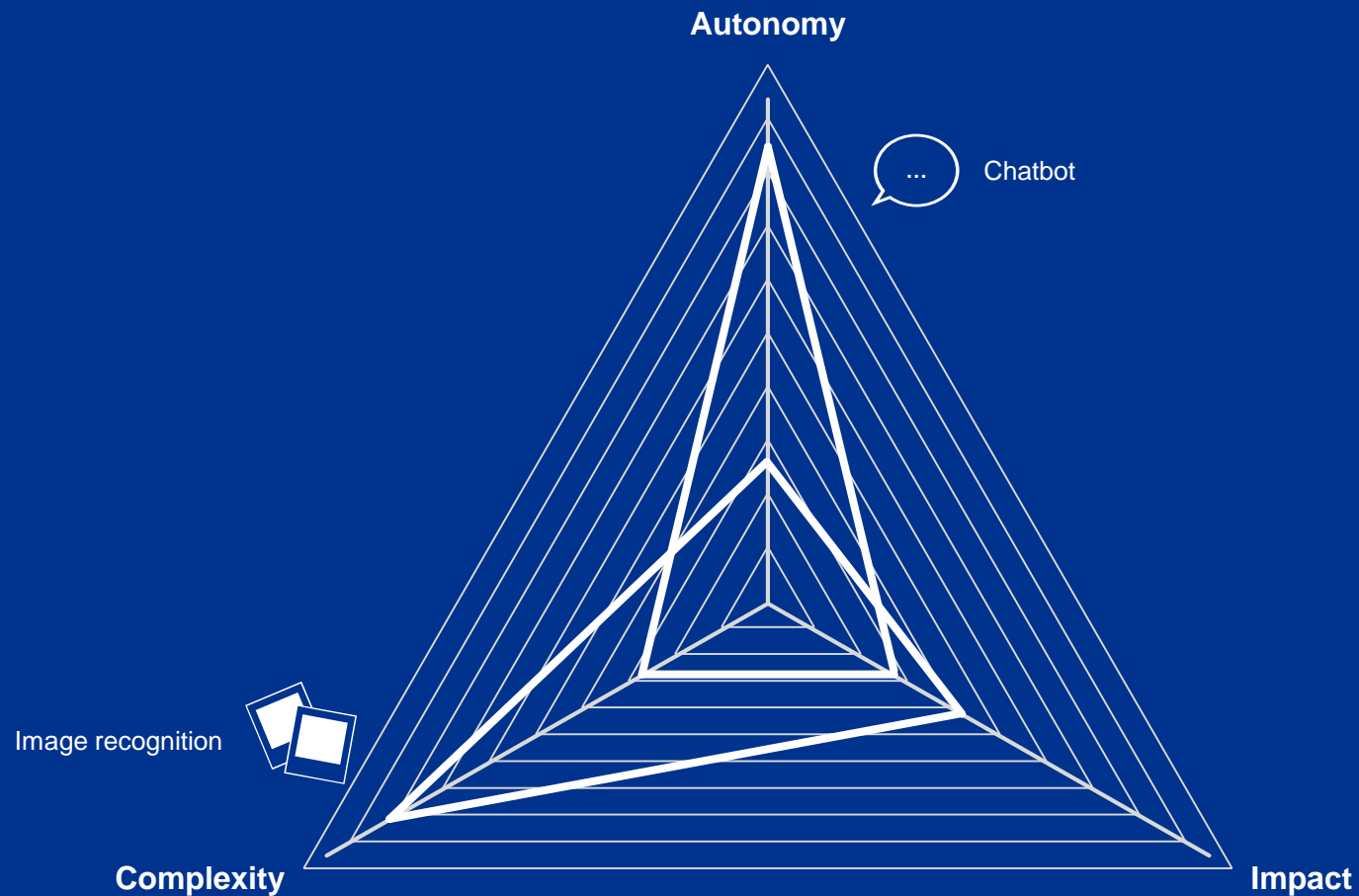
AI Systems have vastly different risk profiles



