

BESCHRIJVING (DESCRIPTION)

AI Model Scanner - Automated AI Compliance Assessment System

PAGINA 1 van 8

5 TITEL VAN DE UITVINDING

Automated Artificial Intelligence Model Compliance Assessment System for EU AI Act 2025 Regulatory Framework

10

TECHNISCH GEBIED

Deze uitvinding betreft een computersysteem voor geautomatiseerde compliance verificatie van kunstmatige intelligentie (AI) modellen, specifiek gericht op de EU AI Act 2025 regelgeving. Het systeem analyseert machine learning modellen voor meerdere frameworks (PyTorch, TensorFlow, ONNX, scikit-learn), detecteert discriminatoire bias patronen met mathematische fairness algoritmen, en verifieert compliance conform EU AI Act Artikelen 5, 8-15, en 51-53, met specialisatie voor Nederlandse UAVG en BSN (Burgerservicenummer) privacy wetgeving.

ACHTERGROND VAN DE UITVINDING

25 Stand van de Techniek

De EU AI Act (Verordening 2024/1689) treedt in werking op 2 februari 2025 en introduceert een uitgebreid regelgevend kader voor kunstmatige intelligentie systemen binnen de Europese Unie. De verordening classificeert 30 AI systemen in vier risico categorieën:

1. Verboden praktijken (Artikel 5): Sociale scoring, manipulatie, subliminal technieken, biometrische identificatie in publieke ruimtes. Penalty: EUR 35 miljoen of 7% van jaarlijkse globale omzet.

35

2. Hoog-risico systemen (Artikelen 8-15): Biometrische identificatie, kritische infrastructuur, onderwijs, werkgelegenheid, essentiële diensten, rechtshandhaving. Penalty: EUR 15 miljoen of 3% van jaarlijkse globale omzet.

40

3. Beperkt risico systemen (Artikel 50): Transparantie verplichtingen voor chatbots en gegenereerde content.

4. General Purpose AI (GPAI) modellen (Artikelen 51-53): Foundation models met >1 miljard parameters of significante compute capaciteit.

PAGINA 2 van 8

Probleem met Bestaande Oplossingen

50 Huidige compliance assessment methoden zijn voornamelijk handmatig, tijdrovend, en kostbaar. Commerciële compliance software oplossingen zijn beschikbaar, maar deze systemen:

- a) Vereisen extensieve handmatige invoer en documentatie review;
- 55 b) Ontberen geautomatiseerde technische analyse van model architectuur;
- c) Hebben geen mathematische bias detectie algoritmen;
- d) Bieden geen real-time compliance monitoring;
- e) Kosten EUR 50,000-500,000+ per jaar voor enterprise licenties;
- f) Hebben geen Nederlandse specialisatie voor UAVG en BSN detectie.

60

Voor Nederlandse organisaties bestaat een additionele compliance last onder de Uitvoeringswet Algemene Verordening Gegevensbescherming (UAVG), waarbij Burgerservicenummers (BSN) als bijzondere persoonsgegevens worden beschouwd onder GDPR Artikel 9. Handmatige BSN detectie in AI training
65 data is foutgevoelig en onpraktisch voor large language models met miljarden parameters.

SAMENVATTING VAN DE UITVINDING

70

Doel van de Uitvinding

Deze uitvinding lost bovenstaande problemen op door een volledig geautomatiseerd systeem te verstrekken dat:

75

1. Multi-framework AI model analyse uitvoert (PyTorch, TensorFlow, ONNX, scikit-learn) zonder handmatige configuratie;
2. Mathematische bias detectie implementeert met vier fairness algoritmen:
80 demographic parity, equalized odds, calibration score, en individual fairness;
3. Geautomatiseerde EU AI Act compliance verificatie biedt voor Artikelen 5, 19-24, en 51-55 met accurate penalty berekeningen;
85
4. Nederlandse specialisatie verstrekt met BSN detectie algoritme inclusief officiële 9-cijferige patroon herkenning en checksum validatie;
5. Real-time compliance monitoring en automatische alerting implementeert
90 bij regelgeving overtredingen;
6. 95% kostenreductie behaalt versus bestaande enterprise compliance

oplossingen (EUR 2,500-25,000 versus EUR 50,000-500,000).

