TOSHIBA

WHITEPAPER

Sichere KI-gestützte Dokumentenverarbeitung

Datenschutz, Datenflüsse und Sicherheitsmaßnahmen im e-FileWatcher IDP

Inhalt

1. Zweck und Geltungsbereich

e-FileWatcher IDP ist eine von Toshiba entwickelte Softwarelösung zur automatisierten und KI-gestützten Dokumentenverarbeitung (Intelligent Data Processing). Eingehende Dokumente werden analysiert, klassifiziert und regelbasiert weiterverarbeitet.

Dieses Whitepaper beschreibt, wie e-FileWatcher IDP personenbezogene Daten verarbeitet, welche technischen und organisatorischen Sicherheitsmaßnahmen greifen und wie die Lösung datenschutzrechtlich nach der Datenschutz-Grundverordnung (DSGVO) einzuordnen ist. Es richtet sich an IT- und Datenschutzverantwortliche auf Kundenseite sowie an Vertriebs- und Fachhandelspartner.

Kurzfassung: e-FileWatcher IDP läuft im Kern als On-Premise-Komponente in der Infrastruktur des Kunden. Für die KI-Analyse werden ausschließlich extrahierte Textinhalte an eine von Toshiba betriebene Azure-AI-Foundry-Ressource übermittelt, die in der Region West Europe (EU) mit dem Bereitstellungstyp DataZone Standard (EU-Datengrenze) betrieben wird. Die Verarbeitung verbleibt damit innerhalb der Europäischen Union. Übermittelte Inhalte werden nicht zum Training von KI-Modellen verwendet.

2. Systemüberblick und Funktionsweise

e-FileWatcher IDP verarbeitet Dokumente aus unterschiedlichen Quellen, insbesondere:

- E-Mail-Anhänge
- Scanprozesse
- überwachte Verzeichnisstrukturen
- Druckdatenströme

Die Dokumente werden automatisiert analysiert, relevante Inhalte extrahiert und anhand definierter Regeln klassifiziert. Anschließend erfolgt die Weiterverarbeitung in angebundene Zielsysteme, beispielsweise Dokumenten-managementsysteme (DMS), ERP-Systeme oder nachgelagerte Workflow-Prozesse.

Die fachliche Steuerung, Texterkennung (OCR/Parsing) und Regelverarbeitung erfolgen On-Premise innerhalb der Systemumgebung des Kunden. Lediglich der KI-gestützte Analyseschritt nutzt einen Cloud-Dienst innerhalb von Microsoft Azure (siehe Abschnitt 4).

3. Architektur und Datenfluss

e-FileWatcher IDP ist als hybride Lösung konzipiert: Die Kernkomponente wird On-Premise in der IT-Infrastruktur des Kunden betrieben; für die KI-Analyse wird ein von Toshiba betriebener Azure-Dienst angebunden. Der Verarbeitungsablauf gliedert sich in folgende Schritte:

Wichtig: An den KI-Dienst wird nicht die Originaldatei (z. B. PDF- oder Bilddatei) übertragen, sondern ausschließlich der aus dem Dokument extrahierte Text. Der Umfang dieser Textübertragung ist in Abschnitt 5 beschrieben.

4. KI-Verarbeitung in Microsoft Azure (Azure AI Foundry)

Für die KI-gestützte Analyse setzt e-FileWatcher IDP leistungsfähige Sprachmodelle (Large Language Models) innerhalb von Azure AI Foundry ein. Die Modelle werden über Microsoft Azure bereitgestellt und ausschließlich innerhalb der von Toshiba betriebenen Azure-Ressource genutzt.

Die zugrunde liegenden Modelle stammen zwar ursprünglich von OpenAI, werden jedoch innerhalb der Microsoft-Azure-Umgebung gehostet und betrieben. OpenAI selbst erhält dabei keinen Zugriff auf die verarbeiteten Inhalte. Es gelten ausschließlich die Datenschutz- und Sicherheitszusagen von Microsoft Azure.

Folgende Grundsätze gelten für die KI-Verarbeitung:

- Es werden ausschließlich extrahierte Textinhalte übertragen, keine Originaldateien.
- Die Verarbeitung erfolgt zweckgebunden zur Klassifizierung und Strukturierung der Dokumente.
- Übermittelte Inhalte werden nicht zum Training von KI-Modellen (weder von Microsoft noch von OpenAI) verwendet.
- Die Verarbeitung verbleibt innerhalb der Europäischen Union (siehe Abschnitt 7).

