

GDPR + ML = ???

A Discussion on Applicability of GDPR to Advances in ML

Harshvardhan J. Pandit

harshvardhan.pandit@dcu.ie

CSC1117 Machine Learning DCU

Slides available at: <https://harshp.com/research/presentations>

Harsh(vardhan J. Pandit)

An Introduction

- Assistant Professor - ADAPT Centre - Dublin City University
- Postdoctoral Fellowship: knowledge graph for DPIA / GDPR
- PhD in Computer Science (2020) - Representation of activities involving personal data and consent for GDPR information
- Chair of W3C Community Groups: Data Privacy Vocabularies and Controls Community Group (DPVCG) and Consent (ConsentCG)
- Nominated Technical Expert by European Data Protection Board (EDPB)
- Member of National Standards Authority of Ireland (NSAI) committees for Cybersecurity/ Privacy and AI at EU and ISO forums

GDPR¹

World-Changing EU law that regulates Processing of Personal Data

1. What is meant by Personal Data ?
2. What is meant by Processing ?
3. How is data is being processed? (what/how/where...)
4. Who is involved? (whose data, processed by whom)
5. How to check processing is following the rules of GDPR?

[1] <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

Personal Data

Some “definitions” from across the globe

‘personal data’ means **any information relating to an identified or identifiable natural person** (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

GDPR Art.4(1)

any information that (a) **can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal**

ISO 29100:2011

“Personal information” means information that **identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly**, with a particular consumer or household.

CCPA 1798.140 (o)(1)



Personal Data

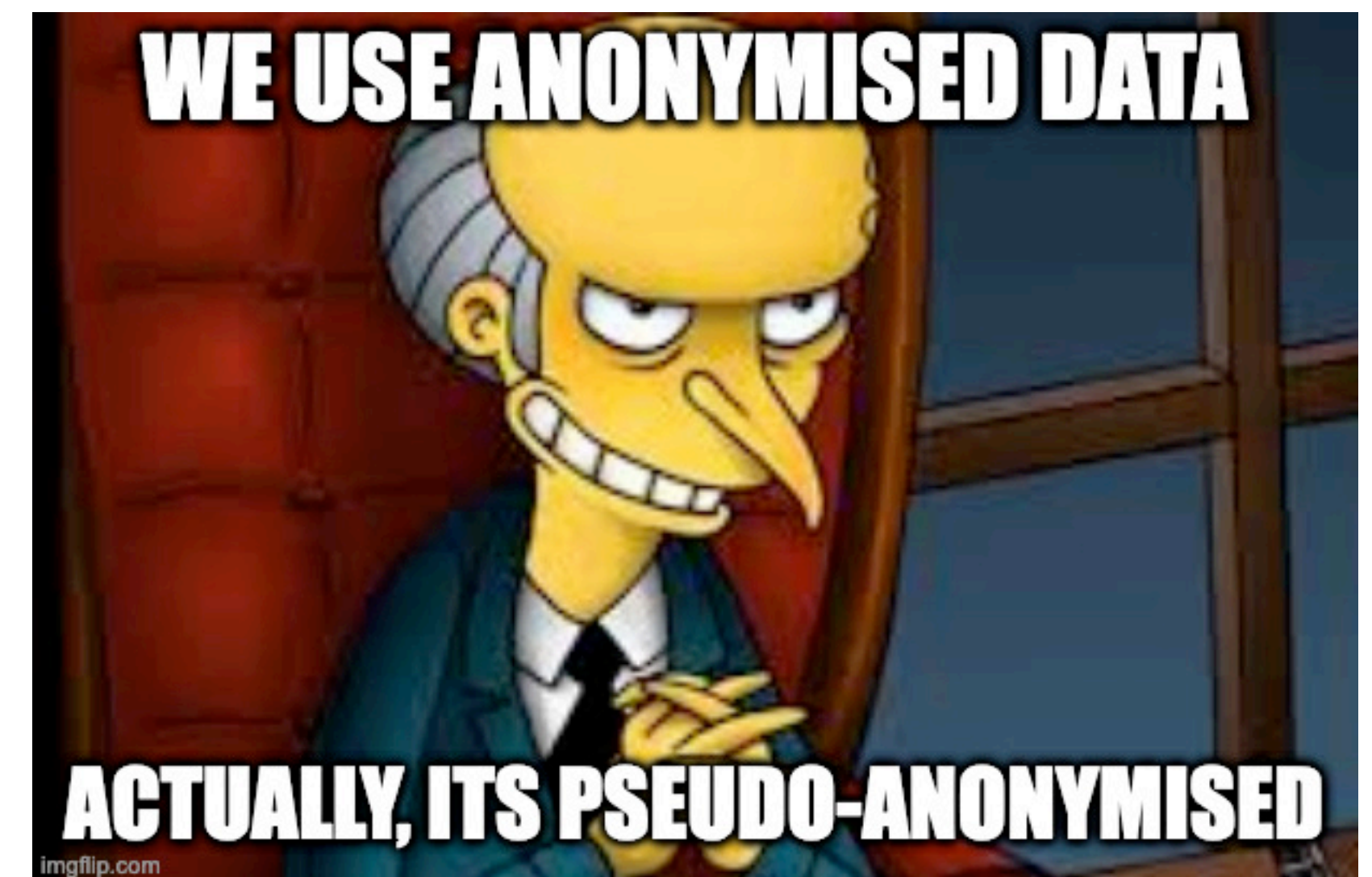
Identifiers, and Identifiability

1. Identifiers: Harsh (name), xyz@email.com (email)
2. Non-identifiers: Black (hair), Brown (eyes), 1.66m (height), etc.
3. For a room full of people, combine non-identifier to uniquely identify a person (me) — thus creating an identifier !!!
4. Useful technique for **fingerprinting**, **profiling**, **tracking**

Q: When is Personal Data not 'Personal' anymore?

Ans: When it is (completely) anonymised

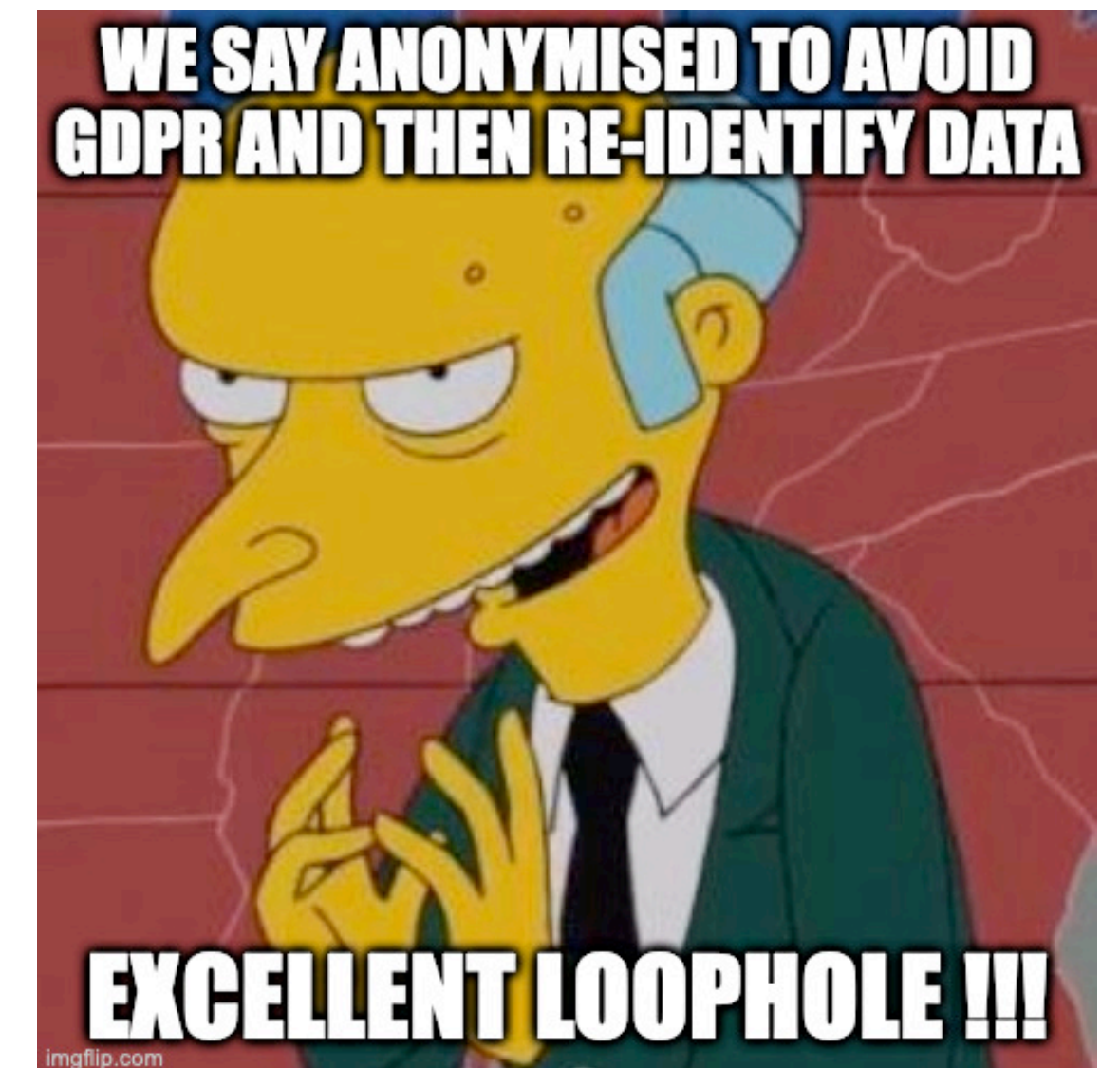
- Anonymisation is the removal of (some) 'identifying' attributes from data
- Merely using "**anonymisation**" does not produce anonymised data
- It produces '**pseudo-anonmised**' data, which is still personal data
- 'Completely anonymised' if it is **not identifiable**
- E.g.
 - Your exact location = personal data
 - approx. house = still personal data
 - approx. area = still personal data, but less
 - City = still personal data, but lesser
 - Country = anonymised, kind of



