

E-Discovery Symposium 2026



**Maakt Retrieval Augmented Generation
AI geschikt voor forensische toepassing?**

From RAG to Riches?

Maakt Retrieval Augmented Generation AI geschikt voor forensische toepassing?



mr. dr. Mark Zoetekouw
Senior Lecturer / Researcher
Leiden University of Applied Sciences
zoetekouw.m@HSLeiden.nl

Hoofdinspecteur (voluntary)
Dutch Police
Noord-Holland Cybercrime Team



All material used in this talk is subject to copyright (auteursrecht) © Zoetekouw 2026.
Should you want to (re)use it in part or in whole, please contact me to discuss.

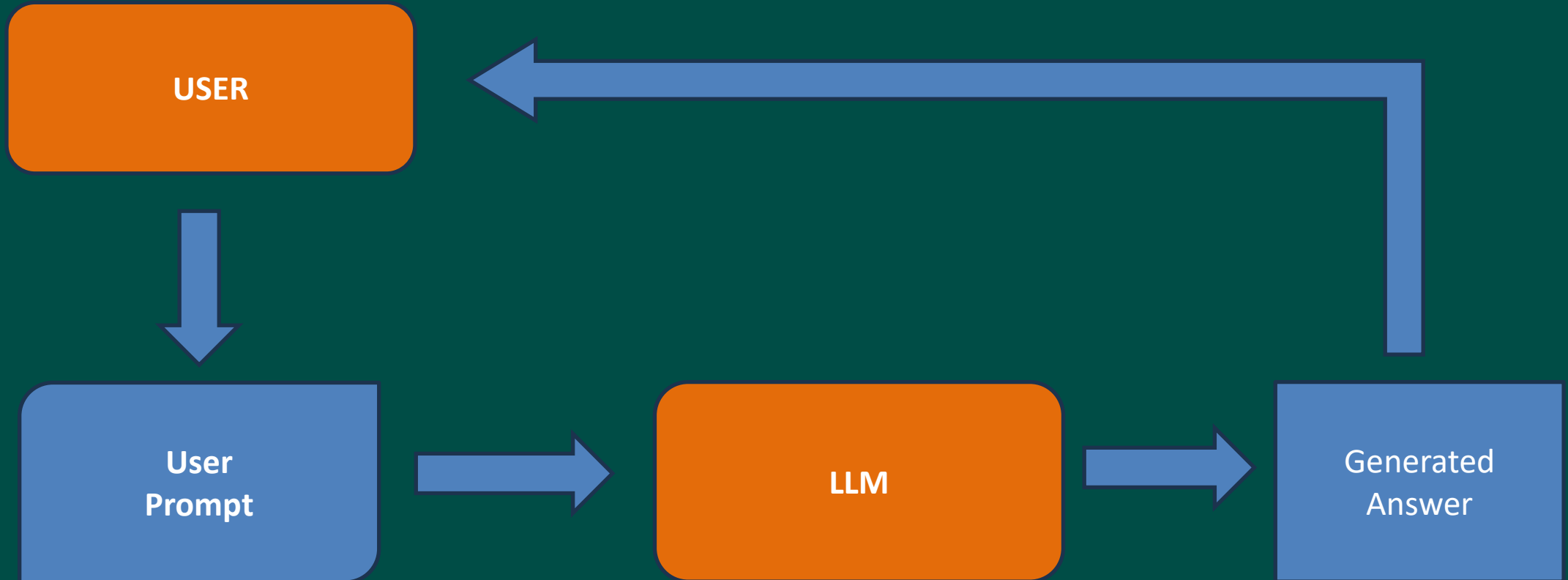
Images used are either free-to-use or generated with AI by me. If anyone thinks (s)he holds rights to a picture used please contact me so we can discuss. I will remove offending images if proper substantiation is made.

Wat is RAG?

(Retrieval Augmented Generation)

