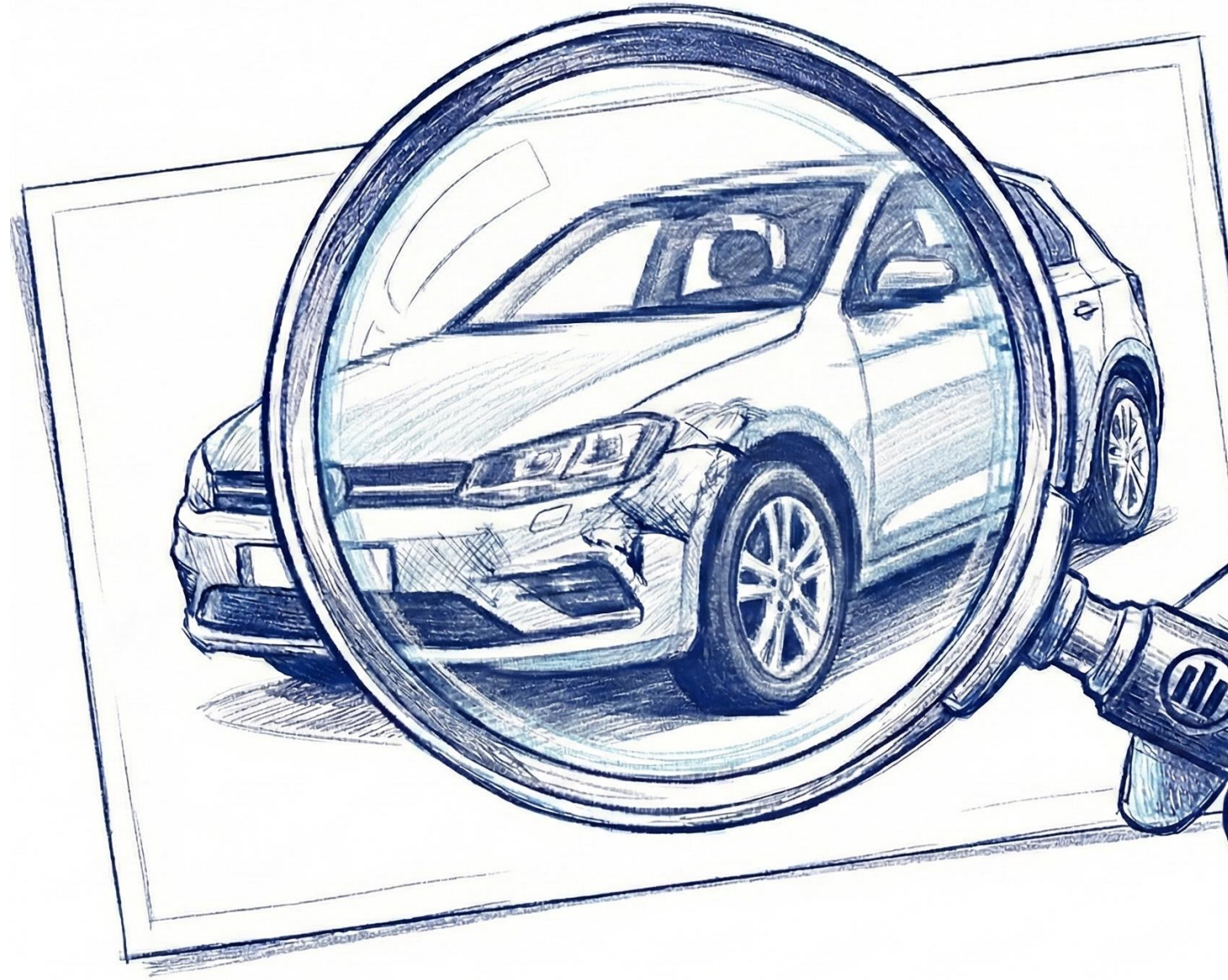# Allianz X TUM
Data Challenge

ÖZALP UFKUN

ÖKSÜZ EFECAN MURAT

SEIBOLD NICOLAS

30.1.2026 – DATA ANALYTICS IN APPLICATION

Agenda

1. Executive Summary

2. Fake Image Production (Our Test Set)

3. The Detection Programs and the Pipeline

4. Our Implementation

5. Recommendations

Executive Summary

- Claims per day 500
- Cost per claim 1.500 €
- Claims (fake) 5
- Claims (real) 495

- Losses per day 7.500 €
- Losses per year 2.737.500 €

- The amount of money and time that the Allianz Group could save by using recognition software justifies potential investments.

