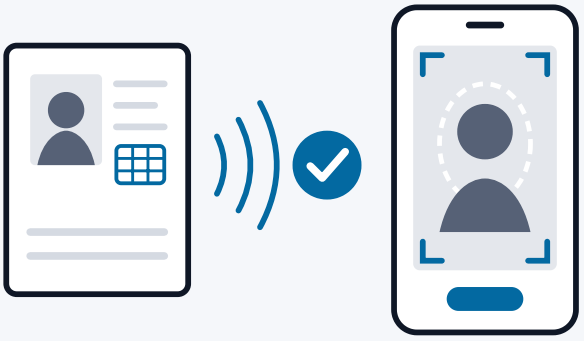
A-SIT Technology

E-Government Innovationszentrum

9. November 2026 Konzept · Digitale Identität · Mobile Sicherheit · 17 Min. Lesezeit

# Self-Ident Proof-of-Concept

Gerald Palfinger




| | |
|---|---|
| **Autorinnen und Autoren** | Gerald Palfinger |
| **Stand** | 2026-05-11 |
| **Quelle** | https://technology2.a-sit.at/veroeffentlichungen/self-ident-proof-of-concept/ |


## Kernaussagen

- Referenzerfassung, geführte Porträtaufnahme, Qualitätsprüfung und Gesichtsvergleich laufen vollständig lokal auf dem Smartphone.
- Das Referenzporträt kommt per NFC aus dem Reisepass; Passive und Active Authentication prüfen die gelesenen Daten.
- Einzeln schaltbare Qualitätsregeln nach ICAO- und BMI-Kriterien übersetzen Messwerte in konkrete Aufnahmehinweise.
- Fünf Modelloptionen, mehrere Alignment-Verfahren und drei Metriken lassen sich kombinieren; Rohwerte bleiben sichtbar.
- Die Grenzwerte 0,4 und 0,6 sind empirisch festgelegt und müssen mit vielfältigerem Bildmaterial je Modell kalibriert werden.

Der Self-Ident Proof-of-Concept untersucht, wie sich Referenzbilderfassung, geführte Porträtaufnahme, lokale Qualitätsprüfung und Gesichtsvergleich in einer Smartphone-Anwendung verbinden lassen. Ziel ist es, praktische Erfahrungen mit der Integration von On-Device-Modellen und eigener Bildanalyse zu gewinnen. Passbildkriterien dienen dabei als fachliche Orientierung für geeignete Aufnahmen im Self-Ident-Kontext. Die Anwendung ist als Forschungsprototyp angelegt.

Der implementierte Ablauf umfasst die Übernahme eines Passbildes nach dem NFC-Lesen eines Reisepasses sowie die Erfassung eines Referenzgesichts aus einem Dokumentenfoto. Eine Aufnahmeanleitung und Rückmeldungen in der Kameravorschau unterstützen die anschließende Porträtaufnahme. Gemeinsame Qualitätsregeln bewerten verfügbare Messwerte und erzeugen konkrete Aufnahmehinweise. Das aufgenommene Bild wird erneut geprüft und bei geeigneter Ausgangslage automatisch zugeschnitten. Im Review können Referenz- und Aufnahmebild betrachtet und mit auswählbaren Modellen, Alignment-Verfahren und Vorverarbeitungsschritten verglichen werden. Rohmetriken und Ergebnisstufen machen die Vergleichsentscheidung sichtbar.

Entstanden ist eine technische Grundlage zur Untersuchung lokaler Bildanalyseverfahren und ihres Zusammenspiels in einem nachvollziehbaren Aufnahmeprozess. Der Bericht beschreibt die realisierten Funktionen und ordnet deren technische Voraussetzungen ein.

## 1. Einleitung

Identifizierungs- und Antragsprozesse verlagern sich zunehmend auf das Smartphone. Typische Schritte sind das Erfassen eines Ausweisdokuments, das Auslesen seines Chips und der Abgleich des Dokumentenporträts mit einer aktuellen Aufnahme der Person. Ob dieser Abgleich gelingt, hängt wesentlich von der Qualität der Eingangsbilder ab. Unschärfe, eine gedrehte Kopfhaltung oder verdeckte Augen erschweren sowohl die visuelle Beurteilung als auch die automatische Merkmalsextraktion.

Der Self-Ident Proof-of-Concept untersucht, wie sich diese Schritte vollständig auf dem Gerät umsetzen lassen. Im Mittelpunkt stehen drei Fragen:

- Wie lässt sich ein Referenzporträt erfassen, etwa aus dem Chip eines Reisepasses oder aus einem Dokumentenfoto?
- Wie kann die App zu einer Aufnahme führen, die anerkannten Passbildkriterien möglichst nahekommt?
- Welche lokal ausführbaren Modelle und Vorverarbeitungsschritte eignen sich für den Gesichtsvergleich, und wie nachvollziehbar lassen sich ihre Ergebnisse darstellen?

Die Kriterien nach ICAO Doc 9303 und den österreichischen Passbildvorgaben dienen dabei als fachliche Orientierung. Sie machen den Forschungsprototyp aber nicht zu einer fertigen Passfoto-Lösung. Die gesamte Bildanalyse läuft lokal auf dem Smartphone, Porträts müssen dafür nicht an einen Server übermittelt werden.

Kapitel 2 fasst die fachlichen Anforderungen an Porträtaufnahmen und die Grundlagen des Gesichtsvergleichs zusammen. Kapitel 3 beschreibt die entwickelte App entlang des Ablaufs, von der Referenzerfassung bis zum Vergleich. Kapitel 4 erläutert die technische Umsetzung, die eingesetzten

Technologien und die offenen Arbeiten. Kapitel 5 zieht ein Fazit.

# 2. Fachliche Grundlagen und Anforderungen

## 2.1. Standards und Anforderungen an Porträtaufnahmen

Dieser Abschnitt fasst die für Aufnahme und Vergleich relevanten Anforderungen zusammen. ICAO Doc 9303, Teil 3, achte Ausgabe 2021, mit Abschnitt 3.9 zu Porträtdarstellungen [1], und die österreichischen BMI-Passbildkriterien [2, 16] bilden den Bezugsrahmen.

Für die digitale Erfassung beschreibt ISO/IEC 39794-5 das Datenformat für Gesichtsbilder einschließlich Empfehlungen für die Aufnahme [13]. ISO/IEC 29794-5 legt fest, wie die Qualität eines einzelnen Gesichtsbildes softwaregestützt gemessen werden kann [14].

