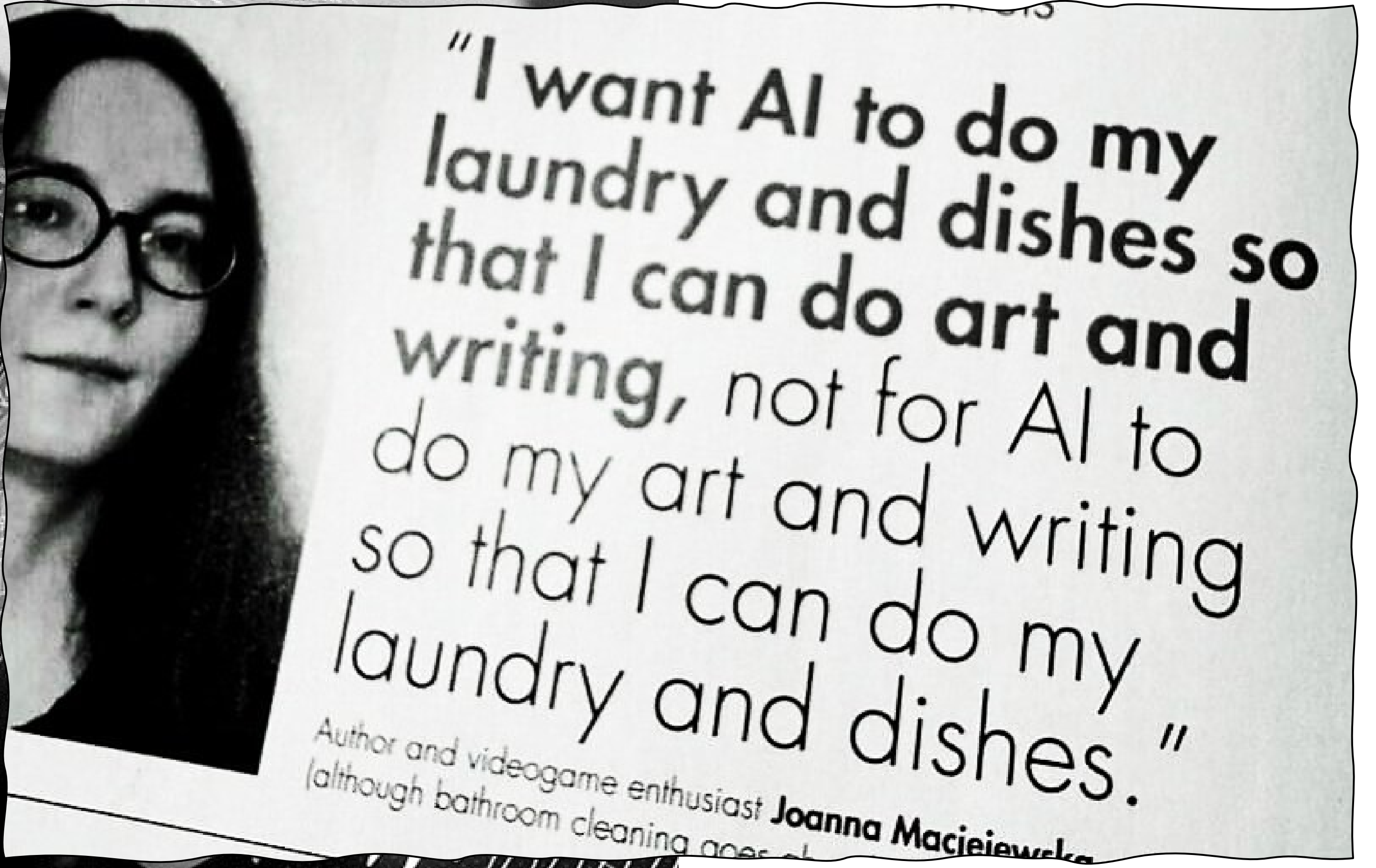
Data *Altruism* *by Design*

Machine Learning, honestly.

ULD Sommerakademie 2024 // Kiel // Tea Mustać

SPIRIT LEGAL®



"I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes."

Author and videogame enthusiast Joanna Maciejewska
(although bathroom cleaning goes

[redacted]

Models know

What you did last summer

A Survey of Privacy Attacks in Machine Learning

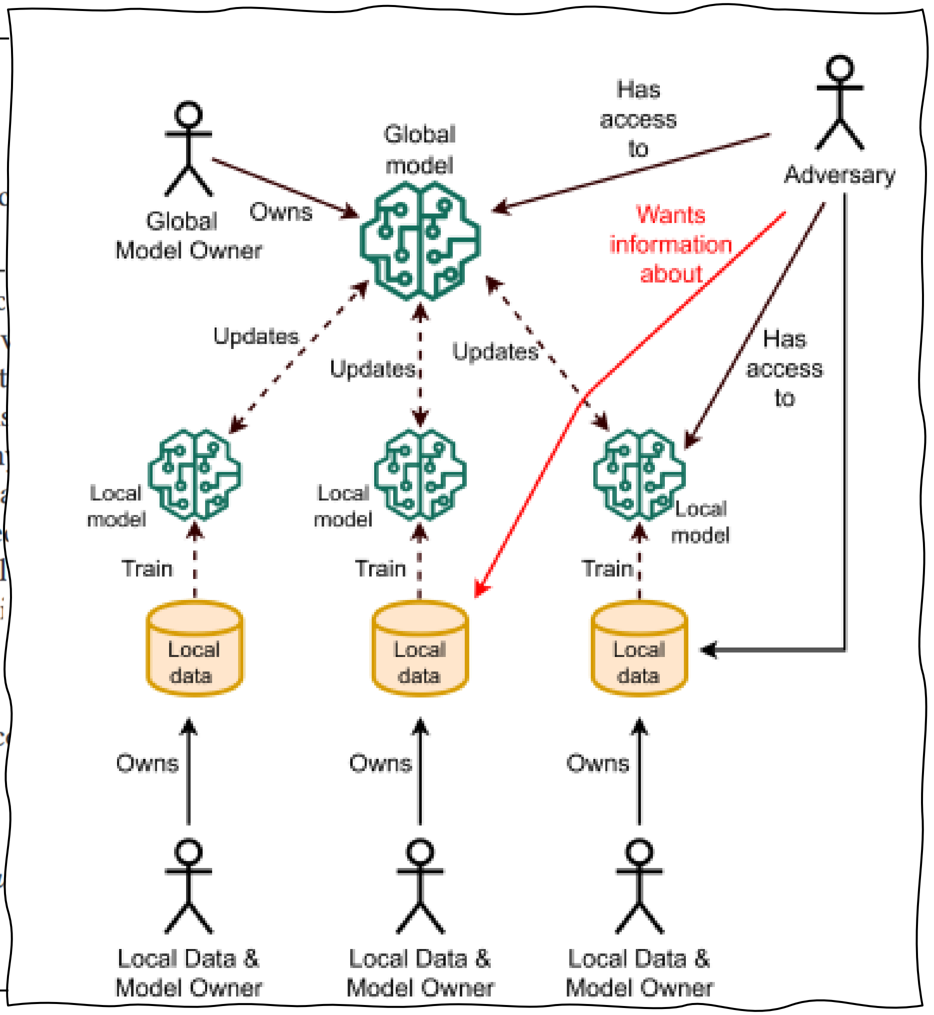
MARIA RIGAKI and SEBASTIAN GARCIA, Czech Technical University in Prague, Czech Republic