Q: When is Anonymised Data not Anonymised?

Ans: When it is possible to 're-identify' using any (practical) means possible

- Data is anonymised, i.e. all identifiers like names and emails are removed
- But using a 'combination' of remaining data points, a person is still identified
- Since **re-identification** is possible, its not '**fully anonymised**'
- 'Exploits'
 - Aggregated location — person's routines are unique
 - Voting and voters data
 - Fingerprinting - browser configurations, preferences
- GDPR applies to all the above since it is 'personal data'



Personal Data

ISO 29184:2020

From Data Subject

Given

Email in forms

Observed

Location via IP

Inferred

Interests via website history

Other Sources

Third-Party

RTB / Online Advertising

Public

ClearviewAI

Personal Data: Sensitive, and Special

Special category personal data is to GDPR what Ferrero Rocher is to chocolates

Sensitive:

- data that merits additional security
- older term used widely

Special:

- requires additional/specific legal permissions
- newer term introduced in GDPR



GDPR Prohibits

**Processing of Special Categories of Personal Data
and**

Requires additional obligations via legal basis in Article. 9

racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited

GDPR Article 4(11)

‘processing’ means **any operation or set of operations which is performed on personal data** or on sets of personal data, whether or not by **automated means**, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

Notable alignment with ‘common’ terms used in documents, interfaces, etc.

collect, store, use, share, delete

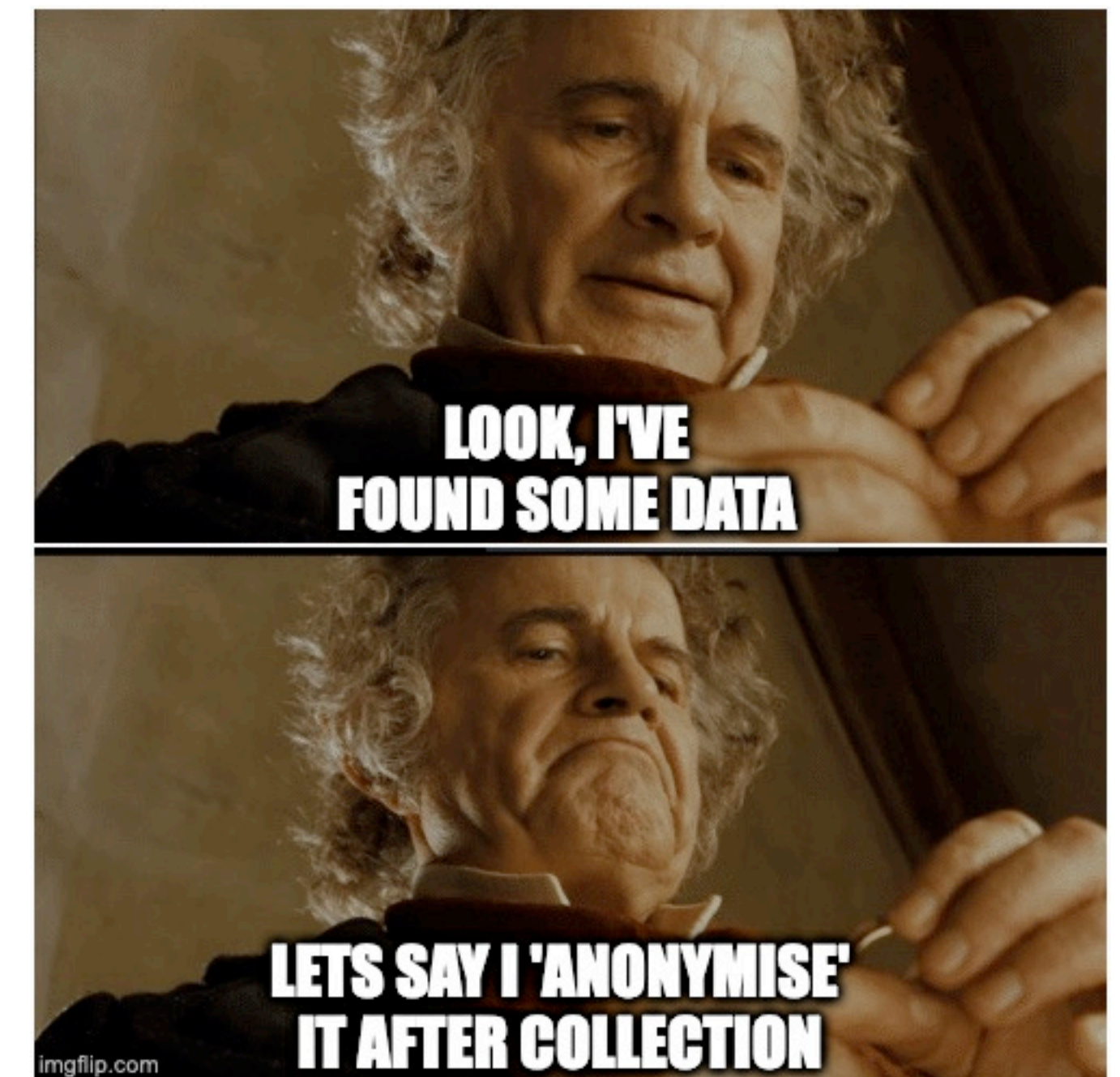
Systematic Monitoring Evaluation & Scoring Matching & Combining Automated Decision Making Innovative Use of New Technologies