The Problem:
Deepfakes and AI-generated images
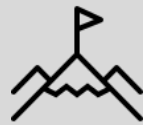
Our Mission:
Detect AI-generated Images

Our Vision:
Find a model that can detect and generate

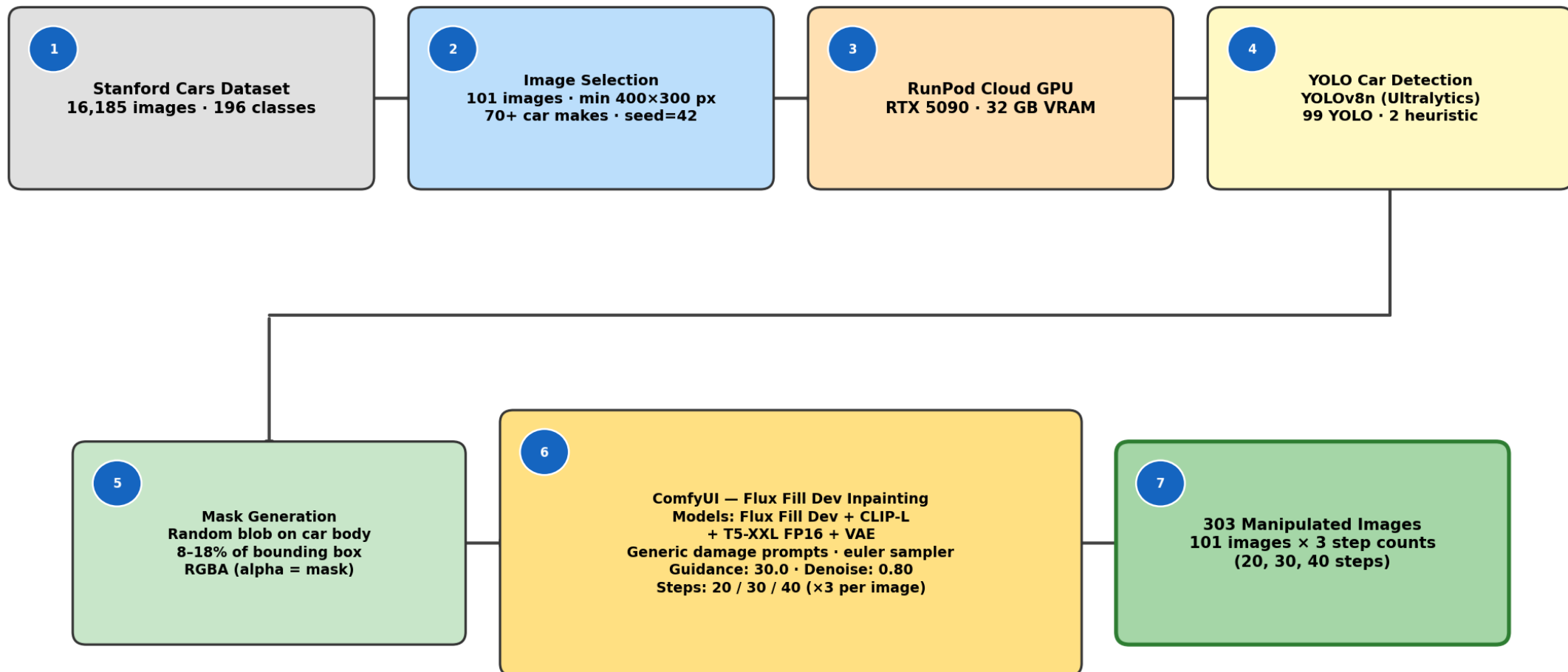**Our Dataset:**
1000 pictures modified with FLUX

Our Tool:
Uses multiple detectors at once

Our Recommendation:
Build a personal tool Or commission one

Dataset Generation Pipeline – Image Flow

1 Stanford Cars Dataset
16,185 images · 196 classes

2 Image Selection
101 images · min 400×300 px
70+ car makes · seed=42

3 RunPod Cloud GPU
RTX 5090 · 32 GB VRAM

4 YOLO Car Detection
YOLOv8n (Ultralytics)
99 YOLO · 2 heuristic

5 Mask Generation
Random blob on car body
8–18% of bounding box
RGBA (alpha = mask)

6 ComfyUI — Flux Fill Dev Inpainting
Models: Flux Fill Dev + CLIP-L
+ T5-XXL FP16 + VAE
Generic damage prompts · euler sampler
Guidance: 30.0 · Denoise: 0.80
Steps: 20 / 30 / 40 (×3 per image)

7 303 Manipulated Images
101 images × 3 step counts
(20, 30, 40 steps)