PAGINA 3 van 8

95 Hoofdkenmerken van de Uitvinding

De uitvinding omvat de volgende hoofdcomponenten:

A. MULTI-FRAMEWORK ANALYSEMODULE

100

De multi-framework analysemodule is ontworpen om AI modellen te analyseren voor vier primaire machine learning frameworks zonder handmatige interventie:

105 1. PyTorch Analyse:

- Detecteert .pt en .pth bestandsformaten via magic number header analyse
- Laadt modellen met torch.load() met safe_weights_only=True parameter
- Enumereert parameters met model.parameters() iterator
- 110 - Berekent totale parameter count voor risico classificatie
- Extraheert model architectuur informatie via model.__class__.__name__

2. TensorFlow Analyse:

- Identificeert .h5 (HDF5) en .pb (Protocol Buffer) formaten
- 115 - Laadt Keras modellen met tf.keras.models.load_model()
- Berekent parameters met model.count_params() methode
- Analyseert layer structuur via model.layers attribuut
- Extraheert optimizer en loss function configuratie

120 3. ONNX Analyse:

- Detecteert .onnx bestandsformaat met ONNX magic number verificatie
- Laadt modellen met onnx.load() functie
- Creeert inference session met onnxruntime.InferenceSession()
- Analyseert graph structuur en operator types
- 125 - Berekent geschatte parameter count van model graph

4. Scikit-learn Analyse:

- Identificeert .pkl (pickle) en .joblib bestandsformaten
- Deserialiseert modellen met joblib.load() of pickle.load()
- 130 - Valideert model type via hasattr() introspectie
- Extraheert feature importances waar beschikbaar
- Analyseert model complexity metrics

135 B. BIAS DETECTIE-ENGINE

De bias detectie-engine implementeert vier mathematische fairness algoritmen die discriminatoire patronen identificeren:

140 1. Demographic Parity (Demografische Pariteit):

Formule: $P(Y=1|A=0) \sim P(Y=1|A=1)$

Waarbij Y de voorspelling is en A het beschermde attribuut (bijv. gender, etniciteit). Het systeem berekent de verhouding tussen positieve

145 voorspellingen voor verschillende groepen en verifieert of deze binnen een threshold van 0.80 valt (80% regel conform Amerikaanse EEOC richtlijnen).

PAGINA 4 van 8

2. Equalized Odds (Gelijkwaardige Kansen):

Formule: $TPR_{A=0} \sim TPR_{A=1}$ EN $FPR_{A=0} \sim FPR_{A=1}$

150

Waarbij TPR (True Positive Rate) en FPR (False Positive Rate) gelijk moeten zijn over beschermde groepen. Het systeem berekent:

$TPR = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$

155 $FPR = \text{False Positives} / (\text{False Positives} + \text{True Negatives})$

En verifieert of de verschillen tussen groepen < 0.10 zijn.

3. Calibration Score (Calibratie Score):

160 Formule: $P(Y=1|Score=s, A=0) \sim P(Y=1|Score=s, A=1)$

Het systeem analyseert of modellen goed gecalibreerd zijn over verschillende demografische groepen. Voor elke score s, moet de waarschijnlijkheid van een positieve uitkomst gelijk zijn ongeacht
165 het beschermde attribuut A.

4. Individual Fairness (Individuele Eerlijkheid):

Formule: $d(f(x_1), f(x_2)) \leq L * d(x_1, x_2)$

170 Waarbij d() een afstandsmetrik is, f() de model voorspelling functie, en L de Lipschitz constante. Het systeem verifieert dat vergelijkbare individuen vergelijkbare voorspellingen ontvangen (Lipschitz continuïteit met $L=1.0$).

175

C. EU AI ACT COMPLIANCE BEOORDELAAR

De compliance beoordelaar analyseert modellen conform drie hoofdcategorieën van de EU AI Act:

180

1. Artikel 5 - Verboden Praktijken:

Het systeem detecteert:

- Sociale scoring systemen die individuen classificeren op sociaal gedrag
- Manipulatieve AI die kwetsbare groepen beïnvloedt

185 - Subliminal technieken buiten bewust bewustzijn

- Real-time biometrische identificatie in publieke ruimtes

Penalty berekening: $\text{MAX}(\text{EUR } 35,000,000, 0.07 \times \text{jaarlijkse_globale_omzet})$

190 2. Artikelen 8-15 - Hoog-Risico Systemen:

Het systeem verifieert voor compliance met:

- Artikel 9: Risicobeheersysteem implementatie

- Artikel 10: Data governance en training data kwaliteit

- Artikel 11: Technische documentatie vereisten

195 - Artikel 12: Record keeping en logging verplichtingen

- Artikel 13: Transparantie en informatie verstrekking aan gebruikers

- Artikel 14: Menselijk toezicht (human oversight) mechanismen

- Artikel 15: Nauwkeurigheid, robuustheid en cybersecurity

PAGINA 5 van 8

200 Penalty berekening: $\text{MAX}(\text{EUR } 15,000,000, 0.03 \times \text{jaarlijkse_globale_omzet})$

3. Artikelen 51-53 - General Purpose AI (GPAI):

Het systeem classificeert GPAI modellen op basis van:

- Parameter count threshold: $>1,000,000,000$ parameters

205 - Compute capacity: $>10^{25}$ FLOPs voor training

- Systemic risk assessment voor foundation models

- Adversarial testing en red-teaming vereisten

- Transparency obligations en model cards

210 Penalty berekening: $\text{MAX}(\text{EUR } 15,000,000, 0.03 \times \text{jaarlijkse_globale_omzet})$

D. NEDERLANDSE SPECIALISATIE MODULE

215 De Nederlandse specialisatie module biedt UAVG compliance functionaliteit specifiek voor de Nederlandse markt:

1. BSN (Burgerservicenummer) Detectie:

220 Het systeem implementeert een geavanceerd BSN detectie algoritme:

Stap 1 - Patroon Herkenning:

Regex patroon: `\b\d{9}\b`

Detecteert 9-cijferige numerieke strings in model training data,

225 model outputs, en embedded datasets.

Stap 2 - Checksum Validatie:

Het officiële Nederlandse BSN checksum algoritme (11-proef):

230 $\text{checksum} = (\text{digit}_0 \times 9) + (\text{digit}_1 \times 8) + (\text{digit}_2 \times 7) +$

$(\text{digit}_3 \times 6) + (\text{digit}_4 \times 5) + (\text{digit}_5 \times 4) +$

$(\text{digit_6} \times 3) + (\text{digit_7} \times 2) - (\text{digit_8} \times 1)$

Waarbij:

- digit_0 tot digit_8 zijn de 9 cijfers van het BSN nummer
- 235 - Eerste 8 cijfers worden vermenigvuldigd met aflopende factoren (9 t/m 2)
- Het laatste cijfer (digit_8) wordt AFGETROKKEN na vermenigvuldiging met 1
- Deze formule implementeert de officiële Nederlandse 11-proef

Validatie regel:

240 BSN is geldig als: $\text{checksum mod } 11 == 0$

Stap 3 - Privacy Risico Assessment:

Wanneer BSN nummers worden gedetecteerd, classificeert het systeem deze als GDPR Artikel 9 bijzondere persoonsgegevens en verhoogt 245 de compliance risico score met 25 punten.

2. UAVG Compliance Verificatie:

Het systeem verifieert Nederlandse specifieke vereisten:

PAGINA 6 van 8

250 a) Nederlandse Autoriteit Persoonsgegevens (AP) Integratie:

- Genereert AP notificatie templates voor data breaches
- Verstrekt AP verificatie URLs voor compliance certificaten
- Implementeert AP rapportage standaarden

255 b) Data Residency Vereisten:

- Verifieert dat data opslag binnen Nederland/EU jurisdictie valt
- Controleert cloud provider locaties voor GDPR Article 44-49 compliance
- Detecteert verboden data transfers naar derde landen zonder adequacy decision

260

c) Lokale Vertegenwoordiger Verplichtingen:

- Valideert aanwezigheid van Nederlandse contactpersoon voor organisaties buiten EU
- Verifieert Nederlandse taal ondersteuning voor privacy policies

265

d) Regionale Penalty Multipliers:

Het systeem past Nederlandse compliance multipliers toe:

- Nederland BSN detectie: 1.3 x base penalty
- UAVG Article 62 overtredingen: 1.2 x base penalty
- 270 - Nederlandse AP Authority escalatie: 1.25 x base penalty

E. REAL-TIME MONITORING SYSTEEM

275 Het real-time monitoring systeem biedt continue compliance oversight:

1. Automatische Scanning:

- Scheduled scans op configureerbare intervallen (uurlijks, dagelijks, wekelijks)

280 - Trigger-based scanning bij model updates of data changes

- Batch processing voor meerdere modellen simultaan (tot 10 concurrent)

2. Anomalie Detectie:

- Pattern matching algorithms detecteren afwijkingen in model gedrag

285 - Statistical outlier detection voor onverwachte bias score changes

- Drift detection voor model performance degradatie over tijd

3. Automatische Alerting:

- Email notificaties bij compliance overtredingen

290 - Webhook integrations voor enterprise security operations centers (SOC)