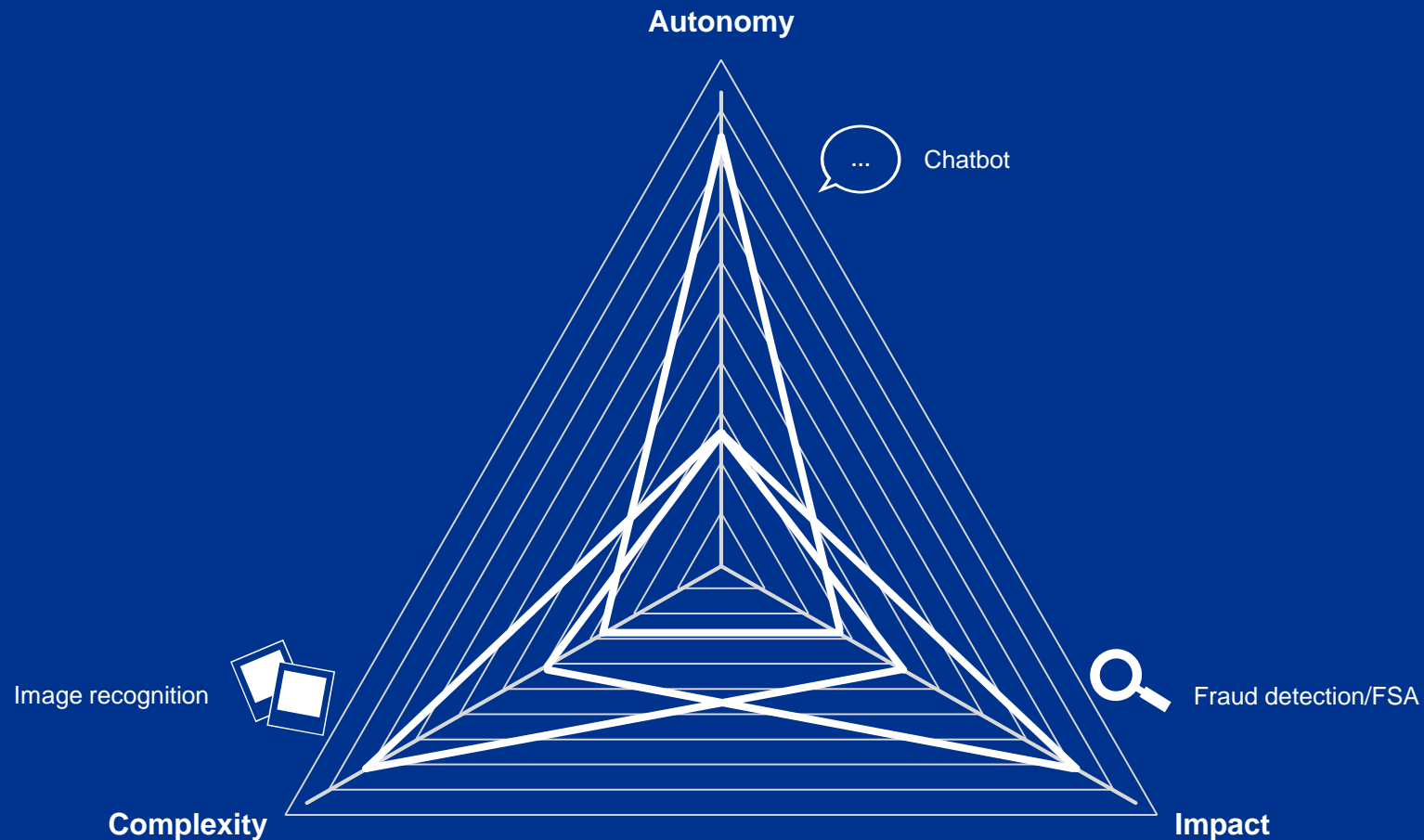
AI Systems have vastly different risk profiles



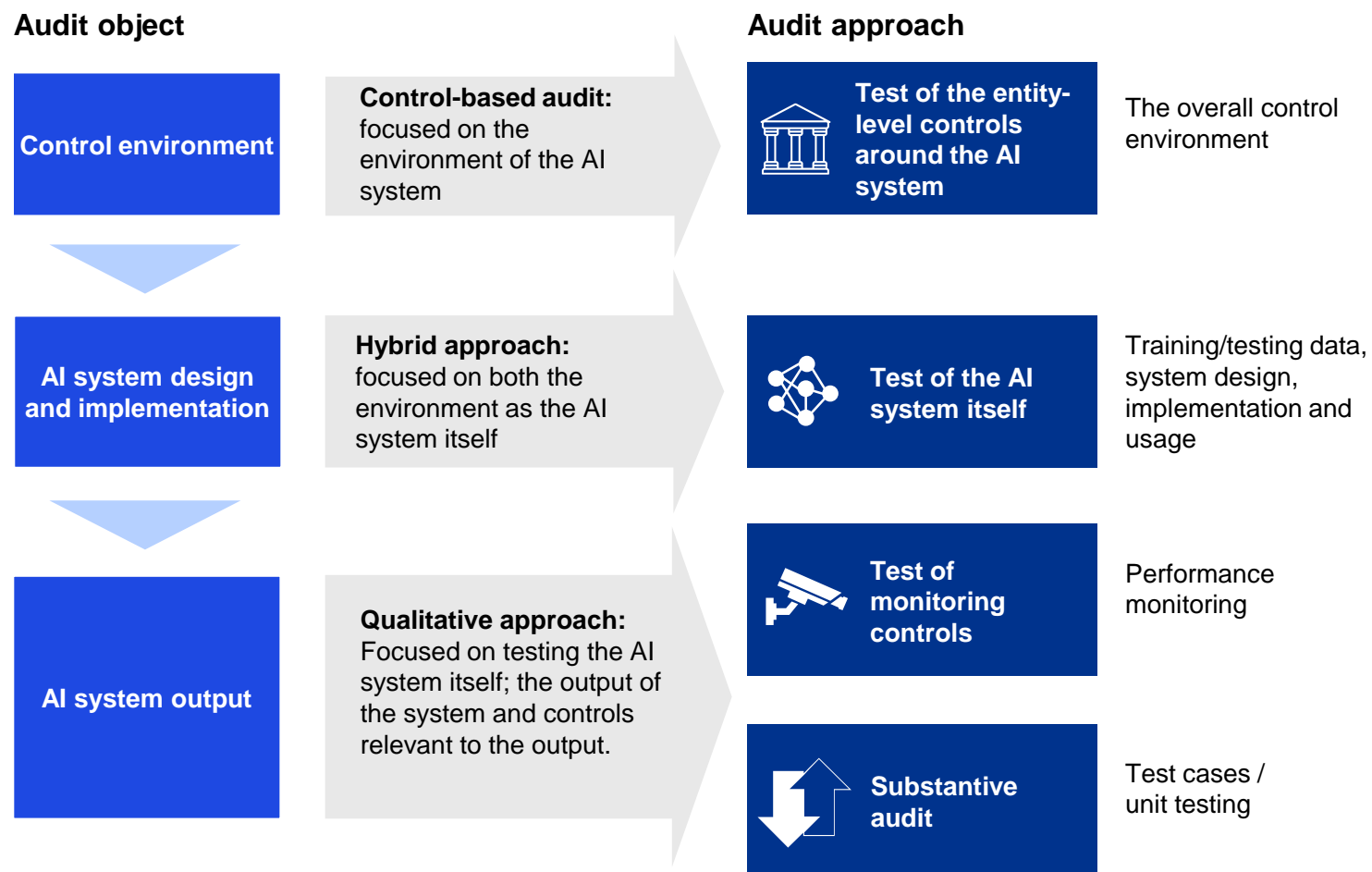
AI Systems have vastly different risk profiles