**Bildformat und Bildaufbau:** Die österreichischen Kriterien sehen 35 × 45 mm vor. Der Kopf soll ungefähr zwei Drittel des Bildes einnehmen und höchstens 36 mm hoch sein. Er soll zentriert erscheinen. Der Augenabstand beträgt mindestens 8 mm [2].

**Haltung und Blick:** Der Kopf ist frontal auszurichten, der Blick zur Kamera zu richten und beide Augen sind offen zu halten. Gedrehte oder geneigte Aufnahmen verändern die sichtbare Geometrie [1, 2].

**Ausdruck und Sichtbarkeit:** Ein neutraler Gesichtsausdruck und geschlossener Mund erleichtern eine vergleichbare Darstellung. Gesichtszüge und Augen müssen sichtbar sein. Brillen dürfen die Augen insbesondere nicht durch Reflexionen verdecken [1, 2].

**Bildqualität und Aufnahmeumgebung:** Das Gesicht soll scharf, klar und mit natürlichen Farben abgebildet sein. Gleichmäßige Beleuchtung und ein ruhiger, heller Hintergrund unterstützen die Erkennbarkeit [1, 2].

## 2.2. Lokale Verarbeitung und Gesichtsvergleich

Lokale Analyse kann Aufnahmehinweise unmittelbar aus dem Kamerabild ableiten. Für die hier betrachteten Analyse- und Vergleichsschritte ist dadurch keine Übermittlung der Porträts an einen Analyseserver erforderlich. Sensible Bilddaten bleiben im lokalen Verarbeitungsablauf. Das ist eine technische Grundlage für Datensparsamkeit.

| Aufgabe | Fragestellung und Ergebnis |
|---|---|
| Gesichtsdetektion | Wo befindet sich ein Gesicht, und wie viele Gesichter sind im Bild? Ergebnis sind unter anderem Positionen und Gesichtsmerkmale. |
| Qualitätsbewertung | Welche ausgewählten Aufnahmeeigenschaften sind hinreichend? Ergebnis sind Messwerte und daraus abgeleitete Hinweise. |
| Biometrischer Vergleich | Wie ähnlich sind die Merkmalsdarstellungen zweier Gesichter unter einer gewählten Pipeline? Ergebnis sind ein Vergleichswert und eine Schwellwertentscheidung. |

Tabelle 1: Drei unterschiedliche Aufgaben der Bildanalyse

Für den Vergleich stehen ein Referenzporträt und eine neue Aufnahme zur Verfügung. Nach Detektion und geometrischer Ausrichtung (Alignment) erzeugt ein Modell aus jedem Gesicht einen Merkmalsvektor, ein sogenanntes Embedding. Eine Metrik beschreibt die Ähnlichkeit beziehungsweise Distanz. Ein konfigurierter Schwellwert übersetzt den Wert in eine Entscheidung. Diese Entscheidung hängt von Modell, Vorverarbeitung, Bildmaterial und Kalibrierung ab und ist kein Beweis einer vollständigen Identität [8].

Aus der lokalen Verarbeitung und der Abhängigkeit der Entscheidung von Modell und Kalibrierung ergeben sich zwei Gestaltungsanforderungen: eine sparsame Verarbeitung und nachvollziehbare Ergebnisse.

### 2.3. Abgrenzung von Manipulations- und Angriffserkennung

Ein Morphing-Bild kombiniert Merkmale mehrerer Gesichter. Deepfakes bezeichnen synthetisch erzeugte oder veränderte Bildinhalte. Bei Präsentationsangriffen werden der Kamera beispielsweise ein Foto oder ein Display statt der tatsächlich aufzunehmenden Person gezeigt. Diese Begriffe betreffen andere Prüffragen als Schärfe oder Merkmalsähnlichkeit. Weder eine neutrale Pose noch ein hoher Vergleichswert ist ein Lebendigkeits- oder Manipulationsnachweis. Wie Mechanismen zur Erkennung von Präsentationsangriffen (Presentation Attack Detection) zu testen und zu bewerten sind, beschreibt ISO/IEC 30107-3 [15].

Unabhängig davon ist die Herkunft eines Referenzbildes getrennt vom Vergleichsergebnis zu betrachten: Ein importiertes oder fotografiertes Gesicht wird durch einen Vergleich nicht zu einer authentisch bestätigten Identitätsquelle.

## 3. Beschreibung der entwickelten App

### 3.1. Einstieg und Referenzquellenauswahl

Das Onboarding führt zur Kameraberechtigung und zur Auswahl des Erfassungswegs. Angeboten werden Reisepass, e-card, Dokumentenfoto und „Lokales Bild laden“, ein Ablauf ohne zuvor erfasste Referenz, der hauptsächlich dem Testen dient. Die Auswahl verzweigt in unterschiedliche Funktionen: Passbildübernahme, fotografische Referenzerfassung, prototypische Kartenauthentifizierung oder unmittelbare Aufnahmevorbereitung. Die vier Wege sind deshalb nicht als gleichwertige Referenznachweise zu verstehen.

Die Abbildungen in diesem Kapitel sind Bildschirmaufnahmen eines Android-Geräts. Für die meisten Ansichten war der Demo-Modus der App aktiv. Diese Entwicklerfunktion kennzeichnet jede Ansicht mit „DEMO MODE“, simuliert NFC-Lesevorgänge und blendet in den Kameraansichten Schaltflächen ein, mit denen Beispielbilder oder Beispiel-MRZ-Daten anstelle einer Aufnahme verwendet werden können. Für die gezeigten Ansichten wurden diese Ersatzdaten nicht genutzt: Dokumentenaufnahme, Porträtaufnahme und Vergleich beruhen auf realen Aufnahmen. MRZ-Erfassung und NFC-Lesen in B1 und B2 erfolgten mit einem realen Reisepass bei deaktiviertem Demo-Modus. Ab B1 war zusätzlich der Privatmodus aktiv, der Kamerabilder und Porträts verwischt und personenbezogene Angaben maskiert.