As machine learning becomes more widely used, the need to study its implications in security and privac
becomes more urgent. Although the body of work in privacy has been steadily growing over the past fev
years, research on the privacy aspects of machine learning has received less focus than the security aspect
Our contribution in this research is an analysis of more than 45 papers related to privacy attacks agains
machine learning that have been published during the past seven years. We propose an attack taxonom
together with a threat model that allows the categorization of different attacks based on the adversaria
knowledge, and the assets under attack. An initial exploration of the causes of privacy leaks is presente
as well as a detailed analysis of the different attacks. Finally, we present an overview of the most commonl
proposed defenses and a discussion of the open problems and future directions identified during our analysi

CCS Concepts: • **Computing methodologies** → **Machine learning**; • **Security and privacy**;

Additional Key Words and Phrases: Privacy, machine learning, membership inference, property inferenc
model extraction, reconstruction, model inversion

ACM Reference format:
Maria Rigaki and Sebastian Garcia. 2023. A Survey of Privacy Attacks in Machine Learning. *ACM Compu*
Surv. 56, 4, Article 101 (November 2023), 34 pages.
https://doi.org/10.1145/3624010



[redacted]

Connected cars

Eyes unseen, yet always near

Privacy of Autonomous Vehicles: Risks, Protection Methods, and Future Directions

Chulin Xie[1], Zhong Cao[2], Yunhui Long[1], Diange Yang[2], Ding Zhao[3], Bo Li[1]
[1]University of Illinois Urbana-Champaign [2]Tsinghua University [3]Carnegie Mellon University
{chulinx2, ylong4,lbo}@illinois.edu {caozhong, ydg}@tsinghua.edu.cn dingzhao@cmu.edu

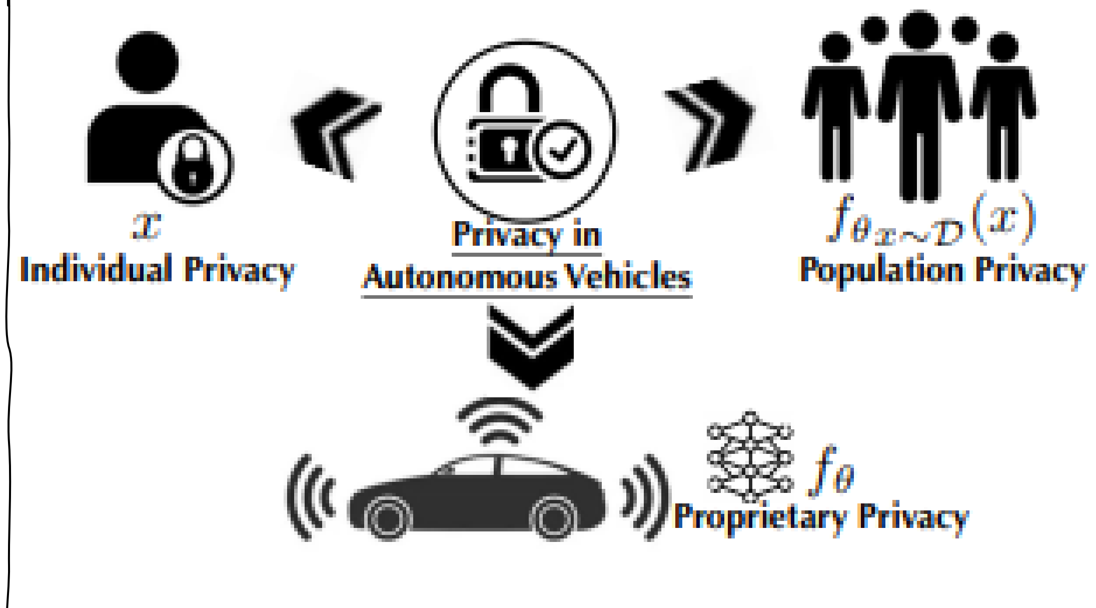


Fig. 1. Overview of privacy in autonomous vehicles.

TABLE II
INDIVIDUAL PRIVACY: RISKS AND PRIVACY ATTACKS

Attack Category	Attack target	Adversary's Knowledge	Attack Method	Data Type
Membership Inference	Classification models	Black-box model, input distribution, model type	shadow model	location, imag tabular data [
		Black-box model	confidence-thresholding	location, imag text, tabular c
		Black-box model	shadow dataset generation	image, tabula
		Black-box model, average training loss	threshold-based	tabular data [
		Black-box model, input distribution, model type	shadow model, robustness evaluation	images, locati
	Generative models	White/Black-box model	confidence-thresholding, GAN	images [79]
		Black-box model	latent encoding	images [80]
	Aggregated statistics	Black-box model, prior observation	game-based procedure	location [81]
Model Inversion	Classification models	White/Black-box model	optimization-based	image, tabular
		Black-box model, public data	inversion model	image [82]
		White-box model, public data, corrupted target input	GAN	image [83]
Model Memorization	Classification models	White/Black-box model	encoding	image, text [
	Generative models	White-box model	random sequences insertion	text [13]