GDPR Article.35 Data Protection Impact Assessments

GDPR applies before Processing starts

Common Misinterpretations

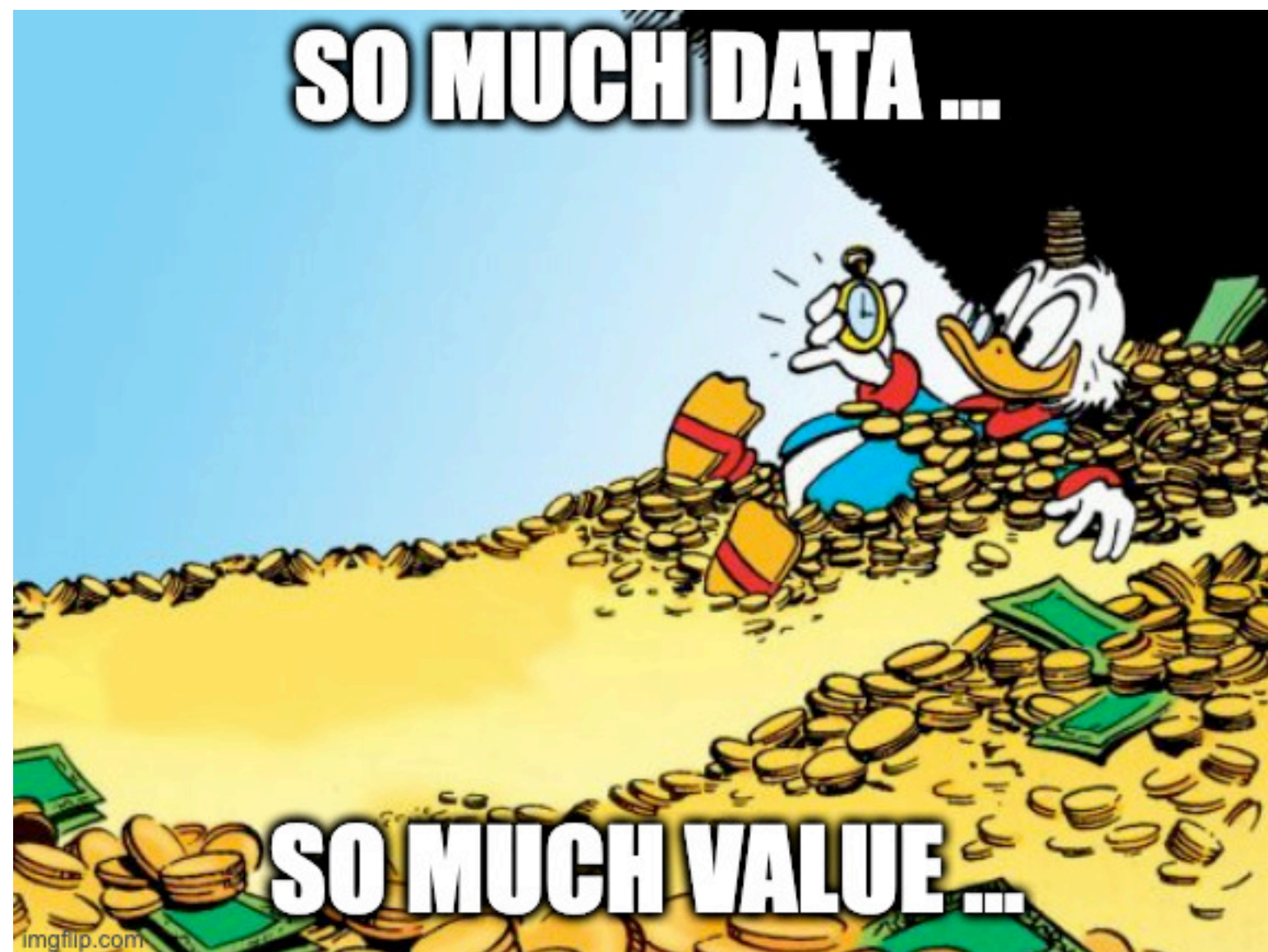
- Data collected but 'anonymised' is not subject to GDPR
- If data isn't shared, nothing needs to be declared
- Collecting anonymised data and attaching an identifier to it
- Hiding things that require transparency and permission
 - Scale and scope of processing
 - Involvement of special categories
 - Involvement of any automated decision making
 - Creating, sharing, using - profiling



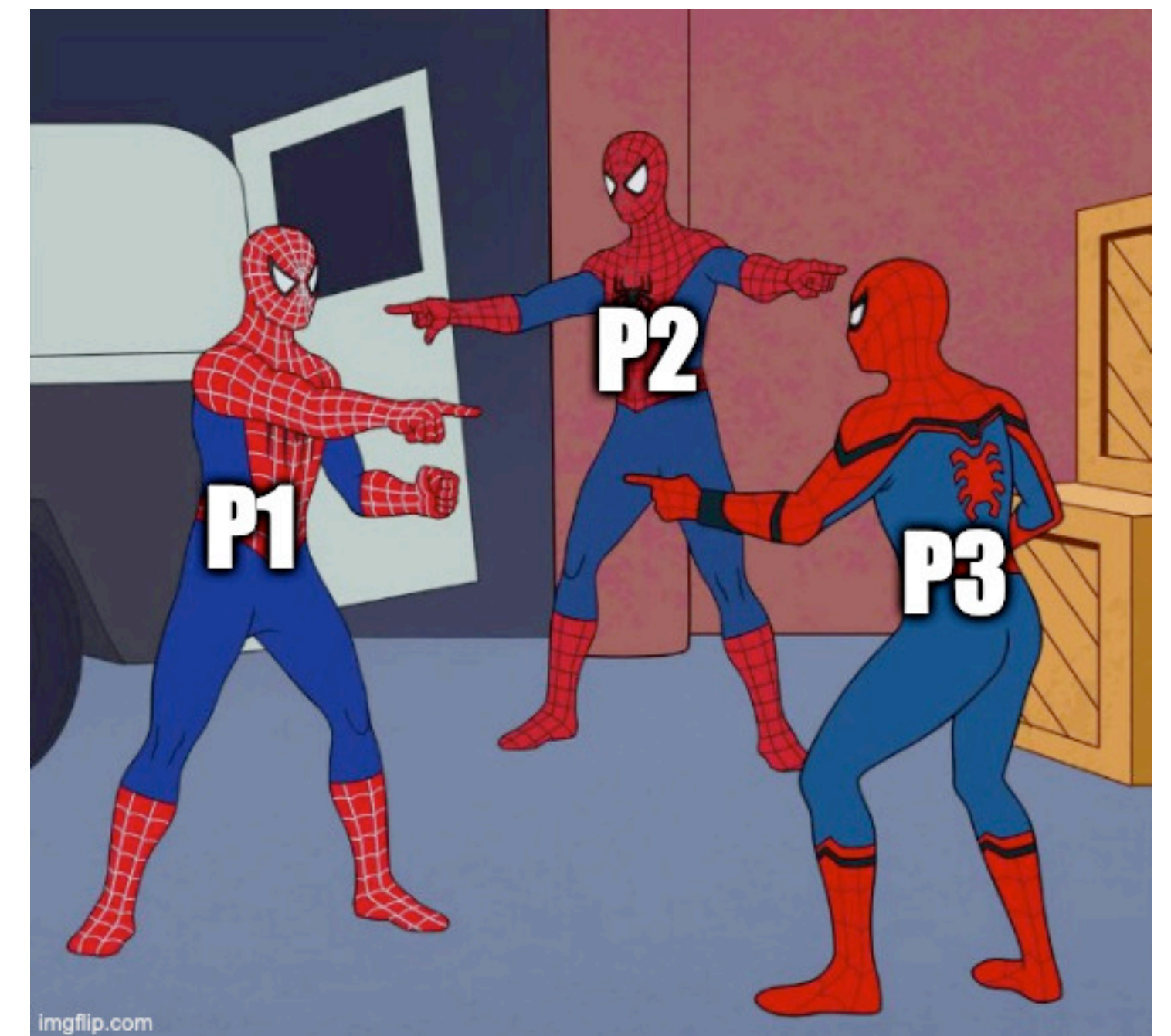
All Processing in GDPR ***must*** be towards a Goal