5. Umfang der Datenübertragung

Zur transparenten Einordnung ist zwischen der Originaldatei und dem extrahierten Text zu unterscheiden:

Originaldatei: Die ursprüngliche Datei (z. B. PDF, TIFF, Bilddatei) verlässt die On-Premise-Umgebung nicht und wird nicht an den KI-Dienst übertragen.

Extrahierter Text: Der durch OCR/Parsing gewonnene Textinhalt wird – soweit für die Klassifizierung erforderlich – an den KI-Dienst übermittelt. Je nach Dokument kann dies den vollständigen Textinhalt des Dokuments umfassen.

Für eine korrekte Klassifizierung und Segmentierung mehrseitiger oder zusammengesetzter Dokumente benötigt das Modell den zusammenhängenden Textkontext. Die übertragenen Inhalte können daher personenbezogene Daten enthalten, sofern diese im Dokumenttext vorhanden sind.

Eine vollständige Filterung personenbezogener Daten vor der Übertragung findet nicht statt, da die fachliche Klassifizierung gerade auf dem inhaltlichen Verständnis des Dokuments beruht.

6. Missbrauchserkennung und Datenaufbewahrung bei Microsoft

Microsoft betreibt für Azure-KI-Dienste standardmäßig ein System zur Missbrauchserkennung (Abuse Monitoring). Dessen Zweck ist es, einen Missbrauch des KI-Dienstes (etwa die Erzeugung schädlicher Inhalte) zu erkennen und zu verhindern. Im Sinne vollständiger Transparenz wird dieser Mechanismus hier offen dargestellt:

- Eine Stichprobe der übermittelten Texte und der erzeugten Antworten kann zur Missbrauchserkennung für bis zu 30 Tage in einem verschlüsselten, kundenbezogen isolierten Datenspeicher vorgehalten werden.
- Dieser Datenspeicher liegt in der von Toshiba gewählten Geografie – im vorliegenden Fall innerhalb der EU.
- Die Auswertung erfolgt zunächst automatisiert. Eine Einsichtnahme durch autorisierte Microsoft-Mitarbeiter erfolgt ausschließlich für Inhalte, die zuvor als potenziell missbräuchlich markiert wurden, und nur über abgesicherte Zugänge mit Einzelfreigabe.
- Bei in der EU bzw. im Europäischen Wirtschaftsraum (EWR) betriebenen Ressourcen sind diese autorisierten Microsoft-Mitarbeiter im EWR ansässig.
- Die übermittelten Inhalte werden nicht zum Training von KI-Modellen verwendet.

7. Datenresidenz und Drittlandtransfer

Die für die KI-Verarbeitung genutzte Azure-Ressource wird in der Region West Europe (EU) mit dem Bereitstellungstyp DataZone Standard betrieben. Daraus ergibt sich:

- Die Verarbeitung der übermittelten Textinhalte erfolgt innerhalb der EU-Datengrenze, also in einem EU-Mitgliedstaat.
- Ruhende Daten – einschließlich des Datenspeichers für die Missbrauchserkennung – verbleiben in der von Toshiba gewählten Geografie innerhalb der EU.

