Lifecycle Management für Künstliche Intelligenz

von Oliver Roth

Bild: Chanita/stock.adobe.com

Wer Embedded KI im Alltag industrieller Fertigung und Automation zum Laufen gebracht hat, kann von ihren Vorteilen und dem Nutzen auch für Endkunden profitieren. Damit dies langfristig so bleibt, sollte das oft vernachlässigte Lifecycle Management berücksichtigt werden. Der sichere und erfolgreiche Betrieb von KI erfordert kontinuierliche Aufmerksamkeit für Themen wie Fleet Management, Observability und Federated Learning.

In einer Hinsicht sind KI-Modelle geradezu erschreckend menschlich: Sie verändern sich im Lauf der Zeit (Stichwort Modelldrift, siehe auch Teil 2 dieser Serie), und das nicht zwingend zu ihrem Vorteil. Die Verwaltung vieler im Feld eingesetzter KI-Geräte stellt Entwicklungsleiter und Produktmanager deshalb vor komplexe Aufgaben. Damit Embedded-KI-Systeme zuverlässig funktionieren und die Qualität ihrer Entscheidungen oder Empfehlungen erhalten bleibt oder verbessert wird, müssen sie kontinuierlich überwacht und aktualisiert werden. Bei einer Vielzahl von KI-Geräten im Feld ist dafür ein professionelles Fleet Management unerlässlich. Es stellt sicher, dass die KI-Modelle aller Geräte auf dem aktuellen Stand bleiben, die Hardwarekomponenten weiter funktionieren und sämtliche Sicherheitsstandards eingehalten werden. Die zentrale Verwaltung von Updates, Konfigurationsänderungen und das Monitoring der KI-Systeme müssen dazu strategisch und operativ miteinander verknüpft sein. Bei Automatisierungskomponenten war dies bisher meist nicht nötig, hier ist sie aus technischer und gesetzlicher Sicht unerlässlich.

Grossenbacher Systeme bietet hierfür bereits eine Lösung: das ‚GUT!Tool', eine integrierte Plattform für Device- und Software Management, die zentrale Aspekte des CRA (Cyber Resilience Act) berücksichtigt. Solche Plattformen bilden eine notwendige, aber nicht ausreichende Grundlage für ein ganzheitliches KI-Ökosystem.

Web-Tipp

Dieser Beitrag ist der 3. Teil der Artikelserie „Der Weg zur Embedded KI". Lesen Sie Teil 1 und 2 online:

- 1: https://bit.ly/3Zq0mEG
- 2: https://bit.ly/4kOUigX

Fleet Management – Sicherheit für Datenfluss und Updates

Damit Updates gemanagt und verteilt werden können, ist ein klar definierter und sicherer Datenfluss zwischen den einzelnen Geräten im Feld und der Zentrale nötig. Dabei sind Datenintegrität und eine zuverlässige Authentifizierung entscheidend, um Manipulationen oder Cyberangriffe zu verhindern. Gleichzeitig muss die Kommunikation effizient und skalierbar bleiben. Geeignete Fleet-Management-Tools müssen daher folgende essenzielle Funktionen bereitstellen:

- **Over-the-Air-Updates (OTA):** für ein zentrales Rollout von Software-Releases, Konfigurationen oder neuen KI-Modellen.
- **Authentifizierte Datenübertragung:** für zertifikatbasierte Kommunikation mit mindestens PKI-ähnlicher Infrastrukturen.
- **Konfigurationsmanagement:** für die schnelle Anpassung der Geräteeinstellungen je nach Anwendungsfall.

Mithilfe dieser Funktionen lassen sich regulatorische Vorgaben des CRA umsetzen und Verbesserungen im Gerätebestand einpflegen, indem gezielt regelmäßige Updates der Embedded-Software eingespielt werden.

Observability – Überwachung ist erforderlich

Fleet Management ist ohne Observability ineffektiv, da orientierungslos. Denn erst Observability überwacht detailliert den

Betrieb und Zustand der KI-Geräte und liefert so die notwendige Datengrundlage, um eine Diagnose der KI-Aktivitäten vor Ort zu erstellen und eine KI-Flotte im Feld gezielt und effizient zu steuern. Dies geschieht auf Basis des Datenflusses, den das Fleet-Management-System ermöglicht und regelt. Entscheidend sind klare, eindeutig definierte Indikatoren (KPIs), die typischerweise mit Dashboards dargestellt werden:

- **KI-Modellperformance:**
  Prediction Accuracy: Wie genau trifft das KI-Modell seine Vorhersagen?
  Confidence Levels: Verteilung der Vertrauenswerte (Confidence Score) der Modelle.
  Anomalie-Erkennungsquote: Anzahl erkannter Anomalien pro Zeitintervall oder ähnlichem
- **Hardware-Zustand:**
  CPU/GPU-Auslastung: Lastprofile der Hardware-Komponenten.
  Temperatur- und Leistungsprofile: Überwachung von Gerätetemperaturen und Energieverbrauch.
  Speicherauslastung: freier und belegter Speicherplatz zur präventiven Wartung.