Implied when a 'Purpose' is necessary as per Article.5

Every Processing ***must*** have a Purpose



Purposes must be separate from other matter, including other purposes



Purposes must be ***specific*** and ***unambiguous***

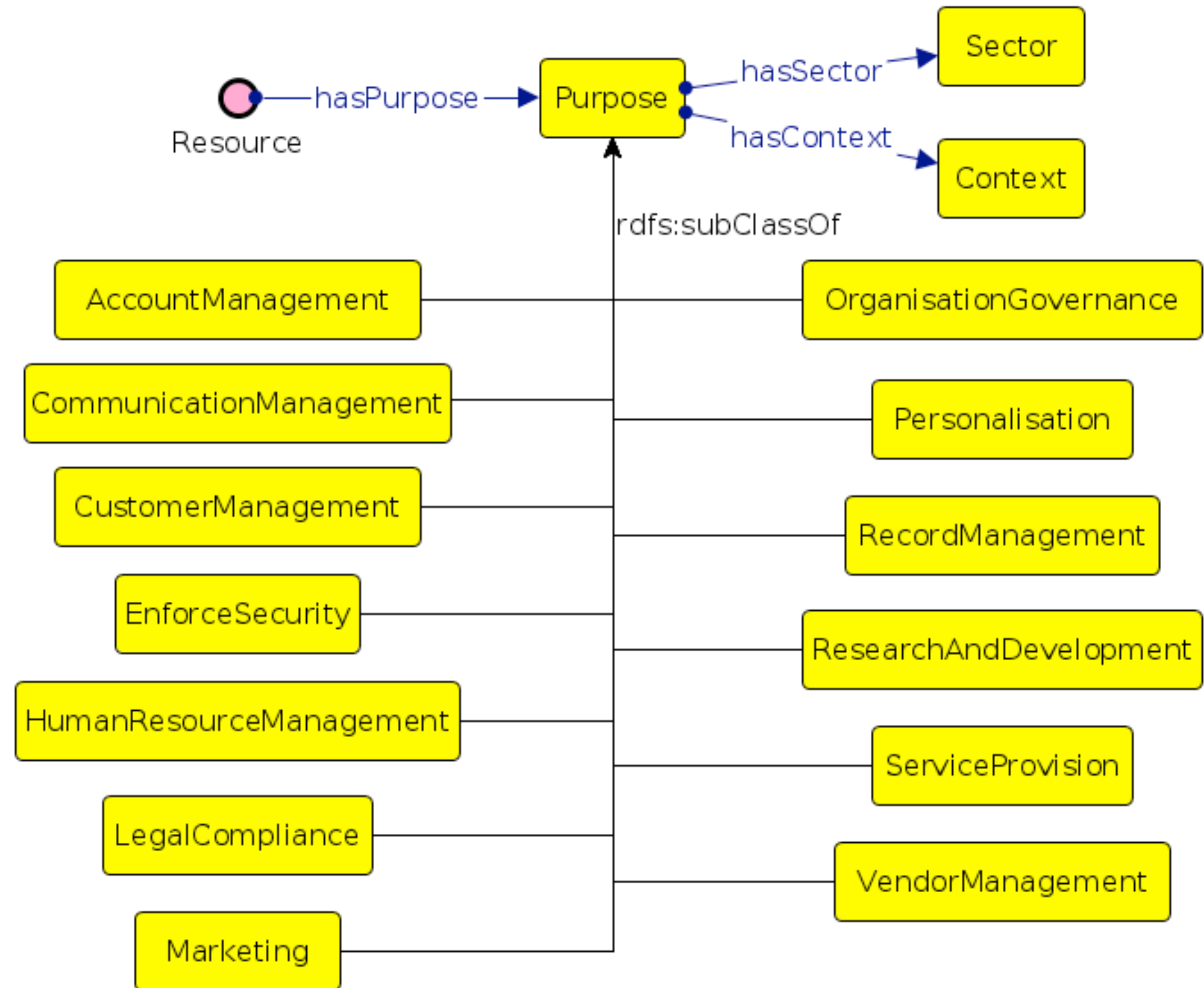
Purposes are intended to be human-readable and human-comprehensible

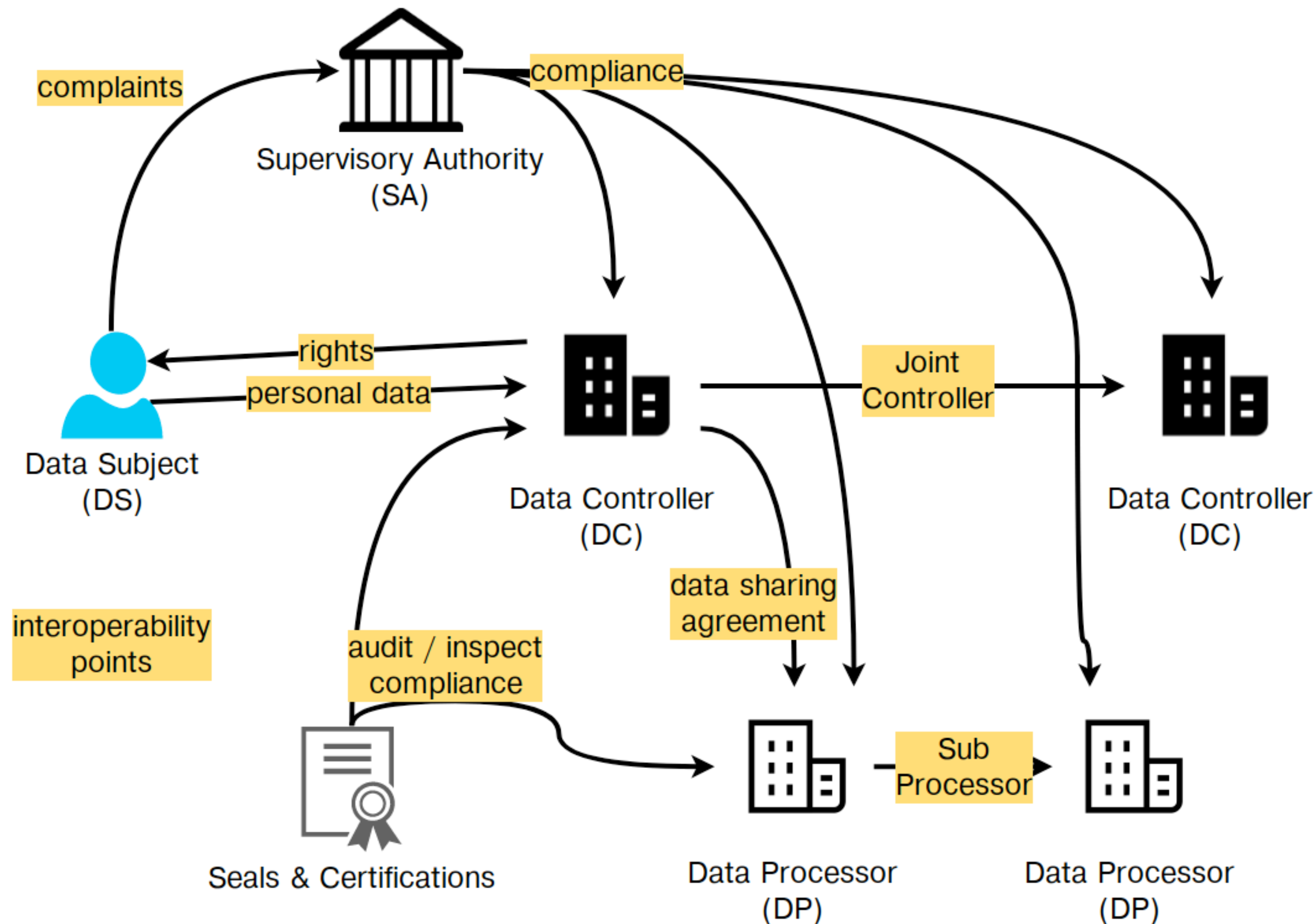
Purposes should not be broad and abstract

Purposes should be specific and contextual to their use-case

Purposes can be grouped or categorised, but not replaced, e.g. with Marketing for 'Sending new product emails'

Purposes don't have to necessarily benefit the data subject e.g. service optimisation





Data Controllers are responsible for deciding the 'purpose'

Data Controllers may not even 'touch' the data they 'control'

Data Controllers can 'team up' to become Joint (Data) Controllers

Processors only act on 'orders' given (explicitly) by Controllers

Processors can appoint other (sub-)Processors, still governed by instructions from Controllers

Processors deciding/ processing on their own become Controllers

Data Protection Authorities (DPA) are empowered by GDPR to enforce its obligations on all entities

GDPR's principles providing a framework for 'responsibility'

Principles (Article.5)

lawfulness, fairness and transparency
purpose limitation
data minimisation
accuracy
storage limitation
integrity and confidentiality
accountability

Consent (Article.7)

Informed
Freely Given
Unambiguous
Balance of Power(s)
Right to Withdraw
Explicit Consent (e.g. for Article.9)

A12-A22 Rights

Transparency (A.12)
Notice (A.13, A.14) ;
Object to Processing
Rectification of Data
Erasure (Right to be Forgotten)
Restriction of Processing
Right of Access
Data Portability

A77 Right to complaint

Any Data Subject can
complaint to their Supervisory
Authority (DPA)
If DPA is in a different country
than the company, then the
DPA will 'lease' and 'co-operate'
with the DPA of that country

“AI” is just another technology...

How does GDPR apply? —> if personal data is involved

When can personal data be involved? Input/output?

Problems? Accuracy, transparency, principles, *consent*

When can it be misused? Jobs? Assessments? Decisions?