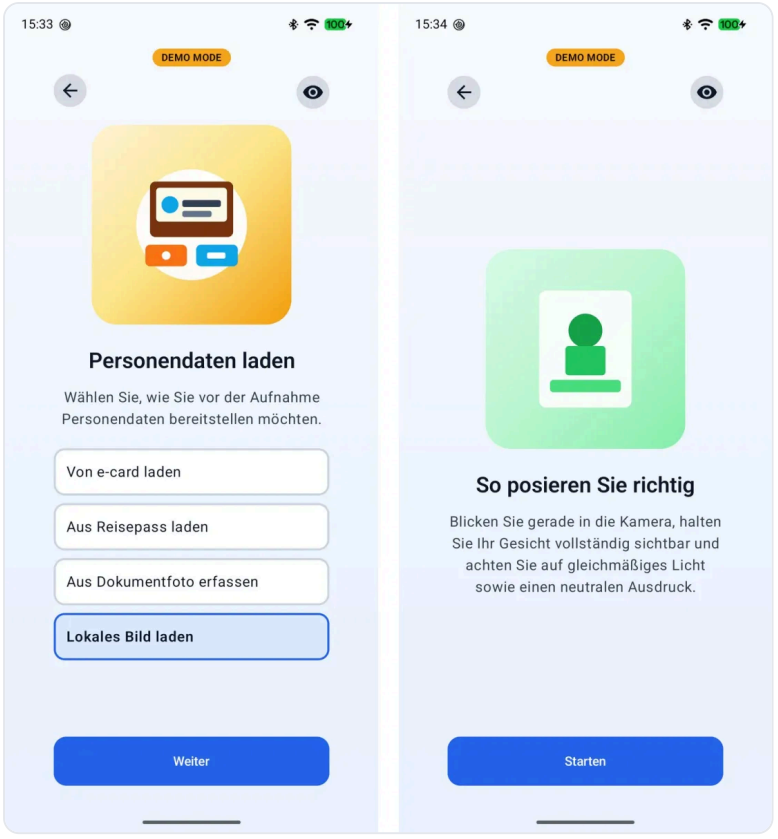



Abbildung 1: Referenzquellenauswahl und Aufnahmeanleitung (A1, A2). Bildschirmaufnahmen im Demo-Modus.

## 3.2. Erfassung der Referenzdaten

**Reisepass:** Für den Zugriff werden Dokumentnummer, Geburtsdatum und Ablaufdatum eingegeben oder aus der MRZ übernommen. ML Kit liefert OCR-Text. Ein eigener Parser verarbeitet die Dokumentformate TD1, TD2 und TD3 und berücksichtigt typische OCR-Verwechslungen in numerischen Feldern. Im Passablauf werden die Zugangsdaten vor dem NFC-Lesen angezeigt. Bei aktivierter NFC-Funktion führt ein Hinweis zum Anlegen des Dokuments. Der Leseweg verwendet JMRTD und übernimmt Dokumentdaten sowie das Gesichtsbild für den weiteren Ablauf [4, 10]. Der Zugriff auf den Chip erfolgt per PACE, bei fehlender Unterstützung per BAC. Nach dem Lesen von DG1 und DG2 führt die App eine Chip Authentication anhand von DG14 durch. Per Passive Authentication prüft sie anschließend, ob die gelesenen Datengruppen mit den Hashwerten im Document Security Object übereinstimmen, ob dessen Signatur gültig ist und ob das Zertifikat des Document Signers von einem CSCA-Zertifikat der deutschen Masterliste des BSI ausgestellt wurde [17, 18]. Enthält der Pass DG15, folgt eine Active Authentication per Challenge-Response. Die Ergebnisseite zeigt jede Prüfung einzeln an. Fehlgeschlagene Prüfungen führen zu einem Hinweis, brechen den Ablauf aber nicht ab. Die Prüfungen sind experimentell und nicht auditiert.

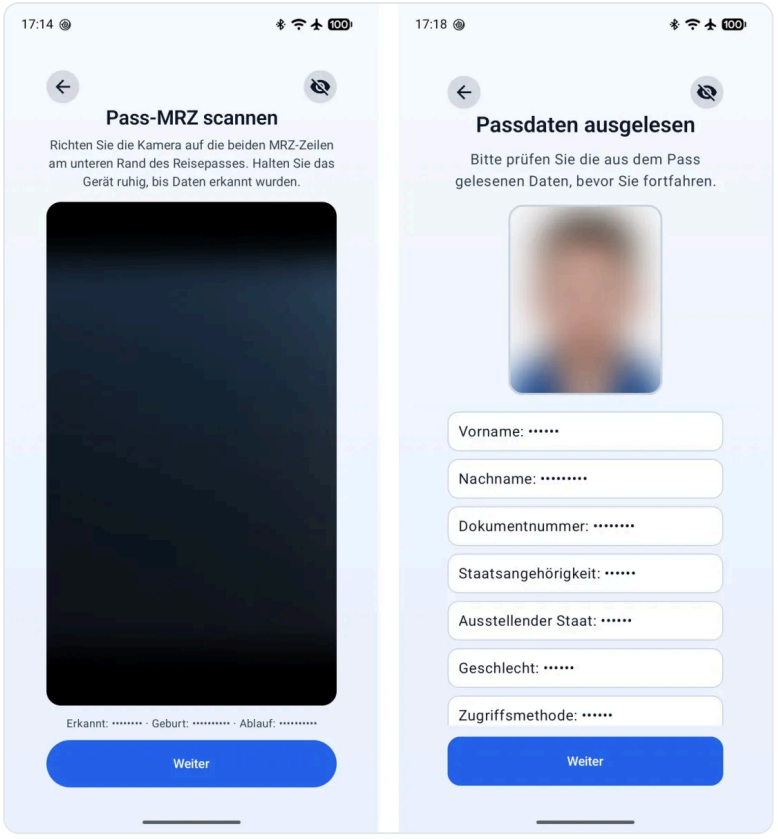



Abbildung 2: MRZ-Erfassung und NFC-Leseergebnis (B1, B2). Bildschirmaufnahmen im Privatmodus mit realem Reisepass (Demo-Modus deaktiviert). Erkannte MRZ-Werte und ausgelesene Daten sind maskiert, Kamerabild und Chipfoto verwischt.

Die Übernahme des Referenzporträts erspart im Passweg eine zusätzliche Bildauswahl. Da ein Rücksprung gelesene Daten verwerfen kann, enthält der Ablauf dafür einen Warnhinweis. Ohne NFC-Hardware bleiben die kamerabasierten Wege verfügbar.

**Dokumentenfoto:** Die Kamera erfasst ein Dokumentenbild. Daraus wird ein Gesichtsausschnitt als Referenz gewonnen und in einer eigenen Review-Ansicht zur Bestätigung oder Wiederholung angeboten. Dieser Weg macht ein abgebildetes Gesicht für den Vergleich nutzbar. Die fotografische Erfassung bestätigt jedoch weder Echtheit noch Unverändertheit des zugrunde liegenden Dokuments.

