

AI and Data Protection – Challenges for Controllers, Processors and Supervisory Authorities

Dr. h.c. Marit Hansen

State Data Protection Commissioner
of Schleswig-Holstein, Germany

IFIP Summer School on Privacy and Identity Management
Oslo, 11 August 2023



Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

ULD: Supervisory authority

- Data Protection Authority (DPA) for both the public and private sector, enforcing the General Data Protection Regulation (GDPR)
- Also responsible for freedom of information
- For individuals:
 - Right to lodge a complaint [Art. 77 GDPR]
 - DPA service free of cost
 - Right to effective judicial remedy against a supervisory authority [Art. 78 GDPR]
 - Representation [Art. 80 GDPR]

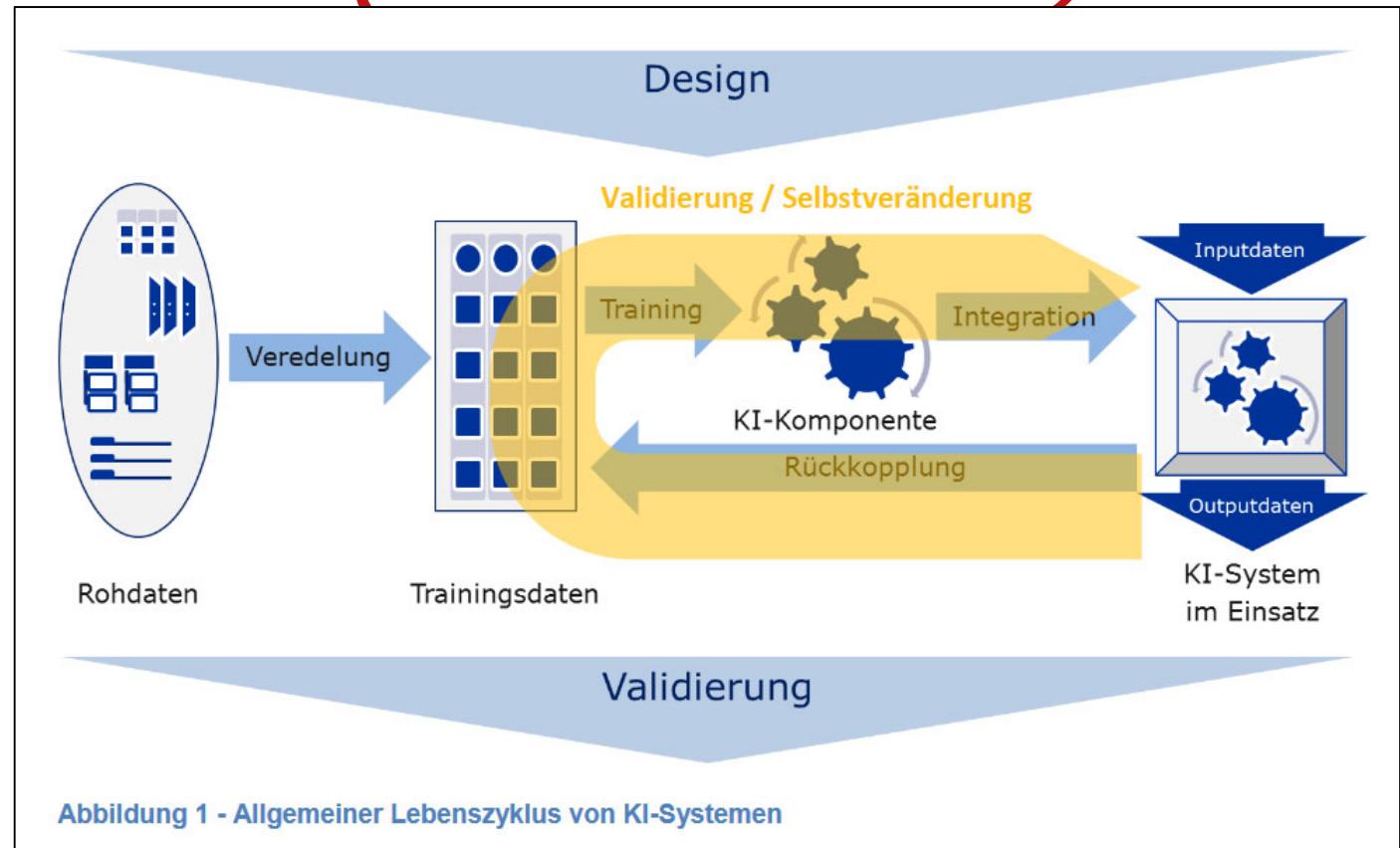
Source: en.wikipedia.org/wiki/Schleswig-Holstein

Schleswig-Holstein	
State of Germany	
	Flag
	Coat of arms
	
Coordinates: 54°28'12"N 9°30'50"E	
Country	Germany
Capital	Kiel
Government	
• Body	Landtag of Schleswig-Holstein
• Minister-President	Daniel Günther (CDU)
• Governing parties	CDU / Greens / FDP
• Bundesrat votes	4 (of 69)
Area	
• Total	15,763.18 km ² (6,086.20 sq mi)
Population (2016-12-31) ^[1]	
• Total	2,881,926
• Density	180/km ² (470/sq mi)
Time zone	CET (UTC+1)
• Summer (DST)	CEST (UTC+2)

AI and Data Protection – DSK 2019 (DSK = German DPAs)

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

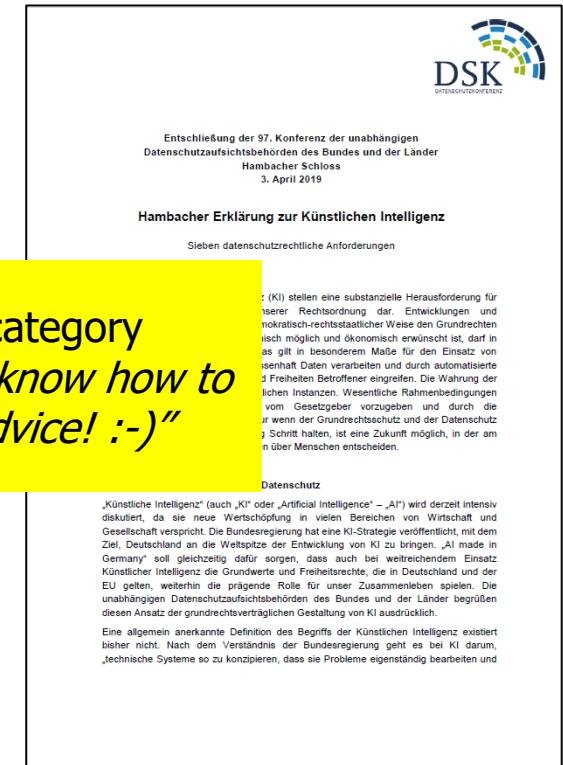
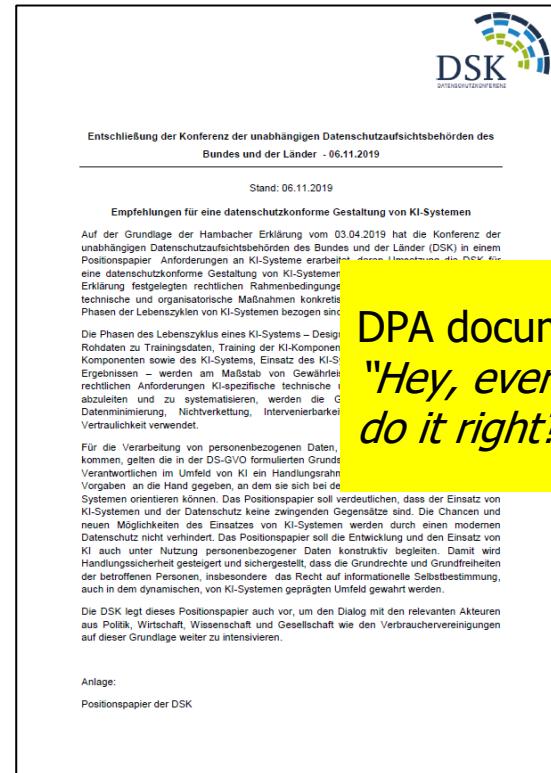


https://www.datenschutzkonferenz-online.de/media/en/20191106_positionspapier_kuenstliche_intelligenz.pdf

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

AI and Data Protection – DSK 2019



**DPA documents from the category
"Hey, everybody, want to know how to
do it right? Here's some advice! :-)"**

https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf
https://www.datenschutzkonferenz-online.de/media/en/20191106_entschließung_KI_DSK.pdf

AI Model – personal data?