Example Images with Inpanting

Example Images with Inpanting

The Ai-images are hard to detect

Performance of Detection Models on their own

Overall Model Comparison

Accuracy: (TP + TN) / Total Images

AUC: Area Under the Curve

F1: Harmonic Mean of Precision and Recall

Performance of Detection Models on their own

Score Distributions by Category
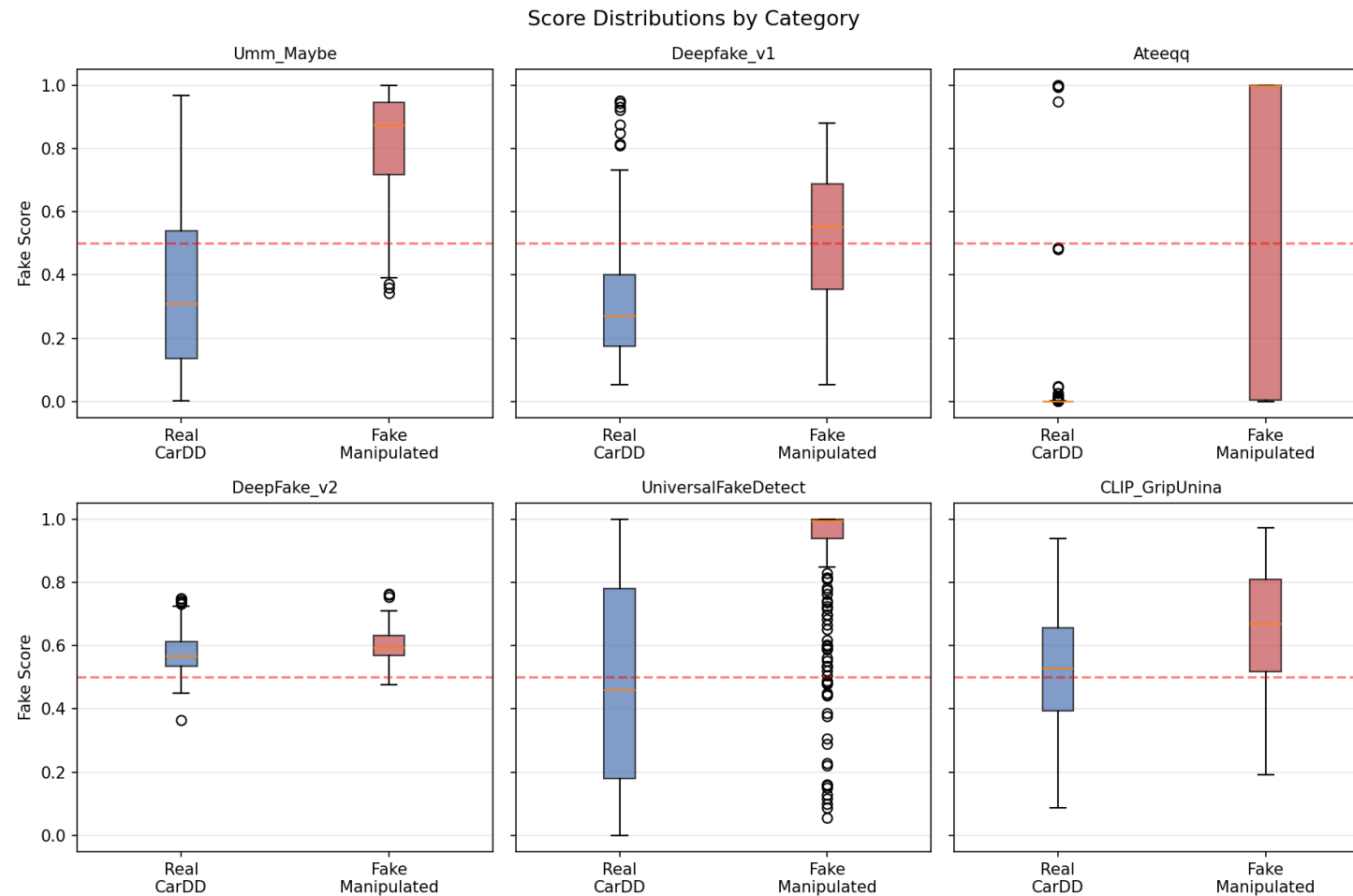
The Detection Models make different mistakes

Error Analysis: Do Models Make Different Mistakes?

False Negatives per Model (Fakes Missed): Ateeqq 115, Umm_Maybe 26, UniversalFakeDetect 23. All 3 miss: 1