Een korte,
niet AL te technische, Primer

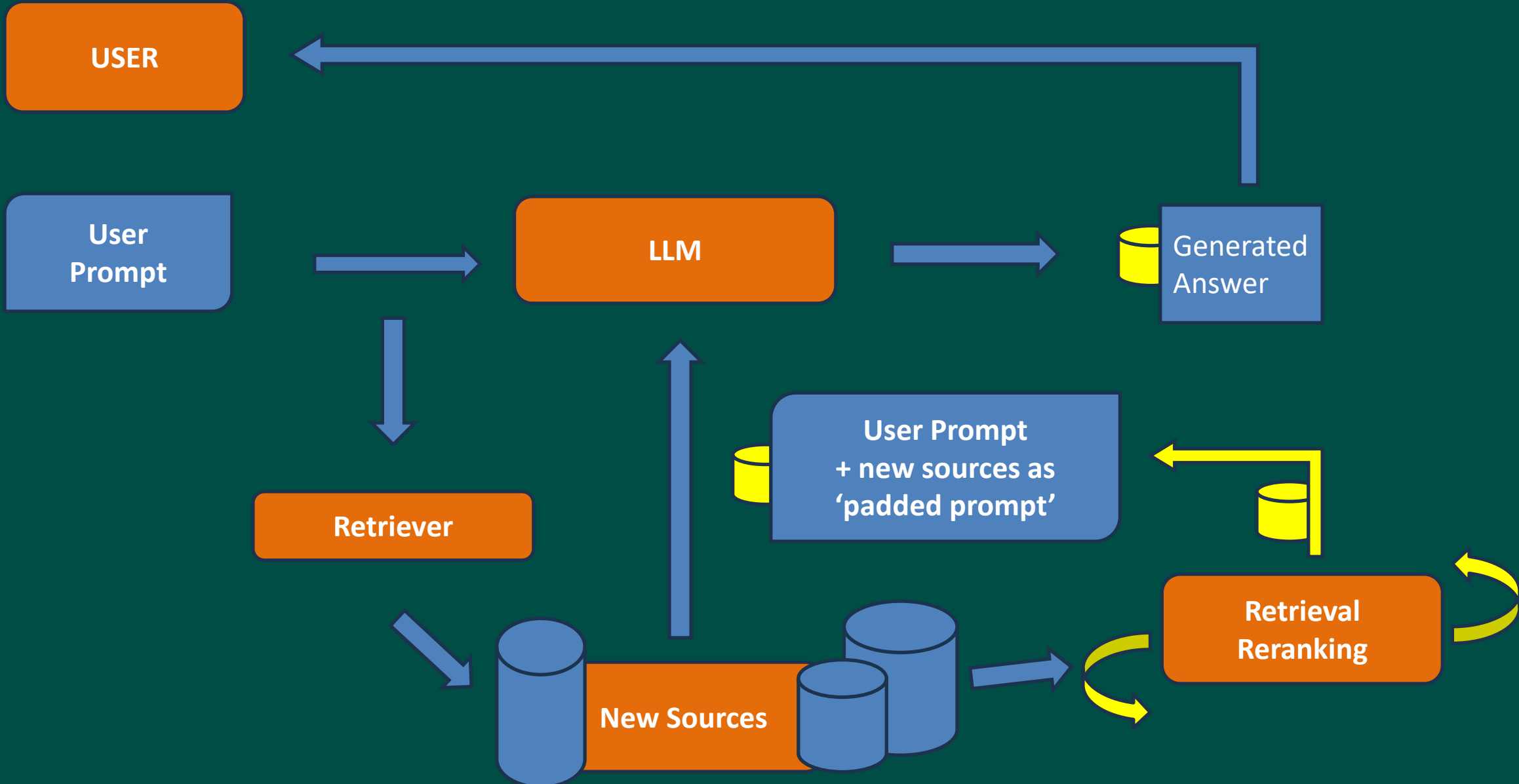
Een 'simple' LLM



Wat is RAG?

- **RAG = uitbreiding standaard LLM**
- **RAG = LLM + nieuwe info, zonder hertraining**
- **RAG = toevoeging 'bijbronnen' via prompt'padding'**