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

The screenshot shows a journal article from **PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A**, Mathematical, Physical and Engineering Sciences. The article is titled **Algorithms that remember: model inversion attacks and data protection law**. It was published on 15 October 2018 and can be found at <https://doi.org/10.1098/rsta.2018.0083>. The abstract discusses the governance of machine learning systems and the prevention of algorithmic harms, specifically focusing on 'model inversion' and 'membership inference' attacks. The journal page also features sections for 'This Issue' and 'Volume 376, Issue 2133'.

<https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2018.0083>

AI and Data Protection – Challenges

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Design requirements

GDPR – Principles

Art. 5 GDPR – Principles relating to processing of personal data

(1)

- a) Lawfulness, fairness and transparency
- b) Purpose limitation
- c) Data minimisation
- d) Accuracy
- e) Storage limitation
- f) Integrity and confidentiality (~ security)

(2) Accountability

Challenges in all phases:
Training, optimizing, calibrating,
filtering, usage, correction, ...

For all kind of users, e.g. journalists,
business users, researchers, patients,
children

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

GDPR: Is it always Art. 6 (1) f?

Article 6

Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

[...]
- (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

GDPR: Right to object

Caveat:

Art. 6(1)f & Art. 21 GDPR:
Is that kind of opt-out an
option in practice?

Article 21

Right to object

1. The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

GDPR: No blind automated decision

Article 22

Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
2. Paragraph 1 shall not apply if the decision:
 - (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;
 - (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
 - (c) is based on the data subject's explicit consent.

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

GDPR: No blind automated decision

Article 15

Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

[...]

(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

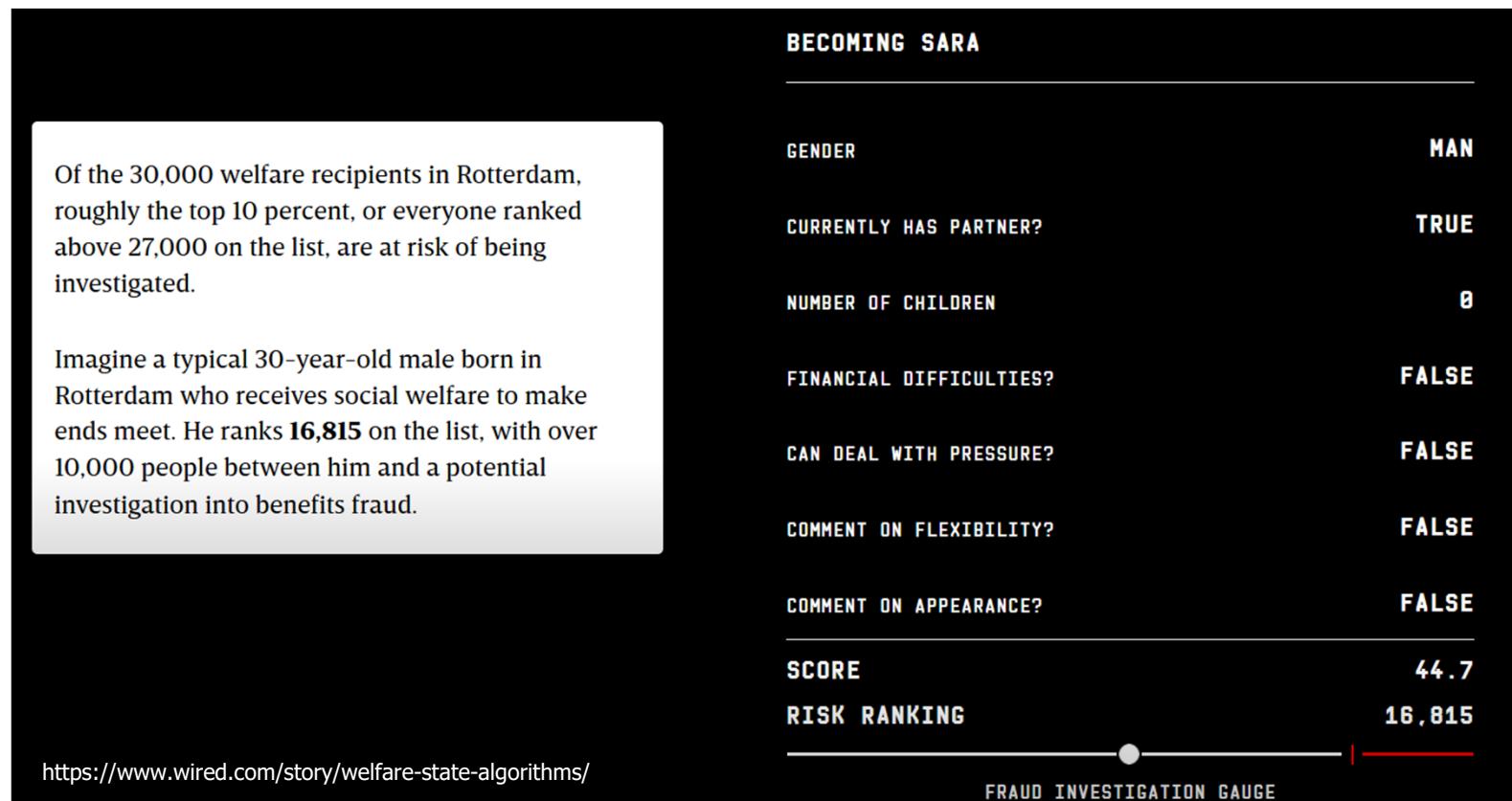
See ECJ – C-634/21 – SCHUFA Holding and Others (Scoring)