False Positives per Model (Reals Wrongly Flagged): Ateeqq 10, Umm_Maybe 114, UniversalFakeDetect 175. All 3 share: 2
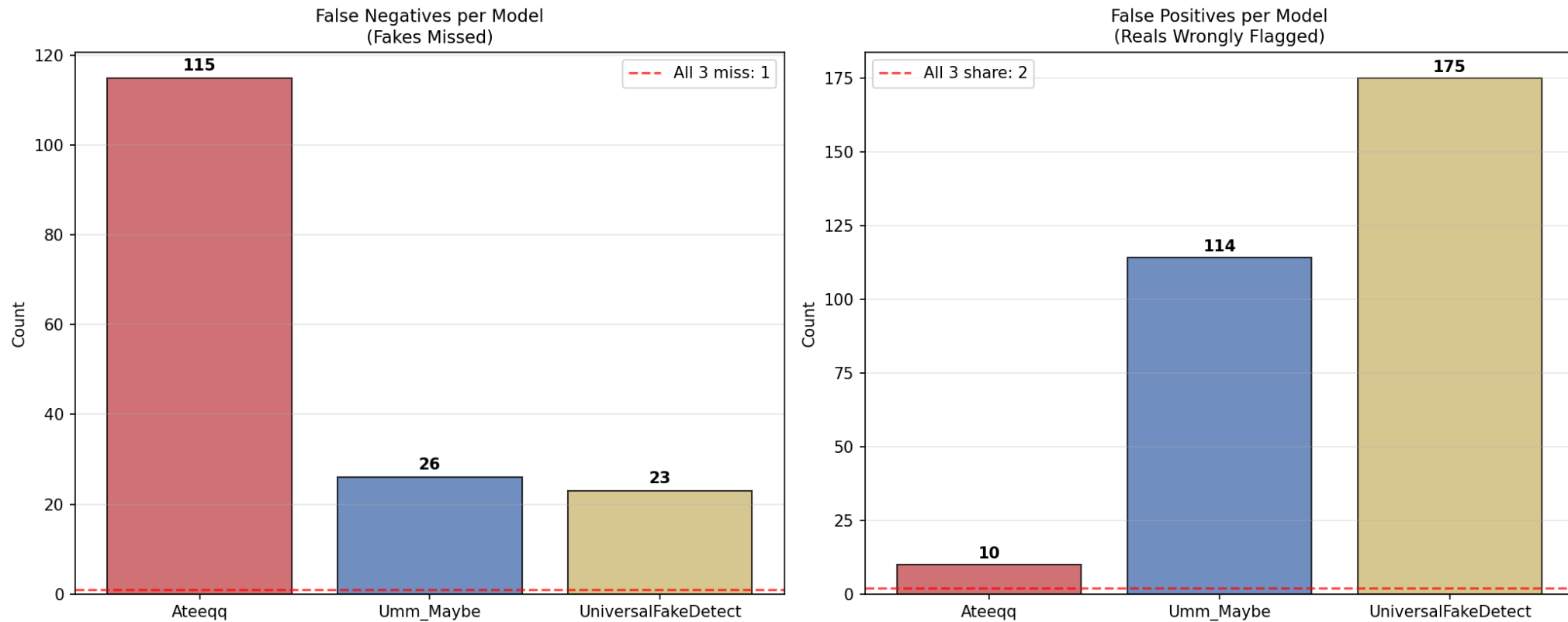
The Detection Pipeline – Two Stage Approach

Staged Detection Pipeline — Image Flow

Input: 677 images
(374 real, 303 manipulated)

Stage 1: Ateeqq
(precision: 94.9%)

198 → FAKE: 198
(188 TP, 10 FP)

479 pass

Stage 2: Umm_Maybe + UniversalFakeDetect
(479 remaining images)

Both agree → SUSPICIOUS: 164
(105 TP, 59 FP)

Both say real → PASS: 141
(1 missed fakes)

