# **ALLES KI ODER WAS?**

Wie man Künstliche Intelligenz im Unternehmen schnell einführt und gewinnbringend einsetzt

Industrial-AI / Embedded-AI

Sprecher: Dr.-Ing. Pierre Gembaczka



Unternehmer Verband.



# **Sprecher**



### Dr.-Ing. Pierre Gembaczka

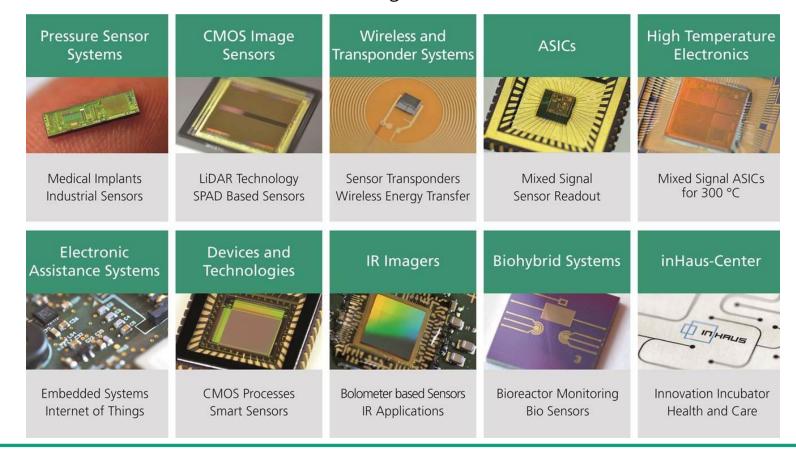
- Studium
  - Physikalische Technik
- Seit 2011 am Fraunhofer IMS
  - Gestartet als Doktorand: Implantierbare Drucksensoren
  - Erfinder von AlfES (Start 2016)
- Aktuelle Position:
  - Program Manager: Industrial AI
  - Product Manager: AlfES



# Fraunhofer IMS - Technologien

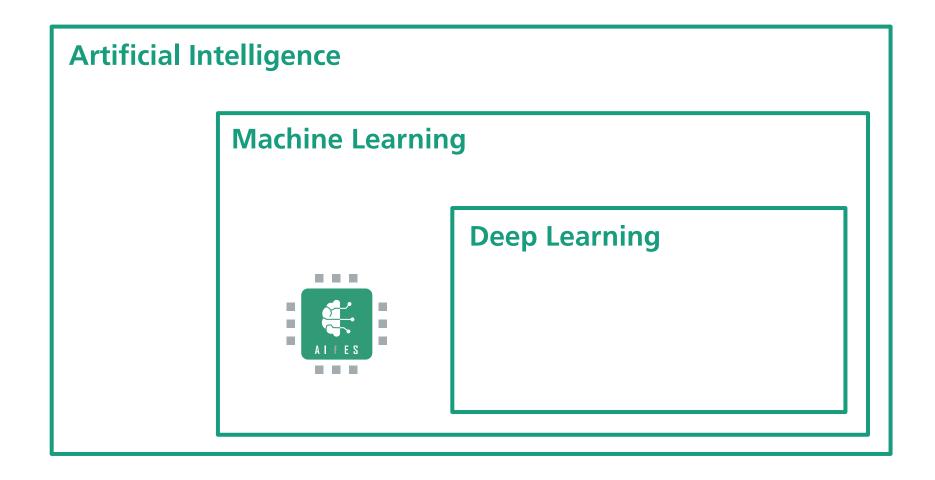


Fraunhofer-Institut für Mikroelektronische Schaltungen



# Begriffsdefinition





### **Aktuelle KI - Software**



#### **Durch TensorFlow & Co wird KI vereinfacht**

- Künstliche neuronale Netze (KNN) sind schnell und einfach am PC erstellt
- Erfordert für erste Ergebnisse nur geringes Fachwissen
- Die Mustererkennung in Bildern steht oft im Fokus
  - Andere Sensoren werden stark vernachlässigt
- Cloud-basierte KI für die Nutzung auf einem eingebetteten System
- Das KI Model kann fast nur auf dem PC genutzt werden (Python)
  - Eine Portierung auf andere Plattformen ist schwierig
  - Hindert Unternehmen bei der Produktentwicklung KI einzusetzen

# Ist größer wirklich besser?



### Deep learning und Big Data werden langfristig ein Problem

- Das Systemverständnis gerät in den Hintergrund
  - Ein tiefes Netz macht das schon…
- Das Big Data Datengrab
  - Hohe Bandbreite und Kosten für Speicher / Verwaltung
- Datenschutz in Gefahr durch Cloud-basierte KI
  - Sensible Rohdaten werden übertragen → Amazon Echo
- Deep learning benötigt teure und spezielle Hardware (GPU)
- Kein Lernen im Feld
- Keine universelle Lösung für Mikrocontroller und eingebettete Systeme

