

Edge ML

Aktuelle Herausforderungen und Ansätze

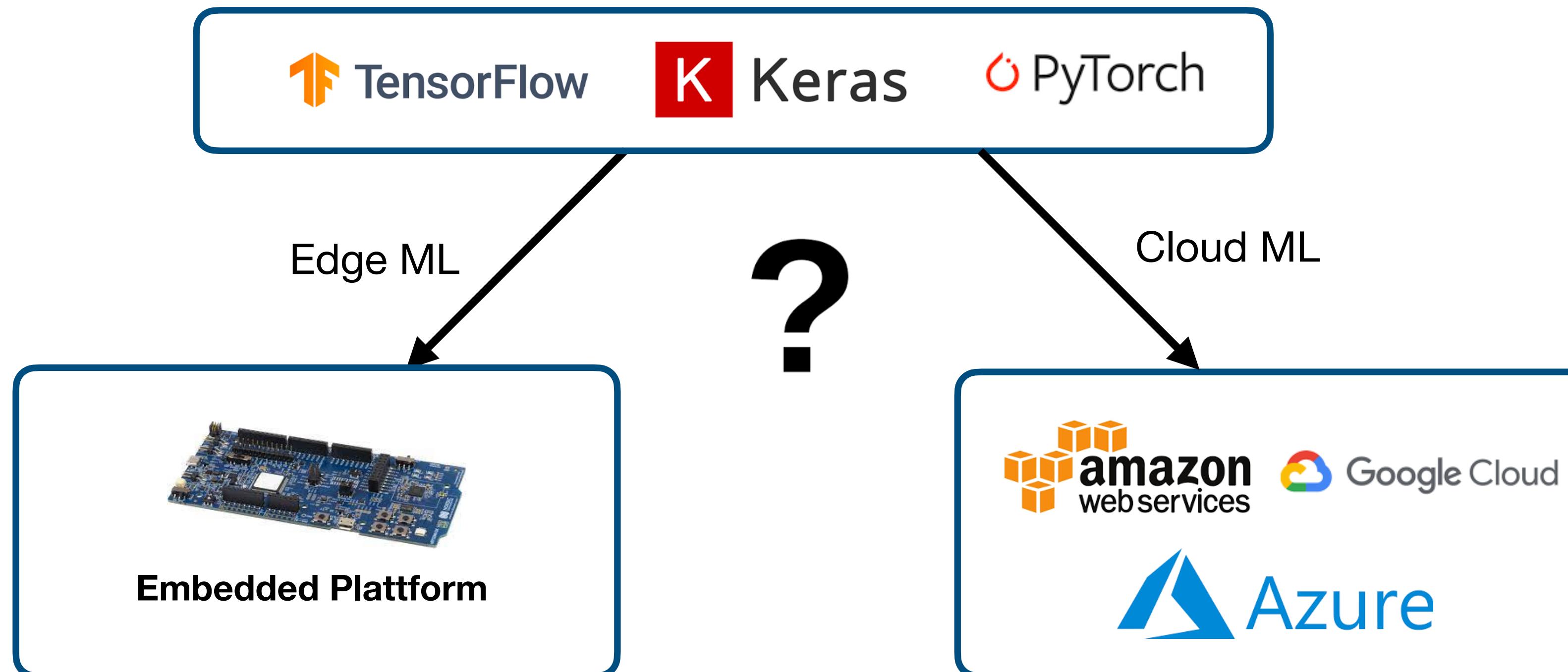
Andre Sprenger (Photonic Insights), 18.10.2022