Diese KPIs werden idealerweise mittels einer Datenvisualisierung in individuell entwickelten Dashboards aggregiert, mit dem Ziel, eine Überwachung in Echtzeit und schnelle Reaktionen zu ermöglichen.

Erst durch die Kombination von Fleet Management und Observability können Embedded-KI-Systeme effizient, robust und skalierbar betrieben werden. Die Skalierung von KI-Flotten bringt jedoch spezifische Herausforderungen mit sich. Dazu zählt zum einen der Datenschutz, denn eine zentrale Speicherung sensibler Produktionsdaten kommt kaum in Frage. Zum anderen betrifft dies die Themen Netzwerklatenz und Bandbreite, wobei Letztere nach wie vor erstaunlich oft ein knappes Gut ist. Der effiziente Austausch von Modellupdates ist also entscheidend, insbesondere bei großen, verteilten Gerätegruppen. Effiziente Protokolle (MQTT, OPC UA, HTTPS mit Kompression) und Edge Computing können das Latenzproblem entschärfen. Vor dem Hintergrund der mit der Skalierung verbundenen Herausforderungen kommt schließlich das Konzept des ‚Federated Learning' ins Spiel, da es die Übertragung der prinzipbedingt umfangreicheren Rohdaten überflüssig macht und so angemessenen Datenschutz für den eigenen ‚Datenschatz' gewährleistet.

Federated Learning als Lösung

Federated Learning ermöglicht durch das zentrale Aggregieren von dezentral weitergelernten KI-Modellen genau das, was für industrielle Anwendungen benötigt wird, in denen Datenschutz und Datenhoheit essenziell sind. Im Zusammenhang mit Fleet Management ergeben sich daraus neue Potenziale:

- **Lokales Training & globale Aggregation:** Jedes Embedded-Gerät optimiert lokal seine KI-Instanz anhand realer Betriebsdaten; nur Modellparameter werden übertragen und zentral aggregiert. Die Frameworks ‚Flower', ‚TensorFlow Federated' oder ‚Nvidia Clara' ermöglichen die technische Realisierung.
- **Integration in Fleet Management Tool:** Modellupdates, z.B. mit Flower Framework erzeugte aggregierte Parameter, können über entsprechende Plattformen direkt automatisiert verteilt werden. So entsteht eine nahtlose Verknüpfung von Modelltraining, Observability und Update-Rollouts.
- **Prozessor- und Hardwareauswahl:** Leistungsfähige Plattformen wie NXP i.MX8M oder i.MX93/95 bieten optimale Unterstützung für lokale KI-Trainingsprozesse und effiziente Integration in Federated-Learning-Systeme.

Federated Learning kann die Modellqualität verbessern und den Datenaustausch gleichzeitig reduzieren und anonymisieren. Anwender können so Prozesse weiterentwickeln. Anbieter freuen sich über ein kontinuierliches Geschäftsmodell und enge Kundenbeziehungen.

Fleet Management, Observability und Federated Learning als strategische Einheit

Die Kombination aus effizientem Fleet Management, tiefgehender Observability und fortschrittlichem Federated Learning ist ein Ansatz, um Embedded KI langfristig nutzbar zu machen. Industrieunternehmen benötigen dafür ein geeignetes Fleet-Management-Tool, ein passendes Federated-AI-Framework und Erfahrung im praktischen Einsatz.

Für Unternehmen, die den Einsatz von KI nachhaltig, sicher und erfolgreich gestalten möchten, kann die Zusammenarbeit mit Partnern folglich auf Dauer sinnvoll sein, wenn diese Erfahrung im Bereich Embedded KI und Federated Learning einbringen. Embedded KI entwickelt sich zunehmend zu einem Wettbewerbsfaktor, dessen sorgfältige Implementierung sich wirtschaftlich auszahlt. ag

Oliver Roth
ist CEO von Grossenbacher Systeme und der Amalthea Gruppe, zu der unter anderem auch Sabo Mobile IT gehört.