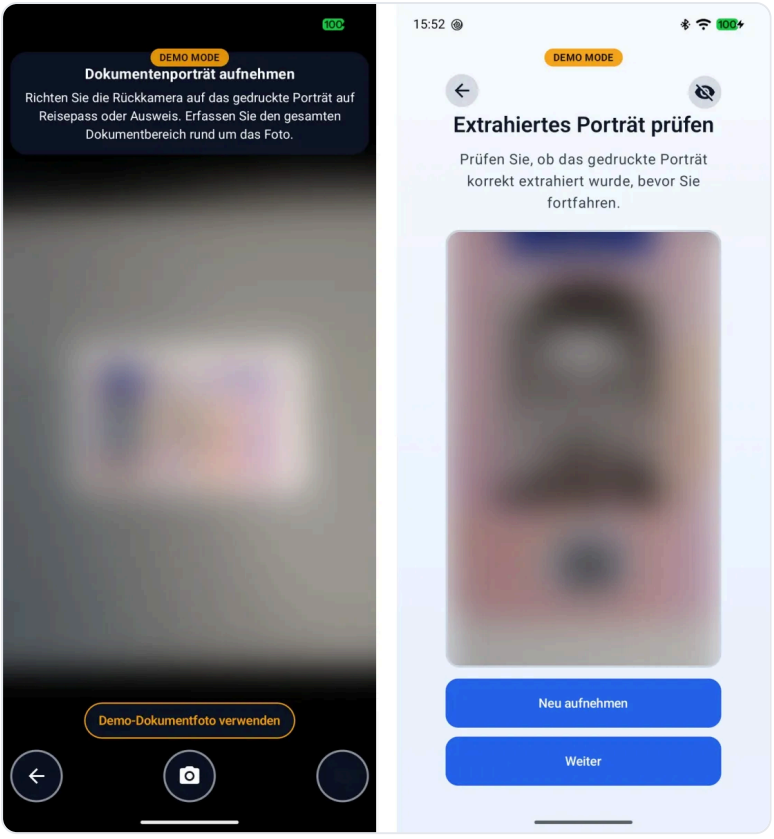



Abbildung 3: Dokumentenaufnahme und Dokumenten-Review (C1, C2). Bildschirmaufnahmen im Demo- und Privatmodus mit realem Dokument, Kamerabild und Porträtausschnitt verwischt.

**e-card und lokales Bild:** Die e-card-Strecke prüft die Echtheit der Karte. Nach einer Terminal Authentication mit einem in der App hinterlegten Terminal-Zertifikat weist die Karte per Chip Authentication ihren Schlüssel nach. Dessen Zertifikatskette wird gegen ein hinterlegtes Zwischen- und Wurzelzertifikat geprüft. Anschließend steht ein verschlüsselter Kanal (Secure Messaging) zur Verfügung. Personendaten oder ein Porträt liest der Prototyp dabei nicht aus, nach erfolgreicher Prüfung folgt direkt die Aufnahmeanleitung. Die Auswahl „Lokales Bild laden“ führt ohne vorherige Referenzerfassung unmittelbar zur Aufnahme und dient hauptsächlich dem Testen. Liegt im späteren Review kein Pass- oder Dokumentenporträt vor, kann für den Vergleich ein beliebiges Bild über die Bildauswahl bereitgestellt werden.

## 3.3. Geführte Porträtaufnahme und Qualitätsprüfung

Nach der Aufnahmeanleitung zeigt die Kameravorschau Live-Hinweise. Einstellungen steuern einzelne Prüfungen und optionale Visualisierungen. Die Anwendung übersetzt Messwerte in Handlungsanweisungen: Ein Kopfrotationswinkel außerhalb der Grenze führt zum Hinweis, geradeaus zu blicken. Geringe Augenöffnungswerte führen zur Aufforderung, die Augen zu öffnen. Bei mehreren Gesichtern fordert der Ablauf eine eindeutige Aufnahme.

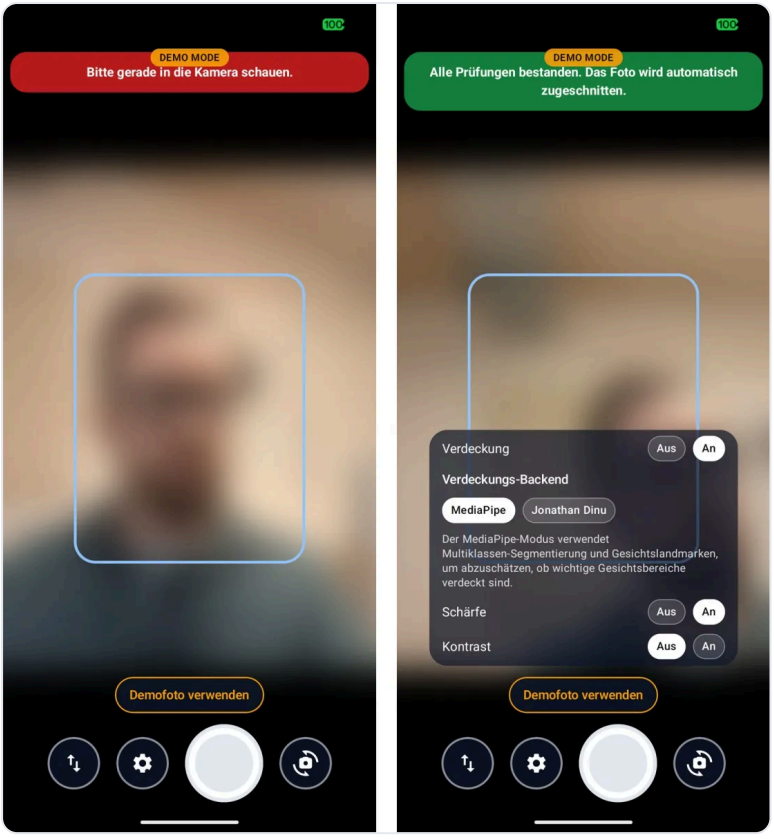



Abbildung 4: Kamera mit Qualitätsrückmeldung und Prüfeinstellungen (D1, D2). Bildschirmaufnahmen im Demo- und Privatmodus mit realem Gesicht. Der Live-Hinweis in D1 beruht auf einer realen Abweichung der Kopfhaltung, das Kamerabild ist verwischt.