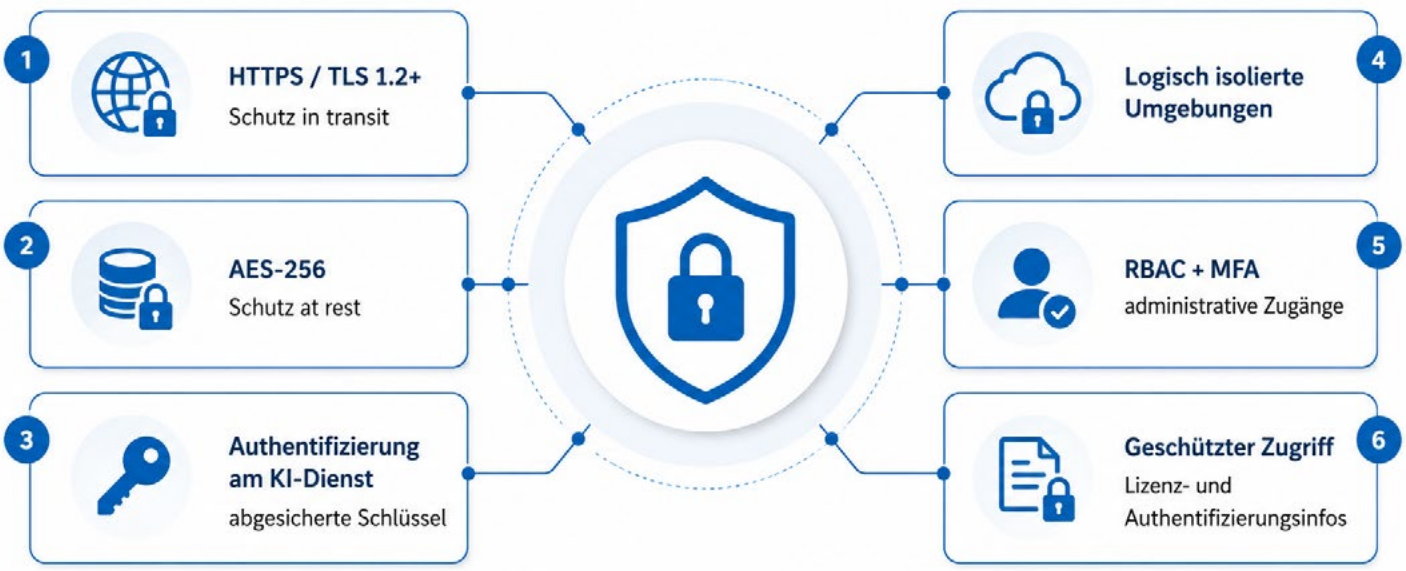
Microsoft ist ein US-amerikanisch geprägtes Unternehmen. Für etwaige Konstellationen, in denen ein Drittlandbezug entstehen könnte, ist Microsoft unter dem EU-US Data Privacy Framework zertifiziert; ergänzend gelten die EU-Standardvertragsklauseln über das Microsoft Products and Services Data Protection Addendum (DPA) https://www.microsoft.com/licensing/docs/view/Microsoft-Products-and-Services-Data-Protection-Addendum-DPA?lang=14 (aktuell Stand Mai 2026).
In der hier beschriebenen EU-DataZone-Konfiguration verbleiben Verarbeitung und Speicherung jedoch innerhalb der EU, sodass im Regelbetrieb kein Drittlandtransfer der verarbeiteten Inhalte stattfindet.

8. Technische und organisatorische Maßnahmen (Art. 32 DSGVO)

Die Kommunikation zwischen e-FileWatcher IDP und den angebundenen Diensten erfolgt über abgesicherte Schnittstellen. Folgende technische und organisatorische Maßnahmen sind vorgesehen:

- Verschlüsselte Datenübertragung über HTTPS (TLS 1.2 oder höher) – Schutz „in transit“.
- Verschlüsselung ruhender Daten innerhalb der Cloud-Infrastruktur mittels AES-256 – Schutz „at rest“.
- Authentifizierung gegenüber dem KI-Dienst über abgesicherte Schlüssel; Verarbeitung in logisch isolierten Umgebungen.
- Rollenbasierte Zugriffskonzepte (RBAC) sowie Mehrfaktor-Authentifizierung für administrative Zugänge.
- Geschützter Zugriff auf Lizenz- und Authentifizierungsinformationen.

Die beschriebenen Sicherheitsmaßnahmen sind fester Bestandteil unserer Standardlösung und werden bei allen Kunden einheitlich umgesetzt. Der einheitliche Standard sorgt für ein konsistentes und nachvollziehbares Sicherheitsniveau; kundenspezifische Anpassungen sind nicht vorgesehen. Die Maßnahmen sind im Rahmen der jeweiligen datenschutzrechtlichen Dokumentation entsprechend zu berücksichtigen.



9. Speicherung, Protokollierung und Löschung

Die primäre Datenhaltung erfolgt ausschließlich innerhalb der Systeme des Kunden. Der e-FileWatcher IDP speichert Daten im Rahmen der konfigurierten Workflows und Zielsysteme.

Im Kontext der KI-Verarbeitung erfolgt über den in Abschnitt 6 beschriebenen Mechanismus hinaus keine dauerhafte Speicherung der verarbeiteten Inhalte im KI-Dienst. Zur Nachvollziehbarkeit der Verarbeitungsprozesse werden im System Protokolle (Logs) erstellt. Umfang, Aufbewahrungsdauer und Zugriff auf diese Protokolle werden im Rahmen der Projektkonfiguration festgelegt; enthalten Logs personenbezogene Daten, unterliegen sie denselben Löschregeln wie die übrigen Verarbeitungsdaten.