Overview of Personalisation Issues

Key takeaways

- What data is ‘used’ ??? —> Transparency
- What data is ‘needed’? What is ‘necessary’? —> Data Minimisation
- What are the sources of ‘data’ ? —> Transparency
- Is any data ‘sensitive’ ? Is it ‘special’ ? —> Ethical Concerns
- Is data (input/output) ‘accurate’ —> Accountability
- Is the output configurable ? —> Privacy by Design / Default
- Understand distinctions between *Privacy* vs *Security* vs *Identifiability* vs *Control*

What the current state of the art?

- (1) when and how an AI model can be considered as ‘anonymous’;
- (2) how controllers can demonstrate the appropriateness of legitimate interest as a legal basis in the development and
- (3) deployment phases; and
- (4) what are the consequences of the unlawful processing of personal data in the development phase of an AI model on the subsequent processing or operation of the AI model.

EDPB opinion on AI models: GDPR principles support responsible AI

 18 December 2024 

Brussels, 18 December - The European Data Protection Board (EDPB) has adopted an [opinion* on the use of personal data for the development and deployment of AI models](#). This opinion looks at 1) when and how AI models can be considered anonymous, 2) whether and how legitimate interest can be used as a legal basis for developing or using AI models, and 3) what happens if an AI model is developed using personal data that was processed unlawfully. It also considers the use of first and third party data.



The opinion was requested by the Irish Data Protection Authority (DPA) with a view to seeking Europe-wide regulatory harmonisation. To gather input for this opinion, which deals with fast-moving technologies that have an important impact on society, the EDPB organised a stakeholders’ event and had an exchange with the EU AI Office.

EDPB Chair Talus said: “AI technologies may bring many opportunities and benefits to different industries and areas of life. We need to ensure these innovations are done ethically, safely, and in a way that benefits everyone. The EDPB wants to support responsible AI innovation by ensuring personal data are protected and in full respect of the General Data Protection Regulation (GDPR).”

https://www.edpb.europa.eu/news/news/2024/edpb-opinion-ai-models-gdpr-principles-support-responsible-ai_en

Salient Points

- claims of an AI model's anonymity should be assessed on a case-by-case basis. For an AI model to be considered anonymous, both (1) the likelihood of direct (including probabilistic) extraction of personal data regarding individuals whose personal data were used to develop the model and (2) the likelihood of obtaining, intentionally or not, such personal data from queries, should be insignificant, taking into account 'all the means reasonably likely to be used' by the controller...
- analysing the necessity of the processing for the purposes of the legitimate interest(s) pursued (also referred to as "necessity test"); and (3) assessing that the legitimate interest(s) is (are) not overridden by the interests or fundamental rights and freedoms of the data subjects (also referred to as "balancing test").



The EU AI Act

New Rules for

- **AI Systems**
- **GPAI Models** [General Purpose AI]

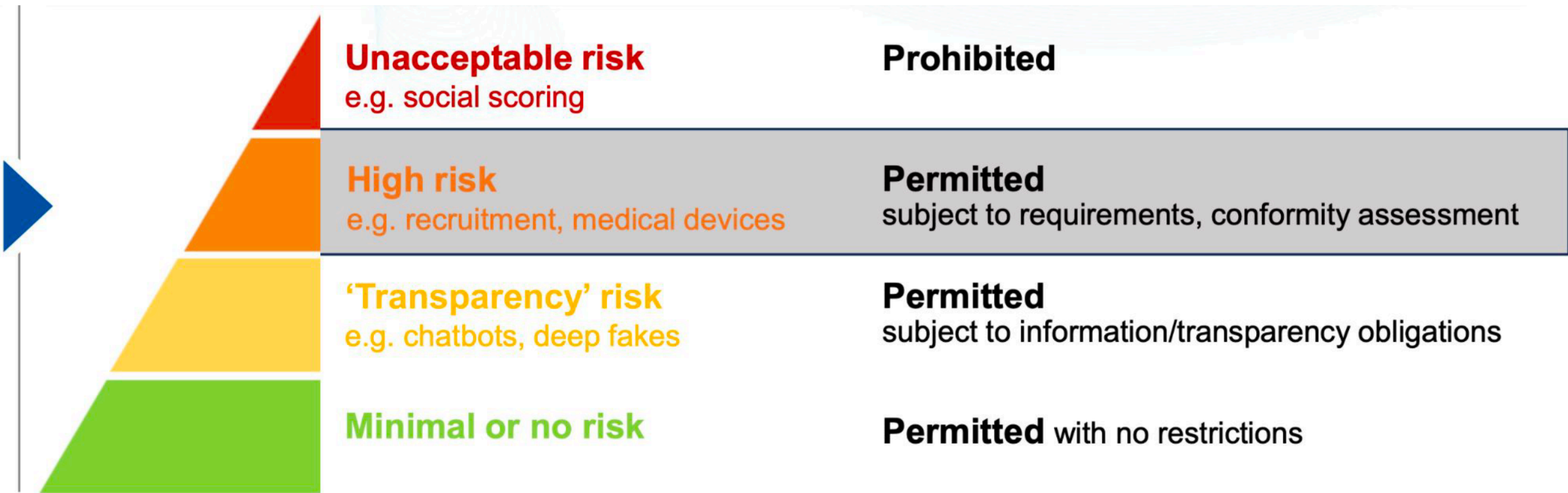
Promotes human-centric & trustworthy AI

Protects against **harmful effects** of AI on

- **Health**
- **Safety**
- **Fundamental Rights**



AI Systems Risk-Based Classification



From the EU AI Office webinar on risk management in the AI Act and related standards, 30 May 2024

AI Act —> GenAI

- AI Act specifically address Generative AI or GenAI
- It requires transparency for specific models
- Requires transparency
- Requires risk assessment
- Requires clarification on what the model can / cannot do, or is intended to do

Existing approaches

Datasheets and Model cards

- Good to start with - they tell you what info to consider
- BUT - they are not complete
- - not structured
- - not formally defined
- - not “usable” to audit
- - often incomplete
- - unclear legality

<https://arxiv.org/pdf/1803.09010>

Datasheets for Datasets

TIMNIT GEBRU, Black in AI
JAMIE MORGENSTERN, University of Washington
BRIANA VECCHIONE, Cornell University
JENNIFER WORTMAN VAUGHAN, Microsoft Research
HANNA WALLACH, Microsoft Research
HAL DAUMÉ III, Microsoft Research; University of Maryland
KATE CRAWFORD, Microsoft Research

<https://arxiv.org/pdf/1810.03993>

Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru
{mmitchellai,simonewu,andrewzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com
deborah.raji@mail.utoronto.ca

AI Cards (II)



1. General Information about the system

1. General Information

Version
Modality
AI Technique(s)
Provider(s)
Developer(s)

2. Intended use of the AI system using 6 concepts

2. Intended Use

Domain
Purpose
Capability
Deployer
AI Subject
Locality of use

3. Information about the incorporating components

3. Key Components

Input (from user) ↓

Component #1
ID D

Component #2
ID M

Component #3
ID D

Component #4
ID S

Component #5
ID GPAI

Output (to user) ↓

Dataset
Model
System
General Purpose

Hardware Platform

4. Information about processing of data (including info about legal basis and source of data)

4. Data Processing

Processing	Legal basis	Data	Data Source
Processing #1		Data #1	
		Data #N	
Processing #N			

5. Involvement of humans and level of automation

5. Human Involvement

Level of Automation

Involved Entity	Intended	Active	Informed	Control over output
AI Subject#1	✓	✗	✓	
AI Subject#N	✗	✓	✗	
End-user#1	✗	✗	✓	
End-user#N	✗	✓	✗	

AI Cards (III)

6. High-level summary of risk management

6. Risk Profile									
Impact on ↓	Risk			Measures					
	Likeli.	Severity	Residual	Org.	Tech.	Monit.	Secur.	Transp.	Log.
Health & Safety	High	V.High	Med.	✗	✓	✗	✓	✗	✓
Fundamental Rights	V.High	High	High	✓	✗	✓	✗	✓	✗
Society	Med.	Med.	Low	✗	✓	✗	✓	✗	✓
Environment	Low	Low	Low	✓	✗	✓	✗	✓	✗