Agenda

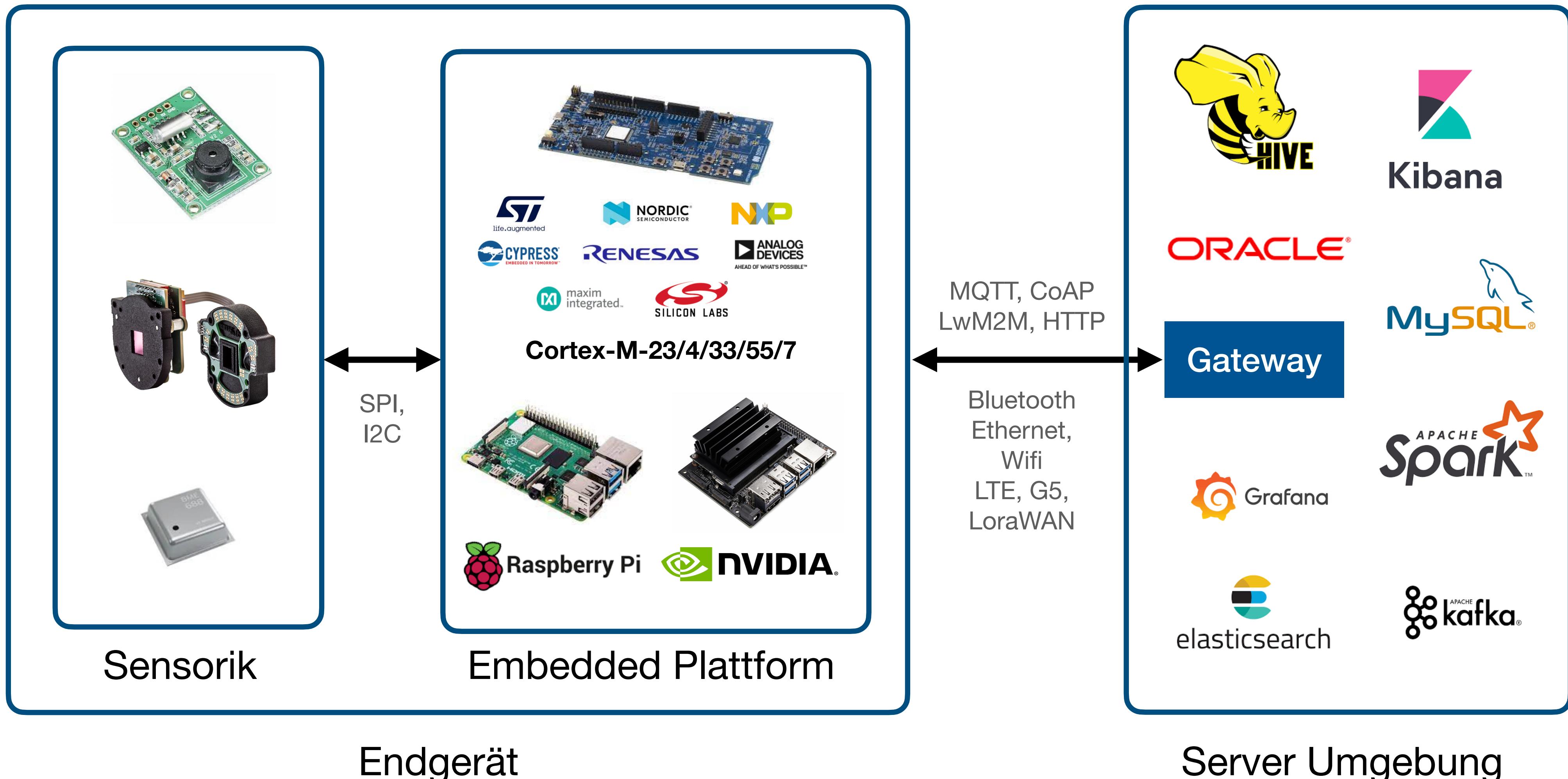
- Motivation
- Anwendungsbeispiele
- Herausforderungen
- Entwicklungsprozess
- Lösungen und SaaS Landschaft
- Zusammenfassung

Definition

Edge ML verlagert die Ausführung von ML Modellen von der Cloud auf die Endgeräte, wo die Daten erfasst werden.