AI Systems have vastly different risk profiles

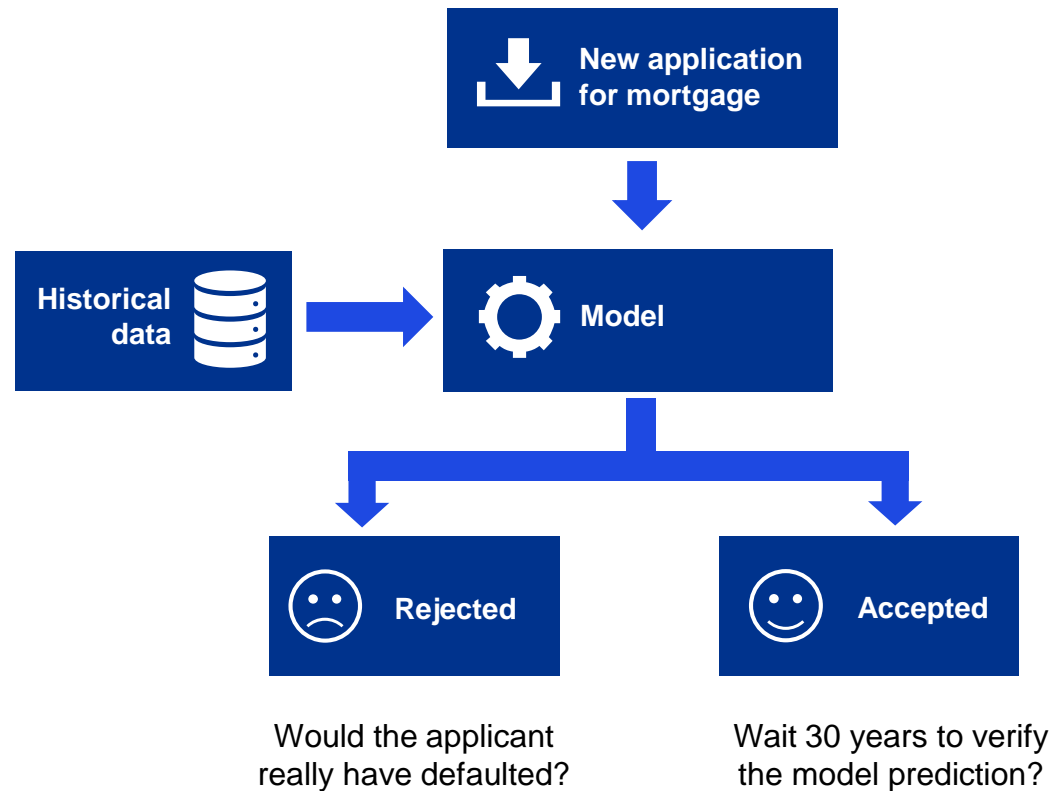


We combine audit procedures into a tailored approach based on relevant risks, norms and feasibility