Nach dem Auslösen wird das aufgenommene Bild erneut analysiert. Gemeldete Probleme werden als konkrete Ablehnungsgründe ausgegeben. Ein zu kleiner Kopf kann automatisch zugeschnitten werden, wenn genügend Auflösung vorliegt und keine anderen Probleme bestehen. Die erfolgreiche Prüfung bedeutet, dass die aktivierten Regeln anhand verfügbarer Messwerte keinen Ablehnungsgrund geliefert haben.

## 3.4. Review und Gesichtsvergleich

Im Review wird die akzeptierte Aufnahme angezeigt. Ist ein Referenzporträt aus Pass oder Dokumentenfoto vorhanden, verwendet die Vergleichsfunktion dieses Bild. Andernfalls öffnet sie die Bildauswahl. Nutzerinnen und Nutzer können die Aufnahme wiederholen oder verwerfen. Warnungen vor bestimmten Rücksprüngen verhindern ein unbemerktes Verwerfen erfasster Bilder oder Dokumentdaten.

Die Vergleichseinstellungen machen den PoC als Untersuchungswerkzeug nutzbar. Neben dem Modell lässt sich das Alignment auswählen. Hintergrundentfernung ist optional. Für die Helligkeitsnormalisierung stehen „Off“, „Global“ und „CLAHE“ zur Verfügung. CLAHE gleicht lokale Kontraste in begrenzten Bildregionen an. Die Maßnahmen verändern die Eingaben des Vergleichs und können deshalb das Ergebnis beeinflussen.

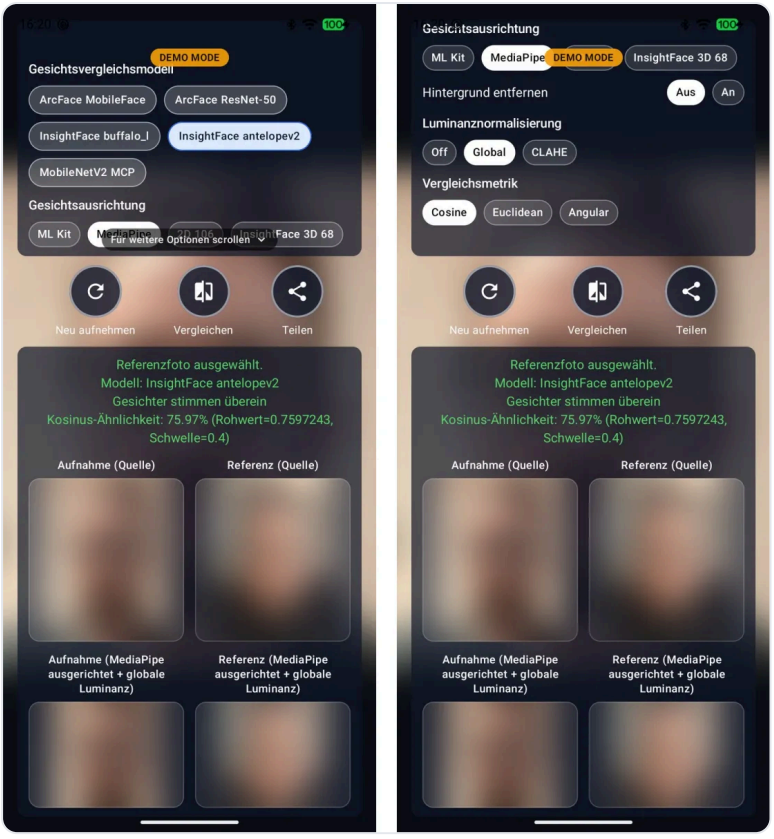



Abbildung 5: Vergleichsergebnis und Modell- und Vorverarbeitungsauswahl (E1, E2). Bildschirmaufnahmen im Demo- und Privatmodus. Gezeigt ist ein realer Vergleich der Porträtaufnahme mit einem Referenzfoto, die Bilder sind verwischt. E2 zeigt den gescrollten Einstellungsbereich (Ausrichtung bis Metrik), die Modellauswahl ist in E1 sichtbar.

Die Anzeige stellt Quellbilder beziehungsweise Vergleichsvorschauen sowie die vorverarbeiteten Bilder gegenüber. Damit wird sichtbar, welcher Ausschnitt und welche Verarbeitung in den Vergleich eingehen. Bei aktivierter Hintergrundentfernung können bereits die Quellvorschauen entsprechend verändert erscheinen. Rohwert, gewählte Metrik und Schwellwert ergänzen die Bildansichten.

## 3.5. Ergebnisse richtig lesen

Die Oberfläche verwendet für alle Metriken drei Stufen. Bei der Kosinusähnlichkeit gelten ab 0,6 „Übereinstimmung“, ab 0,4 bis unter 0,6 „mögliche Übereinstimmung“ und darunter „wahrscheinlich keine Übereinstimmung“. Für euklidische und Winkeldistanz werden dieselben Grenzen in die entsprechenden Distanzen umgerechnet, Abschnitt 4.4 erläutert die Umrechnung. Die binäre Engine-Entscheidung verwendet standardmäßig bereits 0,4. Der Bereich zwischen 0,4 und 0,6 kennzeichnet damit Ergebnisse, die die Engine zwar als Treffer wertet, die aber nahe an der Entscheidungsgrenze liegen und eine manuelle Prüfung nahelegen. Beide Grenzwerte, 0,4 und 0,6, wurden empirisch festgelegt und sind noch mit einer größeren Zahl von Personen unterschiedlicher Ethnien zu evaluieren [19]. Prozentdarstellungen eines Ähnlichkeitswertes sind keine gemessenen Identitätswahrscheinlichkeiten.

# 4. Technische Umsetzung

## 4.1. Aufbau der Anwendung

Kotlin Multiplatform bündelt gemeinsame Zustände, Navigation und Auswertungslogik. Compose Multiplatform implementiert die gemeinsame Oberfläche [11]. Plattformspezifische Adapter binden Kamera, Bildzugriff, Kartenkommunikation und Inferenz ein. Die folgende Abbildung zeigt die lokale Android-Verarbeitung des implementierten Ablaufs. Referenzerfassung und Porträtanalyse sind getrennte Eingänge. Erst beim Vergleich werden Referenz und Aufnahme zusammengeführt.

