BULLETPROOF.AI

Security for Machine Learning

October 2019

BULLETPROOF.AI

Security solutions for AI, machine learning and automated statistical decisions

AI Models make critical business decisions in split seconds, every second of the day

How Secure is your Machine Learning System?

AI Disrupts Finance

- Immediate decisions, anytime
- Better decisions & pricing drive competition
- New markets

- **Immediate convenience**

Attacks on Machine Learning

Confidentiality Attacks

Attacker may be able to **copy the model** and to **extract** the data used to train the model.

Evasion Attacks

Attacker may **discover and exploit existing vulnerabilities** in the model in order to **manipulate** the decision.

Poisoning Attacks

Attacker may strategically **influence the training** of the model in order to **manipulate** the model decision.

Security

- **Input Forgery & Manipulation**
  - Forged documents
    - Payslips
    - Bank Statements
    - POs & Invoices

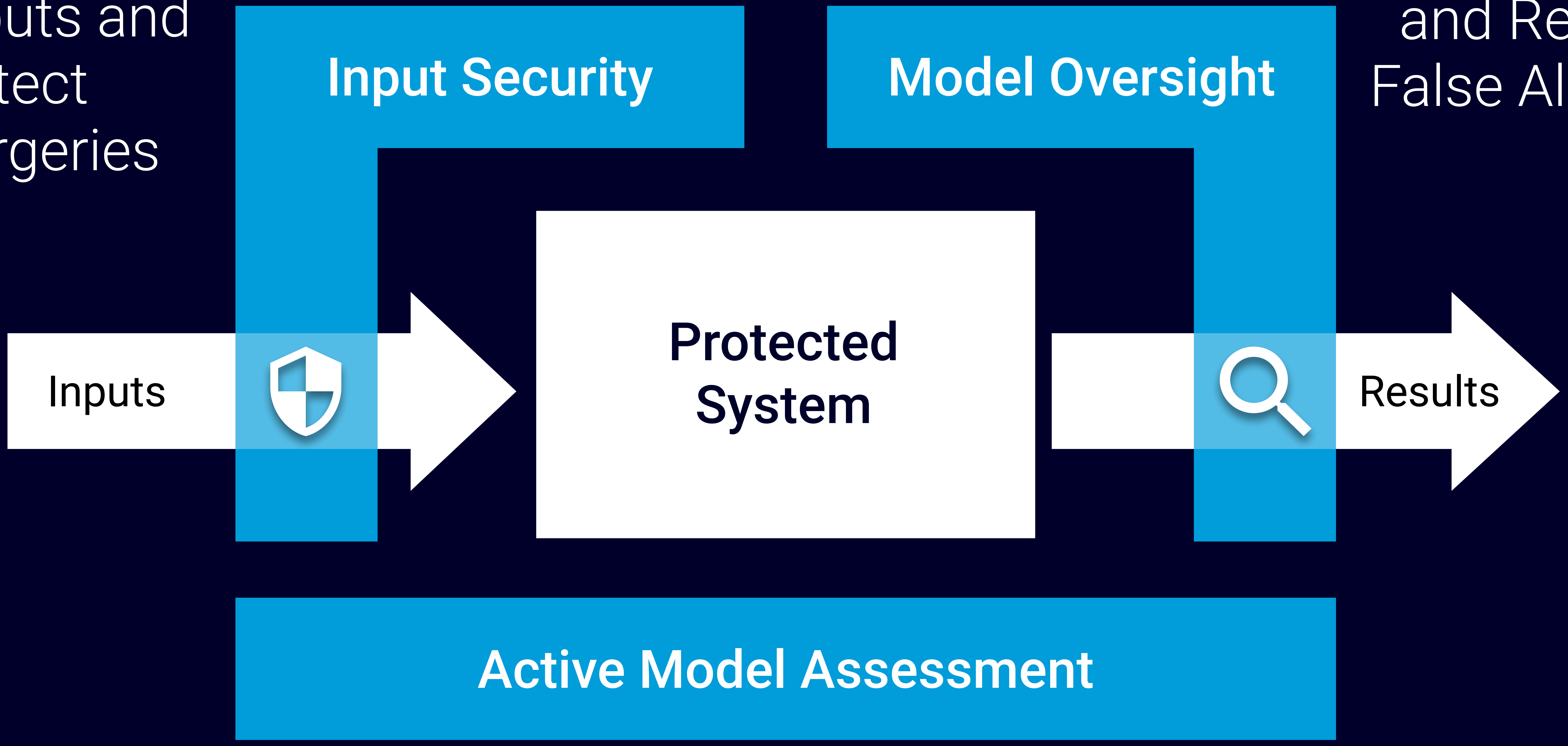
Risk

- **Fraud Detection** & **Anti-Money Laundering**
  - False Positive Reduction
  - Efficacy assessment, hardening and monitoring
- **Manipulation of Loan Decisions**
  - Evasion by misrepresentation
  - Poisoning by strategic behaviour