Disagree → REVIEW: 174
(9 fakes, 165 reals)
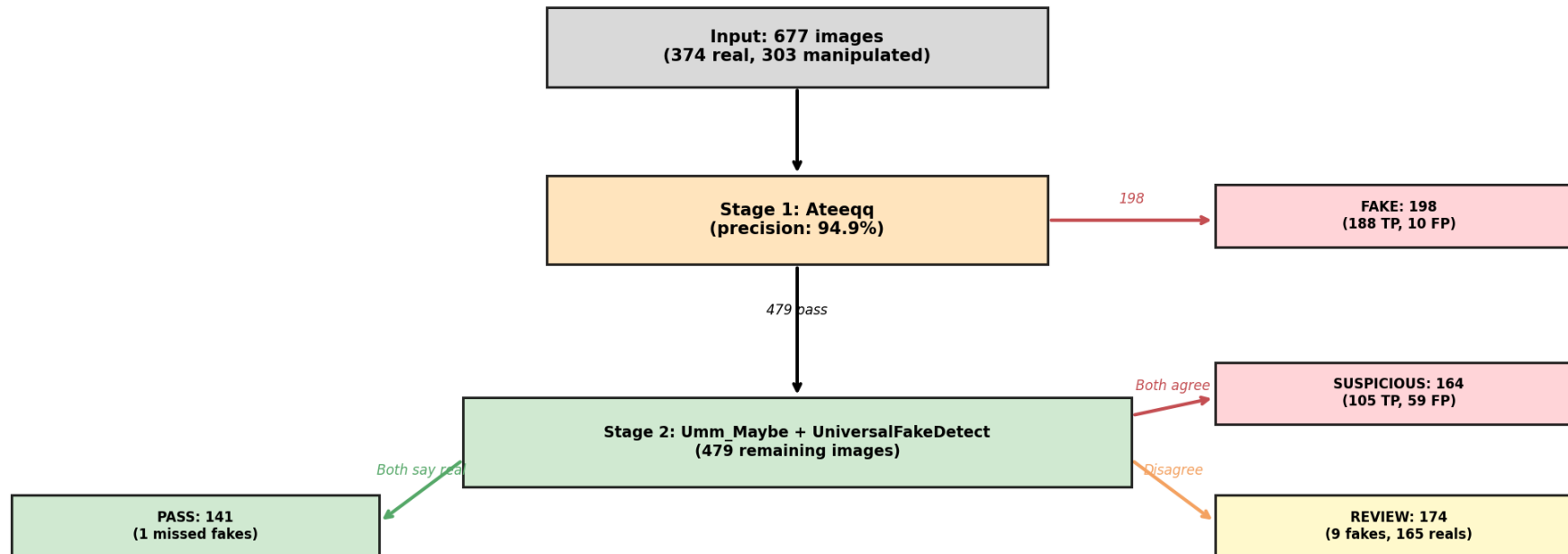
Pipeline Result: 293/303 fakes detected (96.7%), 69 false positives (18.4%)
+ 9 more fakes recoverable via manual review of 174 borderline cases

Optimized Thresholds Strategy — Image Flow

Input: 677 images
(374 real, 303 manipulated)

Step 1: Apply Per-Model Optimized Thresholds

Ateeqq
threshold: 0.05

Flags: 212
(200 TP, 12 FP)

Umm_Maybe
threshold: 0.58

Flags: 340
(265 TP, 75 FP)

UniversalFakeDetect
threshold: 0.93

Flags: 271
(233 TP, 38 FP)

Step 2: Majority Vote
2 of 3 models must agree → FAKE

Vote Distribution

FAKE

FAKE

REAL

REAL

3 votes
(all agree)

122 images
122 fakes, 0 reals

2 votes
(majority)