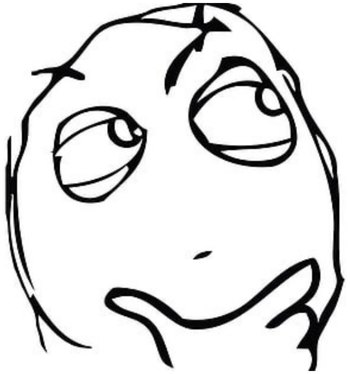
INTERNATIONAL STANDARD

ISO/IEC 5338

First edition
2023-12

Information technology — Artificial intelligence — AI system life cycle processes

Technologies de l'information — Intelligence artificielle — Processus de cycle de vie des systèmes d'IA



[redacted]

Data is the heartbeat of everything

Artificial Intelligence Life Cycle

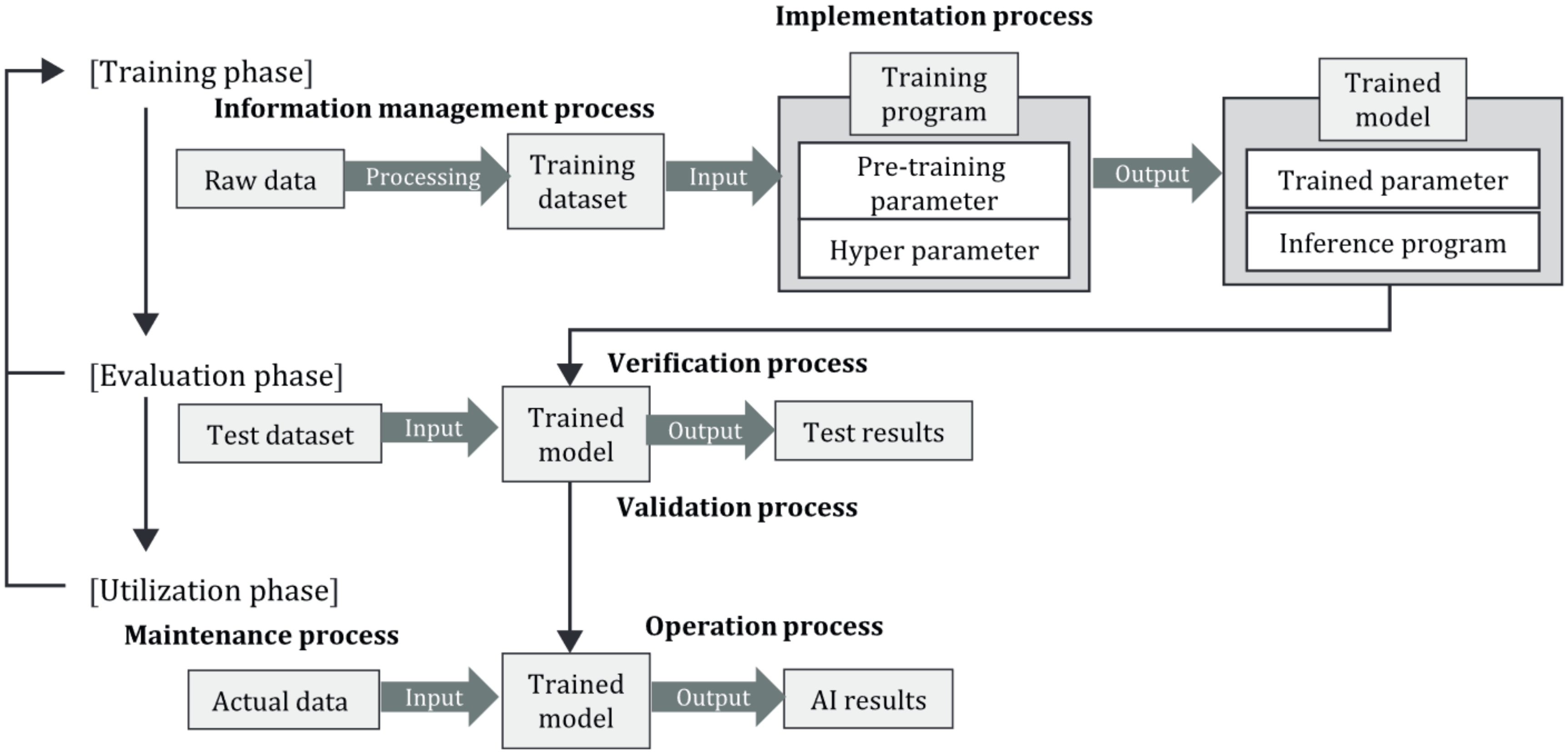


Figure A.1 — The flow of AI-specific processes

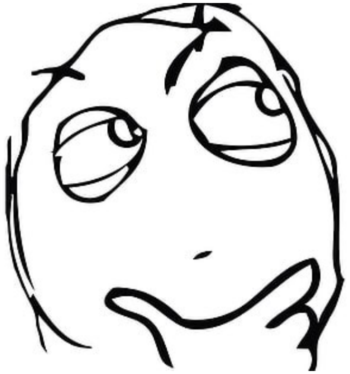
INTERNATIONAL STANDARD

ISO/IEC 5338

First edition
2023-12

Information technology — Artificial intelligence — AI system life cycle processes

Technologies de l'information — Intelligence artificielle — Processus de cycle de vie des systèmes d'IA



[redacted]