Our Product

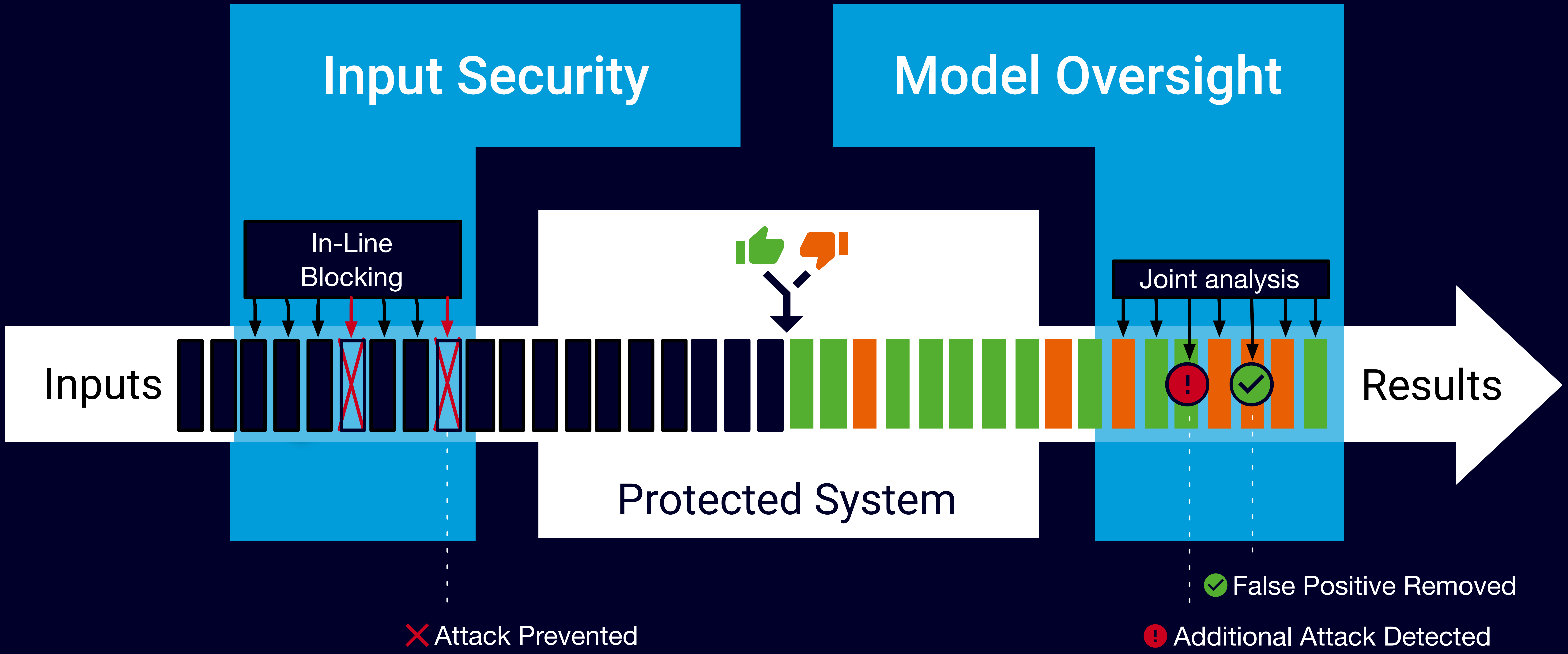
Prevent Bad Inputs and Detect Forgeries

Find Attacks and Reduce False Alarms



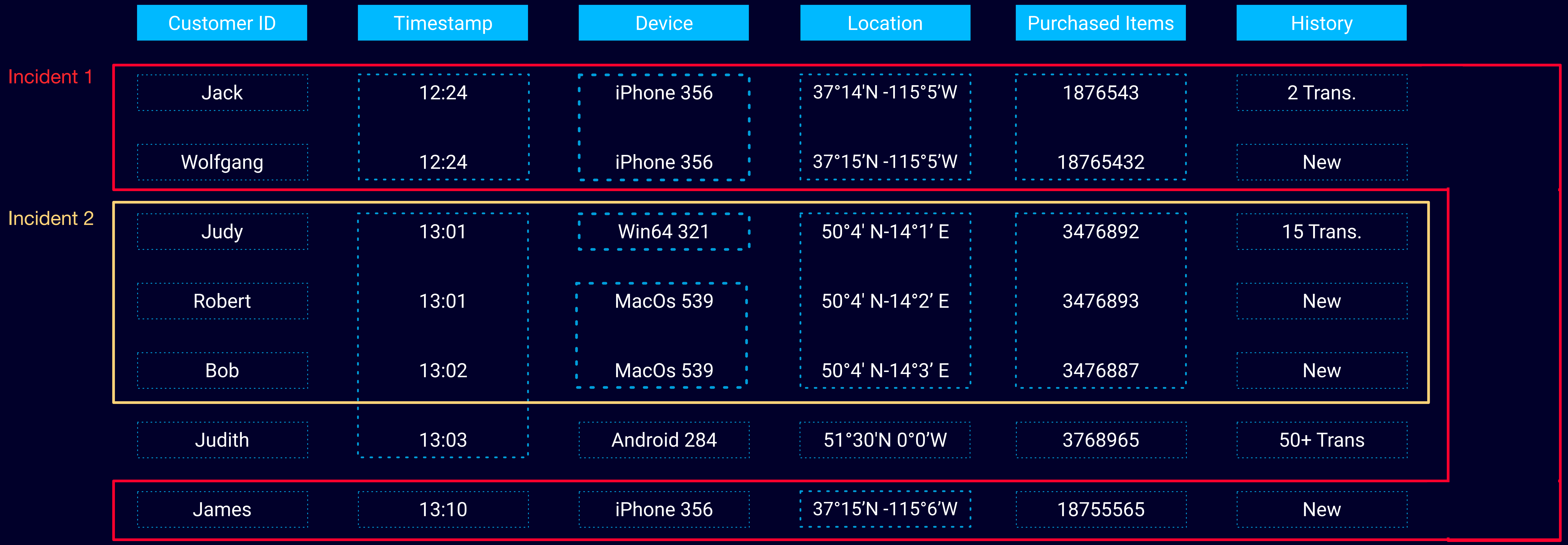
Find Vulnerabilities in AI Models

In-Line Deployment



We find attacks by reasoning about each transaction in the context of similar transactions.

Bulletproof AI Core Advantage



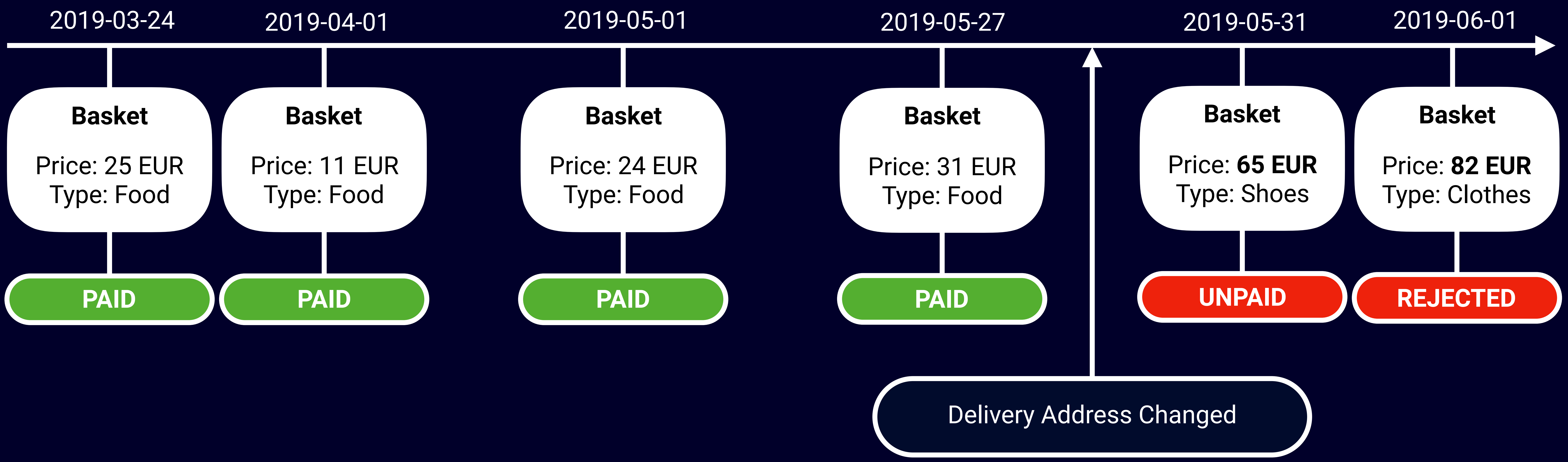
	Customer ID	Timestamp	Device	Location	Purchased Items	History
Incident 1	Jack	12:24	iPhone 356	37°14'N -115°5’W	1876543	2 Trans.
	Wolfgang	12:24	iPhone 356	37°15’N -115°5’W	18765432	New
Incident 2	Judy	13:01	Win64 321	50°4' N-14°1’ E	3476892	15 Trans.
	Robert	13:01	MacOs 539	50°4' N-14°2’ E	3476893	New
	Bob	13:02	MacOs 539	50°4' N-14°3’ E	3476887	New
	Judith	13:03	Android 284	51°30'N 0°0’W	3768965	50+ Trans
	James	13:10	iPhone 356	37°15’N -115°6’W	18755565	New

We achieve better results by **associating seemingly independent transactions** together to detect more attacks and tell the whole story.

Our differentiation is in the ability to model and **associate transactions at scale** and with an in-line-appropriate latency.