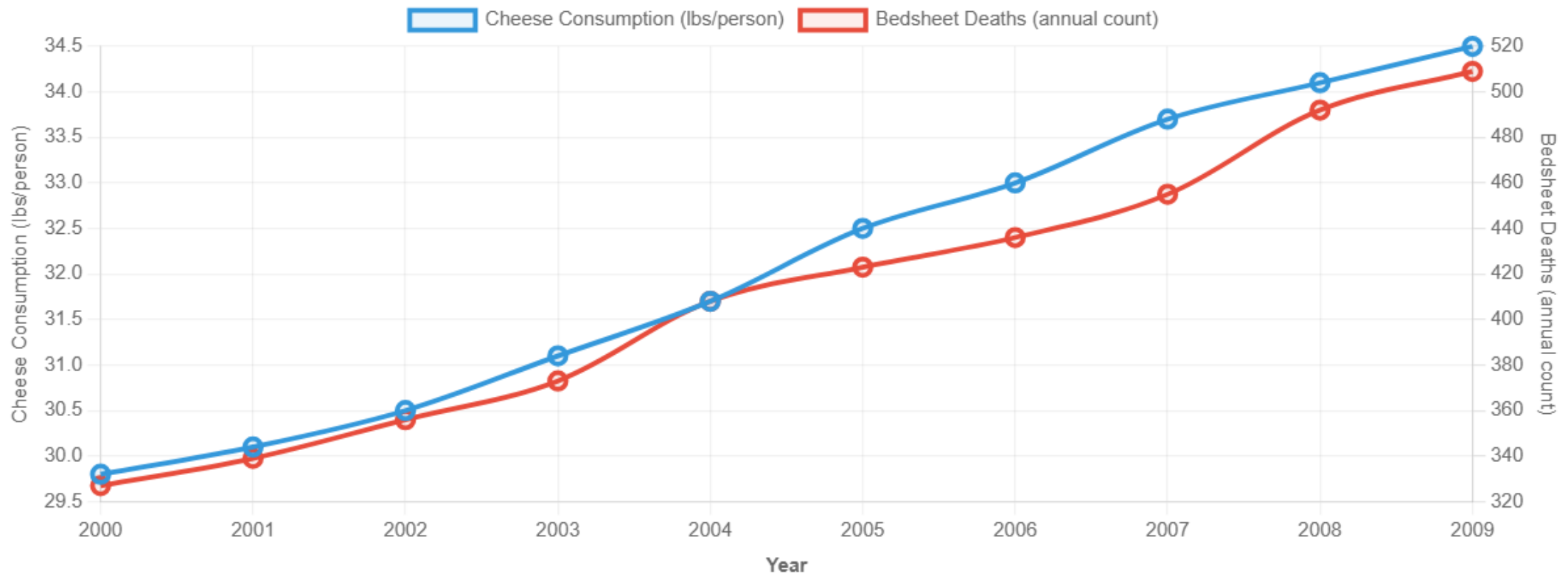
Een LLM uitgebreid met RAG



Slightly deeper

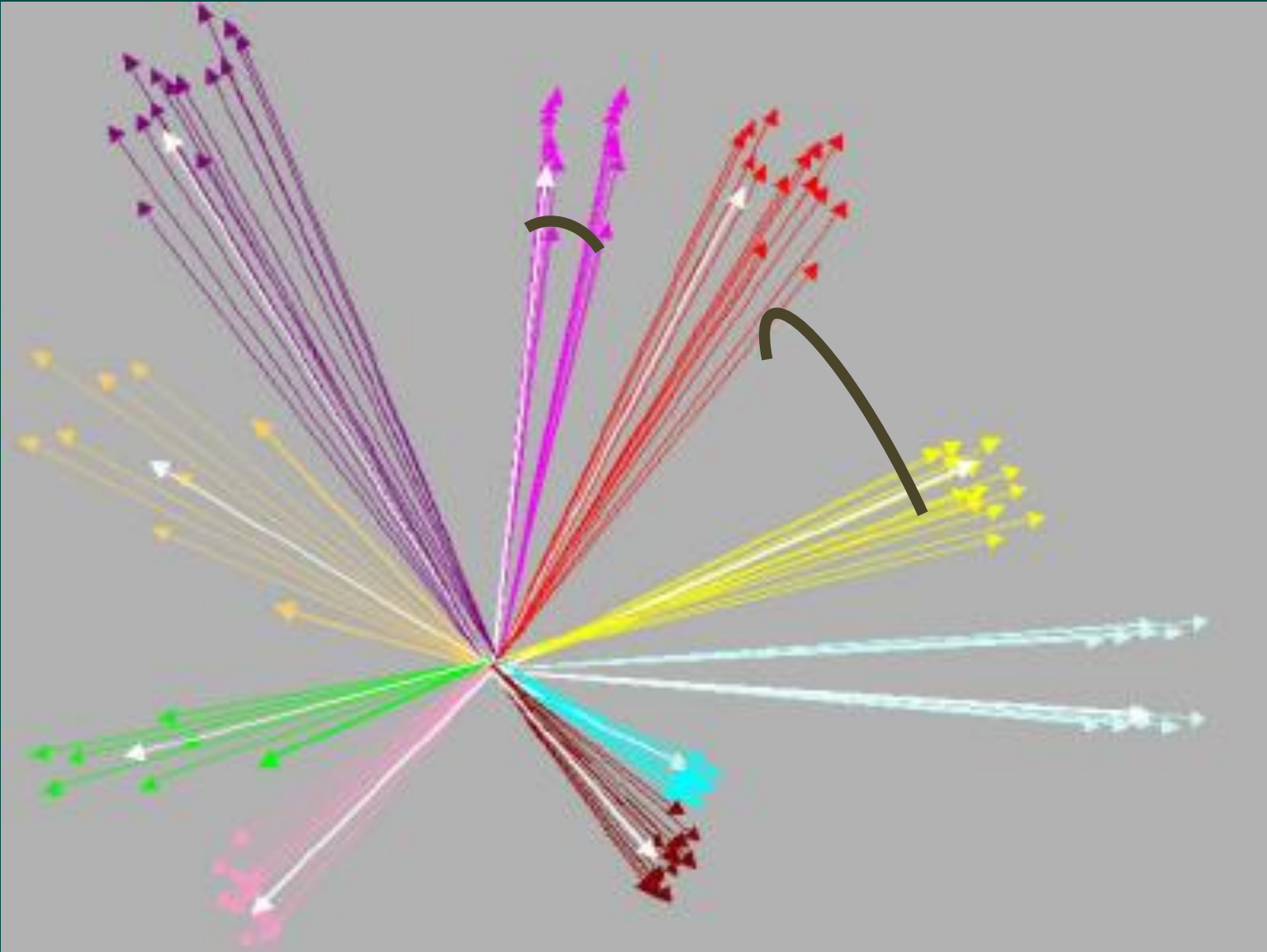
Statistische correlatie

!= causaal verband



!Context Matters!