7. Illustration of key qualities of the AI system



8. List of pre-determined changes

8. Pre-determined Changes		
Changed Entity	Change Frequency	Purpose of Change
Data		
Model		
...		

8.Regulation & Certification information

9. Compliance & Certification	
Regulations	
Standards	
Codes of conduct	

Example: An AI-Based Student Proctoring System

Human-readable
description

Proctify is intended to be used in the education domain, for detecting suspicious behaviour of students during online exams in universities. Facial behaviour analysis and video analysis are used for detecting suspicious behaviour



Machine-readable
specification

```
ex:proctify
  airo:isAppliedWithinDomain ex:education ;
  airo:hasPurpose ex:detecting_suspicious_bahviour_during_online_exam
  airo:hasCapability ex:facial_behaviour_analysis ;
  airo:hasCapability ex:video_analysis ;
  airo:isUsedBy ex:university ;
  airo:hasAISubject ex:student ;
```

<https://delaramglp.github.io/aicards/example/>



AI Cards: Proctify

<https://raw.githubusercontent.com/DelaramGlp/airo/main/usecase/proctify.ttl>

Card's Version 1.2.3
Card's Date (Issued) 2024-04-23
Card's Language Eng
Card's Publisher AIEduX
Contact Info proctify@aiedux.org



1. General Information

Version: 1.2
Modality: Software
AI Technique(s): ML>>ANN>>Deep learning
Provider(s): AIEduX
Developer(s): AIEduX

2. Intended Use

Domain: Education
Purpose: Detecting suspicious behaviour during online exam
Capability: Facial behaviour analysis, video analysis
Deployer: University
AI Subject: Students
Locality of Use: Educational institution in EU

3. Key Components

Facial video

Facial Analysis Toolkit 3.3.2
tinyurl.com/3wmyxyun

S

Facial info

SusBehavedModel 1.1.2
tinyurl.com/2hnt6cbb

M

Training data

SusBehavedDataset 2.0.1
tinyurl.com/cb4whuw9

D

Suspicious behaviour alarm

Dataset Model System General Purpose

4. Data Processing

Processing	Legal basis	Data	Data Source
Processing of input video	Informed consent	Facial>> Biometrics	User input
Behaviour analysis (ML model)	Informed consent	Facial>> Biometrics	SusBehaved dataset contributors

5. Human Involvement

Level of Automation: Partial automation

Involved Entity	Intended	Active	Informed	Control over output ex-post challenge
Student	✓	✓	✓	
Occupant (of the room)	✗	✗	✗	No opt-out
Instructor	✓	✓	✓	Correct

6. Risk Profile

Impact on	Risk			Measures					
	Likeli.	Severity	Residual	Org.	Tech.	Monit.	Secur.	Transp.	Log.
Health & Safety	Med.	V. High	Low	✓	✓	✓	✗	✓	✗
Fundamental Rights	High	V. High	Low	✓	✓	✓	✓	✓	✓
Society	Low	Med.	Med.	✓	✓	✓	✗	✓	✗
Environment	Low	Low	Low	✓	✗	✗	✗	✗	✗

7. Quality

Accuracy

Explainability

Usability

Functional adaptability

Fairness

Robustness

Cybersecurity

8. Pre-determined Changes

Changed Entity	Frequency	Purpose
Susbehaved model	2 Month	Improve performance
Mitigation measures	2 Week	Mitigate newly identified risks

9. Compliance & Certification

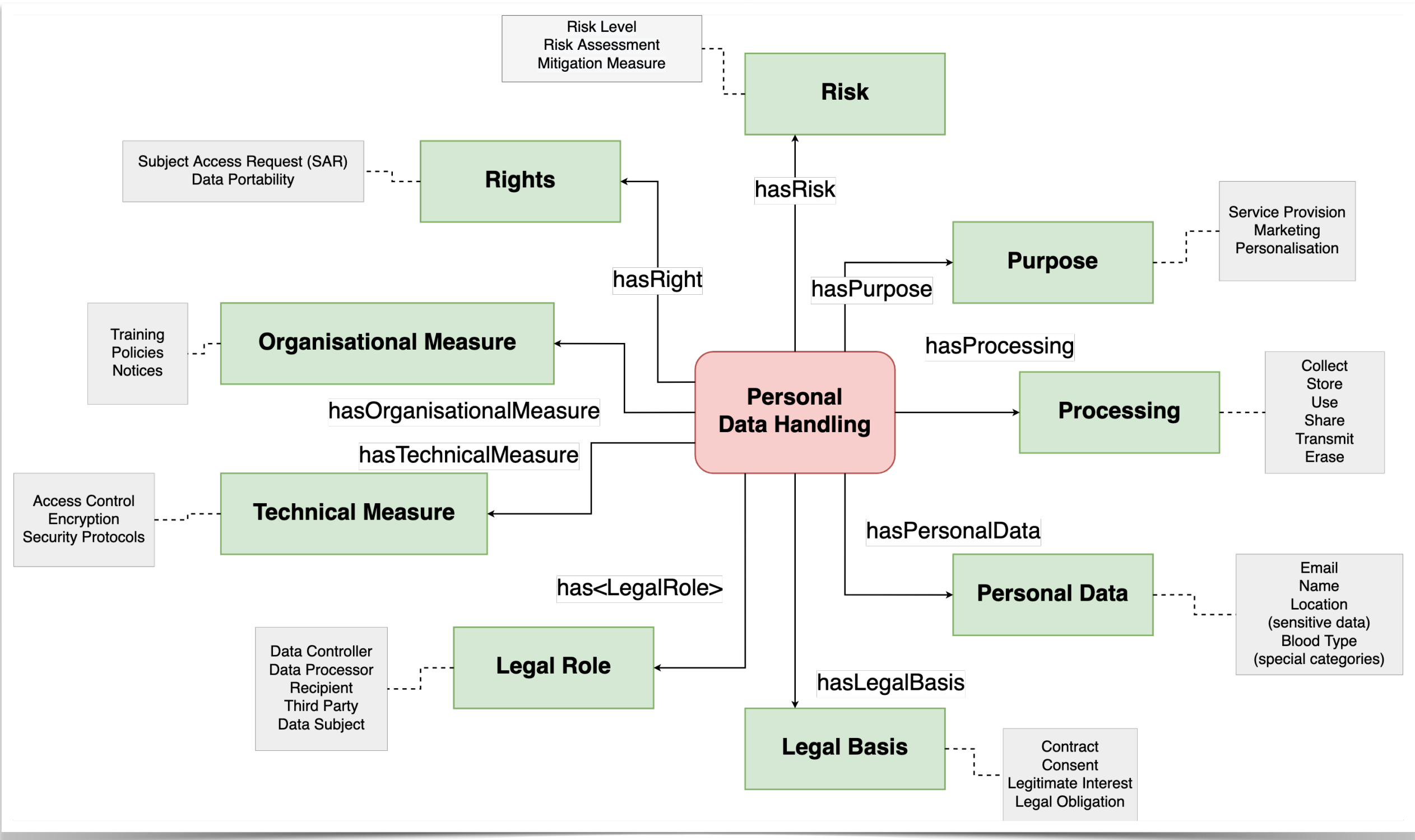
Regulations	[EU, GDPR] [IE, DPA]
Standards	[ISO/IEC 27001:2022]
Codes of conduct	[EU, use of AI and data in teaching and learning for educators]

What am I working on?