Detecting Account Takeovers

Features: Previous Baskets | Time series | Identity Change | Device Change



2019-03-24

Basket

Price: 25 EUR
Type: Food

PAID

2019-04-01

Basket

Price: 11 EUR
Type: Food

PAID

2019-05-01

Basket

Price: 24 EUR
Type: Food

PAID

2019-05-27

Basket

Price: 31 EUR
Type: Food

PAID

Delivery Address Changed

2019-05-31

Basket

Price: **65 EUR**
Type: Shoes

UNPAID

2019-06-01

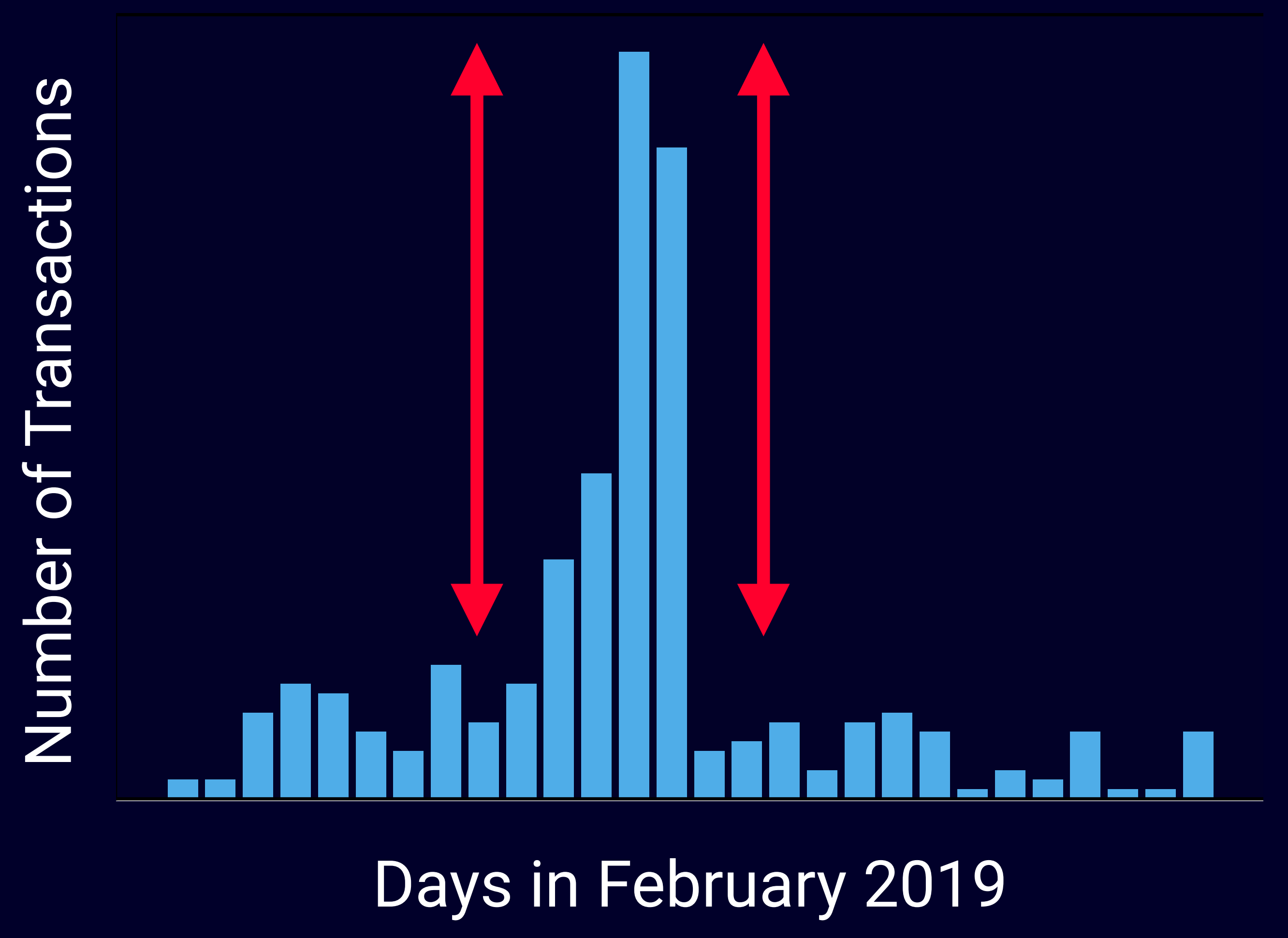
Basket

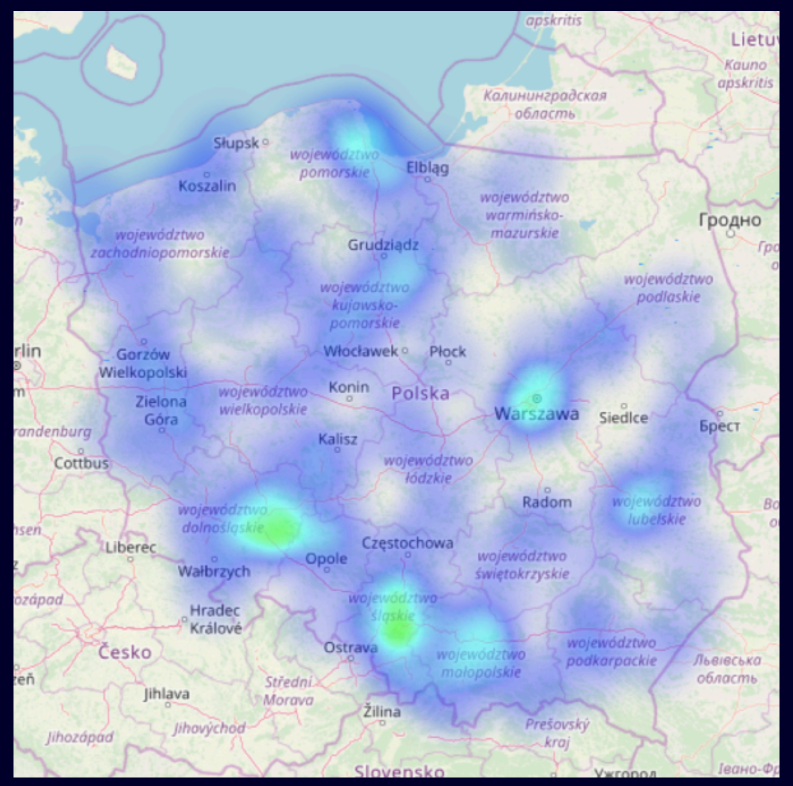
Price: **82 EUR**
Type: Clothes

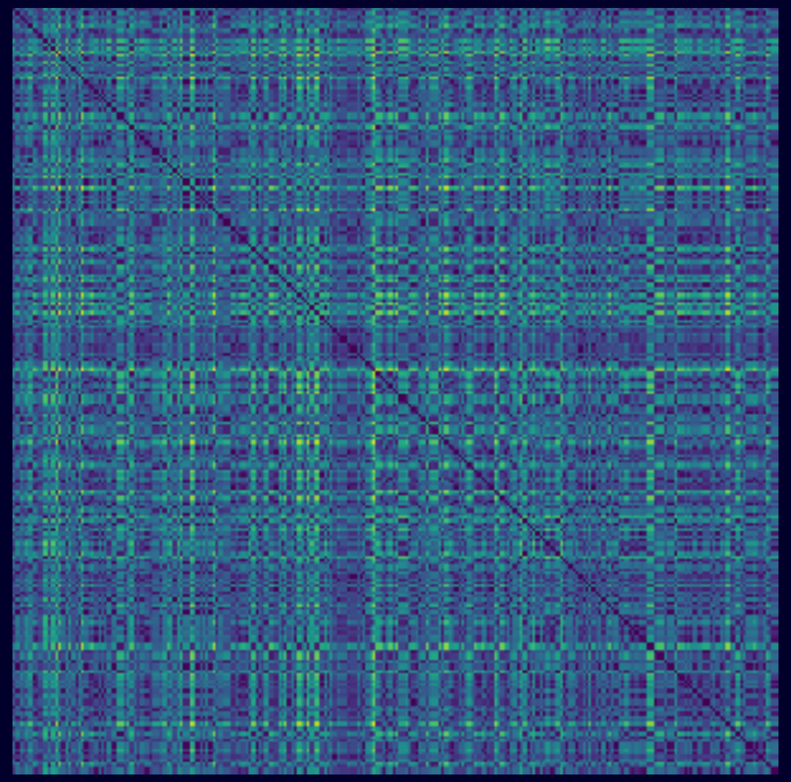
REJECTED

Reduction of False Positives

Buying Flowers for Valentine's Day