<https://www.tylervigen.com/spurious-correlations>





Enough

with the **tech** stuff

and on with the **law**

Voor het gemak

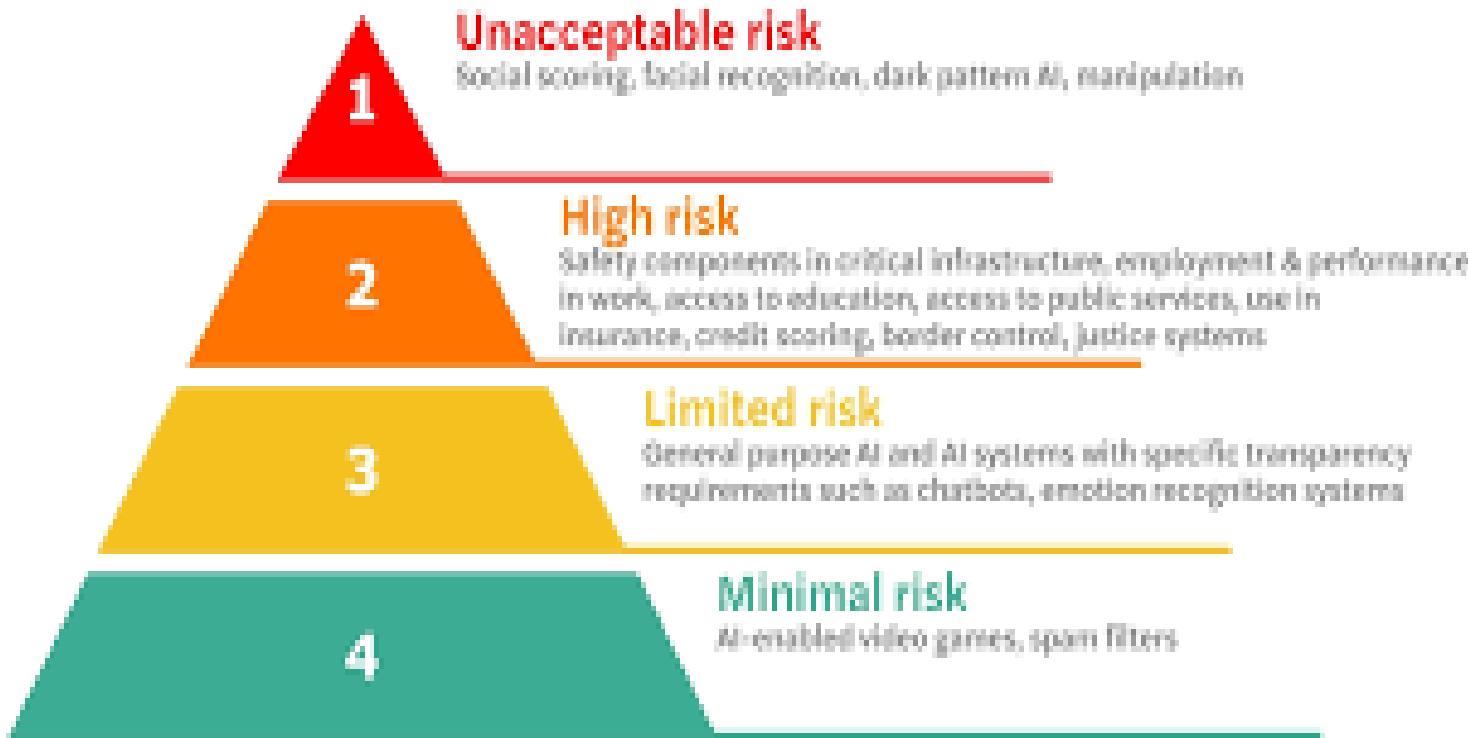
Neem aan dat AI in law enforcement

High Risk

is

(AI Act, article 6(2) jo.

Annex II onder 6 a tot en met e)



AI act obligations

There are **MANY**

Recital 46

High-risk AI systems should **only be** placed on the Union market, put into service or **used** if they **comply** with certain **mandatory requirements**.