Data is the heartbeat of everything

Artificial Intelligence Life Cycle

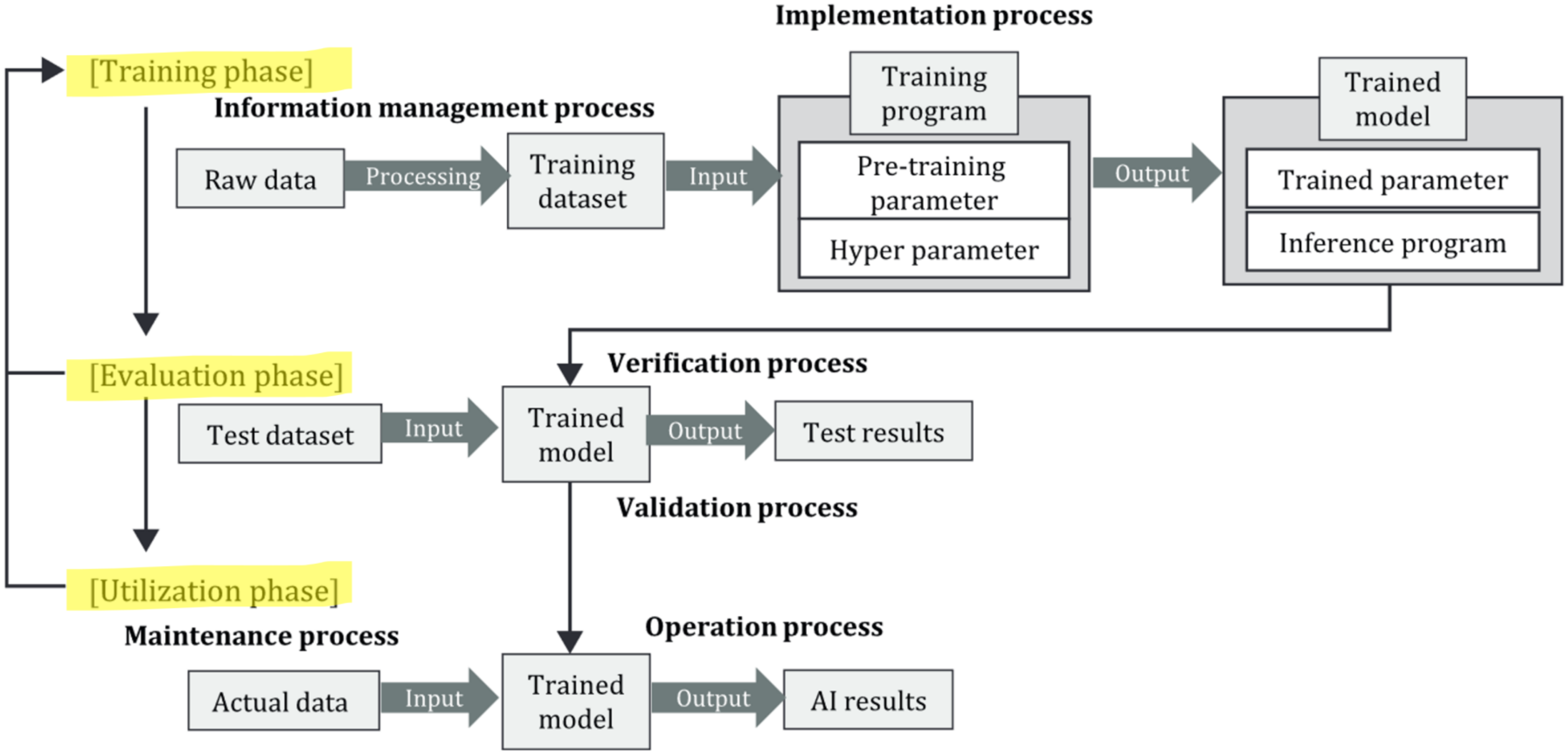
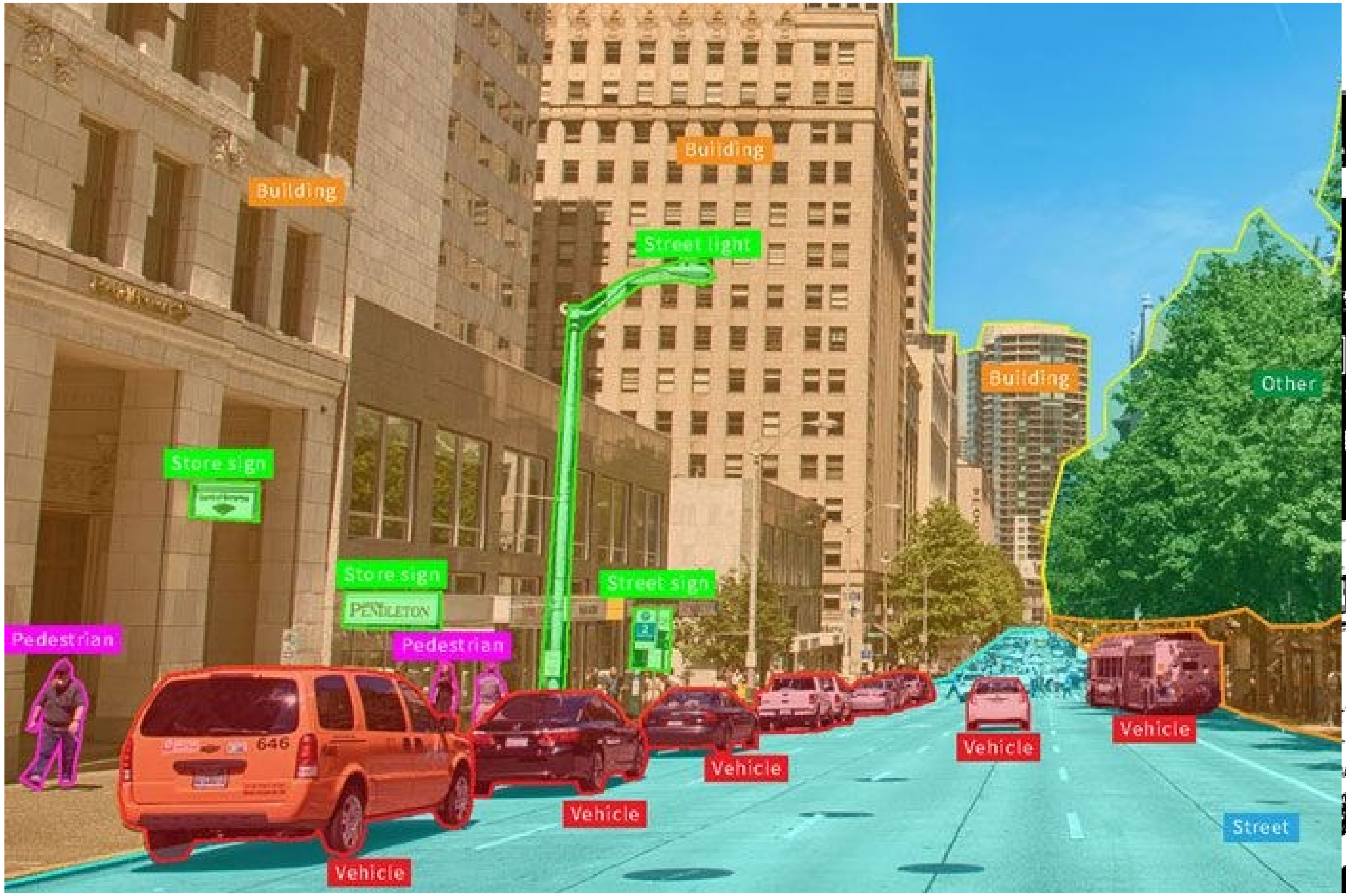