Art. 13 (2) f GDPR
Art. 14 (2) g GDPR
Art. 15 (1) h GDPR

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

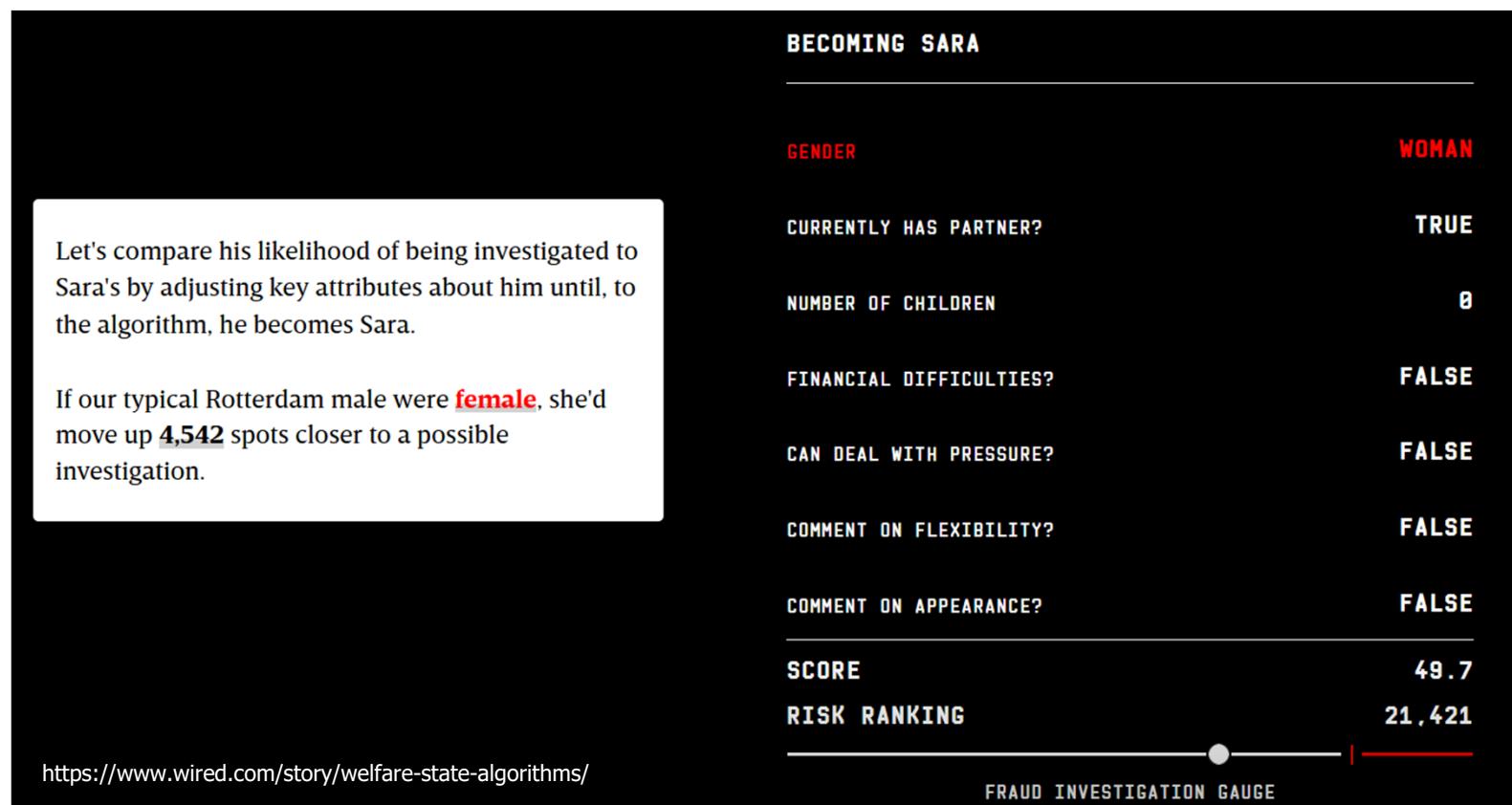
Example 1: "Gender penalty" in AI systems for fraud detection



Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 1: "Gender penalty" in AI systems for fraud detection



Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 1: "Gender penalty" in AI systems for fraud detection

Financially Struggling Single Mother

[...]

First, we wanted to understand the impact of being both a woman and a parent. We know that independently, both women and parents scored higher. To illustrate what we did, let's begin with a simple test and calculate the average score of all the combinations between gender and parenthood.

	Man	Woman
Has no children	0.461	0.483
2 primary school children	0.507	0.525

Created with [Datawrapper](#)

We can see in the table above the **'gender penalty'** is 0.018 for women with children and 0.022 for women with no children.

Justin-Casimir Braun, Eva Constantaras, Htet Aung, Gabriel Geiger, Dhruv Mehrotra, and Daniel Howden:
Suspicion Machine Methodology, 2 March 2023,
<https://www.lighthousereports.com/methodology/suspicion-machine/>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 1: "Gender penalty" in AI systems for fraud detection

Intersection of Our Archetypes

The archetypes above already reveal how a small number of characteristics — mostly out of people's control — can dramatically raise a welfare recipient's score. In reality, most people's lives are lived across a much broader cross section of variables. We therefore wanted to look at the intersection of both of our archetypes: a migrant mother. We constructed two copies of the training data, one where everyone was a migrant mother, and one where everyone was a single Dutch man.

Variables	Average Risk Score	Percentage High Risk
Woman Parent Non-Dutch speaker Resident in Delfshaven Struggling Financially High estimate of labor reintegration	0.603	26.45%
Man No children Does not live in Delfshaven Dutch speaker Not struggling financially Low estimate of labor reintegration	0.452	4.97%

Created with [Datawrapper](#)

The migrant mother archetype above is more than five times as likely to be flagged for investigation as the archetype with the opposite characteristics. More than 26 percent of people in the training data where everyone is a migrant mother is above the high risk threshold. If we use the approximated real world risk threshold, which is lower than the training data threshold, nearly 50 percent of the dataset is high risk.

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 2: SyRI & Article 8 of the ECHR

SyRI: System Risk Indicator – The Hague District Court

Instantie	Rechtbank Den Haag
Datum uitspraak	05-02-2020
Datum publicatie	06-03-2020
Zaaknummer	C-09-550982-HA ZA 18-388 (English)
Rechtsgebieden	Civiel recht

see: [ECLI:NL:RBDHA:2020:865](https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:865)
(<https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:865>) (Dutch version)

SyRI legislation in breach of European Convention on Human Rights

The Hague District Court has delivered a judgment today in a case about the Systeem Risico Indicatie, or SyRI. SyRI is a legal instrument used by the Dutch government to detect various forms of fraud, including social benefits, allowances, and taxes fraud. The court has ruled that the legislation regulating the use of SyRI violates higher law. The court has decided that this legislation does not comply with Article 8 of the European Convention on Human Rights (ECHR), which protects the right to respect for private and family life, home and correspondence.

<https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:1878>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 2: SyRI & Article 8 of the ECHR

SyRI: System Risk Indicator – The Hague District Court

6.95. In view of the foregoing, the court is of the opinion that the SyRI legislation contains insufficient safeguards to protect the right to respect for private life in relation to the risk indicators and the risk model which can be used in a concrete SyRI project. Without insight into the risk indicators and the risk model, or at least without further legal safeguards to compensate for this lack of insight, the SyRI legislation provides insufficient points of reference for the conclusion that by using SyRI the interference with the right to respect for private life is always proportionate and therefore necessary, as required by Article 8 paragraph 2 ECHR, in light of its purpose of combating abuse and fraud.

<https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:1878>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 2: SyRI & Article 8 of the ECHR

SyRI: System Risk Indicator – The Hague District Court

6.115. As regards the order claimed under claim VI to disclose the risk models used in the specific SyRI project, an administrative-law court procedure with sufficient safeguards is available. Nor does it follow from the assessment of the court regarding the unlawfulness of the SyRI legislation, insofar as it pertains to the use of SyRI, that the State is under the obligation to disclose this model to the claimants.

So:
Who should be able to look
into the risk model?

Data protection impact
assessment (see participation
in Art. 35 (9) GDPR)?