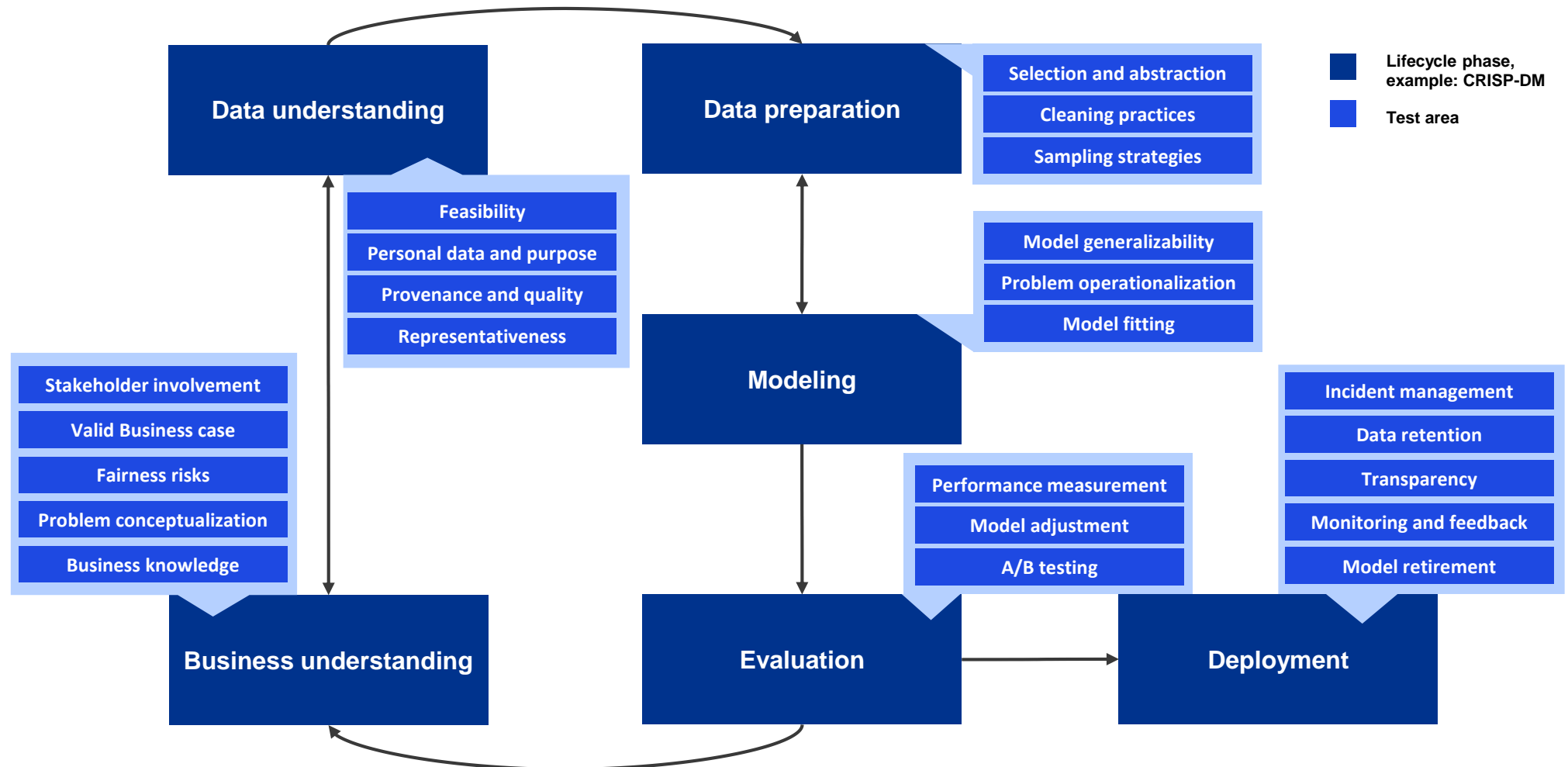


Substantive procedures: the problem of only validating output

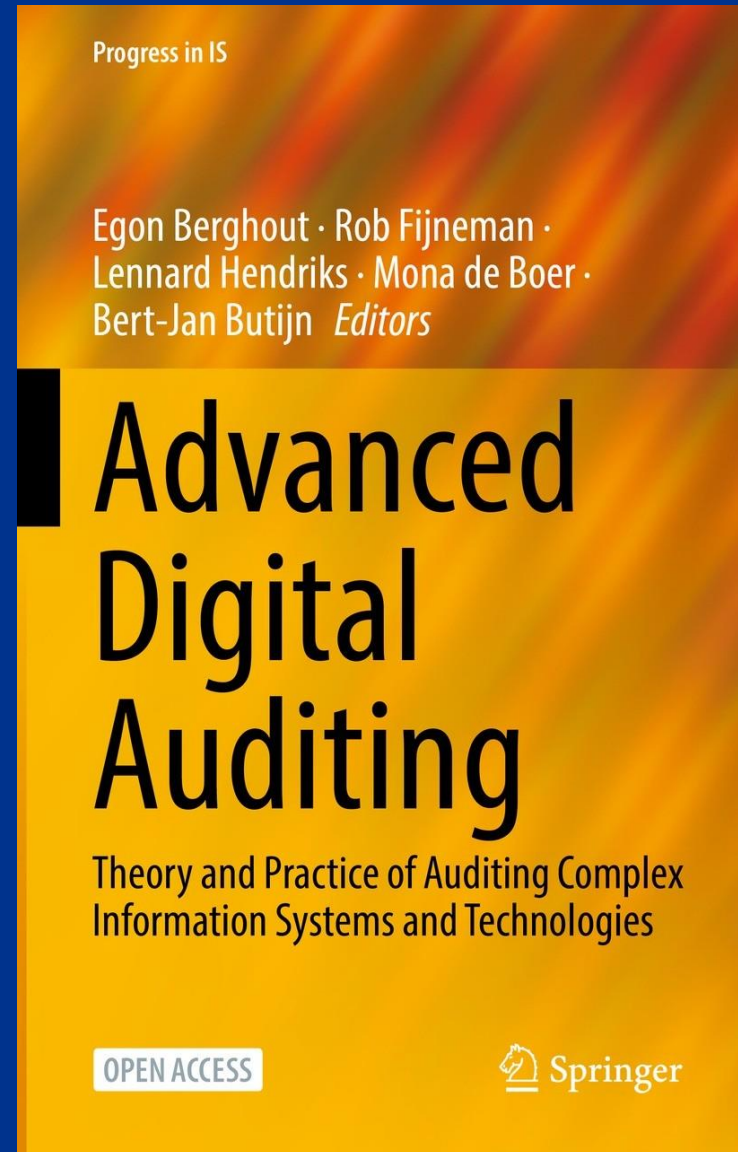
Example: how to properly assess whether your model has accurately predicted loan default probability?



Based on the chosen audit procedures, we test various aspects of the AI system along its cyclic life



Want to know more?
Must-read!