Figure A.1 — The flow of AI-specific processes

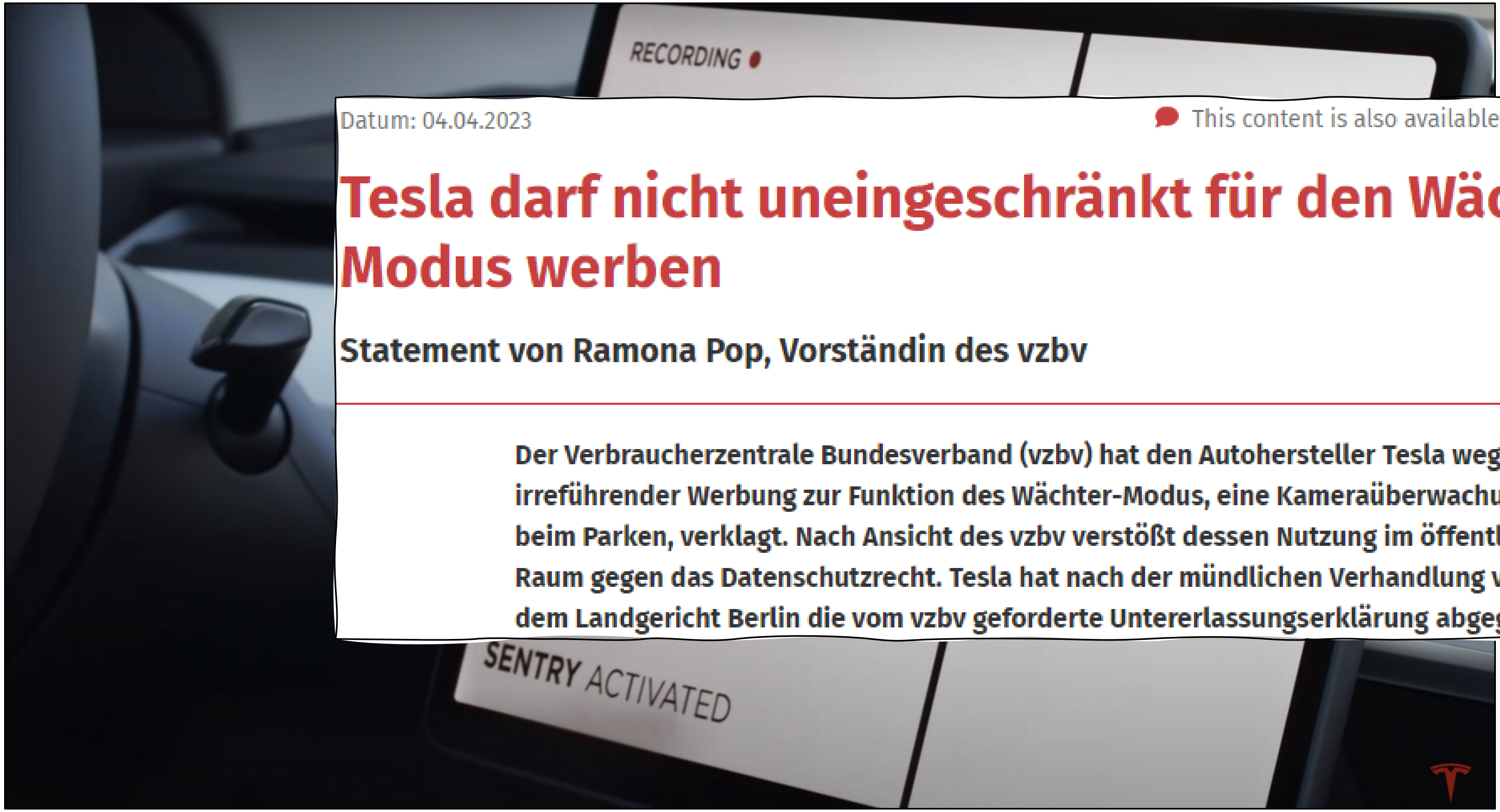
From life’s chaos, data brings order

Through bounding boxes & segmentation



Data moves at the speed of light

The law struggles to catch it



Datum: 04.04.2023 This content is also available in English. Drucken

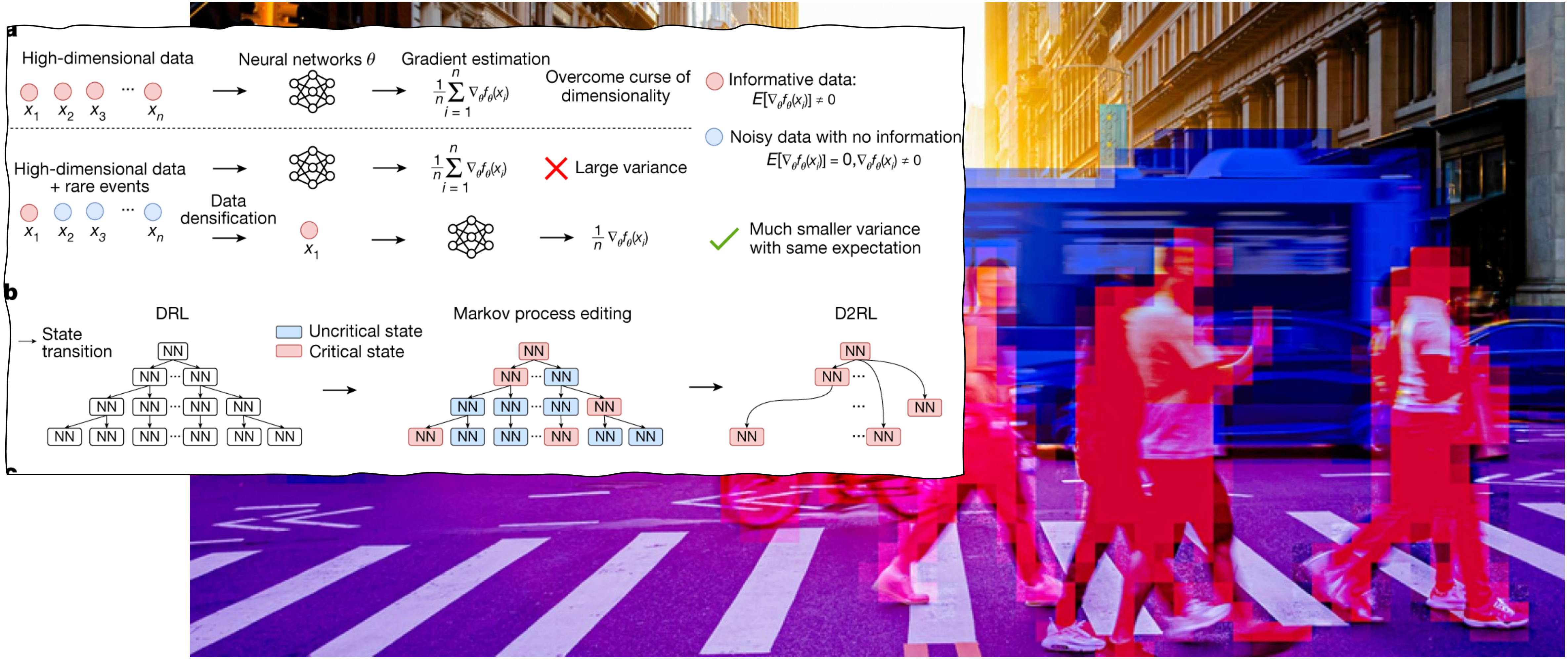
Tesla darf nicht uneingeschränkt für den Wächter-Modus werben

Statement von Ramona Pop, Vorständin des vzbv

Der Verbraucherzentrale Bundesverband (vzbv) hat den Autohersteller Tesla wegen irreführender Werbung zur Funktion des Wächter-Modus, eine Kameraüberwachung beim Parken, verklagt. Nach Ansicht des vzbv verstößt dessen Nutzung im öffentlichen Raum gegen das Datenschutzrecht. Tesla hat nach der mündlichen Verhandlung vor dem Landgericht Berlin die vom vzbv geforderte Untererlassungserklärung abgegeben.

Neural networks feast on data

To reinforce stability and safety



Feng, S., Sun, H., Yan, X. *et al.* Dense reinforcement learning for safety validation of autonomous vehicles. *Nature* **615**, 620–627 (2023).

[redacted]

Autonomous driving’s fairness

Depends on inclusive data

Bias Behind the Wheel: Fairness Analysis of Autonomous Driving Systems

XINYUE LI, Peking University, China
ZHENPENG CHEN, Nanyang Technological University, Singapore
JIE M. ZHANG, King's College London, United Kingdom