Die Umsetzung von Löschfristen erfolgt innerhalb der Systemumgebung des Kunden über konfigurierbare Workflows entsprechend der jeweiligen Aufbewahrungsrichtlinien. Die Verantwortung für die konkrete Ausgestaltung und Einhaltung der Löschfristen liegt beim Kunden als verantwortlicher Stelle.

10. Rollen und Verantwortlichkeiten (Auftragsverarbeitung)

Im Rahmen der KI-gestützten Verarbeitung gelten die folgenden datenschutzrechtlichen Rollen:

Da die KI-Verarbeitung über eine von Toshiba betriebene Azure-Ressource erfolgt, ist zwischen Kunde und Toshiba ein Auftragsverarbeitungsvertrag (AVV) gemäß Art. 28 DSGVO abzuschließen. Microsoft wird darin als Unterauftragsverarbeiter offengelegt; das Verhältnis zwischen Toshiba und Microsoft ist über das Microsoft Data Protection Addendum geregelt. Der Kunde wird über den Einsatz des Unterauftragsverarbeiters informiert und kann seine entsprechenden Rechte wahrnehmen.

11. Wahrung von Betroffenenrechten

Da die maßgebliche Datenhaltung innerhalb der Systeme des Kunden erfolgt, werden Anfragen zu Betroffenenrechten (z. B. Auskunft, Berichtigung oder Löschung) über diese Systeme bearbeitet.

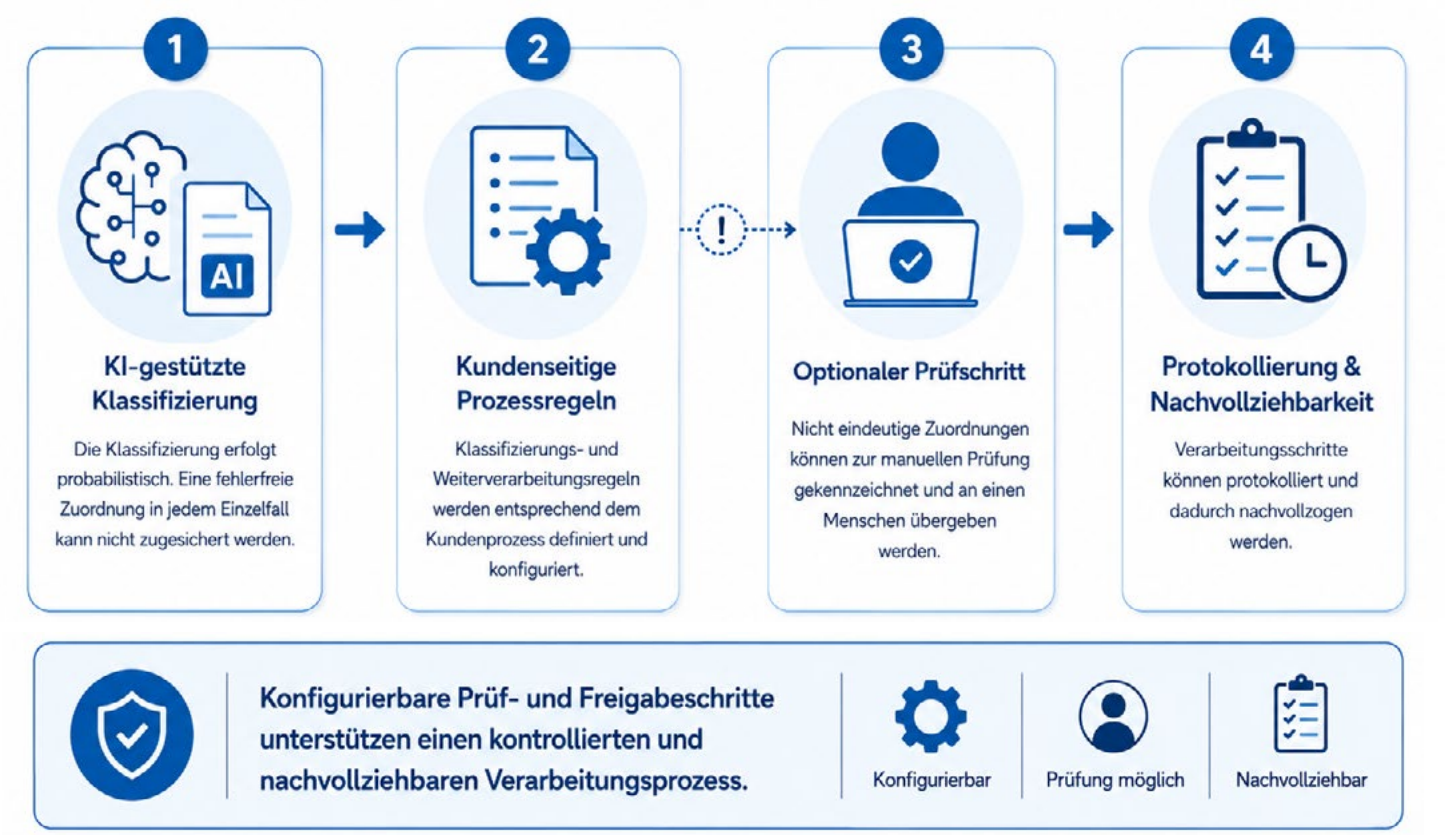
e-FileWatcher IDP unterstützt die strukturierte Verarbeitung und Nachvollziehbarkeit von Dokumentenflüssen, stellt jedoch keine eigenständige Funktion zur gezielten Suche oder Extraktion personenbezogener Daten bereit. Die Umsetzung entsprechender Anforderungen erfolgt im Zusammenspiel mit den angebundenen Systemen und Prozessen des Kunden.