162 images
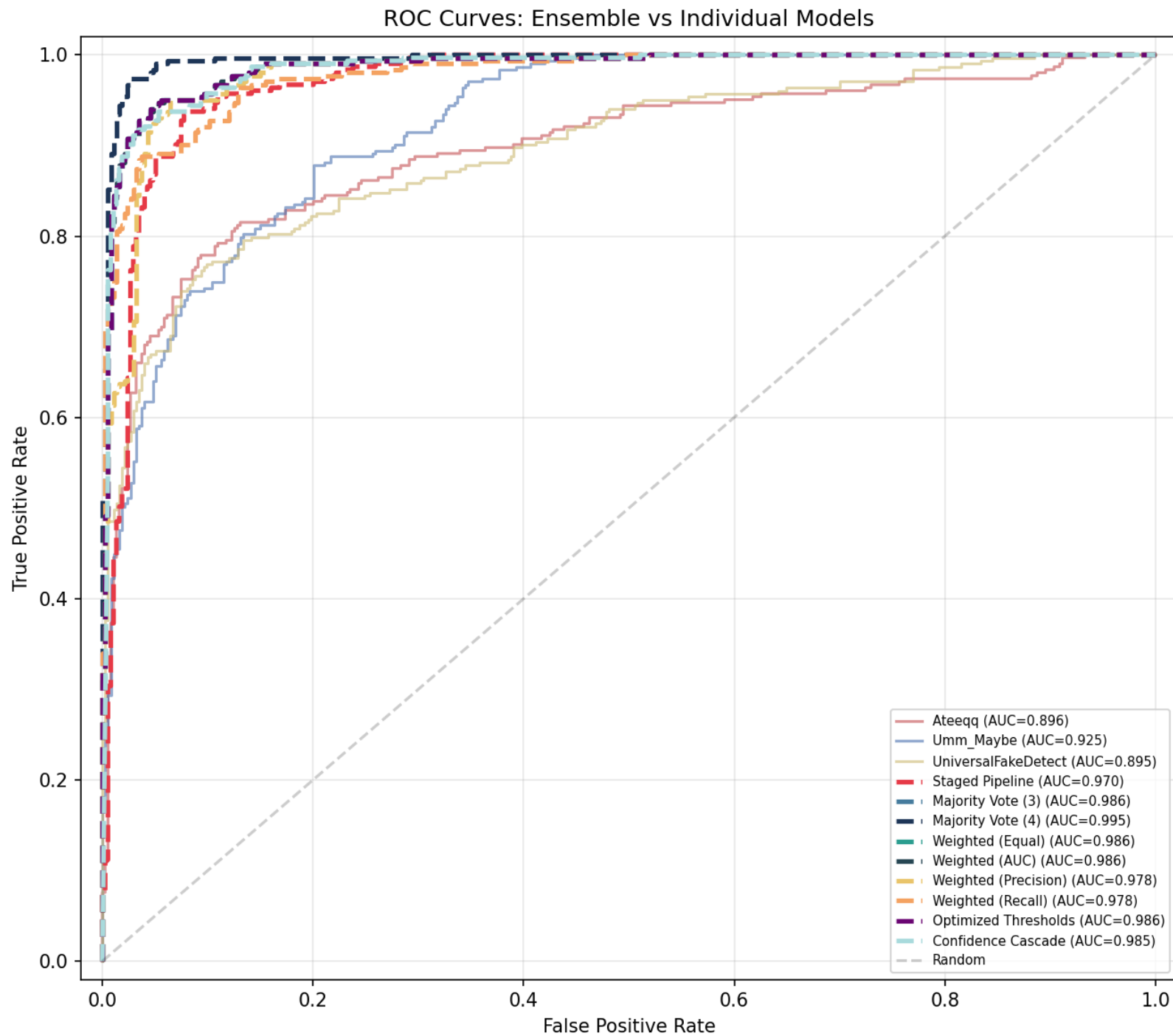**154 fakes, 8 reals**

1 vote
(minority)

133 images
24 fakes, 109 reals

0 votes
(unanimous real)

260 images
3 fakes, 257 reals

Classified FAKE: 284
(276 TP, 8 FP)

Classified REAL: 393
(366 TN, 27 missed fakes)

Result: 276/303 fakes detected (91.1%), 8/374 false positives (2.1%)
Accuracy: 94.8% | F1: 0.9404 | Thresholds: Ateeqq=0.05, Umm_Maybe=0.58, UFD=0.93

ROC Curves: Ensemble vs Individual Models

Building our own Hybrid model

Ensemble Strategies vs Individual Models

F1 Score (higher is better)

Model	F1 Score
Ateeqq	0.750
Umm_Maybe	0.798
UniversalFakeDetect	0.739
Deepfake_v1	0.676
CLIP_GripUnina	0.628
Staged Pipeline	0.881
Majority Vote (3)	0.886
Majority Vote (4)	0.884
Weighted (Equal)	0.928
Weighted (AUC)	0.926
Weighted (Precision)	0.778
Weighted (Recall)	0.860
Optimized Thresholds	0.940
Confidence Cascade	0.909

Detection Rate / Recall (higher is better)

Model	Detection Rate / Recall
Ateeqq	62.0%
Umm_Maybe	91.4%
UniversalFakeDetect	92.4%
Deepfake_v1	61.1%
CLIP_GripUnina	76.6%
Staged Pipeline	96.7%
Majority Vote (3)	96.0%
Majority Vote (4)	80.5%
Weighted (Equal)	95.0%
Weighted (AUC)	95.0%
Weighted (Precision)	66.0%
Weighted (Recall)	98.0%
Optimized Thresholds	91.1%
Confidence Cascade	97.4%
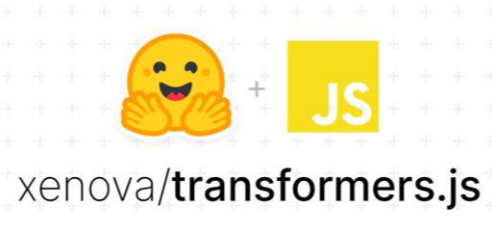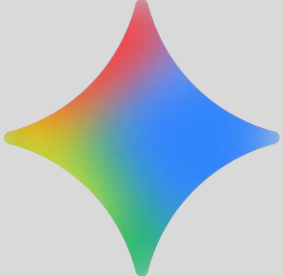
False Positive Rate (lower is better)