<https://link.springer.com/book/10.1007/978-3-031-11089-4>

Every organization will become an AI organization, or go the way of the dinosaur

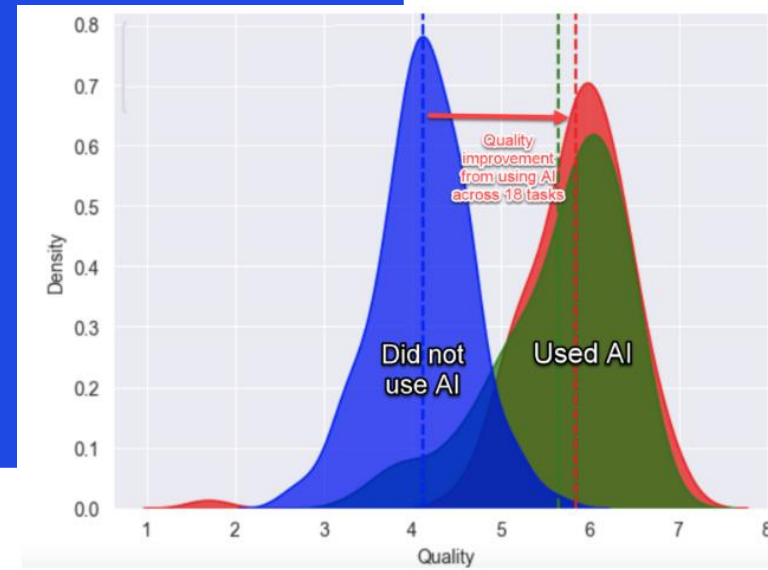


Enterprise workers gain 40 percent performance boost from GPT-4, Harvard study finds

VentureBeat

Microsoft's Plan To Infuse AI And ChatGPT Into Everything

Forbes



Difference in performance among BCG consultants, comparing those who used AI versus those who didn't. (Image Credit: Navigating the Jagged Technological Frontier)

AI significantly increases productivity

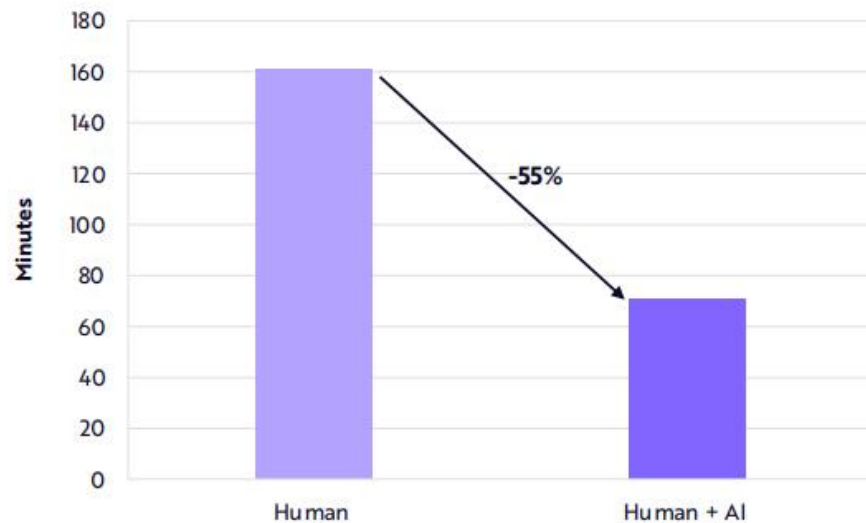


AI Is Increasing The Productivity Of Knowledge Workers

Coding Assistants

Software engineers completed a coding task in less than half the time with AI coding assistant [GitHub Copilot](#).

Time to Complete Coding Tasks: 2022*



Generative Image Models

According to our research, AI can create a graphic design for just \$0.08** in minutes – a *di minimis* cost compared to \$150 for human labor.



Human

Cost \$150

Time 5 Hours



Generative AI

Cost \$0.08

Time < 1 Minute



AI significantly increases productivity



AI Could Lead To A 10-Fold Increase In Coding Productivity

Based on a 70% annualized drop in training costs and feedback loops, AI coding assistants like Copilot could increase the output of software engineers **-10-fold** by 2030.