12. Qualitätssicherung und Prozessgestaltung

Die KI-gestützte Klassifizierung arbeitet probabilistisch. Eine fehlerfreie Zuordnung in jedem Einzelfall kann daher nicht zugesichert werden. e-FileWatcher IDP bietet Funktionen, mit denen der Verarbeitungsprozess entsprechend den Anforderungen des Kunden ausgestaltet und abgesichert werden kann.

- Klassifizierungs- und Weiterverarbeitungsregeln werden kundenseitig definiert und konfiguriert.
- Nicht eindeutige Zuordnungen können – abhängig von der jeweiligen Prozesskonfiguration – zur manuellen Prüfung gekennzeichnet werden.
- Verarbeitungsschritte können protokolliert und dadurch nachvollzogen werden.

Für nicht eindeutig erkannte Dokumente empfiehlt sich ein manueller Prüfschritt vor der weiteren Verarbeitung. Die konkrete Ausgestaltung und Kontrolle des Verarbeitungsprozesses obliegt dem Kunden als Verantwortlichem.



13. Zusammenfassung

e-FileWatcher IDP kombiniert moderne KI-Technologie mit einer datenschutzorientierten Systemarchitektur.

Wesentliche Merkmale sind:

- hybride Architektur mit On-Premise-Kern und EU-gehosteter KI-Verarbeitung;
- Übertragung ausschließlich extrahierter Texte, keine Originaldateien;
- Verarbeitung innerhalb der EU (Region West Europe, DataZone Standard);
- transparente Darstellung der Microsoft-Missbrauchserkennung;
- kein Einsatz der Inhalte zum Training von KI-Modellen;
- klar definierte Rollen mit Auftragsverarbeitung nach Art. 28 DSGVO.

Damit besteht eine belastbare und transparente Grundlage für den datenschutzkonformen Einsatz von e-FileWatcher IDP in sensiblen Unternehmensumgebungen.

Über Toshiba Tec

Toshiba Tec Germany Imaging Systems GmbH ist Teil der weltweit operierenden Toshiba Tec Corporation, die in verschiedenen Bereichen der High-Tech-Industrie tätig ist.

Toshiba Tec Corporation ist ein führender Anbieter von Produkten im Bereich der Informationstechnologie mit vielfältigen Einsatzmöglichkeiten in Industrie, Logistik und Handel sowie im Gesundheitswesen und Dienstleistungssektor. Mit Hauptsitz in Japan und über 70 Niederlassungen weltweit unterstützt die Toshiba Tec Corporation Organisationen dabei, neue Wege bei der Erstellung, Aufzeichnung, Verteilung, Verwaltung und Verbreitung von Informationen zu gehen.

Für weitere Informationen wenden Sie sich gerne an uns:

Toshiba Tec Germany Imaging Systems GmbH
Carl-Schurz-Str. 7
41460 Neuss
Germany

Telefon +49 2131-1245-0
Fax +49 2131-1245-402
Website www.toshibatec.de

Bleiben Sie auf dem Laufenden und verpassen Sie keine Updates – folgen Sie uns auf LinkedIn für die neuesten Insights, Lösungen und Innovationen!