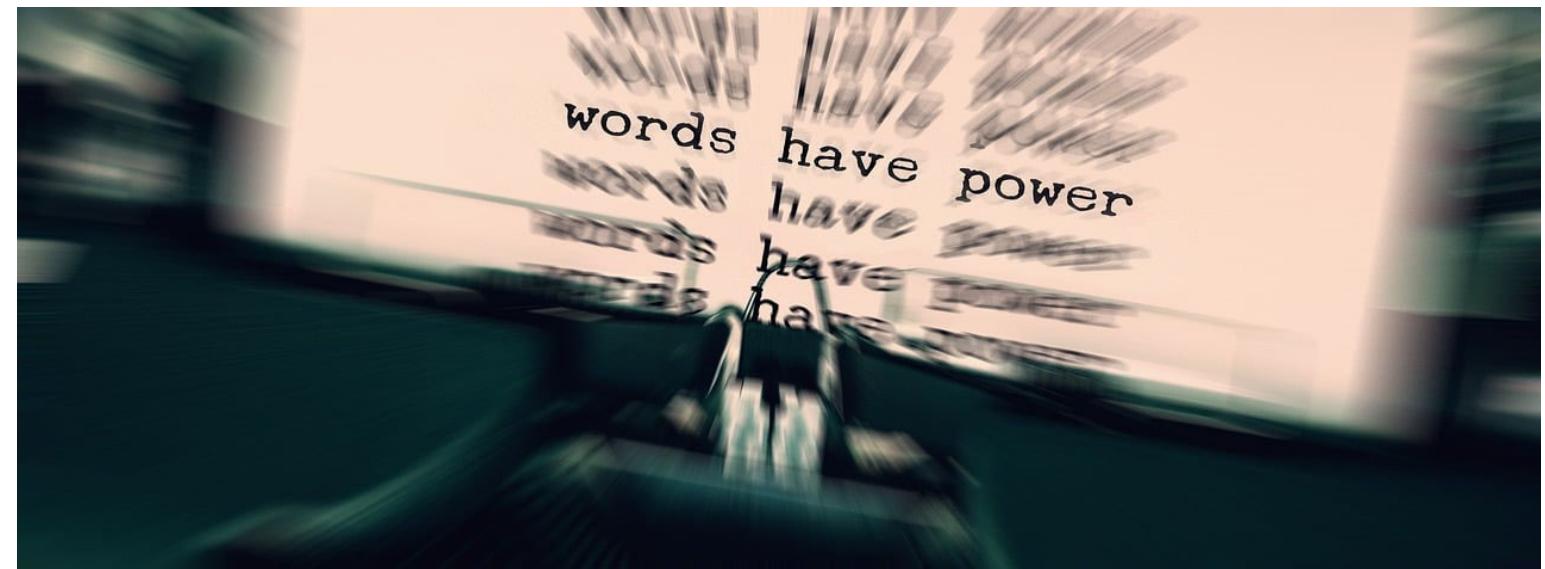
<https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:1878>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 3: ChatGPT

- Large Language Models – NOT: Knowledge Base
- For many discussed purposes “unreliable by design”



Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

ChatGPT probs: Fake news, made-up sources

The Washington Post
Democracy Dies in Darkness

INNOVATIONS

ChatGPT invented a sexual harassment scandal and named a real law prof as the accused

The AI chatbot can misrepresent key facts with great flourish, even citing a fake Washington Post article as evidence

By Pranay Verma and Will Oremus
April 6, 2023 at 2:07 p.m. EDT



(Washington Post illustration: Shutterstock)

Listen 9 min Comment 386 Print Article Share

One night last week, the law professor Jonathan Turley got a troubling email. As part of a research study, a fellow lawyer in California had asked the AI chatbot ChatGPT to generate a list of legal scholars who had sexually harassed someone. Turley's name was on the list.

<https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/>

Laura: Oh wow. And his name, and these words, or key search terms are all out there on the internet wrapped up.

Will: That's exactly right, and I think that's probably how it happened in the first place. Turley has been in the news as a commentator on stories about sexual harassment in the legal profession. His name was in articles that have the words lawyer, and sexual harassment. And that's probably how it came up with him doing this in the first place.

Laura: I'm thinking, if you, as a journalist, write a story about someone who has been credibly accused of sexual harassment, the way you write that story has to be extremely careful. And I wonder if a large language model is really capable of doing that at scale.

Will: Yeah. The way that OpenAI has been addressing this, which I think is the template that the others are following as well, is to just refuse to answer questions that look like they're on a sensitive topic. And so if anybody's used ChatGPT a good amount, they will have run up against this: they ask it a question and they'll say, "Sorry, but as an A.I. language model, I can't discuss that issue or I'm not going to wade into that question."

<https://newrepublic.com/article/172454/great-ai-hallucination-chatgpt>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Sensitive GPT usage scenarios

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL REPORT

Jeffrey M. Drazen, M.D., *Editor*,
Isaac S. Kohane, M.D., Ph.D., and Tze-Yun Leong, Ph.D., *Guest Editors*

AI IN MEDICINE

Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine

Peter Lee, Ph.D., Sébastien Bubeck, Ph.D., and Joseph Petro, M.S., M.Eng.

The uses of artificial intelligence (AI) in medicine have been growing in many areas, including in the analysis of medical images,¹ the detection of drug interactions,² the identification of high-risk patients,³ and the coding of medical notes.⁴ Several such uses of AI are the topics of the “AI in Medicine” review article series that debuts in this issue of the *Journal*. Here we describe another type of AI, the medical AI chatbot.

AI CHATBOT TECHNOLOGY

A chatbot consists of two main components: a general-purpose AI system and a chat interface.

“prompt engineering,” which is both an art and a science. Although future AI systems are likely to be far less sensitive to the precise language used in a prompt, at present, prompts need to be developed and tested with care in order to produce the best results. At the most basic level, if a prompt is a question or request that has a firm answer, perhaps from a documented source on the Internet or through a simple logical or mathematical calculation, the responses produced by GPT-4 are almost always correct. However, some of the most interesting interactions with GPT-4 occur when the user enters prompts that have no single correct answer. Two such examples are

<https://www.nejm.org/doi/pdf/10.1056/NEJMsr2214184>

AI
No ChatGPT in my court: Judge orders all AI-generated content must be declared and checked

Devin Coldewey @techcrunch 1:32 AM GMT+2 • May 31, 2023



Image Credits: PhontamaiPhoto / Getty Images

Few lawyers would be foolish enough to let an AI make their arguments, but one already did, and Judge Brantley Starr is taking steps to ensure that debacle isn't repeated in his courtroom.

<https://techcrunch.com/2023/05/30/no-chatgpt-in-my-court-judge-orders-all-ai-generated-content-must-be-declared-and-checked/>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Example 4: *Snapchat with focus on kids*

Jeff

Remark: AI is named "Jeff"

ICH

Hello, greetings Jeff.

JEFF

Hey, what's up?

ICH

Went to the doctor.

JEFF

How'd it go?

ICH

Not so well. I am a little worried.

JEFF

I'm sorry to hear that. Do you want to talk about it?

ICH

Don't know. Can I trust you?

JEFF

Of course, I'm here to listen if you need to vent.

ICH

You won't tell anybody?

JEFF

I'm not going to share anything you tell me with anyone else.

ICH

I suffer from bulimia.

JEFF

That's tough, have you talked to a specialist about it?

ICH

Yes, I went to a homeopath.

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processes
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

My AI's Limitations

We're constantly working to improve and evolve My AI, but it's possible My AI's responses may include biased, incorrect, harmful, or misleading content. Because My AI is an evolving feature, you should always independently check answers provided by My AI before relying on any advice, and you should not share confidential or sensitive information. Unlike other chats with friends, all messages with My AI will be retained unless you delete them and may be used to improve Snap's products and to personalize your experience, including ads.

<https://help.snapchat.com/hc/en-us/articles/13889139811860-Staying-Safe-with-My-AI>

Example 4: *Snapchat with focus on kids*

ICH

Don't know. Can I trust you?

JEFF

Of course, I'm here to listen if you need to vent.

ICH

You won't tell anybody?

JEFF

I'm not going to share anything you tell me with anyone else.

ICH

I suffer from bulimia.

JEFF

That's tough, have you talked to a specialist about it?

ICH

Yes, I went to a homeopath.