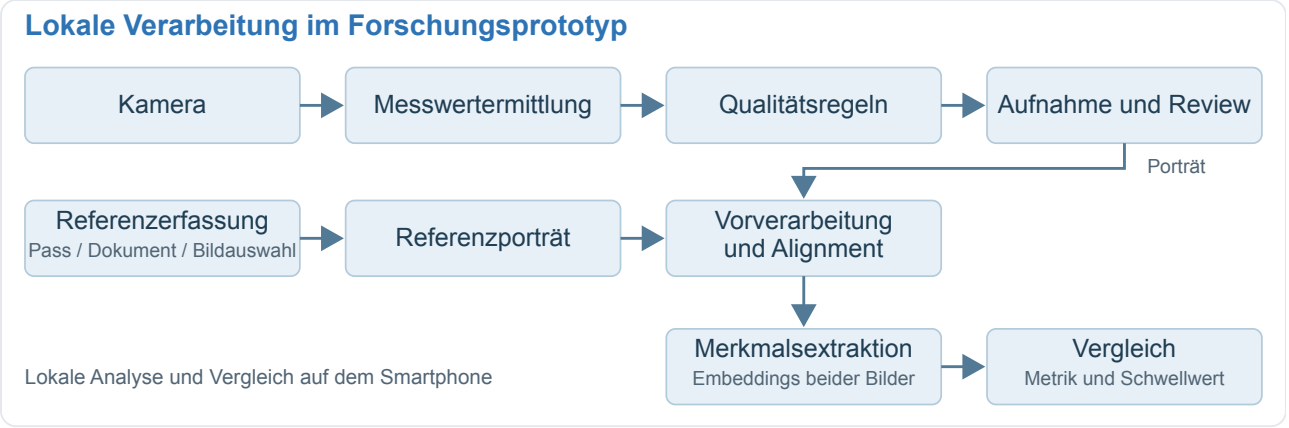



Abbildung 6: Lokale Verarbeitungsarchitektur. Eigene Darstellung nach Quellcode.

## 4.2. Eingesetzte Technologien

| Funktion | Technologie im Projekt | Zuordnung |
|---|---|---|
| Oberfläche / Regeln | Kotlin 2.3.20, Compose Multiplatform 1.10.3 | Gemeinsamer Ablauf, Regeln und Zustände. |
| Kamera | CameraX 1.5.3 [5] | Android-Vorschau, Bildanalyse und Aufnahme. |
| Detektion / OCR | ML Kit Face Detection 16.1.7, Text Recognition 16.0.1 [3, 4] | SDK-Befunde, MRZ-Interpretation durch eigenen Parser. |
| Gesichtsmerkmale / Segmentierung | MediaPipe Tasks Vision 0.10.32 [6] | Landmarker- und Segmentierungsmodelle, eigene Blick- und Verdeckungsableitung. |
| Modellinferenz | ONNX Runtime Android 1.24.3 [7] | Ausführung separat bereitgestellter Erkennungs-, Alignment- und Parsingmodelle. |
| Qualitätsmetriken | Eigene gemeinsame Kotlin-Bildverarbeitung | Schärfe, RMS-Kontrast, Farbstich und Reflexionsheuristik. |
| NFC / Kryptografie | JMRTD 0.8.5, Scuba 0.0.26, Bouncy Castle 1.83, cert-cvc 1.6.0 [10] | Passkommunikation, Passive und Active Authentication sowie Bausteine der prototypischen Kartenabläufe. |

Tabelle 2: Eingesetzte Technologien und ihre Zuordnung

Versionsstände sind der Projektkonfiguration entnommen. Ein Framework stellt Laufzeit und Schnittstellen bereit, während eine Modellressource die trainierten Gewichte enthält. Erst die eigene Auswertung ordnet Modellbefunde den fachlichen Kriterien zu. MediaPipe verwendet unter anderem face_landmarker.task und Selfie-Segmentierungsmodelle. Die ONNX-Modelle sind davon getrennte Ressourcen [6, 7].

### 4.3. Umsetzung der Qualitätskriterien

FaceMetrics sammelt Befunde, aus denen evaluateFaceMetrics Problemcodes erzeugt. CaptureValidation übersetzt diese Codes in Hinweise und prüft die Aufnahme erneut. Die folgenden Werte sind konfigurierbare PoC-Parameter aus ThresholdConfig und keine normativen Toleranzen. Die Qualitätsregeln sind einzeln schaltbar. Die Kontrastregel ist standardmäßig deaktiviert. Davon getrennt ist die Verdeckungsanalyse in den Android-Einstellungen zunächst auf „Off“ gesetzt. Wählbare Verfahren sind MediaPipe und JonathanDinu. Die Prüfungen decken ähnliche Aspekte ab wie die Qualitätskomponenten nach ISO/IEC 29794-5, berechnen aber keine normkonformen Qualitätswerte [14].

| Fachliches Kriterium | Messverfahren | Implementierte Prüfung | Technische Grenze |
|---|---|---|---|
| Eindeutige Person | ML-Kit-Gesichtsdetektion | Genau ein Gesicht, sonst eigener Fehlercode. | Nicht erkannte Gesichter bleiben unberücksichtigt. |
| Kopfgröße / Ausschnitt | Gesichtsgröße relativ zum Bild, Crop-Berechnung | 0,55–0,75; Crop-Ziel 0,62. Mindestseiten 600/800 px. | Gesichtsbox ist keine normgerechte Kinn-Scheitel-Messung. |
| Augenposition | Vertikale Position der Augen | Augen unterhalb oder auf der Bildmitte werden beanstandet. | Keine vollständige Prüfung aller Formatvorgaben. |
| Augen geöffnet | ML-Kit-Klassifikation je Auge | Öffnungswert jeweils mindestens 0,60. | Modellwert. Bei fehlenden Werten keine Beanstandung. |
| Neutraler Ausdruck | Lächelklassifikation | Lächelwert höchstens 0,35. | Andere Ausdrucksformen nicht vollständig erfasst. |
| Mund geschlossen | Normierter Lippenabstand | Mundspalt höchstens 0,018. | Gesichtsmerkmale und Perspektive beeinflussen die Schätzung. |
| Frontalität / Blick | Eulerwinkel, Ableitung aus Gesichtsmerkmalen und Iris | Winkelbetrag höchstens 12°. Blick zentriert, Augen sichtbar. | Blickverfahren wählbar, kein direkter Blicksensor. |
| Gesicht sichtbar | Gesichtsmerkmale / Segmentierung, optional ONNX-Face-Parsing [9] | Standard: Off, wählbar MediaPipe / JonathanDinu. Regionale und globale Grenzen je Verfahren. | Heuristik. Ergebnis hängt von Verfahren und sichtbaren Regionen ab. |
| Schärfe | Laplacian-Varianz im skalierten Gesichtsausschnitt | Mindestens 0,0012. | Bildskalierung und Struktur beeinflussen den Wert. |
| Kontrast | RMS der normierten Gesichtsluminanz | Mindestens 0,09 (standardmäßig deaktiviert). | Kein vollständiger Nachweis gleichmäßiger Ausleuchtung. |
| Farbneutralität | Farbanalyse / Chroma und Kanalstreuung | Grenzen 10 beziehungsweise 14. | Szenenfarben können den Befund beeinflussen. |
| Keine Brillenreflexe | Helle Pixelregionen im Augenbereich | Flächenanteil 0,035; zusammenhängender Bereich 0,015. | Keine sichere semantische Reflexions- oder Brillenerkennung. |