AI act obligations

There are many, which you **ALL** have to follow

AI Act – preambule (10)

The hazards of AI systems [...] concern different aspects [...] therefore the **requirements of this Regulation** would **complement** the **existing body** of the **Union** harmonisation **legislation**

Recital 64

The hazards of AI systems [...] concern different aspects [...] therefore the **requirements of this Regulation** would **complement** the **existing body** of the **Union** harmonisation **legislation**.

AI act obligations

There are many, **SOO MANY**

Regulation (EC) No 765/2008 and Regulation (EU) 2019/1020 (product safety) en nog 20 andere regulations

GDPR / AVG

LED / WPG

AI-act annex III, IV, V, VI, VII, VIII, IX

en nog veeeeel meer

AI act obligations

Annex IV

ANNEX IV

Technical documentation referred to in Article 11(1)

The technical documentation referred to in Article 11(1) shall contain at least the following information, as applicable to the relevant AI system:

1. A general description of the AI system including:
 - (a) its intended purpose, the name of the provider and the version of the system reflecting its relation to previous versions;
 - (b) how the AI system interacts with, or can be used to interact with, hardware or software, including with other AI systems, that are not part of the AI system itself, where applicable;
 - (c) the versions of relevant software or firmware, and any requirements related to version updates;
 - (d) the description of all the forms in which the AI system is placed on the market or put into service, such as software packages embedded into hardware, downloads, or APIs;
 - (e) the description of the hardware on which the AI system is intended to run;
 - (f) where the AI system is a component of products, photographs or illustrations showing external features, the marking and internal layout of those products;
 - (g) a basic description of the user-interface provided to the deployer;
 - (h) instructions for use for the deployer, and a basic description of the user-interface provided to the deployer, where applicable;

AI act obligations

Annex IV

2. A detailed description of the elements of the AI system and of the process for its development, including:
 - (a) the methods and steps performed for the development of the AI system, including, where relevant, recourse to pre-trained systems or tools provided by third parties and how those were used, integrated or modified by the provider;
 - (b) the design specifications of the system, namely the general logic of the AI system and of the algorithms; the key design choices including the rationale and assumptions made, including with regard to persons or groups of persons in respect of who, the system is intended to be used; the main classification choices; what the system is designed to optimise for, and the relevance of the different parameters; the description of the expected output and output quality of the system; the decisions about any possible trade-off made regarding the technical solutions adopted to comply with the requirements set out in Chapter III, Section 2;
 - (c) the description of the system architecture explaining how software components build on or feed into each other and integrate into the overall processing; the computational resources used to develop, train, test and validate the AI system;
 - (d) where relevant, the data requirements in terms of datasheets describing the training methodologies and techniques and the training data sets used, including a general description of these data sets, information about their provenance, scope and main characteristics; how the data was obtained and selected; labelling procedures (e.g. for supervised learning), data cleaning methodologies (e.g. outliers detection);
 - (e) assessment of the human oversight measures needed in accordance with Article 14, including an assessment of the technical measures needed to facilitate the interpretation of the outputs of AI systems by the deployers, in accordance

Etc.

AI act obligations

Annex V

ANNEX V

EU declaration of conformity

The EU declaration of conformity referred to in Article 47, shall contain all of the following information:

1. AI system name and type and any additional unambiguous reference allowing the identification and traceability of the AI system;
2. The name and address of the provider or, where applicable, of their authorised representative;
3. A statement that the EU declaration of conformity referred to in Article 47 is issued under the sole responsibility of the provider;
4. A statement that the AI system is in conformity with this Regulation and, if applicable, with any other relevant Union law that provides for the issuing of the EU declaration of conformity referred to in Article 47;
5. Where an AI system involves the processing of personal data, a statement that that AI system complies with Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680;
6. References to any relevant harmonised standards used or any other common specification in relation to which conformity is declared;
7. Where applicable, the name and identification number of the notified body, a description of the conformity assessment procedure performed, and identification of the certificate issued;
8. The place and date of issue of the declaration, the name and function of the person who signed it, as well as an indication for, or on behalf of whom, that person signed, a signature.

AI act obligations

Annex VI

ANNEX VI

Conformity assessment procedure based on internal control

1. The conformity assessment procedure based on internal control is the conformity assessment procedure based on points 2, 3 and 4.
 2. The provider verifies that the established quality management system is in compliance with the requirements of Article 17.
 3. The provider examines the information contained in the technical documentation in order to assess the compliance of the AI system with the relevant essential requirements set out in Chapter III, Section 2.
 4. The provider also verifies that the design and development process of the AI system and its post-market monitoring as referred to in Article 72 is consistent with the technical documentation.
-

AI act obligations

Annex VII

ANNEX VII

Conformity based on an assessment of the quality management system and an assessment of the technical documentation

1. Introduction

Conformity based on an assessment of the quality management system and an assessment of the technical documentation is the conformity assessment procedure based on points 2 to 5.