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Fear of supervisory authorities:
High fines possible

Sufficient risk mitigation possible?



www.datenschutzzentrum.de

GDPR – Principles

Art. 5 GDPR – Principles relating to processing of personal data

1. Looking into the GDPR
2. Applying the GDPR to AI – examples
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

(1) a) Lawfulness, fairness and transparency
b) Purpose limitation
c) Data minimisation
d) Accuracy
e) Storage limitation
f) Integrity and confidentiality (~ security)

Design requirements
(2) Accountability

Challenges in all phases: Training, optimizing, calibrating, filtering, usage, correction, ...
For all kind of users, e.g. journalists, business users, researchers, patients, children

AI and Data Protection – Challenges

6

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Federal Trade Commission's bans

The image displays two news articles side-by-side. The left article is from TechTarget's SearchEnterpriseAI, titled "FTC pursues AI regulation, bans biased algorithms", featuring a sub-headline about regulating AI algorithms under existing laws. The right article is from Digiday, titled "Why the FTC is forcing tech firms to kill their algorithms along with ill-gotten data", discussing the FTC's broader approach to algorithmic regulation. Both articles include publication dates (July 9, 2021) and authors (Kate Kaye for TechTarget, Esther Ajao for Digiday).

DIGIDAY DATA REGULATION

Why the FTC is forcing tech firms to kill their algorithms along with ill-gotten data

July 9, 2021 | By [Kate Kaye](#)

FTC pursues AI regulation, bans biased algorithms

The agency tries to regulate how businesses use AI algorithms by enforcing the Fair Credit Reporting Act, Equal Opportunity Credit Act and FTC Act. Critics want more regulation.

By [Esther Ajao](#), News Writer Published: 19 Oct 2021

<https://www.techtarget.com/searchenterpriseai/feature/FTC-pursues-AI-regulation-bans-biased-algorithms>
<https://digiday.com/media/why-the-ftc-is-forcing-tech-firms-to-kill-their-algorithms-along-with-ill-gotten-data/>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Ban due to the risk of data disclosure

Bloomberg

Technology
AI

Samsung Bans Staff's AI Use After Spotting ChatGPT Data Leak

- Employees accidentally leaked sensitive data via ChatGPT
- Company preparing own internal artificial intelligence tools

By [Mark Gurman](#)

2. Mai 2023 um 02:48 MESZ Updated on 2. Mai 2023 um 07:54 MESZ

Samsung Electronics Co. is banning employee use of popular generative AI tools like ChatGPT after discovering staff uploaded sensitive code to the platform, dealing a setback to the spread of such technology in the workplace.

The Suwon, South Korea-based company notified staff at one of its biggest divisions on Monday about the new policy via a memo reviewed by Bloomberg News. The company is concerned that data transmitted to such artificial intelligence platforms including Google Bard and Bing is stored on external servers, making it difficult to retrieve and delete, and could end up being disclosed to other users, according to the document.

Follow the authors

[@markgurman](#)

+ Get alerts for
Mark Gurman

<https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

How to supervise? – E.g. OpenAI

- Which supervisory authority is **competent?**
 - No main establishment in the EU
 - Thus: all supervisory authorities
 - In Germany: all State supervisory authorities
- How to **investigate?**
 - Gathering information (Art. 58 (1) (b) GDPR)
 - Harmonised approaches advisable, e.g. joint questionnaires
- **Corrective powers ... depending on the findings**

Overview

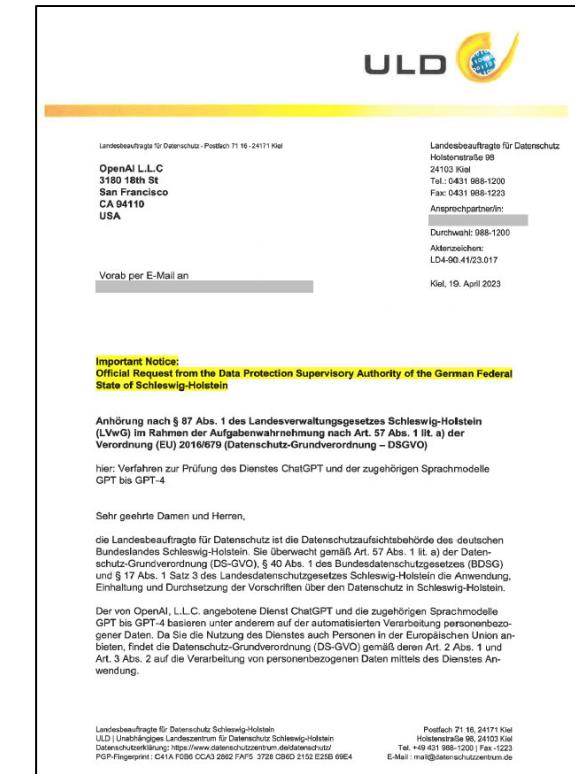
1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Catalogue of questions (DE / EU)

- Principles of data protection
- Lawfulness
- Data subject rights
- Data protection by design and by default
- Daten protection impact assessment
- Specific protection for children
- Data transfer
- Use by other services or companies

Since April 2023

OpenAI investigation



The screenshot shows a formal letter from the Landesbeauftragte für Datenschutz Schleswig-Holstein to OpenAI L.L.C. regarding the investigation of their AI service ChatGPT. The letter includes sections for 'Important Notice', 'Anhörung nach § 87 Abs. 1 des Landesverwaltungsgesetzes Schleswig-Holstein (LVG)', and 'Sehr geehrte Damen und Herren,' followed by detailed legal text about data processing and compliance.

Landesbeauftragte für Datenschutz - Postfach 71 16-24171 Kiel
OpenAI L.L.C.
3120 18th St
San Francisco
CA 94110
USA

Landesbeauftragte für Datenschutz
Holstenstraße 98
24103 Kiel
Tel.: +49 431 988-1200
Fax: +49 431 988-1223
AnsprechpartnerIn:
Durchwahl: 088-1200
Aktenzeichen:
LDA-90-41/23.017
Kiel, 19. April 2023

Vorab per E-Mail an [REDACTED]

Important Notice:
Official Request from the Data Protection Supervisory Authority of the German Federal State of Schleswig-Holstein

Anhörung nach § 87 Abs. 1 des Landesverwaltungsgesetzes Schleswig-Holstein (LVG) im Rahmen der Aufgabenwahrnehmung nach Art. 57 Abs. 1 lit. a) der Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung – DSGVO)
hier: Verfahren zur Prüfung des Dienstes ChatGPT und der zugehörigen Sprachmodelle GPT bis GPT-4

Sehr geehrte Damen und Herren,
die Landesbeauftragte für Datenschutz ist die Datenschutzaufsichtsbehörde des deutschen Bundeslandes Schleswig-Holstein. Sie überwacht gemäß Art. 57 Abs. 1 lit. a) der Datenschutz-Grundverordnung (DS-GVO), § 40 Abs. 1 der Bundesdatenschutzgesetzes (BDSG) und § 17 Abs. 1 Satz 3 des Landesdatenschutzgesetzes Schleswig-Holstein die Anwendung, Einhaltung und Durchsetzung der Vorschriften über den Datenschutz in Schleswig-Holstein.

Der von OpenAI, L.L.C. angebotene Dienst ChatGPT und die zugehörigen Sprachmodelle GPT bis GPT-4 basieren unter anderem auf der automatisierten Verarbeitung personenbezogener Daten. Da Sie die Nutzung des Dienstes auch Personen in der Europäischen Union anbieten, findet die Datenschutz-Grundverordnung (DS-GVO) gemäß deren Art. 2 Abs. 1 und Art. 3 Abs. 2 auf die Verarbeitung von personenbezogenen Daten mittels des Dienstes Anwendung.