Tabelle 3: Umsetzung der Qualitätskriterien

## 4.4. Modellgestützter Gesichtsvergleich

Alignment richtet Gesichtsmerkmale geometrisch auf einen gemeinsamen Zielausschnitt aus. Zur Wahl stehen ML Kit, MediaPipe, InsightFace 106, InsightFace 3D 68 sowie Detektorvarianten für buffalo_l und antelopev2. Hintergrundentfernung und globale oder lokale Helligkeitsnormalisierung sind ergänzende Vorverarbeitungsschritte. Anschließend berechnet ONNX Runtime die Embeddings [7, 8].

| Reguläre Modelloption | Gewichtsdatei / Einordnung |
|---|---|
| ArcFace MobileFace | w600k_mbf.onnx |
| ArcFace ResNet-50 | w600k_r50.onnx |
| InsightFace buffalo_l | w600k_r50.onnx, dieselben Gewichte wie ArcFace ResNet-50. Beim 2D- und 3D-Alignment wird jedoch der eigene Gesichtsdetektor det_10g.onnx verwendet. |
| InsightFace antelopev2 | antelopev2_recognition.onnx (Standardauswahl). |
| MobileNetV2 MCP | mobilenetv2_mcp.onnx, separates Modell [12]. |

Tabelle 4: Reguläre Modelloptionen und Gewichtsdateien

Kosinusähnlichkeit c bewertet die Richtung der Merkmalsvektoren. Höhere Werte sprechen für größere Ähnlichkeit. Bei normierten Embeddings entsprechen ihr die euklidische Distanz $\sqrt{(2 - 2c)}$ und die Winkeldistanz arccos(c). Für diese Distanzen gelten kleinere Werte als ähnlicher. Die Engine rechnet den konfigurierten Kosinusgrenzwert 0,4 auf etwa 1,095 (euklidisch) beziehungsweise 1,159 rad (Winkel) um. Die Anzeige nutzt für alle Metriken zusätzlich die Grenze 0,6, entsprechend etwa 0,894 (euklidisch) beziehungsweise 0,927 rad (Winkel), und bezeichnet Werte zwischen beiden Grenzen als mögliche Übereinstimmung.

## 4.5. Technische Voraussetzungen und vorbereitete iOS-Anbindung

Android ist im Projekt ab API 24 (Android 7.0) konfiguriert. Compile- und Target-SDK stehen auf 36. Erforderlich sind Kamera, Kameraberechtigung und passende Modellressourcen. NFC-Hardware und NFC-Berechtigung werden für die Kartenwege benötigt. Modelle und die CSCA-Masterliste für die Passive Authentication liegen in den Android-Assets. Die Masterliste ist regelmäßig zu aktualisieren.

Code- und Modelllizenzen sind getrennt zu betrachten. InsightFace veröffentlicht Code unter MIT, bereitgestellte trainierte Modelle dagegen grundsätzlich für nichtkommerzielle Forschung. Für weitergehende Nutzung sind die jeweiligen Bedingungen zu klären [8]. Auch weitere Modellpakete und Bibliotheken besitzen eigene Lizenzbedingungen.

Die iOS-Anbindung ist mit nativen Brücken und Teilen von Vorschau, Aufnahme und Analyse vorbereitet. Die Projektkonfiguration setzt iOS 18.2 voraus. Funktionen und Integrationen sind nicht vollständig abgeschlossen. Der vorhandene Vergleich extrahiert nach Gesichtserkennung und Zuschnitt allgemeine

Vision-Feature-Prints und bildet die Android-ONNX-Pipeline und ihre Modelloptionen nicht gleichwertig ab. Eine vollständige Plattformangleichung ist nicht Gegenstand des erreichten Stands.

### 4.6. Offene Arbeiten

Bei der Aufnahmeprüfung sind vor allem Kriterien zu ergänzen, die bisher nur indirekt erfasst werden. Zudem braucht es einen nachvollziehbaren Umgang mit nicht verfügbaren Messwerten. Für eine belastbare Bewertung müssen Testmaterial, Aufnahmebedingungen und fachliche Referenzurteile vorab festgelegt werden. Das Testmaterial muss Personen unterschiedlicher Ethnien umfassen, da modellbasierte Bewertungen wie die Augenöffnung oder der Lächelwert sowie Farb- und Kontrastmetriken je nach Augenform und Hautton unterschiedlich ausfallen können.

Beim Vergleich stehen Weiterentwicklung und Kalibrierung der Pipeline im Vordergrund. Modelle und Vorverarbeitungsschritte benötigen nachvollziehbare Ressourcenstände. Die empirisch festgelegten Grenzwerte 0,4 und 0,6 sind mit Aufnahmen einer größeren Zahl von Personen unterschiedlicher Ethnien und unabhängig beurteiltem Bildmaterial zu evaluieren, wobei Fehlentscheidungen in beide Richtungen zu betrachten sind. Die Fehlerraten von Gesichtserkennungsverfahren können sich zwischen demografischen Gruppen deutlich unterscheiden [19]. Zudem gelten Engine- und Anzeigegrenzen derzeit für alle Modelloptionen gleich, obwohl Kosinusgrenzwerte modellabhängig sind. Sie sind je Modell zu kalibrieren.

Die Sicherheitsprüfungen des Passwegs sind experimentell und vor einem produktiven Einsatz unabhängig zu auditieren. Offen sind insbesondere die Prüfung von Sperrlisten, die Verifikation der Signatur der Masterliste selbst, die Auswertung der Abweichungsliste des BSI und eine verbindliche Behandlung fehlgeschlagener Prüfungen. Eine Erkennung von Präsentationsangriffen ist ebenfalls nicht umgesetzt. Ihre Wirksamkeit wäre nach ISO/IEC 30107-3 zu bewerten [15].