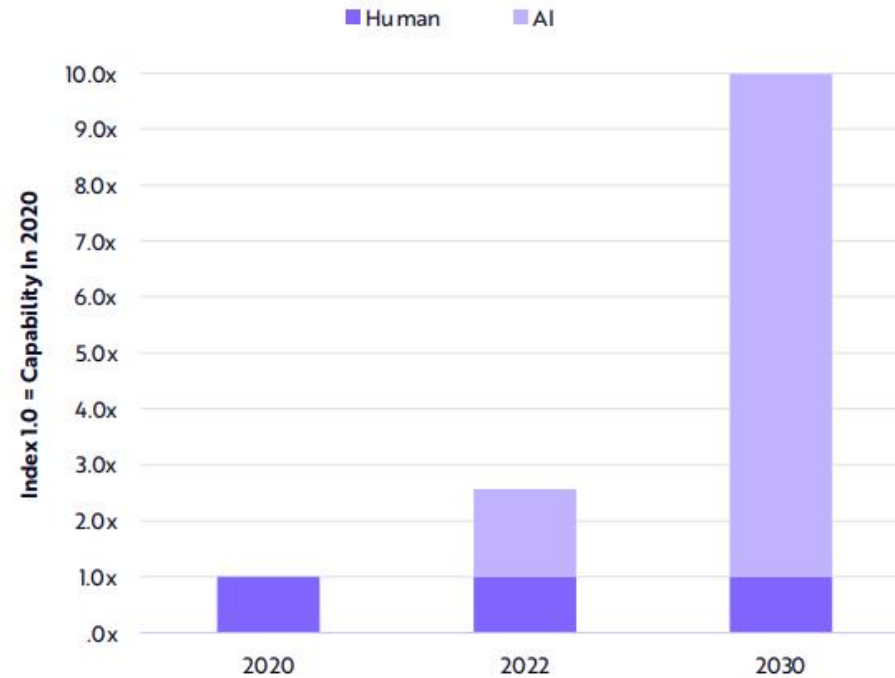
Github Copilot Example

Human Input

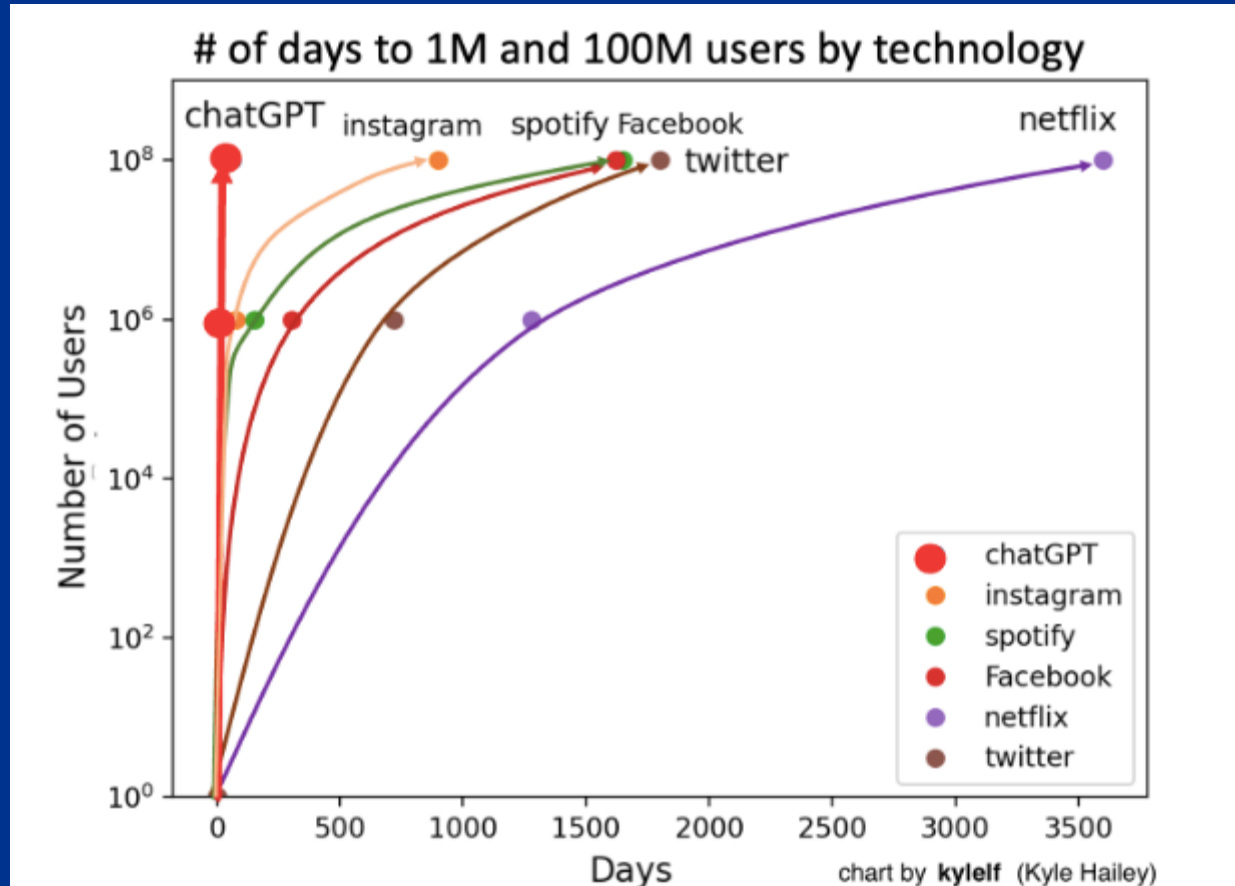
```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch('http://text-processing.com/api/sentiment/', {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17 }
```

Copilot

Output of Human + AI: Coding Tasks



The success of ChatGPT is its accessibility



*“Trust is good,
control is better.”
- Lenin*

Only 8% of CIO's has a vision about AI..

COMPUTABLE
Alles over zakelijke ICT in België

Zoeken

Home | Artikelen | Nieuws | Management

Gartner lanceert ai-wake-up call voor cio's

6 november 2023 14:13 | [Luc Blyaert](#)

Topic Management

[f](#) [t](#) [in](#) [✉](#)



Gartner vindt dat cio's de komende twaalf tot 24 maanden werk moeten maken van een duidelijke strategie voor artificiële intelligentie. 'Generatieve ai is geen nieuwe technologie en ook geen businesstrend. Het is een fundamentele shift in hoe mensen en machines met elkaar omgaan', vertelde vice president analyst Mary Mesaglio tijdens de openingstoespraak van het Gartner Symposium in Barcelona. Daar zijn bijna 6.500 cio's en it-leiders aanwezig.