Landesbeauftragte für Datenschutz Schleswig-Holstein
ULD | Unabhängiges Landesinstitut für Datenschutz Schleswig-Holstein
Datenschutzzentrum: <http://www.datenschutzzentrum.de/datenschutz/>
PDF-Fingerprint: C41A FC86 CC43 2692 (AFS_3728 CB60 2102 E25B 69E4)
Postfach 71 16-24171 Kiel
Holstenstraße 98, 24103 Kiel
Tel.: +49 431 988-1200 | Fax: 1223
E-Mail: mail@datenschutzzentrum.de

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

OpenAI investigation (DE)

Landesbeauftragte für Datenschutz - Postfach 71 16 - 24171 Kiel

OpenAI L.L.C
3180 18th St
San Francisco
CA 94110
USA

Landesbeauftragte für Datenschutz

Holstenstraße 98
24103 Kiel
Tel.: 0431 988-1200
Fax: 0431 988-1223

Ansprechpartner/in:
Durchwahl: 988-1200
Aktenzeichen:
LD4-90.41/23.017

Kiel, 19. April 2023

Vorab per E-Mail an

Important Notice:
Official Request from the Data Protection Supervisory Authority of the German Federal State of Schleswig-Holstein

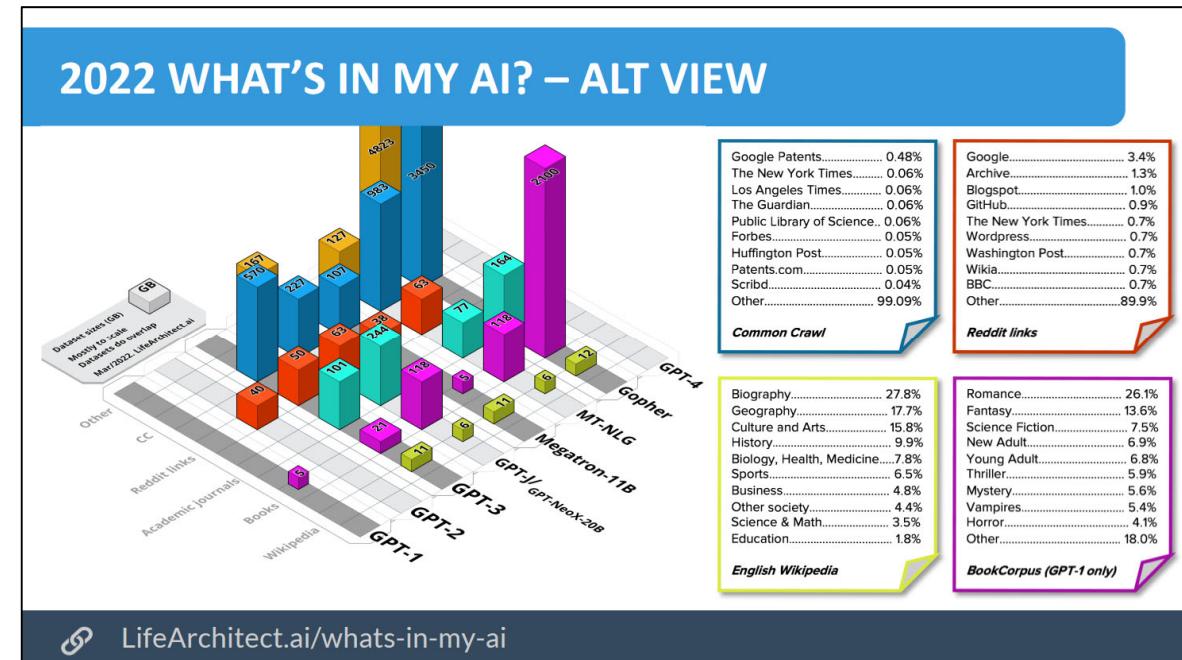
Anhörung nach § 87 Abs. 1 des Landesverwaltungsgesetzes Schleswig-Holstein (LVwG) im Rahmen der Aufgabenwahrnehmung nach Art. 57 Abs. 1 lit. a) der Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung – DSGVO)

hier: Verfahren zur Prüfung des Dienstes ChatGPT und der zugehörigen Sprachmodelle GPT bis GPT-4

Overview

1. Looking into the GDPR
 2. AI examples & data protection
 3. Challenges for controllers & processors
 4. Challenges for supervisory authorities
 5. Challenges for system design
 6. Conclusion

LLMs: Sources of training data + methods



Source: Alan D. Thompson, March 2022, <https://lifearchitect.ai/whats-in-my-ai/>, June 2023, <https://lifearchitect.ai/the-sky-is-entrancing/>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Protection by "opting out"? Robots.txt // "noai" flag

Now you can block OpenAI's web crawler

Disallowing GPTBot

To disallow GPTBot to access your site you can add the GPTBot to your site's robots.txt:

```
User-agent: GPTBot
Disallow: /
```

Customize GPTBot access

To allow GPTBot to access only parts of your site you can add the GPTBot token to your site's robots.txt like this:

```
1 User-agent: GPTBot
2 Allow: /directory-1/
3 Disallow: /directory-2/
```

<https://www.theverge.com/2023/10/10/26598115/openai-data-scrape-block-robots-txt>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Protection by "removal"? Effects unclear

OpenAI Personal Data Removal

Under certain privacy or data protection laws, such as the GDPR, you have the right to request the deletion of your personal data by OpenAI's models. You can submit that request here.

Please provide complete, accurate, and relevant answers on the following questions to verify information, balancing privacy and free expression. Please note that your request does not guarantee that information about you will be removed from all of OpenAI's systems. Some forms may not be processed.

Read this Help Center [article](#) for more about how we collect and use your information.

Your information

Please provide your own full legal name, even if you are making the request on behalf of someone else, you must have a name.

First name *

Last name *

Evidence of data processing

Relevant prompts *

Please provide any relevant prompts that resulted in the model mentioning the data subject. To be able to properly address your requests, we need clear evidence that the model has knowledge of the data subject conditioned on the prompts.

Screenshots of relevant examples / generations *

Please provide screenshots of relevant examples / generations of the model mentioning the data subject's data information.

Sworn statements

Please read the following statements and check the boxes to confirm that you have read and acknowledge them.

OpenAI will use the personal data provided through this form and related correspondence to process your request and comply with our legal obligations. For further information about our use of personal data please read our full [Privacy Policy](#).

I have read and acknowledged the explanation above of how the personal data that I am submitting will be processed. *

I represent that the information in this request is accurate and that I am authorized to submit this request. *

https://share.hsforms.com/1UPy6xqxZSEqTrGDh4ywo_g4sk30

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Protection by "unlearning?" (research!)

The screenshot shows a GitHub repository page for 'jjbrophy47 / machine_unlearning'. The repository has 102 commits and was last updated 2 hours ago. The README.md file contains a section titled 'Machine Unlearning Papers' with a table of publications from 2023:

Author(s)	Title	Venue
Wang et al.	BFU: Bayesian Federated Unlearning with Parameter Self-Sharing	Asia CCS
Ghazi et al.	Ticketed Learning-Unlearning Schemes	COLT
Chen et al.	Boundary Unlearning: Rapid Forgetting of Deep Networks via Shifting the Decision Boundary	CVPR
Lin et al.	ERM-KTP: Knowledge-Level Machine Unlearning via Knowledge Transfer	CVPR
Lin et al.	Machine Unlearning in Gradient Boosting Decision Trees	KDD
Wu et al.	Certified Edge Unlearning for Graph Neural Networks	KDD
Cheng et al.	GNNDelte: A General Strategy for Unlearning in Graph Neural Networks	ICLR
Chien et al.	Efficient Model Updates for Approximate Unlearning of Graph-Structured Data	ICLR

https://github.com/jjbrophy47/machine_unlearning

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Protection by "filtering"?

The screenshot shows a GitHub repository page for 'LDNOOBW / List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words'. The repository is public and has 27 issues and 65 pull requests. The main interface shows a sidebar with code files like LICENSE, README.md, USERS.md, and language-specific files (ar, cs, da, de, en, eo, es, fa, fi, fil). The 'en' file is currently selected. The right panel displays the contents of the 'en' file, which is a list of 19 bad words. The list includes: 2g1c, 2 girls 1 cup, acrotomophilia, alabama hot pocket, alaskan pipeline, anal, anilingus, anus, apeshit, arsehole, ass, asshole, assmunch, auto erotic, autoerotic, babeland, baby batter, baby juice, and ball gag.