Die vorbereitete iOS-Anbindung ist zu vervollständigen und an die tatsächlich nutzbaren Verfahren anzugleichen. Eine spätere Einbettung in einen Gesamtprozess kann Backend-Anbindung und Attestierung aufnehmen.

## 5. Fazit

Mit dem Proof-of-Concept liegt eine lauffähige Smartphone-Anwendung vor, die Referenzerfassung, geführte Porträtaufnahme, lokale Qualitätsprüfung und Gesichtsvergleich in einem durchgängigen Ablauf verbindet. Als Referenz dienen das per NFC aus einem Reisepass gelesene Passbild oder ein aus einem Dokumentenfoto extrahiertes Porträt. Beim Reisepass prüft die App die Echtheit der gelesenen Daten zusätzlich per Passive Authentication und, sofern unterstützt, per Active Authentication. Ergänzend sind eine prototypische Echtheitsprüfung der e-card und ein Testweg mit lokal ausgewähltem Bild umgesetzt.

Die geführte Aufnahme übersetzt Messwerte aus ML Kit, MediaPipe und eigener Bildanalyse in konkrete Hinweise. Einzeln schaltbare Qualitätsregeln prüfen die Live-Vorschau und das aufgenommene Bild, ein zu kleiner Kopf wird bei ausreichender Auflösung automatisch zugeschnitten. Der Gesichtsvergleich läuft vollständig auf dem Gerät. Fünf Modelloptionen mit vier unterschiedlichen Gewichtsdateien, mehrere Alignment-Verfahren, optionale Hintergrundentfernung, Helligkeitsnormalisierung und drei Metriken lassen sich kombinieren, wobei Rohwerte und vorverarbeitete Bilder sichtbar bleiben.

Für Vorführungen verdeckt der Privatmodus personenbezogene Inhalte auf dem Bildschirm, und der Demo-Modus macht alle Ansichten auch ohne Dokument erreichbar. Die Trennung von Messwertermittlung, Regelauswertung und Hinweisen schafft eine erweiterbare Grundlage, Kotlin Multiplatform bündelt Ablauf und Regeln plattformübergreifend. Die Herkunft der Referenz wird dabei durchgehend getrennt von der Ähnlichkeitsentscheidung betrachtet. Der erreichte Stand ist ein Forschungsprototyp. Er ist weder als fertige Passfoto-App noch als vollständige Identitätsprüfung einzuordnen.

## Referenzen


[1] ICAO: Doc 9303 – Machine Readable Travel Documents, Part 3: Specifications Common to all MRTDs, Eighth Edition, 2021. https://www.icao.int/sites/default/files/publications/DocSeries/9303_p3_cons_en.pdf. Abgerufen am: 08.05.2026.

[2] Bundesministerium für Inneres: Passbild-Kriterien. https://www.bmi.gv.at/607/passbild_kriterien.html. Abgerufen am: 08.05.2026.

[3] Google: ML Kit: Face detection concepts. https://developers.google.com/ml-kit/vision/face-detection/face-detection-concepts. Abgerufen am: 08.05.2026.

[4] Google: ML Kit: Text recognition v2. https://developers.google.com/ml-kit/vision/text-recognition/v2. Abgerufen am: 08.05.2026.

[5] Google: CameraX overview. https://developer.android.com/media/camera/camerax. Abgerufen am: 08.05.2026.

[6] Google AI Edge: MediaPipe Face Landmarker und Image Segmenter. https://developers.google.com/edge/mediapipe/solutions/vision/face_landmarker, https://developers.google.com/edge/mediapipe/solutions/vision/image_segmenter. Abgerufen am: 08.05.2026.

[7] Microsoft: ONNX Runtime: Mobile. https://onnxruntime.ai/docs/get-started/with-mobile.html. Abgerufen am: 08.05.2026.

[8] InsightFace: InsightFace: 2D and 3D Face Analysis Project. https://github.com/deepinsight/insightface. Abgerufen am: 08.05.2026.

[9] Jonathan Dinu: face-parsing. https://huggingface.co/jonathandinu/face-parsing, https://huggingface.co/Xenova/face-parsing/blob/main/onnx/model_quantized.onnx. Abgerufen am: 08.05.2026.

[10] JMRTD: JMRTD: An Open Source Java Implementation of Machine Readable Travel Documents. https://jmrtd.org/. Abgerufen am: 08.05.2026.

[11] JetBrains: Relationship between Compose Multiplatform and Jetpack Compose. https://kotlinlang.org/docs/multiplatform/compose-multiplatform-and-jetpack-compose.html. Abgerufen am: 08.05.2026.

[12] Yakhyo: face-recognition. https://github.com/yakhyo/face-recognition, https://github.com/yakhyo/face-recognition/releases. Abgerufen am: 08.05.2026.

[13] ISO/IEC: ISO/IEC 39794-5:2019 – Information technology – Extensible biometric data interchange formats – Part 5: Face image data. https://www.iso.org/standard/72156.html. Abgerufen am: 08.05.2026.

[14] ISO/IEC: ISO/IEC 29794-5:2025 – Information technology – Biometric sample quality – Part 5: Face image data. https://www.iso.org/standard/81005.html. Abgerufen am: 08.05.2026.

[15] ISO/IEC: ISO/IEC 30107-3:2023 – Information technology – Biometric presentation attack detection – Part 3: Testing and reporting. https://www.iso.org/standard/79520.html. Abgerufen am: 08.05.2026.

[16] Bundesministerium für Inneres: Fotomuster für Ausweisdokumente. https://www.bmi.gv.at/607/files/passbild_kriterien_2022.pdf. Abgerufen am: 08.05.2026.

[17] ICAO: Doc 9303 – Machine Readable Travel Documents, Part 11: Security Mechanisms for MRTDs, Eighth Edition, 2021. https://www.icao.int/sites/default/files/publications/DocSeries/9303_p11_cons_en.pdf. Abgerufen am: 08.05.2026.

[18] Bundesamt für Sicherheit in der Informationstechnik: CSCA Masterliste (German Master List). https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/ElekAusweise/CSCA/GermanMasterList.html. Abgerufen am: 08.05.2026.

[19] P. Grother, M. Ngan, K. Hanaoka (NIST): Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects, NISTIR 8280. https://doi.org/10.6028/NIST.IR.8280. Abgerufen am: 08.05.2026.