FEDE
YINC
XUAN

This p
tonom
demog
extens
and 3,
undet
variou
strate
female
both e
fairnes
code,

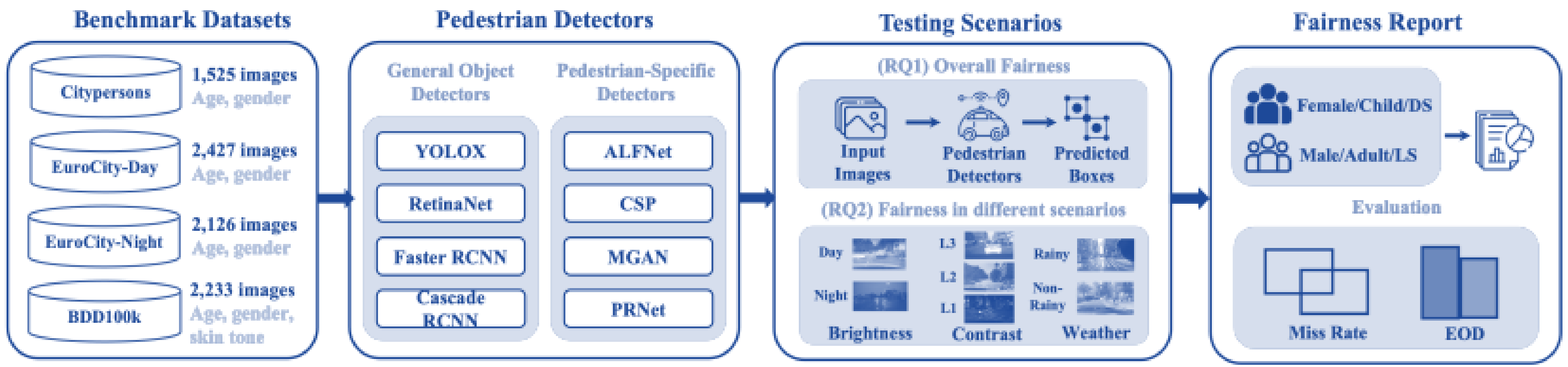


Fig. 1. Overview of our experimental settings.

[redacted]

Diverse data saves lives

Without it, systems fail

Table 4. Number of labeled pedestrian instances per dataset.

	Gender		Age		Skin tone	
Dataset	Male	Female	Adult	Child	Light	Dark
CityPersons	2,357	1,822	4,568	233	-	-
EuroCity-Day	1,726	1,646	4,498	100	-	-
EuroCity-Night	1,265	1,318	4,165	68	-	-
BDD100k	3,457	2,479	6,293	190	2,724	789
Overall	**8,805**	**7,265**	**19,524**	**591**	**2,724**	**789**

	Age					
	Day-time			**Night-time**		
Detectors	**MR Adult**	**MR Child**	**EOD (Age)**	**MR Adult**	**MR Child**	**EOD (Age)**
YOLOX	10.80%	40.77%	-29.97%	17.81%	54.93%	-37.12%
RetinaNet	12.73%	41.92%	-29.19%	18.97%	61.97%	-43.01%
Faster RCNN	4.50%	25.00%	-20.50%	7.34%	33.80%	-26.46%
Cascade RCNN	4.52%	25.58%	-21.05%	6.81%	33.80%	-26.99%
ALFNet	26.52%	49.42%	-22.90%	65.10%	83.10%	-18.00%
CSP	28.64%	46.92%	-18.28%	63.42%	76.06%	-12.64%
MGAN	27.87%	43.46%	-15.59%	49.72%	69.01%	-19.30%
PRNet	34.80%	56.92%	-22.13%	72.60%	92.96%	-20.36%
Average	**18.80%**	**41.25%**	**-22.45%**	**37.72%**	**63.20%**	**-25.48%**

[redacted]

Diverse data saves lives

Without it, systems fail

Table 4. Number of labeled pedestrian instances per dataset.