2. Overview

The approved quality management system for the design, development and testing of AI systems pursuant to Article 17 shall be examined in accordance with point 3 and shall be subject to surveillance as specified in point 5. The technical documentation of the AI system shall be examined in accordance with point 4.

3. Quality management system

3.1. The application of the provider shall include:

- (a) the name and address of the provider and, if the application is lodged by an authorised representative, also their name and address;
- (b) the list of AI systems covered under the same quality management system;
- (c) the technical documentation for each AI system covered under the same quality management system;
- (d) the documentation concerning the quality management system which shall cover all the aspects listed under Article 17;
- (e) a description of the procedures in place to ensure that the quality management system remains adequate and effective;

AI act obligations

Annex VIII

ANNEX VIII

Information to be submitted upon the registration of high-risk AI systems in accordance with Article 49

Section A — Information to be submitted by providers of high-risk AI systems in accordance with Article 49(1)

The following information shall be provided and thereafter kept up to date with regard to high-risk AI systems to be registered in accordance with Article 49(1):

1. The name, address and contact details of the provider;
2. Where submission of information is carried out by another person on behalf of the provider, the name, address and contact details of that person;
3. The name, address and contact details of the authorised representative, where applicable;
4. The AI system trade name and any additional unambiguous reference allowing the identification and traceability of the AI system;
5. A description of the intended purpose of the AI system and of the components and functions supported through this AI system;
6. A basic and concise description of the information used by the system (data, inputs) and its operating logic;
7. The status of the AI system (on the market, or in service; no longer placed on the market/in service, recalled);
8. The type, number and expiry date of the certificate issued by the notified body and the name or identification number of that notified body, where applicable;
9. A scanned copy of the certificate referred to in point 8, where applicable;
10. Any Member States in which the AI system has been placed on the market, put into service or made available in the Union;
11. A copy of the EU declaration of conformity referred to in Article 47;
12. Electronic instructions for use; this information shall not be provided for high-risk AI systems in the areas of law enforcement or migration, asylum and border control management referred to in Annex III, points 1, 6 and 7;
13. A URL for additional information (optional).

AI act obligations

Annex VIII

Section B — Information to be submitted by providers of high-risk AI systems in accordance with Article 49(2)

The following information shall be provided and thereafter kept up to date with regard to AI systems to be registered in accordance with Article 49(2):

1. The name, address and contact details of the provider;
2. Where submission of information is carried out by another person on behalf of the provider, the name, address and contact details of that person;
3. The name, address and contact details of the authorised representative, where applicable;
4. The AI system trade name and any additional unambiguous reference allowing the identification and traceability of the AI system;
5. A description of the intended purpose of the AI system;
6. The condition or conditions under Article 6(3) based on which the AI system is considered to be not-high-risk;
7. A short summary of the grounds on which the AI system is considered to be not-high-risk in application of the procedure under Article 6(3);
8. The status of the AI system (on the market, or in service; no longer placed on the market/in service, recalled);
9. Any Member States in which the AI system has been placed on the market, put into service or made available in the Union.

AI act obligations

Annex VIII

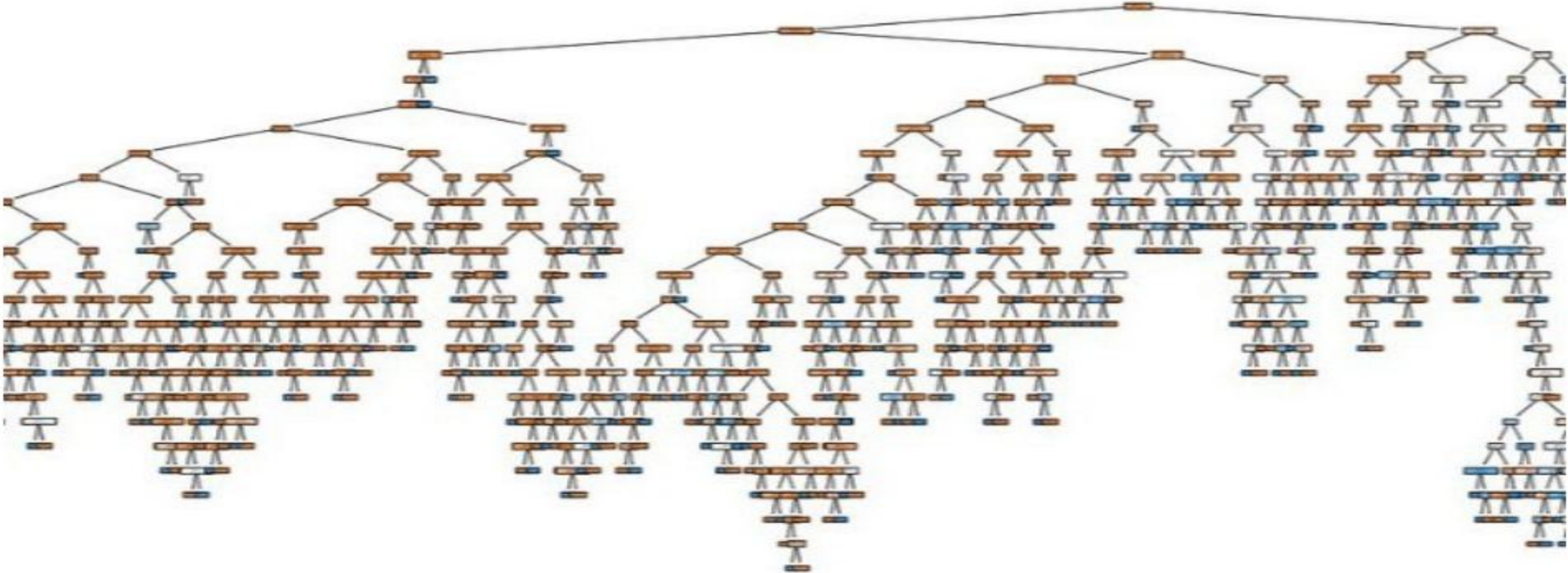
Section C — Information to be submitted by deployers of high-risk AI systems in accordance with Article 49(3)