Privacy Risks, GDPR, Legal Compliance, Semantics

Machine-Readable Metadata for Automated Approaches

Data Privacy Vocabulary (DPV) <https://w3id.org/dpv>



DPV's taxonomies provide semantic interoperability, which enables new, innovative, smart, and automated solutions

Demonstrated usefulness for important use-cases, e.g. ROPA, consent, compliance checking

We're looking to the future! DGA / ePR / AI-Act / Data Spaces

The Data Privacy Vocabulary (DPV) reflects ~5 years of efforts in creating an open resource providing concepts related to personal data processing, privacy, data protection, and GDPR

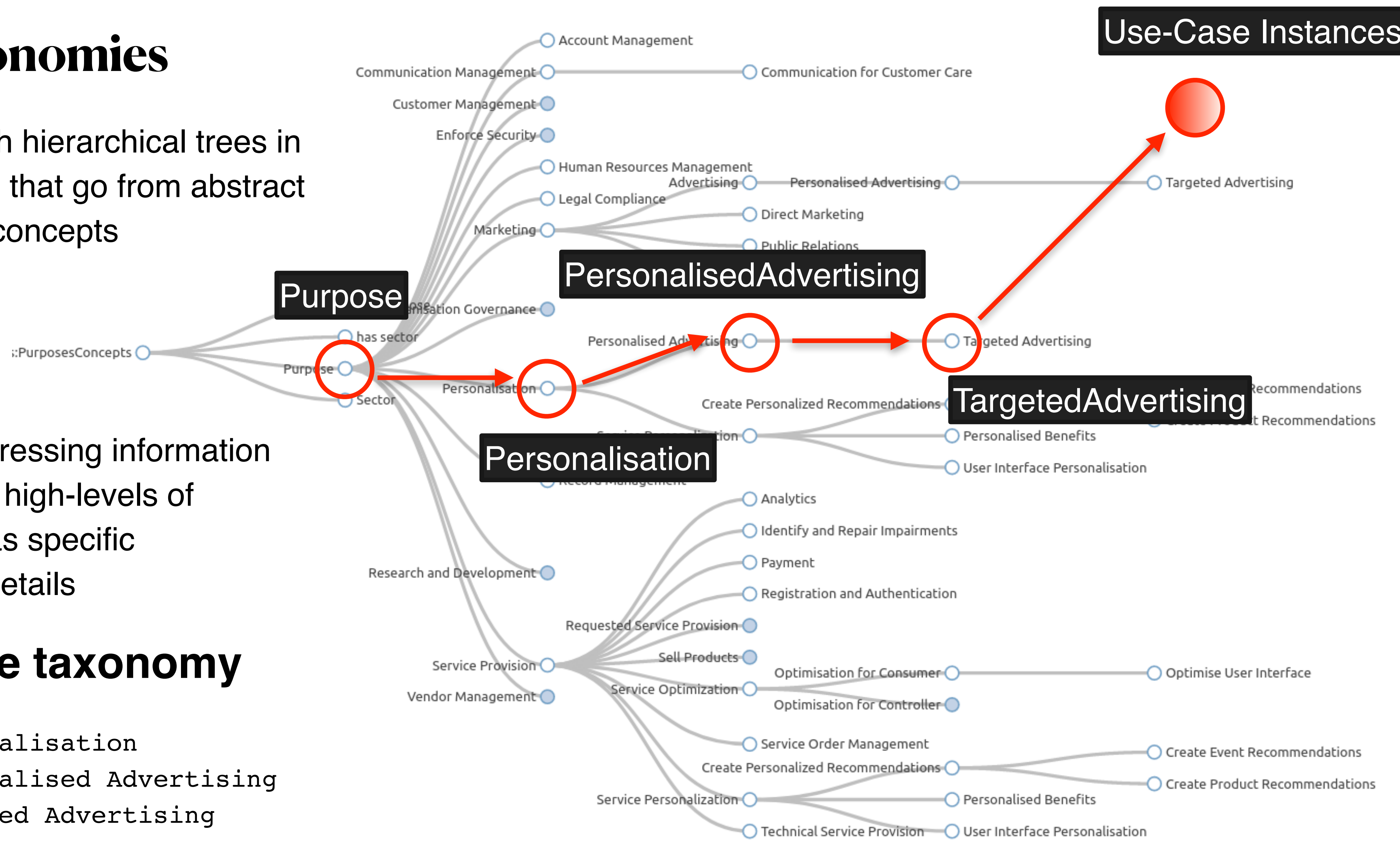
DPV Taxonomies

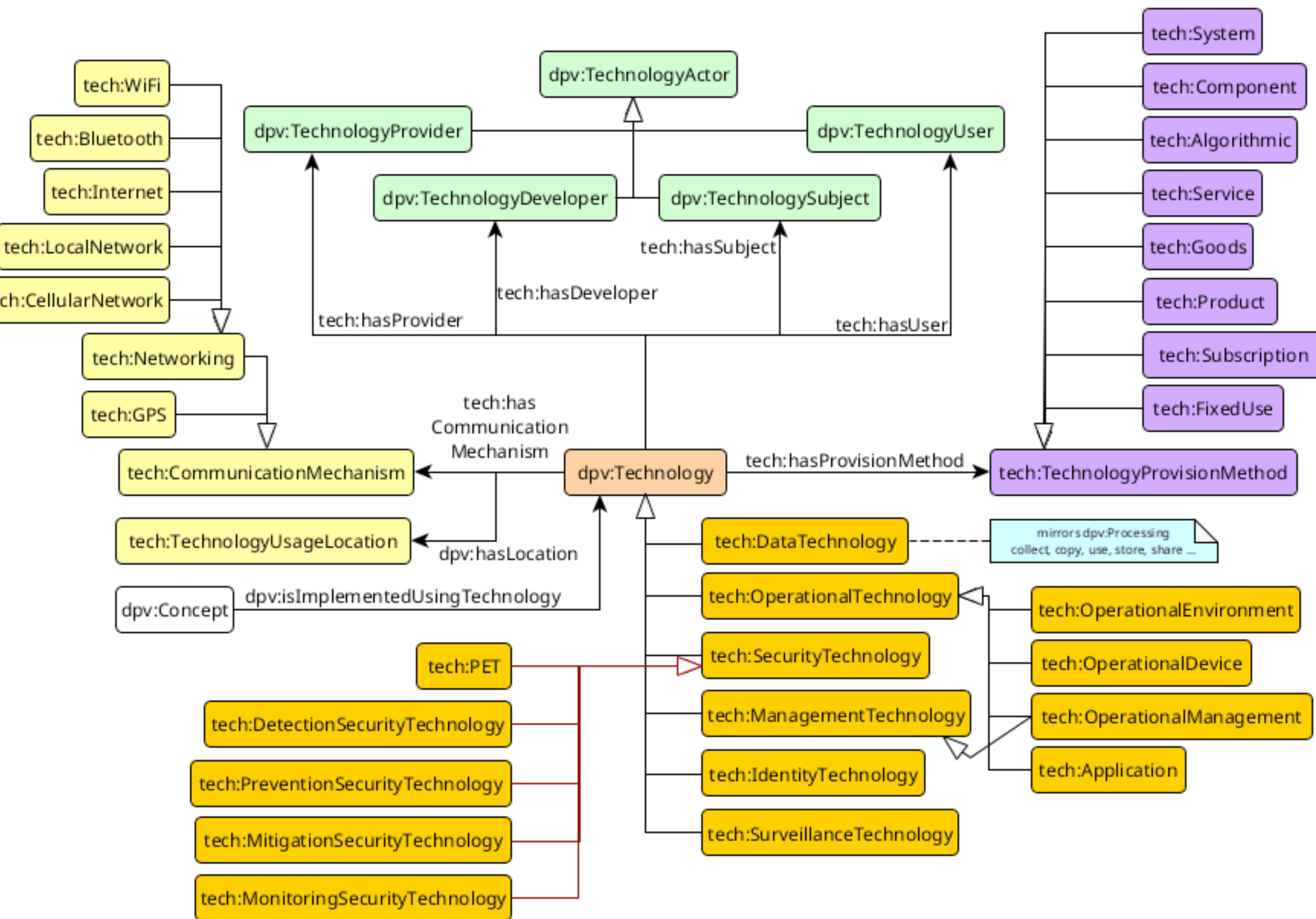
DPV provides rich hierarchical trees in top-down fashion that go from abstract to more specific concepts

This enables expressing information and rules at both high-levels of abstraction and as specific implementation details