Zij was verrast dat iets meer dan de helft van de ondervraagde ceo's aangaf dat ze de **cio** zien als de leidende figuur bij de ontwikkeling van artificiële intelligentie. 'Dit is een ongelooflijke kans voor de cio om het verschil te kunnen maken.' Ze vreest wel dat allerlei afdelingen zoals marketing, sales of de ceo die rol zal willen opnemen. In juni ondervroeg [Gartner](#) ruim 600 cio's en amper 8 procent liet weten dat de organisatie een visie had op het gebruik van ai. Een op de drie had helemaal geen plannen in die zin.

Gartner maakt een duidelijk verschil tussen 'everyday ai' en 'game changing ai'. In het eerste geval gaat het om efficiëntie en productie. Daar zou ai voor verbetering van 5 tot 20 procent kunnen zorgen, 'wat toch nog vrij beperkt is', aldus Mary Mesaglio. In het tweede geval gaat het vooral om creativiteit en zal het voor een omwenteling van businessmodellen en zelfs hele industrieën zorgen. Dat ai geen goedkope oplossing zal worden is nu al duidelijk, hoewel Gartner hier geen bedragen wil of kan vrijgeven. Het schuift cio's een lijstje met vier to-do's voor: wees klaar voor ai, zorg voor ai-ready principes, maak je data ai-klaar en vooral, zorg ervoor dat je security ai-ready is.

<https://www.computable.be/artikel/nieuws/management/7572603/5440850/gartner-lanceert-ai-wake-up-call-voor-cios.html>

KPMG RAAD – Programma De Glazen Leider

KPMG vorig jaar een programma gestart onder de noemer ‘De glazen leider’. KPMG heeft dit initiatief genomen omdat wij een groeiende angst constateerden onder leiders om beslissingen te nemen waar risico's aan verbonden zijn (hypegiaphobia). Deze trend wordt onder andere gedreven door **moderne technologieën**, waardoor beeldvorming een steeds prominentere, zo niet belangrijkere plek inneemt naast de inhoud.

Dit bestond uit o.a.

- ❑ 9 interviews
- ❑ 6 rondetafelgesprekken met 30 (ervarings)deskundigen
- ❑ 1 whitepaper *De Glazen Leider*
- ❑ 2 conferenties met 80 bestuurders
- ❑ 5 opinieartikelen

KPMG RAAD

KPMG RAAD inspireert en helpt commissarissen, bestuurders en NextGen bij het nog beter invullen van hun rol. Het bestaat uit een magazine en een online platform waar de nieuwste relevante inzichten en ervaringsverhalen van collega's elkaar afwisselen. Daarnaast organiseert KPMG met het RAAD-programma diverse evenementen waar urgente thema's voor in de boardroom worden uitgelicht.



De schadelijke gevolgen



- Hypegiaphobia
- Afhaken talentvolle leiders
- Overregulering
- Doorgeschoten compliance kosten
- **Rem op innovatie**
- Verminderde klantenservice
- Afstoten producten, diensten, markten

**Can we Trust
ChatGPT?**

Large Language Models (LLM) – autocomplete on steroids

Goal: produce a reasonable continuation of the text we have so far

Idea: simply calculate the probability of the next word in a text and repeat 

I am walking
my _____

Large Language Models (LLM) – autocomplete on steroids

Goal: produce a reasonable continuation of the text we have so far

Idea: simply calculate the probability of the next word in a text and repeat 

I am walking
my _____

DOG	90%
CAT	5%
PARROT	1%
...	...

.. but that would have been too easy!

- **Problem 1:** there doesn't exist enough text in the world to deduce all probabilities..
- **Problem 2:** we can't even compute the probabilities of all text that's currently out there..

Quick maths:

The English language contains around **40.000** common words.

Number of possible **2-grams** is 1.6 billion

Number of possible **3-grams** is 60 trillion

..

Essay of 20 words > #particles in the universe.

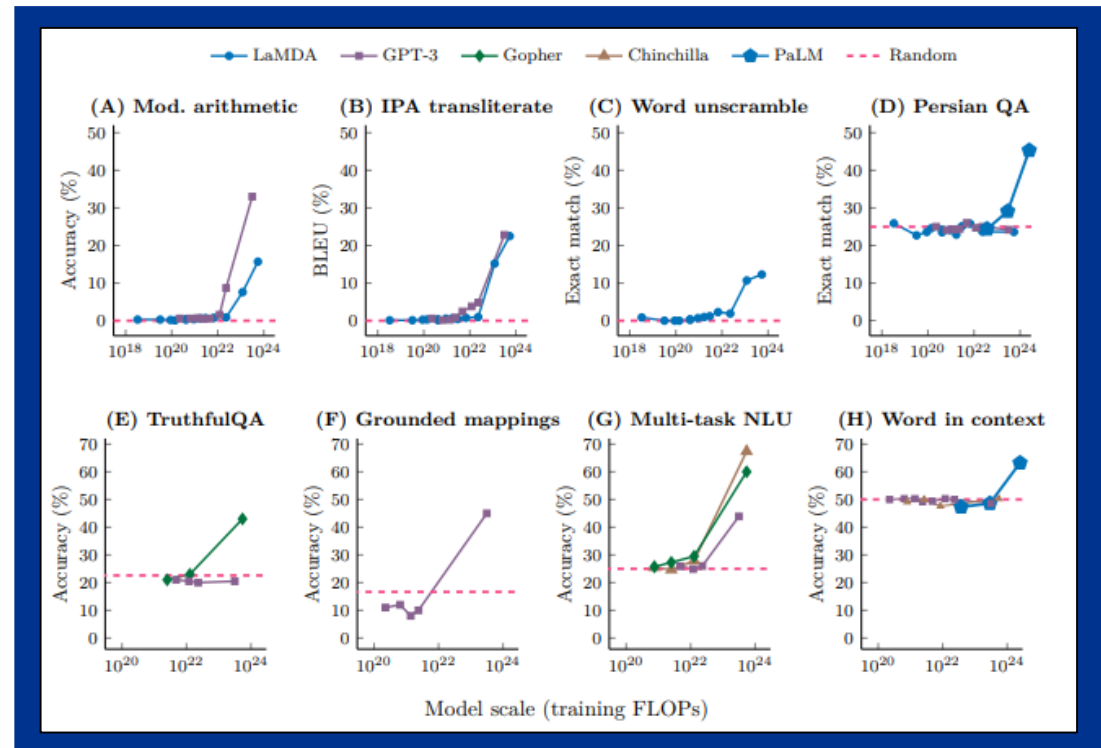
Solution: use a **model** to predict these probabilities, even if we never saw the sequences of words themselves.