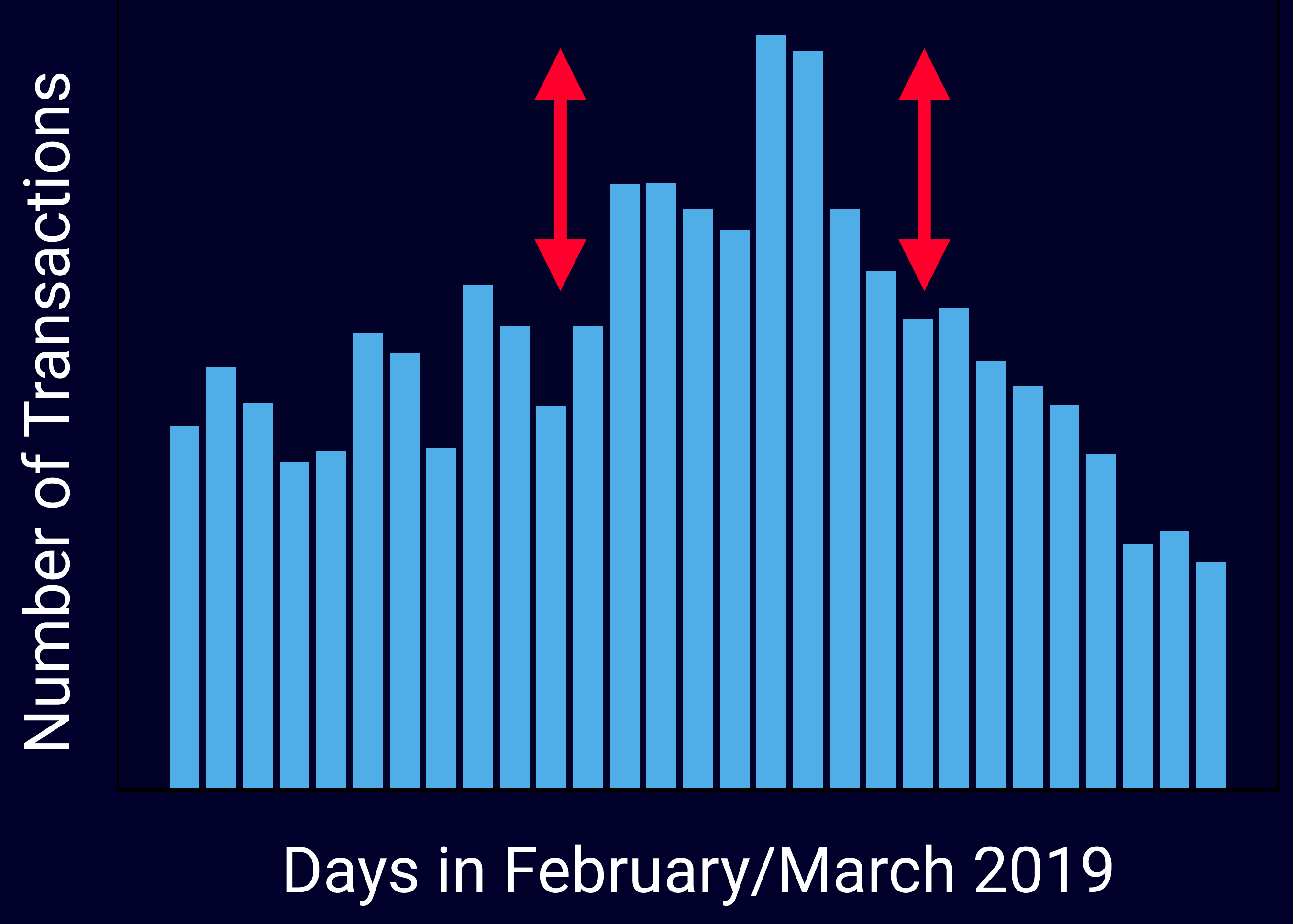


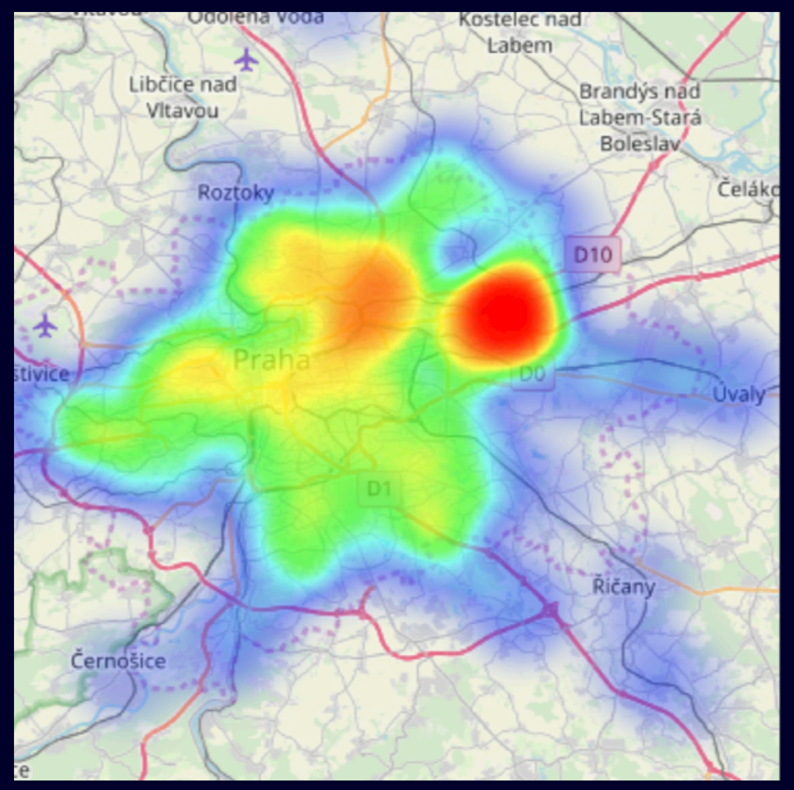
Widespread Locations

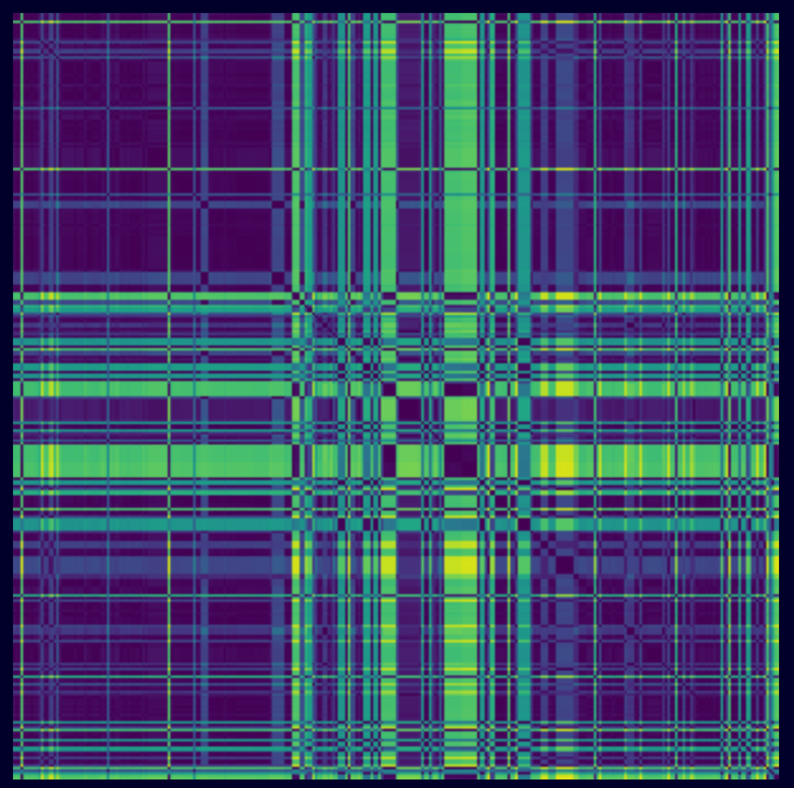
Weak Context Similarities

LEGITIMATE

Online Food Store Fraud







Isolated Locations

Large Context Similarities

FRAUD

Catching Evasive Fraud

Threat/Detectors	Identity-Based		Constraint/Location-Based	
	Sequence Analysis	Context Similarity	Transaction Structuring	Probing Identification
Account Takeover	■ ■ ■	■		
Bust-Out Fraud	■ ■	■ ■ ■		
Generated Identity Synthetic Identity			■ ■ ■	■
Approval Boundary Probing			■	■ ■ ■
False Positive Reduction Occasional Fraud	■ ■	■ ■		