The following information shall be provided and thereafter kept up to date with regard to high-risk AI systems to be registered in accordance with Article 49(3):

1. The name, address and contact details of the deployer;
2. The name, address and contact details of the person submitting information on behalf of the deployer;
3. The URL of the entry of the AI system in the EU database by its provider;
4. A summary of the findings of the fundamental rights impact assessment conducted in accordance with Article 27;
5. A summary of the data protection impact assessment carried out in accordance with Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680 as specified in Article 26(8) of this Regulation, where applicable.

Uitwerkingen van de verplichtingen die de AI act oplegt (schematisch)

AI act, **here**

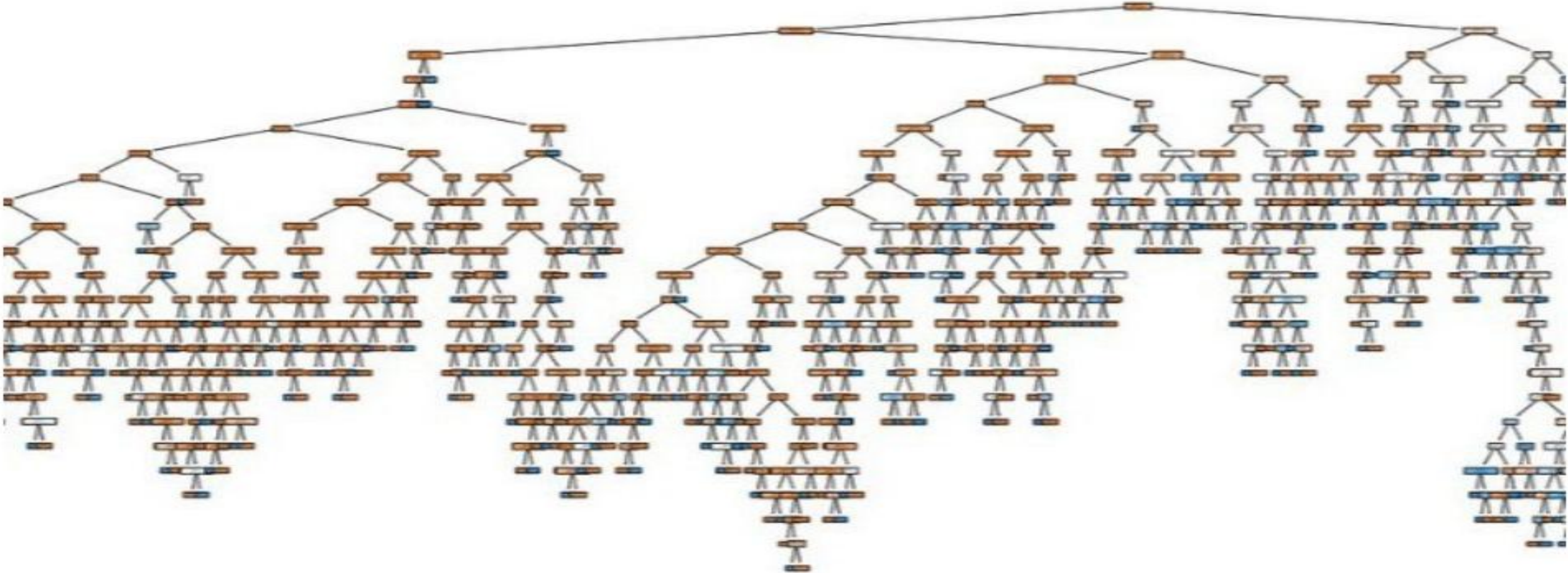


The Users



Uitwerkingen van de verplichtingen die de AI act oplegt (schematisch)

AI act, **here**



The Users

The Users

Begrijpen de gebruikers eigenlijk de tool die ze gebruiken?