	Gender		Age		Skin tone
Dataset	Male	Femal	Adult	Child	
CityPersons	2,357	1,822	4,568	233	
EuroCity-Day	1,726	1,646	4,498	100	
EuroCity-Night	1,265	1,318	4,165	68	
BDD100k	3,457	2,479	6,293	190	
Overall	**8,805**	**7,26**	**19,524**	**591**	

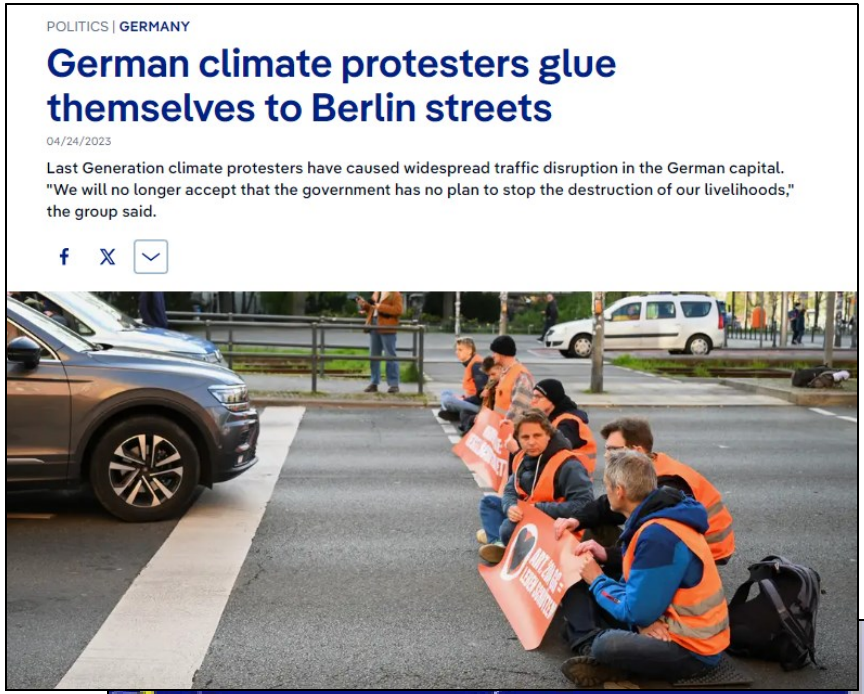
	Age					
	Day-time					
Detectors	**MR Adult**	**MR Child**	**EOD (Age)**			**)D (Age)**
YOLOX	10.80%	40.77%	-29.97%			-37.12%
RetinaNet	12.73%	41.92%	-29.19%			-43.01%
aster RCNN	4.50%	25.00%	-20.50%			-26.46%
scade RCNN	4.52%	25.58%	-21.05%			-26.99%
ALFNet	26.52%	49.42%	-22.90%			-18.00%
CSP	28.64%	46.92%	-18.28%			-12.64%
MGAN	27.87%	43.46%	-15.59%			-19.30%
PRNet	34.80%	56.92%	-22.13%			-20.36%
Average	**18.80%**	**41.25%**	**-22.45%**	**37.72%**	**63.20%**	**-25.48%**

EOD (Age)
-29.97%
-29.19%
-20.50%
-21.05%
-22.90%
-18.28%
-15.59%
-22.13%
-22.45%

[redacted]