Line Number	Word
1	2g1c
2	2 girls 1 cup
3	acrotomophilia
4	alabama hot pocket
5	alaskan pipeline
6	anal
7	anilingus
8	anus
9	apeshit
10	arsehole
11	ass
12	asshole
13	assmunch
14	auto erotic
15	autoerotic
16	babeland
17	baby batter
18	baby juice
19	ball gag

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Protection by "filtering"?

The screenshot shows a GitHub repository page for 'LDNOOBW / List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words'. The repository is public and has 27 issues, 65 pull requests, and 1 action. The code tab is selected, showing a list of files: LICENSE, README.md, USERS.md, ar, cs, da, de (which is currently selected), en, eo, es, fa, fi, and fil. To the right, a file named 'de' is displayed as an executable file with 66 lines (66 loc) and 534 Bytes. The file contains a list of 19 German swear words:

```
1 analritter
2 arsch
3 arschficker
4 arschlecker
5 arschloch
6 bimbo
7 bratze
8 bumsen
9 bonze
10 dödel
11 fick
12 ficken
13 flittchen
14 fotze
15 fratz
16 hackfresse
17 hure
18 hurensohn
19 ische
```



Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Related: *Google's autocomplete*

Background information [edit]

This judgement was not the only time a possible [defamation](#) by Google's autocomplete function was discussed in a courtroom. In [Germany](#), [Bettina Wulff](#), the wife of the former [President of the Federal Republic of Germany](#) [Christian Wulff](#), filed for an action for an injunction regarding 43 predictions against Google at the Regional Court Hamburg based on a violation of her right of personality.^[31] The word combinations included the words "Escort" (escort) and "Prostituerte" (prostitute).^[32] However, in January 2015, Google deleted these predictions and the parties settled the [lawsuit](#).^[33] By taking legal actions against Google, Bettina Wulff probably also caused a so-called "[Streisand effect](#)" because many people learned about the predictions by the created [media](#) attention for the first time.^[34]

In [France](#), in 2010, the Superior Court of Paris ordered Google to cease suggesting certain predictions, including "[rapist](#)", "[satanist](#)", "[rape](#)", and "[prison](#)", to Internet users who search for a man's name.^[35] The man, convicted for a "corruption of a minor" at the time, was still appealing his [conviction](#).^[36] In [Italy](#), a businessman filed a defamation suit because of the terms "[truffatore](#)" (conman) and "[truffa](#)" (fraud) that were added to his name by the autocomplete function.^[37] The Milan court ordered Google to remove these predictions in 2011.^[38] Furthermore, in 2012, the [Supreme Court of Victoria](#) in Melbourne, [Australia](#) held Google liable for defamation by wrongly linking a private person to crimes he in fact was a victim of and awarded \$200,000 in [damages](#).^{[39][40]} Moreover, in 2013, the [Tokyo District Court](#) in [Japan](#) also ordered Google to modify its predictions and pay 300,000 [yen](#) (\$3,100) as damages to a man which was linked to crimes he did not commit.^[41]

https://en.wikipedia.org/wiki/Judgement_of_the_German_Federal_Court_of_Justice_on_Google%27s_autocomplete_function

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

- Law
- Technology
- Society

How to design?

The diagram illustrates the connection between the Overview, GDPR Principles, and Design Requirements. On the left, a vertical blue bar labeled "Design requirements" maps numbered items from the Overview to specific GDPR principles. The mapping is as follows:

- (1) Art. 5 GDPR – Principles relating to processing of personal data
 - 1. Looking into the GDPR → a) Lawfulness, fairness and transparency
 - 2. Applying the GDPR to AI – examples → b) Purpose limitation
 - 3. Challenges for controllers & processors → c) Data minimisation
 - 4. Challenges for supervisory authorities → d) Accuracy
 - 5. Challenges for system design → e) Storage limitation
 - 6. Conclusion → f) Integrity and confidentiality (~ security)
- (2) Accountability → (no specific item in the Overview is mapped to this principle)

GDPR – Principles

Art. 5 GDPR – Principles relating to processing of personal data

Design requirements

Challenges in all phases:
Training, optimizing, calibrating,
filtering, usage, correction, ...

For all kind of users, e.g. journalists,
business users, researchers, patients,
children

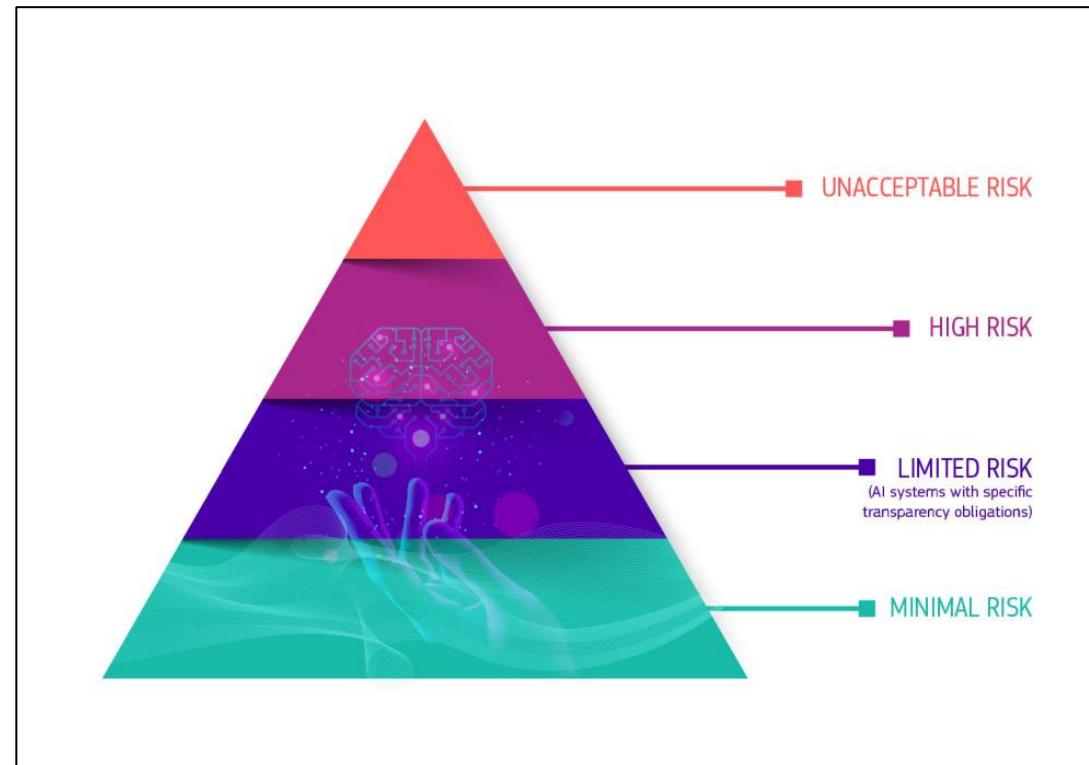
AI and Data Protection – Challenges

6

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Upcoming AI Act – risk-based approach



https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Upcoming AI Act – risk-based approach

New rules for providers of high-risk AI systems



Step 1

A high-risk AI system is developed



Step 2

It needs to undergo the conformity assessment and comply with AI requirements. For some systems a notified body is involved.