Typische IoT Lösung



Vorteile

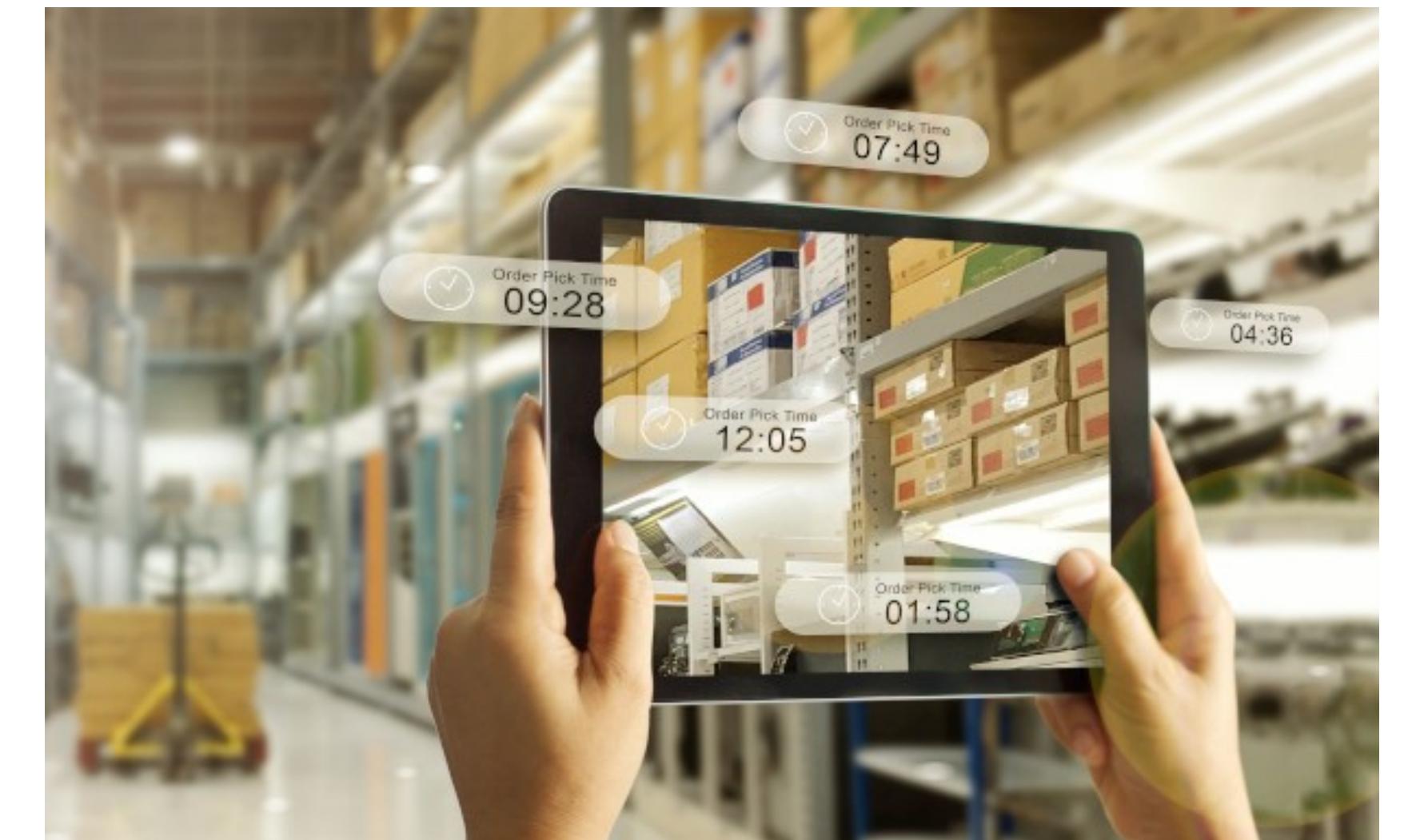
Geschäftsrelevante Ereignisse statt Rohdaten versenden

- Bandbreite und Stromverbrauch minimieren
- Lokale Entscheidungen in Echtzeit
- Keine zuverlässige Cloud Verbindung notwendig
- Volle Auflösung der Daten
- Daten bleiben auf dem Endgerät: höhere Datenschutz