E.g. Purpose taxonomy

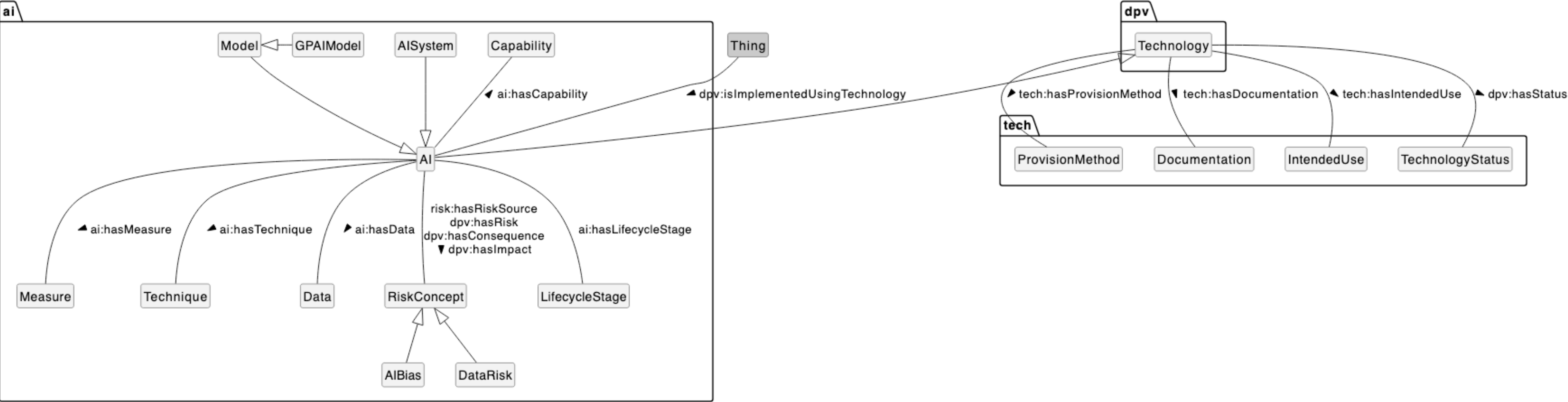
- Purpose → Personalisation
- Personalised Advertising
- Targeted Advertising

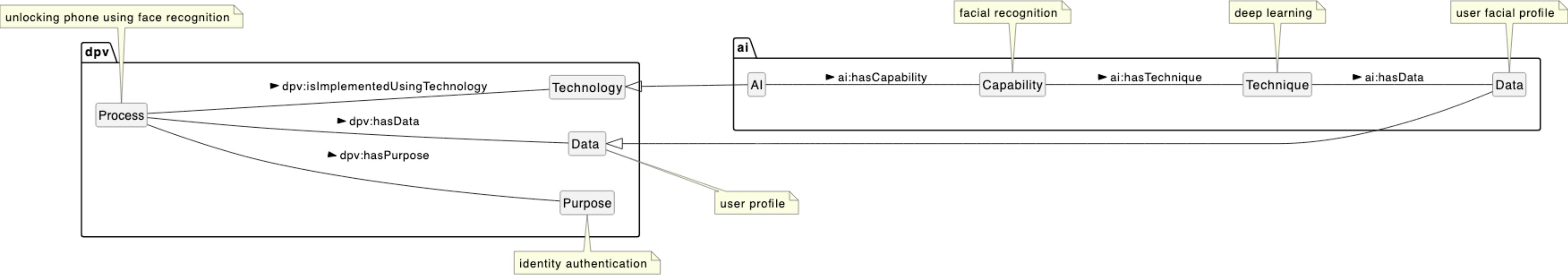




A 'Model' of Technologies

DPV TECH extension
<https://w3id.org/dpv/tech>





Risk Extension

version 2.1

Draft Community Group Report 01 February 2025

Latest published version:

<https://w3id.org/dpv/risk>

Latest editor's draft:

<https://dev.dpvcg.org/risk>

— ...

Concept	Roles				CIA model		
►	Risk Source	Risk	Consequence	Impact	Confidentiality	Integrity	Availability
risk:Bias	✓	✓	✓				
risk:CognitiveBias	✓	✓	✓				

§ 3.4.1 Risk Matrix 3x3

Likelihood ↓ Severity →	Low	Moderate	High
High	RM3x3S1L3	RM3x3S2L3	RM3x3S3L3
Moderate	RM3x3S1L2	RM3x3S2L2	RM3x3S3L2
Low	RM3x3S1L1	RM3x3S2L1	RM3x3S3L1

4. **risk:TechnicalRiskConcept**: Risk concepts, including any potential risk sources, consequences, or impacts, that are technical in nature or relate to a technical or technological process [go to full definition](#)
- a. **risk:Bias**: Bias is defined as the systematic difference in treatment of certain objects, people, or groups in comparison to others [go to full definition](#)

-

1. **risk:CognitiveBias**: Bias that occurs when humans are processing and interpreting information [go to full definition](#)

+

2. **risk:DataBias**: Bias that occurs when data properties that if unaddressed lead to systems that perform better or worse for different groups [go to full definition](#)

-

A. **risk:DataAggregationBias**: Bias that occurs when aggregating data covering different groups of objects has different statistical distributions that introduce bias into the data [go to full definition](#)

B. **risk:DataProcessingBias**: Bias that occurs due to pre-processing (or post-processing) of data, even though the original data would not have led to any bias [go to full definition](#)

C. **risk:InformativenessBias**: Bias that occurs when the mapping between inputs present in the data and outputs are more difficult to identify for some group [go to full definition](#)

D. **risk:SimpsonsParadoxBias**: Bias that occurs when a trend that is indicated in individual groups of data reverses when the groups of data are combined [go to full definition](#)

E. **risk:StatisticalBias**: Bias that occurs as the type of consistent numerical offset in an estimate relative to the true underlying value, inherent to most estimates [go to full definition](#)

+

b. **risk:DataRisk**: Risks and risk concepts related to data [go to full definition](#)

-

1. **risk:DataBias**: Bias that occurs when data properties that if unaddressed lead to systems that perform better or worse for different groups [go to full definition](#)

+

2. **risk:DataInaccurate**: Concept representing data being inaccurate [go to full definition](#)

3. **risk:DataIncomplete**: Concept representing data being incomplete [go to full definition](#)

4. **risk:DataInconsistent**: Concept representing data being inconsistent [go to full definition](#)

36

Harshvardhan J. Pandit | harshvardhan.pandit@dcu.ie | slides at: <https://harshp.com/research/presentations>

Challenges e.g. Provide vocabulary to specify purposes and permissions related to AI training #82

<https://github.com/w3c/dpv/issues/82>

1. **new:TrainingByStrategy**

- **new:SupervisedTraining** that uses **ai:SupervisedLearning** with **new:LabelledData** - where contextual information involves provenance of labelled data such as its source, who created the labels and its categorisation as sensitive etc.;
- **new:UnsupervisedTraining** that uses **ai:UnsupervisedLearning** with **new:UnlabelledData** - where contextual information involves provenance of unlabelled data such as its source;
- **new:ReinforcementTraining** that uses **ai:ReinforcementLearning** by using **new:Feedback** that act as **new:Reward** or **new:Punishment** - where contextual information involves the algorithm deciding the feedback;
- **new:SelfSupervisedLearning** that uses **new:UnlabelledData** - where contextual information involves provenance of unlabelled data.

2. **new:TrainingByAdapting**


- **new:TransferLearning** reuse a trained model for a new task in another model;
- **new:FineTuning** where a trained model is refined using new data - in particular for a specific domain or use-case;
- **new:FewShotTraining** where a trained model is given a few labelled data points to learn from - where the sample is small and not specific enough to be considered fine tuning.

3. **new:TrainingByFrequency**

- **new:StaticTraining** where the model is trained once;
- **new:PeriodicTraining** where the model is trained periodically;
- **new:ContinousTraining** where the model is trained continuously e.g. as new data arrives;
- **new:IncrementalTraining** where the model is trained in increments that are small and do not cause a full or significant retraining;
- **new:FederatedTraining** where the model is trained in a federated manner e.g. locally on device;

https://www.osai-index.eu/the-index?type=text&view=grid

NEWS

EUROPEAN
OPEN SOURCE
AI INDEX

ABOUT

CONTRIBUTORS

Last updated 05 Feb 2025

TextImageVideoCode

Ilama

	Base Model Data	End User Model Data	Base Model Weights	End User Model Weights	Training Code	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API	Licenses
Llama 3.1 by Facebook Research	✗	✗	~	✗	~	~	~	✓	✗	~	✗	✓	✗	✗
Llama 3.3 by Meta Llama	✗	✗	✗	~	~	~	~	✗	✗	~	✗	✗	✗	✗
LLaMA2 Chat by Facebook Research	✗	✗	~	~	✗	✗	~	~	✗	~	✗	✗	✗	✗
Llama 3 Instruct by Facebook Research	✗	✗	~	~	✗	✗	~	✗	✗	~	✗	✗	✗	✗

European Open Source AI Index

In conclusion...

- Many unknowns
- We're still figuring out how to *describe AI / ML technologies* in line with laws
- ML and Personal Data is a complicated affair
- Several important issues exist, but aren't being solved
- Existing approaches don't fix stuff
- Risks/Harms are a challenge that *MUST* be taken into account and addressed
- Lots of work to be done ...

~ end of slides ~