Customer Outcome

85% alert volume reduction for fraud team

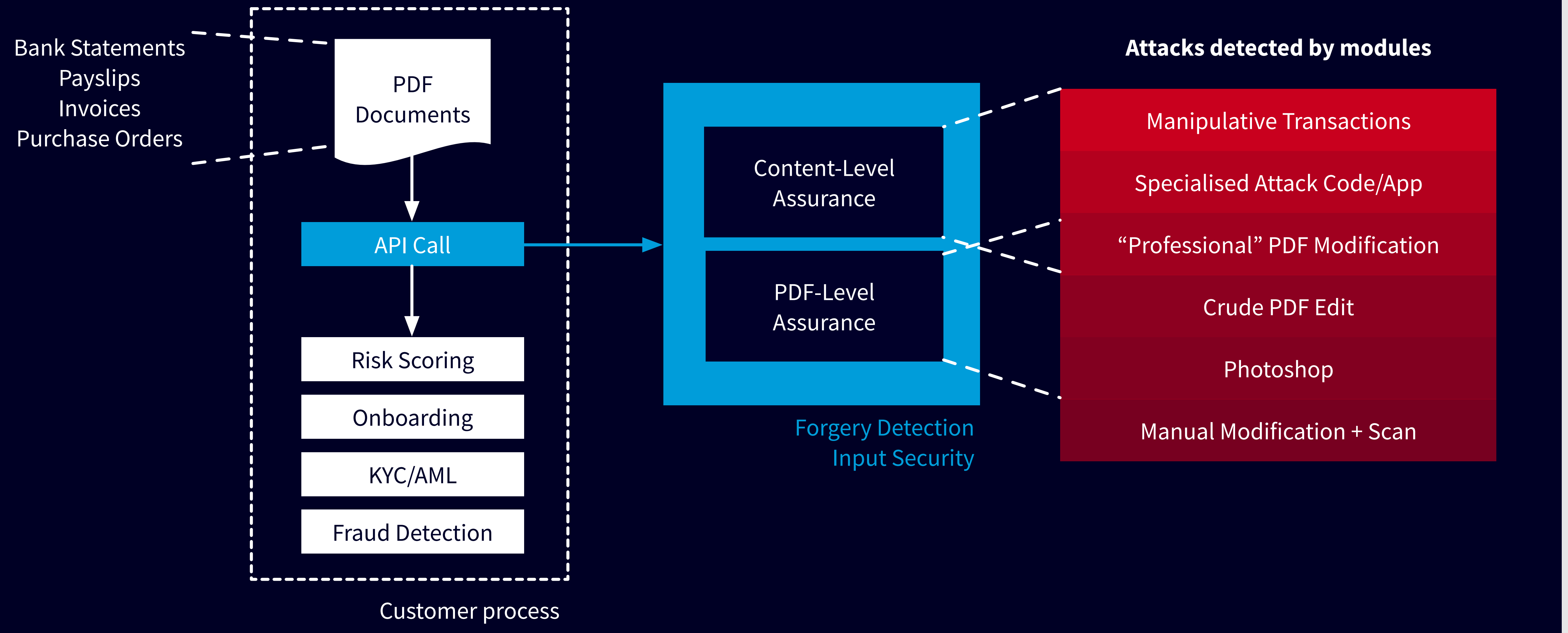
50% of fraud incidents auto-prevented before approval

15% of previously “non-fraud failures” identified as fraud

+ Better robustness against new attacks

Document Forgery Detection

PDF Forgery Detection



Bank Statements
Payslips
Invoices
Purchase Orders

PDF Documents

API Call

Risk Scoring

Onboarding

KYC/AML

Fraud Detection

Customer process

Content-Level Assurance

PDF-Level Assurance

Forgery Detection
Input Security

Attacks detected by modules

Manipulative Transactions

Specialised Attack Code/App

“Professional” PDF Modification

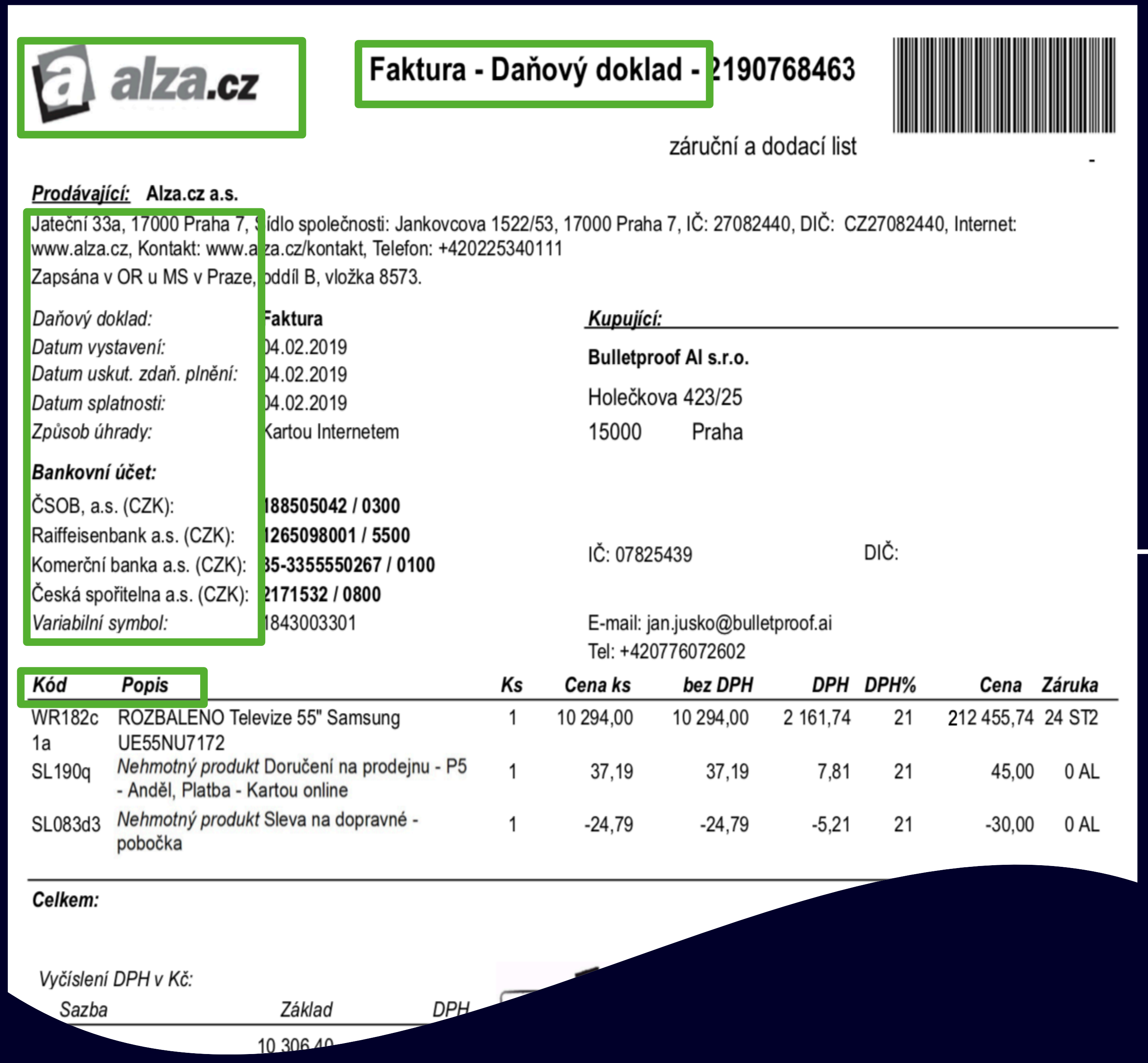
Crude PDF Edit

Photoshop

Manual Modification + Scan

- PDF documents are used in most financial processes:
  - Know Your Customer
  - Consumer Lending
  - Car Financing
  - SMB Lending & Factoring
- Information in the documents is rarely verified or authenticated
- We offer early **automated detection of modified & malicious PDF documents**

1. First Glance



alza.cz

Faktura - Daňový doklad - 2190768463

záruční a dodací list

-

Prodávající: **Alza.cz a.s.**

Jateční 33a, 17000 Praha 7, Sídlo společnosti: Jankovcova 1522/53, 17000 Praha 7, IČ: 27082440, DIČ: CZ27082440, Internet: www.alza.cz, Kontakt: www.alza.cz/kontakt, Telefon: +420225340111
Zapsána v OR u MS v Praze, oddíl B, vložka 8573.

Daňový doklad:	**Faktura**
Datum vystavení:	04.02.2019
Datum uskut. zdaň. plnění:	04.02.2019
Datum splatnosti:	04.02.2019
Způsob úhrady:	Kartou Internetem
Bankovní účet:	
ČSOB, a.s. (CZK):	**188505042 / 0300**
Raiffeisenbank a.s. (CZK):	**1265098001 / 5500**
Komerční banka a.s. (CZK):	**35-3355550267 / 0100**
Česká spořitelna a.s. (CZK):	**2171532 / 0800**
Variabilní symbol:	1843003301

Kupující:

Bulletproof AI s.r.o.
Holečkova 423/25
15000 Praha

IČ: 07825439 DIČ:

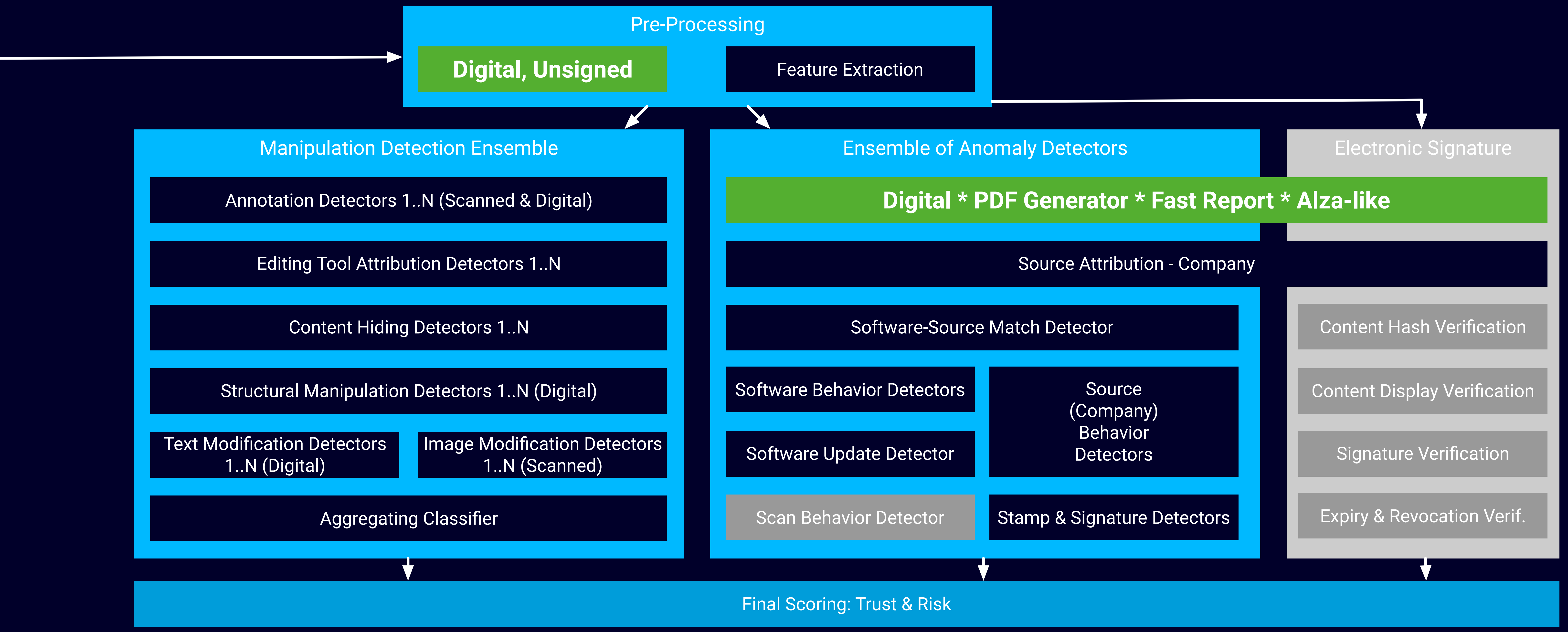
E-mail: jan.jusko@bulletproof.ai
Tel: +420776072602

Kód	*Popis*	*Ks*	*Cena ks*	*bez DPH*	*DPH*	*DPH%*	*Cena*	*Záruka*
WR182c1a	ROZBALENO Televize 55" Samsung UE55NU7172	1	10 294,00	10 294,00	2 161,74	21	212 455,74	24 ST2
SL190q	*Nehmotný produkt* Doručení na prodejnu - P5 - Anděl, Platba - Kartou online	1	37,19	37,19	7,81	21	45,00	0 AL
SL083d3	*Nehmotný produkt* Sleva na dopravné - pobočka	1	-24,79	-24,79	-5,21	21	-30,00	0 AL

Celkem:

Vyčíslení DPH v Kč:

Sazba	*Základ*	*DPH*
	10 306,40	



Pre-Processing

Digital, Unsigned

Feature Extraction

Manipulation Detection Ensemble

Annotation Detectors 1..N (Scanned & Digital)

Editing Tool Attribution Detectors 1..N

Content Hiding Detectors 1..N

Structural Manipulation Detectors 1..N (Digital)

Text Modification Detectors 1..N (Digital)

Image Modification Detectors 1..N (Scanned)

Aggregating Classifier

Ensemble of Anomaly Detectors

Digital * PDF Generator * Fast Report * Alza-like

Source Attribution - Company

Software-Source Match Detector

Software Behavior Detectors

Source (Company) Behavior Detectors

Software Update Detector

Scan Behavior Detector

Stamp & Signature Detectors

Electronic Signature

Content Hash Verification

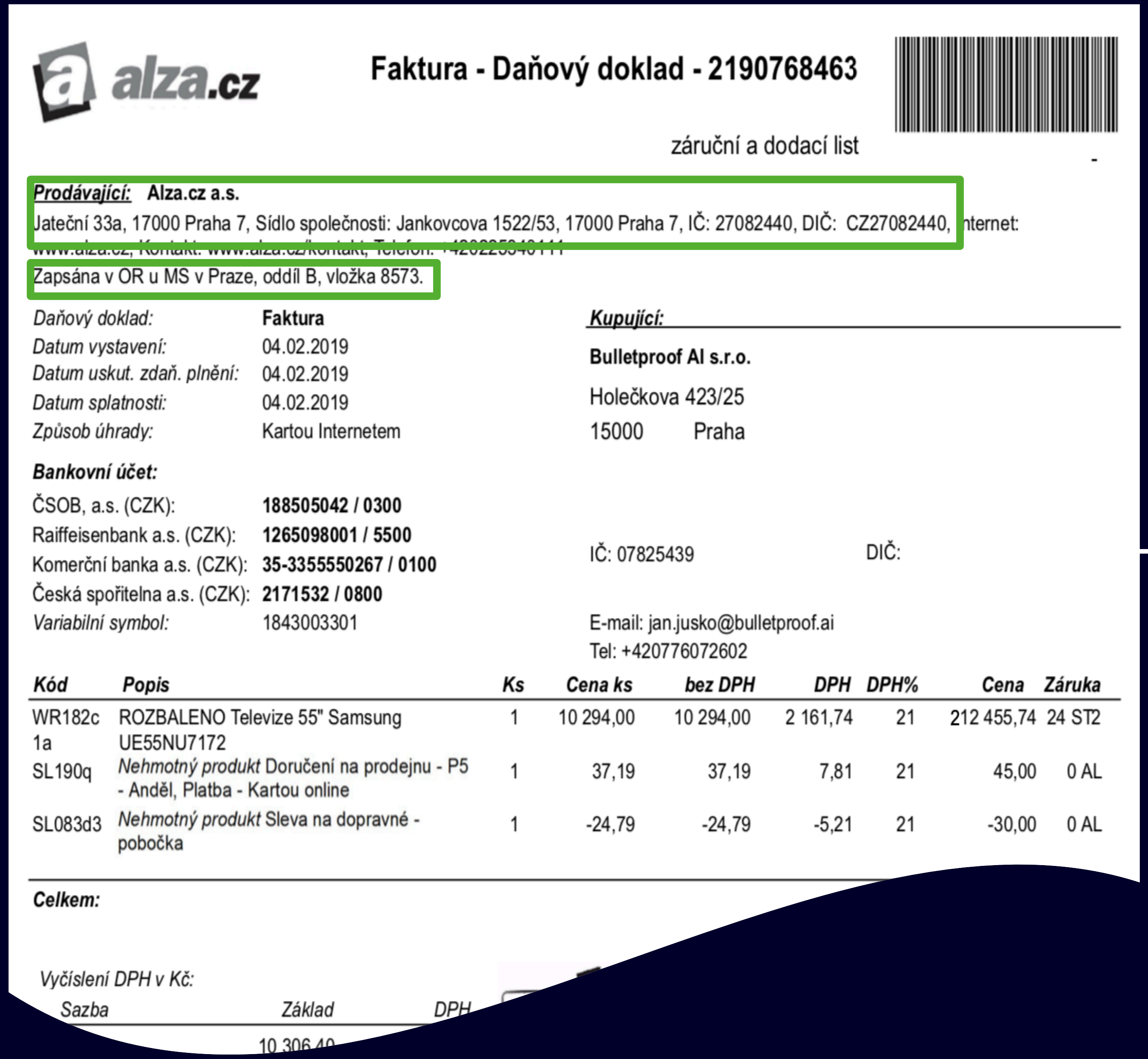
Content Display Verification

Signature Verification

Expiry & Revocation Verif.