Hebben de gebruikers **fatsoenlijke training** ontvangen voor AI gebruik; niet alleen de technische cursus, maar ook de bredere scope.

Volgen de gebruikers ook *daadwerkelijk* **de regels** die met het gebruik worden meegegeven?

The Users

AI act – article 4

Providers and deployers of AI systems **shall** take measures to ensure, to their best extent, a **sufficient level of AI literacy** of their staff and other persons dealing with the **operation and use** of AI systems on their behalf, **taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in**, and considering the persons or groups of persons on whom the AI systems are to be used.

The Users

Recital 20

... benefits from AI systems while protecting fundamental rights... AI literacy **should equip** providers, deployers and affected persons with the **necessary notions** to make **informed decisions**

AI literacy should provide **all relevant actors** in the value chain with the insights required to **ensure the appropriate compliance** and its **correct enforcement**...

The Users



The Users

Recital 20

... benefits from AI systems while protecting fundamental rights... AI literacy **should equip** providers, deployers and effected persons with the **necessary notions** to make **informed decisions**

AI literacy should provide **all relevant actors** in te value chain with the insights required to **ensure the appopriate compliance** and its **correct enforcement**...

+ **RAG** lijkt het 'significant' beter te doen, dan LLM-AI zonder.

+ **RAG** biedt de mogelijkheid op 'nieuwe' bronnen **BIJ** te voegen bij een 'meer generieke' LLM, bijvoorbeeld data uit een opsporingsonderzoek

+ **RAG** biedt de mogelijkheid om 'disconnected & local' te draaien (terwijl er dus wel data kan worden toegevoegd)
Relevant in high-security omgevingen / opsporingsonderzoeken, waar niet gewenst is dat de 'deployer' meekijkt.

— **RAG** is nooit beter dan de toegevoegd bronnen; opsporingsonderzoeken zijn niet ‘op maat gemaakt’ om een LLM effectief te voeden.

— Nederlands blijf een **low-resource** model. Dat brengt beperkingen met zich mee in zowel de LLM training (statistiek is geen conceptbegrip) als de verwerking van prompt based additions.

— **RAG** wordt beperkt door een maximale promptgrootte. Dat kan betekenen dat moet worden geselecteerd in het toegevoegde materiaal (wat weer een bias oplevert)

— Ook **RAG** voorkomt hallucineren niet; alles zal dus door een mens nauwkeurig moeten worden gecontroleerd, maar doen ze dat ook.

Maakt Retrieval Augmented Generation AI geschikt voor forensische toepassing?

NEE, in ieder geval op dit moment nog niet.

Het is WEL de moeite van verder onderzoek waard, maar
zonder garantie dat het 'goed genoeg' wordt.

PLEASE THINK OF OUR POOR LEGAL BRAINS

Draagt jonge onderzoekers een warm hart toe



Wilt u tijdens ergens vandaag, bijvoorbeeld zo in de pauze,
een moment nemen om Timo's vragenlijst in te vullen?

Enquête AI voor de digitale opsporing

CRIME SCENE DO NOT CROSS



SCAN DE CODE EN DEEL JE ERVARINGEN

[*https://rug.eu.qualtrics.com/jfe/form/SV_1Xlyhe1oCG8Efoa?Q_CHL=qr](https://rug.eu.qualtrics.com/jfe/form/SV_1Xlyhe1oCG8Efoa?Q_CHL=qr)

Doel: Deze enquête heeft als doel om in kaart te brengen wat de behoeften en uitdagingen zijn met betrekking tot AI-technologie binnen de digitale opsporing.

Over de enquête: De enquête bestaat uit 12 vragen. Het invullen neemt ongeveer 10 minuten in beslag en is anoniem. Scan de QR-code of vul de url in.

Resultaat: De resultaten zullen worden gepubliceerd in een wetenschappelijk artikel, om de behoefte van digitaal onderzoekers in kaart te brengen.

Opslag van gegevens: Alle gegevens worden verzameld via de Qualtrics-software op een beveiligde account en server van de Rijksuniversiteit Groningen.

Vragen/opmerkingen: Neem contact op met Timo Meconi, promovendus AI voor de digitale opsporing, via t.n.t.meconi@rug.nl.



university of
groningen

STeP

SECURITY, TECHNOLOGY
& E-PRIVACY
RESEARCH GROUP



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid



Hansken
the open digital forensic platform

Draagt jonge onderzoekers een warm hart toe



Wilt u tijdens ergens vandaag, bijvoorbeeld zo in de pauze,
een moment nemen om Timo's vragenlijst in te vullen?