Model	False Positive Rate
Ateeqq	2.7%
Umm_Maybe	30.5%
UniversalFakeDetect	46.8%
Deepfake_v1	15.8%
CLIP_GripUnina	54.5%
Staged Pipeline	18.4%
Majority Vote (3)	16.8%
Majority Vote (4)	1.3%
Weighted (Equal)	8.0%
Weighted (AUC)	8.3%
Weighted (Precision)	2.9%
Weighted (Recall)	24.3%
Optimized Thresholds	2.1%
Confidence Cascade	13.6%
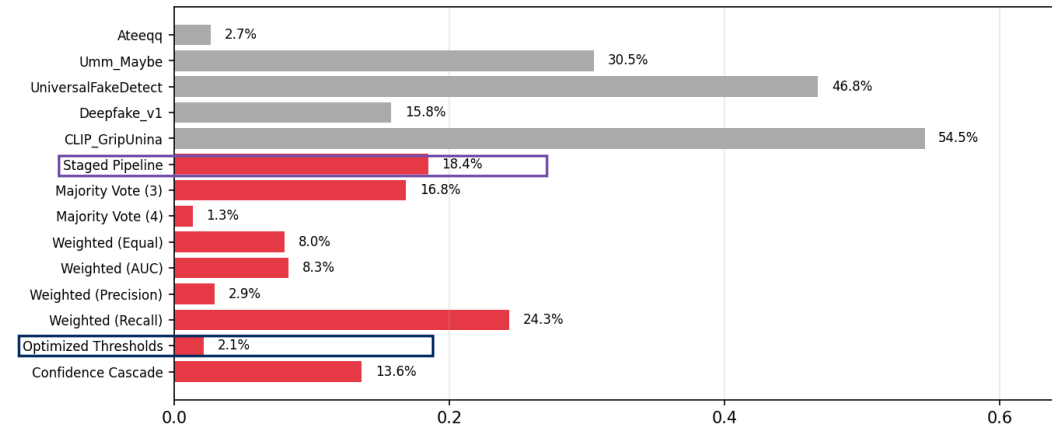
One more thing…

Technology & Architecture

Frontend	Backend	Database Architecture	The AI Core (The "Secret Sauce")
React, TypeScript, Vite	Node.js, express	mongoDB	xenova/transformers.js, pollinations.ai

We will start now with a short live demo of

Impacts on the Allianz Group – Buying a Solution

Buying a Product:			
Post:	**Price**	**Description**	**Cost Estimate**
CAPEX -			
System integration	500.000,00 €	Implementing API to software (SAP)...	250-750T
Security and Technichal Audit	120.000,00 €	Testing the software on actual cases	80-150T
Process Redesign	210.000,00 €	Redesigning the Insurance claim reporting	120-300T
Train Employees	500.000,00 €	E-Workshops, Istrauctions, ...	200-800T
Conformity Assessment EU_AI_ACT	75.000,00 €	EU zertifikate for ai applications	50-100T
Summ Capex	1.405.000,00 €		
OPEX -			
SaaS-Lizenze	375.000,00 €	Depending on requests and amount / Output type	250-500T
Model Retraining Extern	100.000,00 €	Regular updates to the model with new data	100T
Cloud/Compute (Run)	120.000,00 €	Cost for external computing power	50-190T
Support for employees	225.000,00 €	Internal AI Team for questions, problems	150-300T
Support & Monitoring	100.000,00 €	Survailance of Output and metrics	100T
Summ Opex	920.000,00 €		
Total Summ (External product)	2.325.000,00 €		

Number of Claims per Year	166.000
Average Claim ammount	1.500
Percentage of Fake claims	1%
The cost of missed Claims	1.245.000,00 €
Amortization Scenarion 1	4,32
Amortization Scenarion 2	4,79
ROI S1	23,13
ROI S2	20,90

Impacts on the Allianz Group – Building a Solution

Building a Product

Post:	Price	Description	Cost Estimate
CAPEX-			
Development Team	1.050.000,00 €	Salaries for data scientists, ML engineers, backend (a	800-1.500T
Data Acquisition & Labeling	225.000,00 €	Purchase of training data or internal labeling of Allianz	150-300T
Infrastructure Setup	150.000,00 €	Setting up the GPU cluster or cloud environment for trai	100-200T
Security & Compliance Audit	150.000,00 €	Internal auditing and penetration testing (often more e:	150T
Conformity EU AI ACT	100.000,00 €	Technical documentation & risk management for in-h(	100T
Sum CAPEX	1.675.000,00 €		
OPEX-			
Maintenance & DevOps	300.000,00 €	Bug fixing and system maintenance team (2-3 FTE)	300T
Model Retraining	100.000,00 €	Regular updates to the model with new data	100T
Support for employees	225.000,00 €	Internal AI Team for questions, problems	150-300T
Cloud/Compute (Run)	120.000,00 €	Hosting-Cost (cloud solutions) for application	120T
Support & Monitoring	150.000,00 €	Survailance of Output and metrics	150T
Summe OPEX	895.000,00 €		
Total Summ (Internal Product)	2.570.000,00 €		