Final Scoring: Trust & Risk

2. Origin and Source



alza.cz

Faktura - Daňový doklad - 2190768463

záruční a dodací list

-

Prodávající: **Alza.cz a.s.**

Jateční 33a, 17000 Praha 7, Sídlo společnosti: Jankovcova 1522/53, 17000 Praha 7, IČ: 27082440, DIČ: CZ27082440, Internet: www.alza.cz, Kontakt: www.alza.cz/kontakt, Telefon: +420225340111

Zapsána v OR u MS v Praze, oddíl B, vložka 8573.

Daňový doklad:	**Faktura**
Datum vystavení:	04.02.2019
Datum uskut. zdaň. plnění:	04.02.2019
Datum splatnosti:	04.02.2019
Způsob úhrady:	Kartou Internetem

Bankovní účet:

ČSOB, a.s. (CZK):	**188505042 / 0300**
Raiffeisenbank a.s. (CZK):	**1265098001 / 5500**
Komerční banka a.s. (CZK):	**35-3355550267 / 0100**
Česká spořitelna a.s. (CZK):	**2171532 / 0800**
Variabilní symbol:	1843003301

Kupující:

Bulletproof AI s.r.o.

Holečkova 423/25
15000 Praha

IČ: 07825439 DIČ:

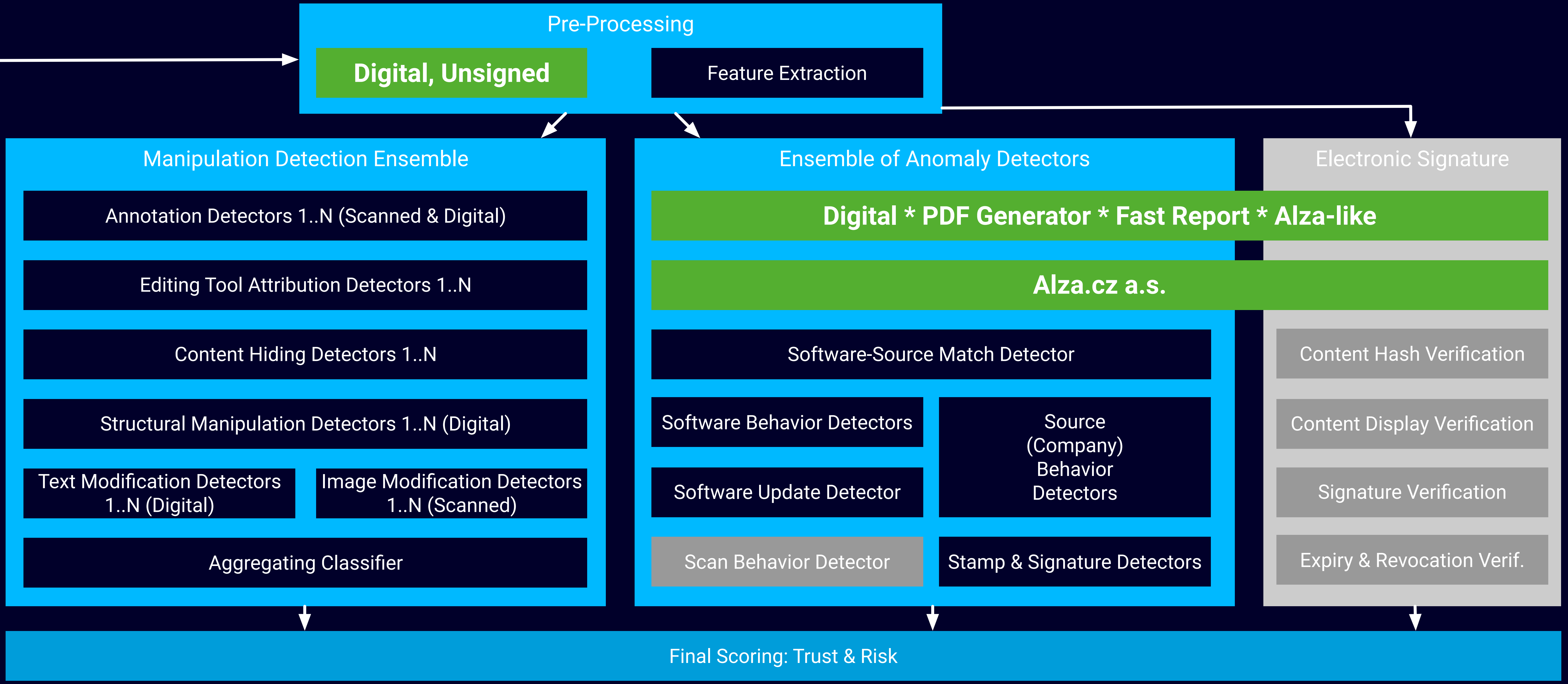
E-mail: jan.jusko@bulletproof.ai
Tel: +420776072602

Kód	*Popis*	*Ks*	*Cena ks*	*bez DPH*	*DPH*	*DPH%*	*Cena*	*Záruka*
WR182c 1a	ROZBALENO Televize 55" Samsung UE55NU7172	1	10 294,00	10 294,00	2 161,74	21	212 455,74	24 ST2
SL190q	*Nehmotný produkt* Doručení na prodejnu - P5 - Anděl, Platba - Kartou online	1	37,19	37,19	7,81	21	45,00	0 AL
SL083d3	*Nehmotný produkt* Sleva na dopravné - pobočka	1	-24,79	-24,79	-5,21	21	-30,00	0 AL

Celkem:

Vyčíslení DPH v Kč:

Sazba	*Základ*	*DPH*
	10 306,40	



Pre-Processing

Digital, Unsigned

Feature Extraction

Manipulation Detection Ensemble

- Annotation Detectors 1..N (Scanned & Digital)
- Editing Tool Attribution Detectors 1..N
- Content Hiding Detectors 1..N
- Structural Manipulation Detectors 1..N (Digital)
- Text Modification Detectors 1..N (Digital)
- Image Modification Detectors 1..N (Scanned)
- Aggregating Classifier

Ensemble of Anomaly Detectors

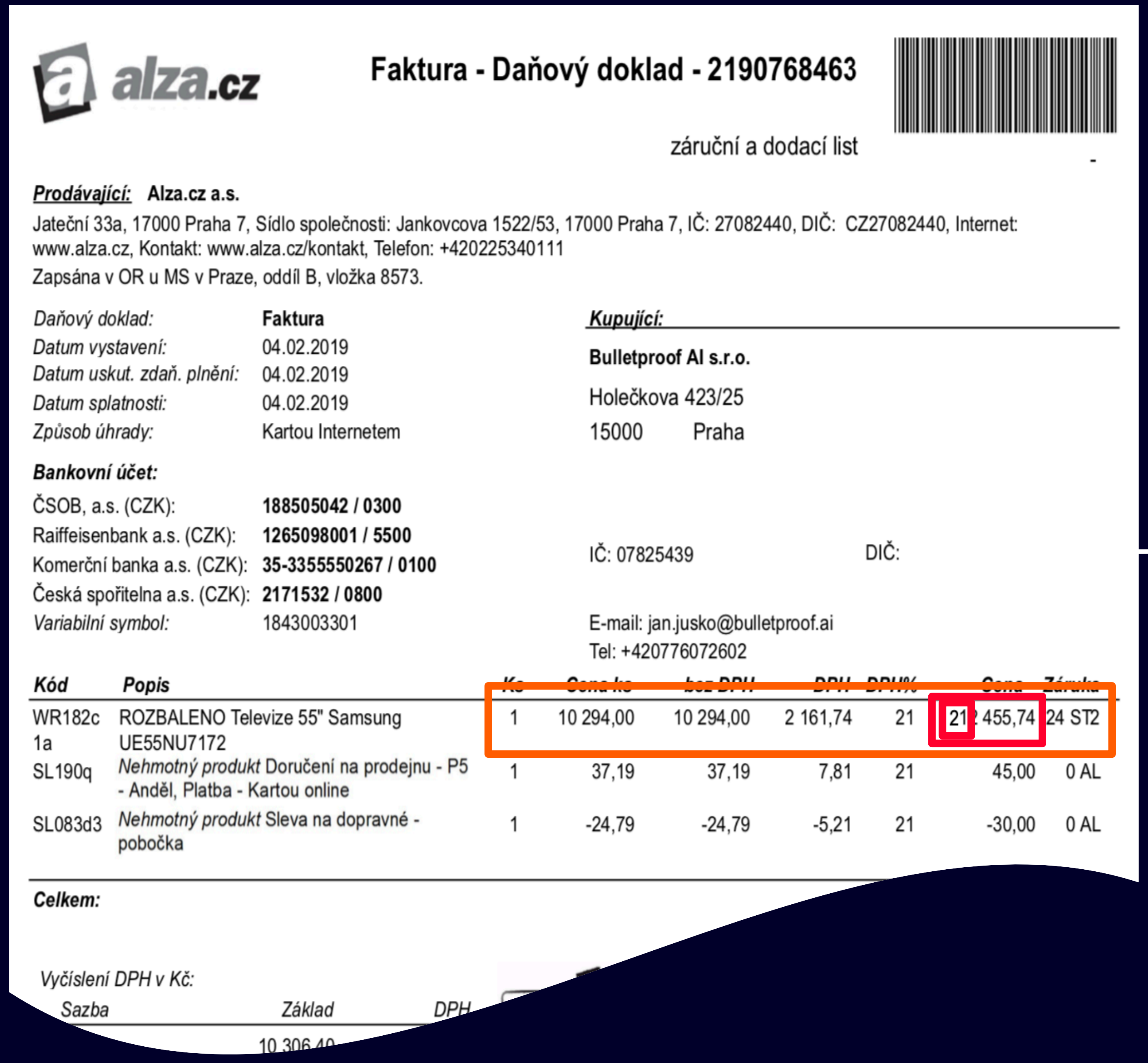
- **Digital * PDF Generator * Fast Report * Alza-like**
- **Alza.cz a.s.**
- Software-Source Match Detector
- Software Behavior Detectors
- Source (Company) Behavior Detectors
- Software Update Detector
- Scan Behavior Detector
- Stamp & Signature Detectors