The model's manifold grows

With each layer of data diversity

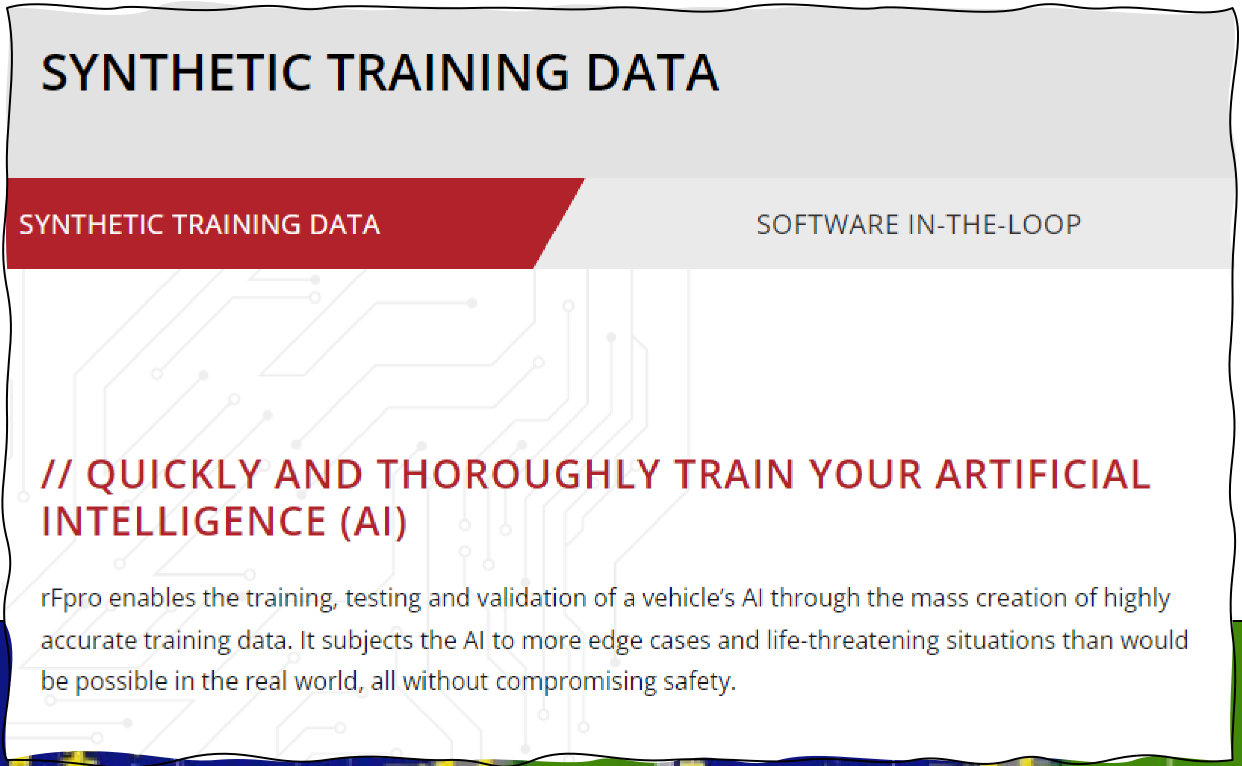


POLITICS | GERMANY

German climate protesters glue themselves to Berlin streets

04/24/2023

Last Generation climate protesters have caused widespread traffic disruption in the German capital. "We will no longer accept that the government has no plan to stop the destruction of our livelihoods," the group said.

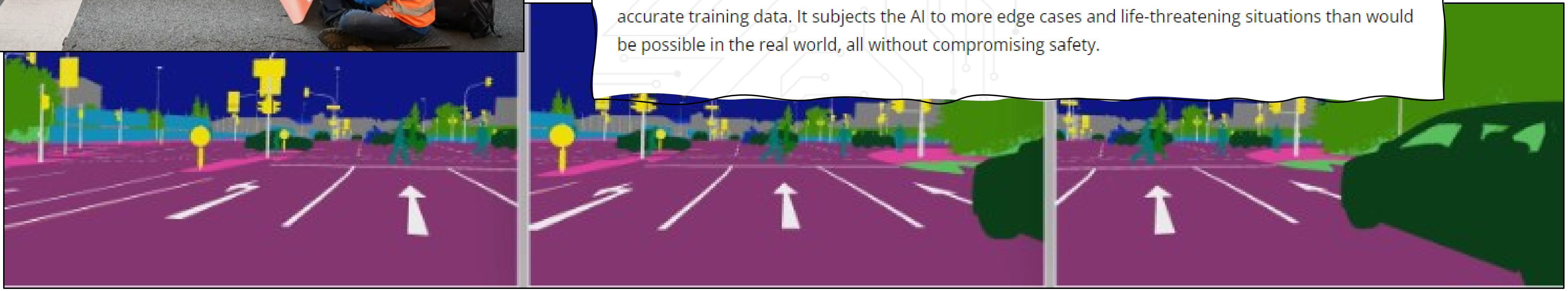


SYNTHETIC TRAINING DATA

SYNTHETIC TRAINING DATA | SOFTWARE IN-THE-LOOP

// QUICKLY AND THOROUGHLY TRAIN YOUR ARTIFICIAL INTELLIGENCE (AI)

rFpro enables the training, testing and validation of a vehicle's AI through the mass creation of highly accurate training data. It subjects the AI to more edge cases and life-threatening situations than would be possible in the real world, all without compromising safety.



[redacted]

Comprehensive, diverse data

The lifeblood of accurate medical AI

Article | Published: 18 May 2020

A deep learning system for differential diagnosis of skin diseases

Yuan Liu, Ayush Jain, Clara Eng, David H. Way, Kang Lee, Peggy Bui, Kimberly Kanada, Oliveira Marinho, Jessica Gallegos, Sara Gabriele, Vishakha Gupta, Nalini Singh, Vivek Hofmann-Wellenhof, Greg S. Corrado, Lily H. Peng, Dale R. Webster, Dennis Ai, Susan R. Carter Dunn & David Coz

Nature Medicine **26**, 900–908

23k Accesses | **377** Citations

Letter | Published: 20 May 2019

End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography

Diego Ardila, Atilla P. Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J. Reicher, Lily Peng, Daniel Tse, ...do, David P. Naidich & Shravya Shetty

this article

tmetric | Metrics

Original Investigation | Innovations in Health Care Delivery

December 13, 2016

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD[1]; Lily Peng, MD, PhD[1]; Marc Coram, PhD[1]; et al

› Author Affiliations | Article Information

JAMA. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

[redacted]

Protectionism

Will not last forever

Der Hamburgische Beauftragte für
Datenschutz und Informationsfreiheit

Diskussionspapier: Large Language Models und personenbezogene Daten

1. Die bloße Speicherung eines LLMs stellt keine Verarbeitung im Sinne des Art. 4 Nr. 2 DSGVO dar. Denn in LLMs werden keine personenbezogenen Daten gespeichert. Soweit in einem LLM-gestützten KI-System personenbezogene Daten verarbeitet werden, müssen die Verarbeitungsvorgänge den Anforderungen der DSGVO entsprechen. Dies gilt insbesondere für den Output eines solchen KI-Systems.

2. Mangels Speicherung personenbezogener Daten im LLM können die Betroffenenrechte der DSGVO nicht das Modell selbst zum Gegenstand haben. Ansprüche auf Auskunft, Löschung oder Berichtigung können sich jedoch zumindest auf Input und Output eines KI-Systems der verantwortlichen Anbieter:in oder Betreiber:in beziehen.

3. Das Training von LLMs mit personenbezogenen Daten muss datenschutzkonform erfolgen. Dabei sind auch die Betroffenenrechte zu beachten. Ein ggf. datenschutzwidriges Training wirkt sich aber nicht auf die Rechtmäßigkeit des Einsatzes eines solchen Modells in einem KI-System aus.

[redacted]

Data Altruism By Design

Carving out a GDPR exception for Machine Learning

- 1.Lawful access
- 2. The storing of the data is not the purpose
- 3. The systems aren’t used to the detriment of the data subjects or in a way that
- 4. Reversed burden of proof
- 5. Point 4 includes implementing appropriate TOMs throughout the AI lifecycle
- An insurance fund to cover risks.

[redacted]

Kontakt

Tea Mustać

tea.mustac@spiritlegal.com

www.spiritlegal.com