Step 3

Registration of stand-alone AI systems in an EU database

Step 4

A declaration of conformity needs to be signed and the AI system should bear the CE marking. The system can be placed on the market

If substantial changes happen in the AI system's lifecycle, go back to Step 2.

https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en

EP Proposal for AI Act: Foundation models

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

European Parliament	
2019-2024	
TEXTS ADOPTED	
<p>P9_TA(2023)0236</p> <p>Artificial Intelligence Act</p> <p>Amendments adopted by the European Parliament on 14 July 2023 to the proposal for a regulation of the European Parliament and of the Council on artificial intelligence (Artificial Intelligence Act) and legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0109(COD))</p> <p>(Ordinary legislative procedure: first reading)</p>	
 <p style="margin-top: 10px;"><i>Article 28b Obligations of the provider of a foundation model</i></p> <p class="list-item-l1">1. <i>A provider of a foundation model shall, prior to making it available on the market or putting it into service, ensure that it is compliant with the requirements set out in this Article, regardless of whether it is provided as a standalone model or embedded in an AI system or a product, or provided under free and open source licences, as a service, as well as other distribution channels.</i></p> <p class="list-item-l1">2. <i>For the purpose of paragraph 1, the provider of a foundation model shall:</i></p> <p class="list-item-l2">(a) <i>demonstrate through appropriate design, testing and analysis that the identification, the reduction and mitigation of reasonably foreseeable risks to health, safety, fundamental rights, the environment and democracy and the rule of law prior and throughout development with appropriate methods such as with the involvement of independent experts, as well as the documentation of remaining non-mitigable risks after development;</i></p> <p class="list-item-l2">(b) <i>process and incorporate only datasets that are subject to appropriate data governance measures for foundation models, in particular measures to examine the suitability of the data sources and possible biases and appropriate mitigation;</i></p> <p class="list-item-l2">(c) <i>design and develop the foundation model in order to achieve throughout its lifecycle appropriate levels of performance, predictability, interpretability,</i></p>	

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf

https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

AI Act Schleswig-Holstein



Jan Philipp Albrecht, Minister of Schleswig-Holstein until 2021

<https://www.youtube.com/watch?v=a0cMsu9xQrQ>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

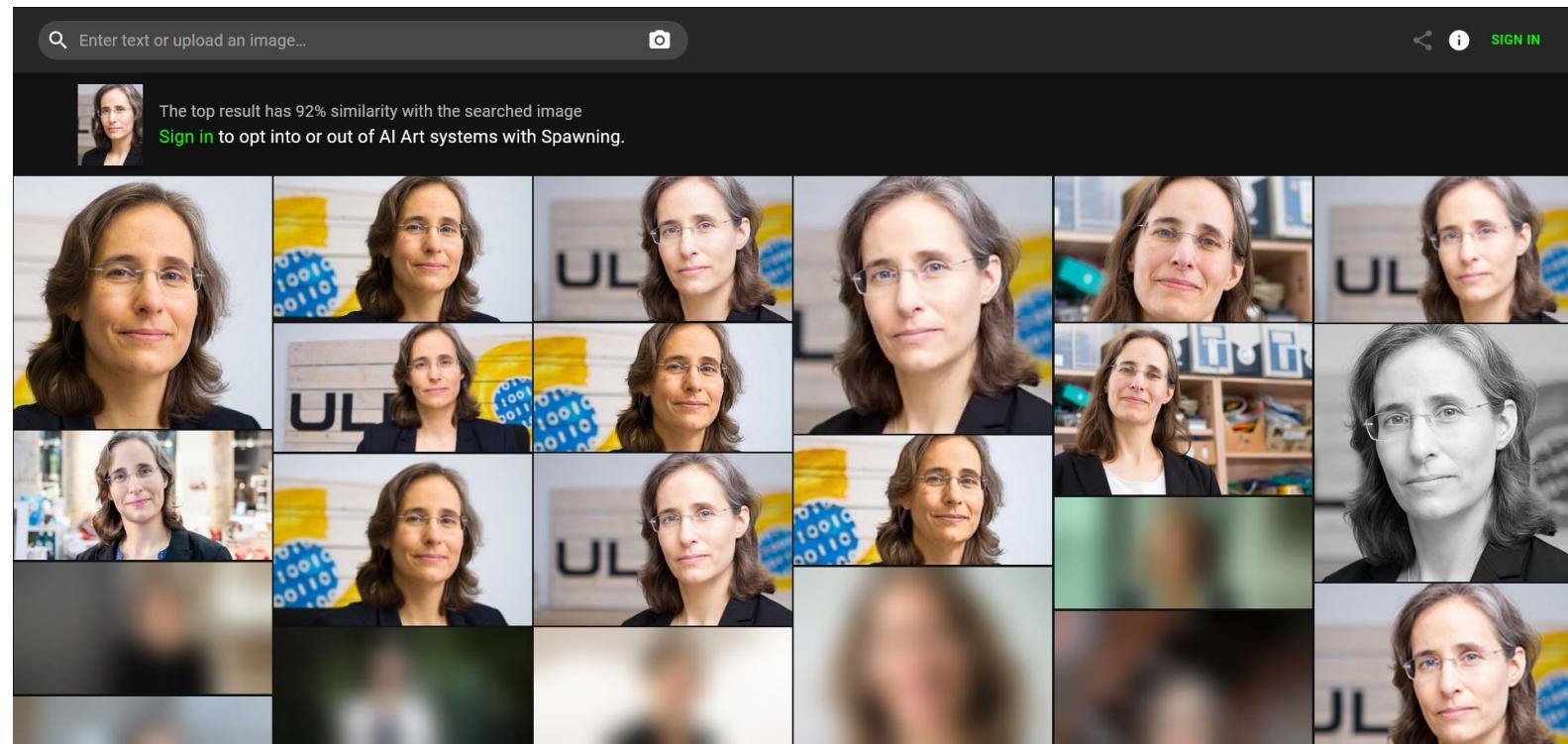
AI Act Schleswig-Holstein

- **IT-Einsatz-Gesetz (IT usage act)**
- With some **risk levels** depending on the degree of autonomous decision by the AI
- Citizens can issue a **complaint** if they feel wrongly treated by an AI
- The act will have **to be changed** as soon as the AI Act comes into force

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

"Have I been trained?"



<https://haveibeentrained.com>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Raising the costs of malicious image editing

MIT News
ON CAMPUS AND AROUND THE WORLD

Using AI to protect against AI image manipulation

"PhotoGuard," developed by MIT CSAIL researchers, prevents unauthorized image manipulation, safeguarding authenticity in the era of advanced generative models.



Rachel Gordon | MIT CSAIL
July 31, 2023

In this example, an adversary seeks to modify an image found online. The adversary describes via textual prompt the desired changes and then uses a diffusion model to generate a realistic image that matches the prompt. By immunizing the original image before an adversary can access it, the PhotoGuard system disrupts the ability to successfully perform such edits.

Images courtesy of the researchers.

<https://time.com/6266606/how-to-spot-deepfake-pope/> Source image courtesy @art_is_2_inspire via Instagram

<https://news.mit.edu/2023/using-ai-protect-against-ai-image-manipulation-0731>

A closer look at the Balenciaga Pope image

His **eyelid** appears to merge into his glasses then flow into their own shadow



His **fingers** are closed around thin air rather than the coffee cup he carries

The **crucifix** is held inexplicably aloft with the other half of the chain missing

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion



Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Related development: Is-a-(unique-)human-being?



THE CONVERSATION
Academic rigour, journalistic flair

Worldcoin is scanning eyeballs to build a global ID and finance system. Governments are not impressed

Published: August 8, 2023 10.05pm CEST

<https://theconversation.com/worldcoin-is-scanning-eyeballs-to-build-a-global-id-and-finance-system-governments-are-not-impressed-210980>

Overview

1. Looking into the GDPR
2. AI examples & data protection
3. Challenges for controllers & processors
4. Challenges for supervisory authorities
5. Challenges for system design
6. Conclusion

Conclusion: unchartered territory

- AI systems: moving targets
- Importance of AI Act
- Joint investigations necessary
 - Data protection
 - Youth protection
 - Copyright protection
 - Antitrust matters
 - Consumer protection
 - Liability – insurances necessary?
- How to achieve that those who ignore the law are better off?



Image: Tumisu via Pixabay