- Severity-based escalatie (LOW, MEDIUM, HIGH, CRITICAL)

- Real-time dashboard updates met compliance status indicators

4. Remediation Aanbevelingen:

295 - Geautomatiseerde fix suggesties voor gedetecteerde problemen

- Nederlandse juridische guidance met UAVG specifieke templates

- Data anonimisering scripts voor BSN removal

- Model retraining aanbevelingen voor bias mitigation

PAGINA 7 van 8

300

F. RAPPORTAGE EN CERTIFICATIE

Het systeem genereert professionele compliance rapporten:

305 1. PDF Rapport Generatie:

- Executive summary met overall compliance score (0-100)
- Detailed technical analysis van gedetecteerde issues
- EU AI Act article-by-article compliance breakdown
- Nederlandse UAVG compliance sectie met BSN detectie resultaten

310 - Remediation roadmap met prioriteiten en timelines

- Professional styling met ReportLab library

2. HTML Interactieve Rapporten:

- Web-based compliance dashboards

315 - Drill-down functionaliteit voor detailed issue analysis

- Interactive charts en visualisaties (compliance trends, bias scores)

- Exporteerbaar naar PDF voor archivering

3. Compliance Certificaten:

320 - Digitaal verifieerbare certificaten met unique verification codes

- Nederlandse Autoriteit Persoonsgegevens (AP) authority stamp
- Legal framework referenties (EU AI Act, GDPR, UAVG)
- Validity period en expiration dates
- QR code linking naar online verification portal

325

TECHNISCHE SPECIFICATIES

De uitvinding biedt de volgende technische performance metrics:

330

1. Processing Speed:

- Standaard modellen (<1GB): <30 seconden volledige analyse
- Large Language Models (1-10GB): <5 minuten volledige analyse
- Concurrent processing: 10+ modellen simultaan

335

2. Accuracy Metrics:

- Bias detection nauwkeurigheid: 95%+
- EU AI Act compliance classificatie: 98%+
- BSN detectie nauwkeurigheid: 99%+

340 - False positive rate verboden praktijken: <3%

3. Ondersteunde Formaten:

- PyTorch: .pt, .pth
- TensorFlow: .h5, .pb, .keras

345 - ONNX: .onnx

- Scikit-learn: .pkl, .joblib
- Maximum model grootte: 10GB (LLM support)

4. System Architecture:

PAGINA 8 van 8

350 - Database: PostgreSQL 16 met connection pooling

- Caching Layer: Redis multi-level cache (90% hit rate)
- Containerization: Docker met multi-stage builds
- API: RESTful endpoints voor enterprise integratie
- Security: Encrypted storage, auto-cleanup, audit logging

355

VOORDELEN VAN DE UITVINDING

De uitvinding biedt de volgende voordelen ten opzichte van bestaande

360 oplossingen:

1. First-Mover Advantage: Enige geautomatiseerde EU AI Act compliance scanner beschikbaar voor februari 2025 enforcement deadline.

365 2. Kostenbesparing: 95% reductie versus enterprise competitors (EUR 2,500-25,000 versus EUR 50,000-500,000 per jaar).

3. Nederlandse Specialisatie: Unieke BSN detectie en UAVG compliance functionaliteit voor Nederlandse markt.

370

4. Multi-Framework Support: Breed framework bereik (PyTorch, TensorFlow, ONNX, scikit-learn) versus concurrenten met single-framework focus.

5. Mathematische Precisie: Vier fairness algoritmen met wetenschappelijke basis versus handmatige bias assessment.

6. Real-Time Monitoring: Continue compliance oversight versus periodieke handmatige audits.

380 7. Penalty Prevention: Detecteert EUR 35 miljoen risico's voordat regelgeving enforcement optreedt.

8. Automated Remediation: Semi-geautomatiseerde fix generatie versus volledig handmatige correcties.

385

INDUSTRIELE TOEPASBAARHEID

De uitvinding is toepasbaar in de volgende industrieën:

390

- Financial Services: Kredietrisico modellen, fraud detectie, KYC/AML
- Healthcare: Diagnostische AI, patient triage systemen, treatment planning
- Human Resources: CV screening, talent assessment, promotion algorithms
- Government: Sociale uitkeringen, belasting fraude detectie, subsidies
- 395 - E-commerce: Product recommendaties, dynamic pricing, customer segmentation
- Education: Studenten assessment, toegangsbeslissingen, curriculum planning