Electronic Signature

- Content Hash Verification
- Content Display Verification
- Signature Verification
- Expiry & Revocation Verif.

Final Scoring: Trust & Risk

3. Detecting the Forgery



alza.cz **Faktura - Daňový doklad - 2190768463**

záruční a dodací list

-

Prodávající: **Alza.cz a.s.**

Jateční 33a, 17000 Praha 7, Sídlo společnosti: Jankovcova 1522/53, 17000 Praha 7, IČ: 27082440, DIČ: CZ27082440, Internet: www.alza.cz, Kontakt: www.alza.cz/kontakt, Telefon: +420225340111

Zapsána v OR u MS v Praze, oddíl B, vložka 8573.

Daňový doklad:	**Faktura**
Datum vystavení:	04.02.2019
Datum uskut. zdaň. plnění:	04.02.2019
Datum splatnosti:	04.02.2019
Způsob úhrady:	Kartou Internetem

Kupující:

Bulletproof AI s.r.o.

Holečkova 423/25

15000 Praha

Bankovní účet:

ČSOB, a.s. (CZK):	**188505042 / 0300**
Raiffeisenbank a.s. (CZK):	**1265098001 / 5500**
Komerční banka a.s. (CZK):	**35-3355550267 / 0100**
Česká spořitelna a.s. (CZK):	**2171532 / 0800**
Variabilní symbol:	1843003301

IČ: 07825439 DIČ:

E-mail: jan.jusko@bulletproof.ai

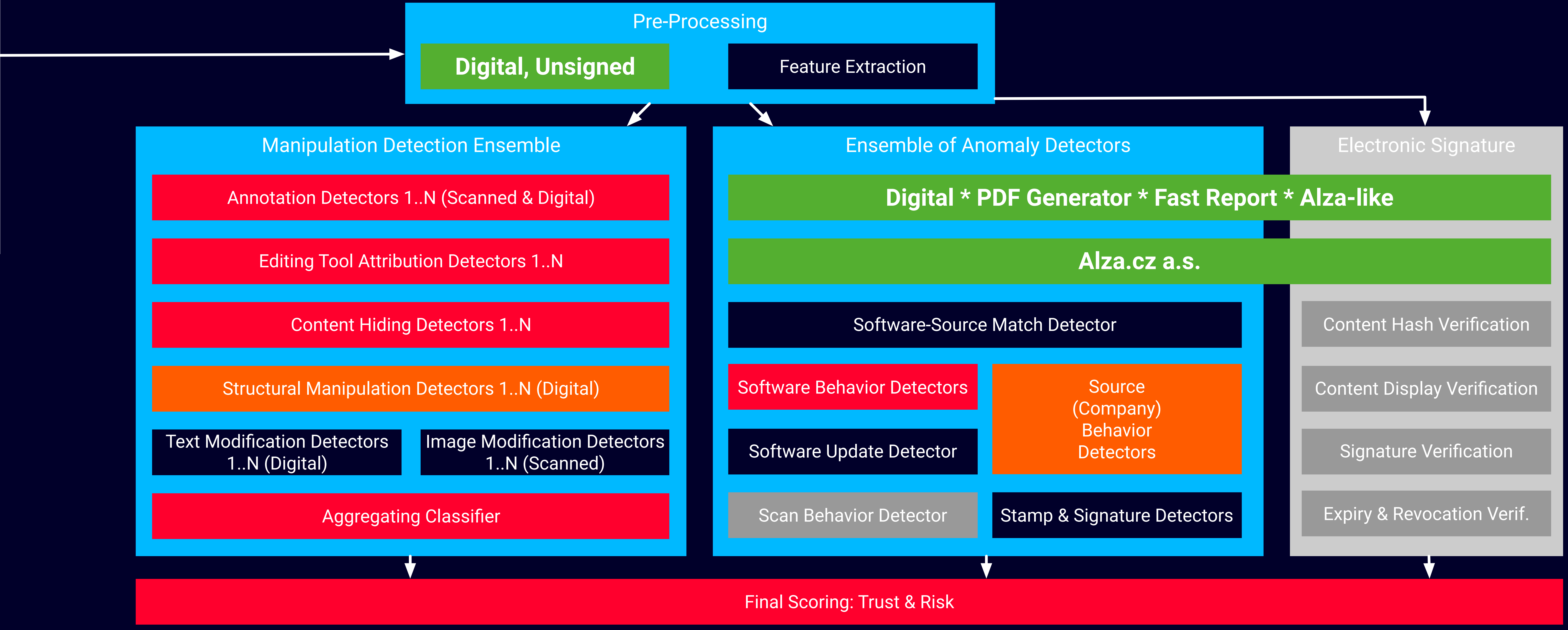
Tel: +420776072602

Kód	*Popis*	*Ks*	*Cena ks*	*bez DPH*	*DPH*	*DPH%*	*Cena*	*Záruka*
WR182c 1a	ROZBALENO Televize 55" Samsung UE55NU7172	1	10 294,00	10 294,00	2 161,74	21	212 455,74	24 ST2
SL190q	*Nehmotný produkt* Doručení na prodejnu - P5 - Anděl, Platba - Kartou online	1	37,19	37,19	7,81	21	45,00	0 AL
SL083d3	*Nehmotný produkt* Sleva na dopravné - pobočka	1	-24,79	-24,79	-5,21	21	-30,00	0 AL

Celkem:

Vyčíslení DPH v Kč:

Sazba	*Základ*	*DPH*
	10 306,40	



Pre-Processing

Digital, Unsigned

Feature Extraction

Manipulation Detection Ensemble

- Annotation Detectors 1..N (Scanned & Digital)
- Editing Tool Attribution Detectors 1..N
- Content Hiding Detectors 1..N
- Structural Manipulation Detectors 1..N (Digital)
- Text Modification Detectors 1..N (Digital)
- Image Modification Detectors 1..N (Scanned)
- Aggregating Classifier

Ensemble of Anomaly Detectors

- **Digital * PDF Generator * Fast Report * Alza-like**
- **Alza.cz a.s.**
- Software-Source Match Detector
- Software Behavior Detectors
- Source (Company) Behavior Detectors
- Software Update Detector
- Scan Behavior Detector
- Stamp & Signature Detectors

Electronic Signature

- Content Hash Verification
- Content Display Verification
- Signature Verification
- Expiry & Revocation Verif.

Final Scoring: Trust & Risk

Demo:

Internal Fraud

The tale of two monitors

LG 27"27MD5KL

EUR 1300

- For the price of 3

LG 27" 27UK850

EUR 550

- You can get 7

- …and end up with 4 monitors

BULLETPROOF.AI

Prague & Brussels

sales@bulletproof.AI, +420 737 113 153