A glimpse of intelligence – emerging properties of ChatGPT

Emerging properties: functionality present in larger models and not in smaller models, which improves with an increase in model size, and which cannot be predicted from smaller models.

- Excellent at text summarization, classification and sentiment analysis
- Generates “human sounding” text
- Can produce programming code
- Displays reasoning capabilities
- Can produce multi-lingual text
- One/few-shot learning
- Can assume a ‘role’
- ...
- Rise of prompting: programming LLM’s

Language is deeper than we think, it captures the **essence** of reasoning (to some degree)



 [\[2206.07682\]](#) Emergent Abilities of Large Language Models

Your hallucinating buddy



- ChatGPT is trying to make sense of its inputs, similar to **dreaming**
- As a result, ChatGPT is able to confidently **bullsh*t** about things that are **wrong** and/or don't exist!
- ChatGPT lacks any actual **knowledge** of what is represented in text (objects/concepts)
- Not every aspect of reality can be captured in human **language** itself! (context needed)
- Some text requires more context than others (Mandarin vs English)

People are starting to realize ChatGPT has the confidence and verbal ability of a **40-year** old, but the world representation of a **4-year** old.

 [\[2302.03494\] A Categorical Archive of ChatGPT Failures](#)

“Asking whether a computer can think is as interesting as asking if a submarine can swim.”

- Edsger W. Dijkstra

**Responsible AI is no
Longer a Luxury, but a
Precondition for Success
in the Digital Era**

Generative AI unlocks opportunities, while bringing forth new challenges

Brings Challenges



How to audit AI?



Reputational damage



Doorgeschoten (non)
compliance kosten



Rem op innovatie



Impacts ESG Score

Unlocks Opportunities



Increased efficiency



Increased quality



Increased fairness



Enhanced risk assessment &
Predictive analyses



Continuous monitoring &
auditing

Blijf in control en laat uw organisatie niet op deze manier in het nieuws komen

NOS op3

Vrijdag 23 juni 2023, 15:32

DUO mag algoritme niet gebruiken totdat meer bekend is over mogelijke discriminatie

Berichtgeving Follow The Money zorgt ervoor dat politie stopt met algoritme

NOS Nieuws • Maandag 17 juli, 17:57

Privacywaakhond: organisaties hebben algoritmes niet goed onder controle

NOS Nieuws • Zaterdag 15 juli, 14:00 • Aangepast zaterdag 15 juli, 14:05


UWV verzamelde illegaal gegevens van uitkeringsgerechtigden

Vragen om uzelf en de organisatie te stellen

1. Heeft de organisatie voldoende overzicht waar AI wordt ingezet binnen de organisatie? Is deze AI besluit ondersteunend of autonoom?
2. Welke (ethische) risico's en bijbehorende beheersingsuitdagingen zijn er met de inzet van AI binnen de organisatie?
3. Welk overkoepelend beleid heeft de organisatie geformuleerd ten aanzien van de (beheerste) inzet van AI, en hoe verhoudt zich dat met de business strategie?
4. Hoe heeft de organisatie de risico's van AI geborgd en opgenomen in risk management? (gedragscode, internal audit, governance, risk-assessment, training, awareness, rollen en verantwoordelijkheden, etc.)
5. Is er een protocol inzake gebruik van ChatGPT en andere openbaar toegankelijke technologieën voor onze medewerkers (- wenselijk/noodzakelijk)?
6. Welke ethiek heeft de organisatie geformuleerd ten aanzien van de ontwikkeling en inzet van AI? (leidende principes, normen, ethische commissie, etc.)
7. Waar is dit onderwerp organisatorisch belegd? Welk orgaan/functie gaat zich AI toerekenen?
8. Worden de costs of control van AI meegewogen in innovatietrajecten?
9. Is er duidelijk begrip van de ROI van AI en innovatie?
10. Hoe blijf ikzelf geïnformeerd? -> themadagen, advies, of stuur mij (Marc) een berichtje! :)

Thank you, let's grab a coffee!



✉ vanmeel.marc@kpmg.nl
🌐 www.marcvanmeel.com
 [linkedin.com/in/marc-van-meel/](https://www.linkedin.com/in/marc-van-meel/)



Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.



kpmg.com/socialmedia



ir. Marc van Meel
Manager

Responsible AI

vanmeel.marc@kpmg.nl

© 2023 KPMG N.V., a Dutch limited liability company and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

Document Classification: KPMG Public