=> Ziel ist die **Autonomie** der Endgeräte

Anwendungsbeispiele Logistik

- Tracking von Ladungen oder Containern
 - Ort und Zustand einer Ladung (in Echtzeit) verfolgen
- “Letter of Credit” Digitalisierung
 - Tracker + Versicherung gegen Zahlungsausfall
 - Bezahlung abhängig von Ort und Zustand einer Ladung auslösen



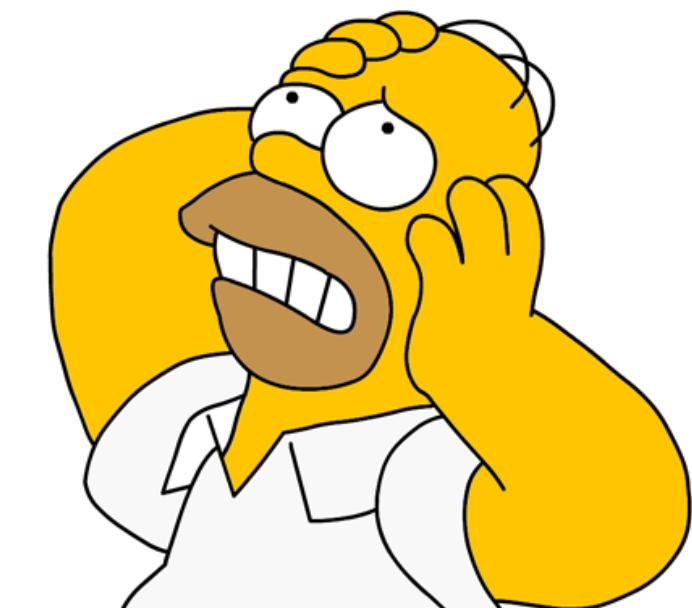
Anwendungsbeispiele Industrie

- Qualitätssicherung von Produktionsprozessen
 - Automatische, visuelle Inspektion von Produkten während der Produktion
- Digitaler Zwilling
 - Digitales Modell einer Maschine oder Anlage
 - Statische + Echtzeitdaten
 - smartBridge: Modell der Köhlbrandbrücke