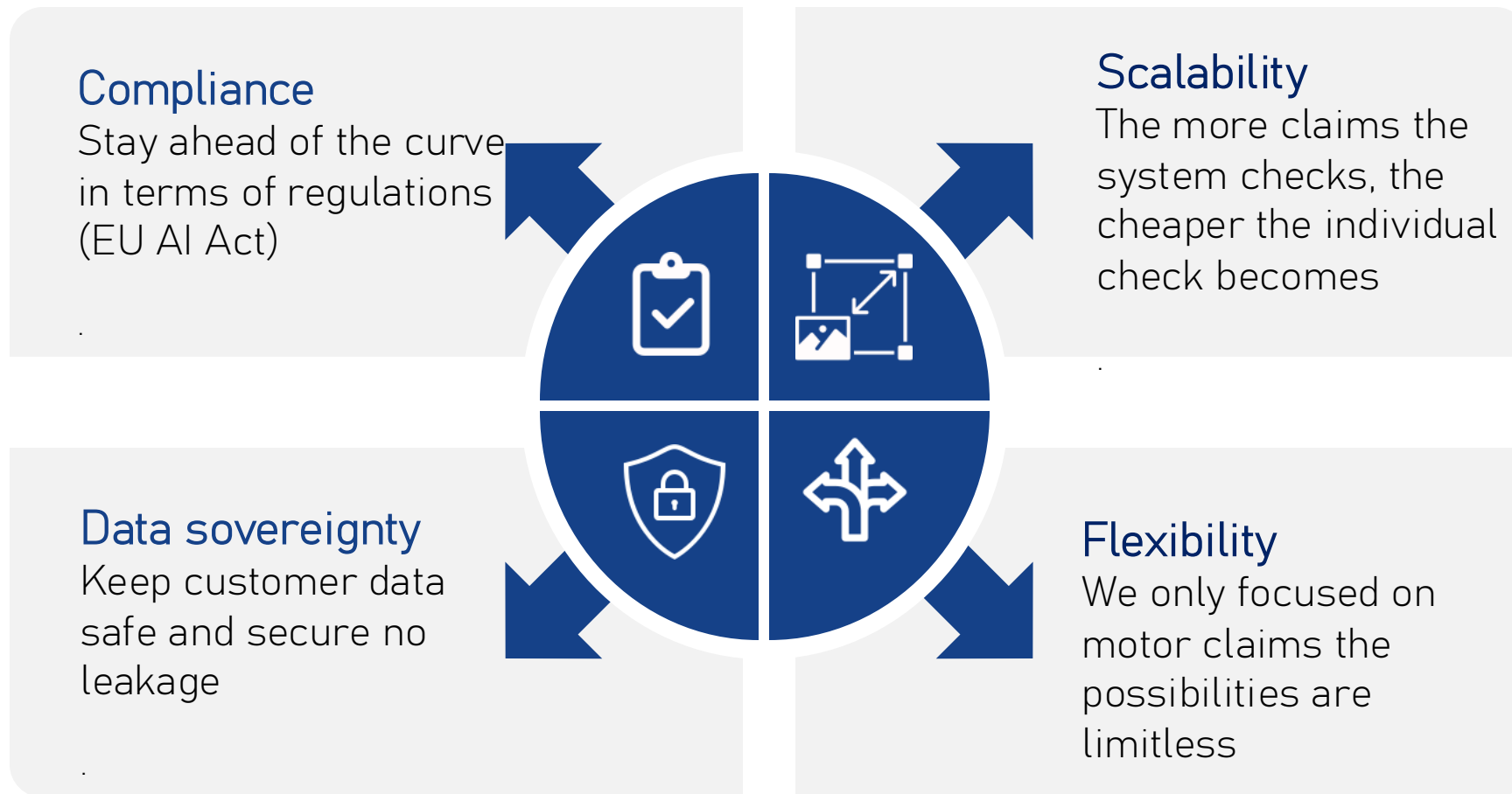|	
Number of Claims per Year	166.000
Average Claim ammount	1.500
Percentage of Fake claims	1%
The cost of missed Claims	1.245.000,00 €
Amortization Scenarion 1	4,32
Amortization Scenarion 2	4,79
ROI S1	23,13
ROI S2	20,90

Impacts on the Allianz Group – Outlook

- **The business case:** Efficiency & fraud prevention

- The (ROI) comes from two sources:

- **Operational efficiency** (time-to-settlement): Automation of preliminary checks – reduces processing time per claim (80%)

- **Claims savings** (fraud detection): identifies attempts at manipulation with a precision that exceeds human capabilities

Time saving potential	Days
Registering a new claim	1
Survey on One Day Service (ODS)	0
Assignment to claim handler	1
Issuing the Work Order (SPK)	3
Verification of payment documents	3
Claim payment process	2
Summ	10
	(prob 15 days)
80% reduction	80%
Introduction of AI average saving in days	8
Summ time for claim handling	2

Even a slight increase in the detection rate among millions of claims will amortize the CAPEX of approximately €1.675 million in a very short time

Our Recommendation –To build a model that fits Allianz

Compliance
Stay ahead of the curve in terms of regulations (EU AI Act)

.

Scalability
The more claims the system checks, the cheaper the individual check becomes

.

Data sovereignty
Keep customer data safe and secure no leakage

.

Flexibility
We only focused on motor claims the possibilities are limitless

.