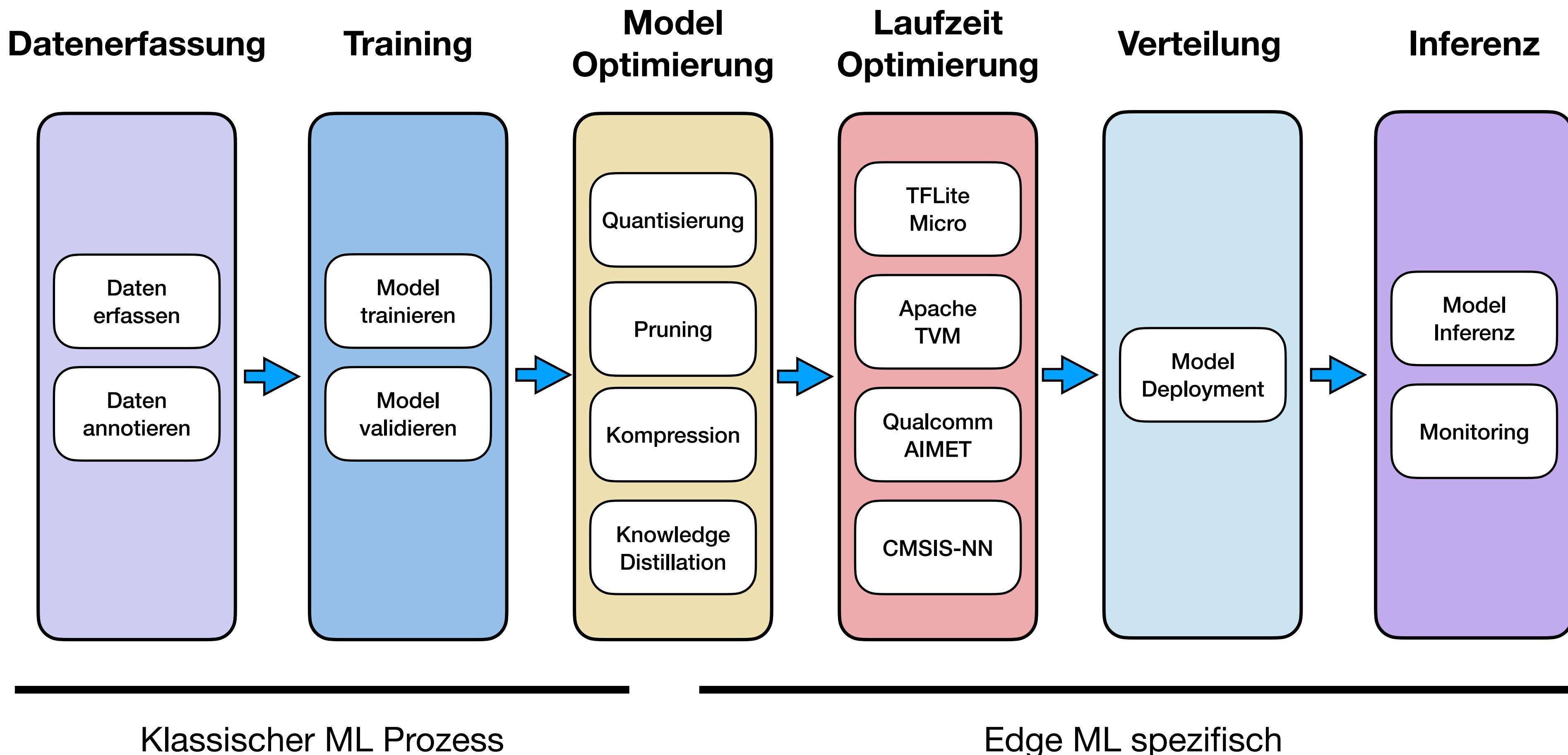


Herausforderungen

- Limitierter Speicher und Rechenleistung auf den Endgeräten
- Modelloptimierung erfordert Wissen über den Anwendungsbereich, Algorithmen und Optimierungsmethoden
- Experimenteller Charakter von Werkzeugen und SaaS Lösungen
- Driftende Datenverteilungen: Umwelteinflüsse, Standorteinflüsse, Sensoren altern
- Management der Endgeräte und Modelle
- Datenerfassung und Annotation



Edge ML Arbeitsablauf



SaaS Landschaft



**EDGE
IMPULSE**



emza
visual sense

Deeplite



Grovety Inc.

NotaAI



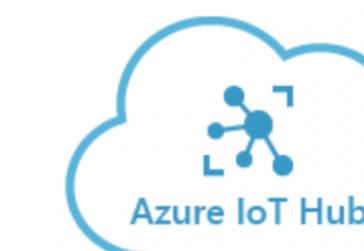
OctoML

Edge ML Dienste



**GOLIATH
TECHNOLOGIES®**

STREAM ANALYZE



HOTG



Memfault

Gerätemanagement

Zusammenfassung

- Geschäftsrelevante Ereignisse statt Rohdaten übertragen
- Edge ML ist mit dem aktuellen Stand der Technik praktikabel
- Werkzeuge / Arbeitsprozesse sind experimental
- Neues Skillset notwendig
- Projektaufwände sind ca. 5-10% ML, der Rest ist Infrastruktur

Zusatzfolie 1: Optimierungsmethoden für ML Modelle

- Quantisierung, Pruning, Kompression
- Model Compiler: Apache TVM, AIMET, Glow
- Optimierte Kernel Implementierungen: CMSIS-NN
- Knowledge Distillation
- Edge ML spezifische Modellarchitekturen
- Architektursuche
 - Optimierung über Modelperformance **und** Resourcenverbrauch

Zusatzfolie 2: On Device Learning

Training von Modellen auf dem Endgerät

- Adressiert hauptsächlich das Feintuning existierender Modelle
- Personalisierung der Endgeräte
- Lösung für driftende Datenverteilungen und unterschiedliche Deployments der Endgeräte
- In aktuellen Frameworks nicht implementiert
- Für Supervised Learning: Wo kommen die Labels her?