### Was bedeutet Embedded-Al für uns?



### Embedded-AI als Lösung für ressourcenbeschränkte Systeme

- Dezentrale, hoch integrierte KI am Ort der Datenerzeugung (Sensor, Komponente, Produkt, Gerät) hat folgende Vorteile:
  - Schnelle Verarbeitung, keine Übertragungsverzögerungen
  - **Erhöhte Sicherheit**, nur relevante, geschützte Daten werden übertragen
  - Erhöhte Zuverlässigkeit durch dezentralisierte Architektur
  - **Einsparung von Ressourcen**, geringere Datenmengen, geringere Gesamtleistung des Prozessors
  - **Energiesparende**, kleine und ressourcenschonende Systeme wie Mikrocontroller

### Wie wir Embedded-Al realisieren



- Plattformunabhängiges KI-Framework
- Ein skalierbarer KI-Mikrocontroller auf der Grundlage der offenen RISC-V-Architektur
- Grey-box-Ansätze und neue Techniken der Merkmalsextraktion
- Neue und optimierte Algorithmen für ressourcenbeschränkte Systeme
- Strategien für verteiltes Lernen auf Mikrocontrollern

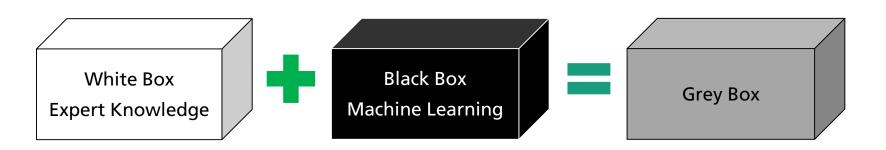




# **Grey-box-Ansätze**



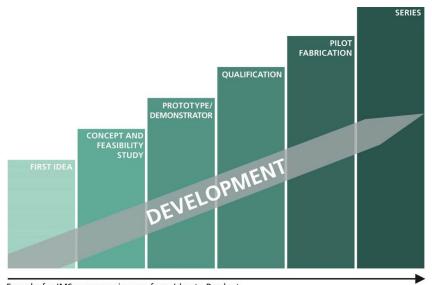
- Die Deep-Learning- und Big-Data-Strategie ist eine komplette Blackbox
  - Derzeit werden z.B. ConvNets zur Extraktion von Merkmalen verwendet
  - Oft kann nicht verstanden werden was überhaupt gelernt wird
- Eine spezialisierte Merkmalsextraktion reduziert den Input drastisch
  - Es werden nur Merkmale verwendet die von Experten ausgewählt wurden
  - Erfordert völlig neue Ansätze zur Merkmalsextraktion
  - Das Modell wird verständlicher und kleiner (grey-box)



# Von der Idee zum Produkt – KI-Dienstleistungen am IMS



- Gemeinsame Entwicklung eines Produktes
  - Konzepte wie der Einsatz von KI ein Produkt verbessern kann
  - Entwicklung von KI Modellen
  - Integration von KI Modellen in die Toolchain des Kunden
  - Upgrade vorhandener Produkte mit KI
  - KI Hard- und Software Codesign
  - Verifikation und Validierung der KI
- Beratung und Schulung
  - Kundenberatung im KI Umfeld
  - AlfES Workshops



Fraunhofer IMS accompanies you from Idea to Product



# Was interessiert die Kunden nach unserer Erfahrung?



- KI in das Produkt bringen ohne die Hardware ändern zu müssen
  - Quellcode einsehbar
  - Nutzbar mit der bekannten Entwicklungsumgebung
  - Lauffähig auf dem bereits genutzten Mikrocontroller
- KI Anwendung mit besonderen Anforderungen
  - Spezielle Hardware nötig
  - Ultra Low Power
  - Hohe Integration

# Soft- und Hardwarelösungen



#### AlfES - Künstliche Intelligenz für eingebettete Systeme

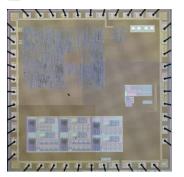
- Open Source KI-Software-Framework, das auf fast jeder Hardware läuft
  - Trainieren eines ANN auf nahezu beliebiger Hardware (programmiert in C, {GCC})
  - Optimierte und neue Algorithmen speziell für ressourcenbeschränkte Systeme
  - Kompatibel zu praktisch jeder Entwicklungsumgebung
  - Kompatibel mit anderen KI-Software-Frameworks (TensorFlow)

#### **AIRISC**

- Ein skalierbarer KI-Mikrocontroller
  - Lizenzfreie und offene RISC-V-Architektur
  - Spezielle Hardwarekomponenten für KI-Anwendungen
  - Optimierbar für die Anwendung









# **AlfES Lizensierung und Partner**

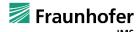


- AlfES wird als Dual License Model angeboten und ist Open Source
  - GNU GPL V3: Private oder Free Open Source Software
  - Kostenpflichtiger Lizenzvertrag: Kommerzieller Einsatz
- Kompatibel zu der Arduino® IDE
  - AlfES und Arduino bereiten eine Partnerschaft vor
  - Direkter Download aus dem Arduino Library Manager
- Weitere Partner
  - Arm Al Partner
  - Open Roberta Lab



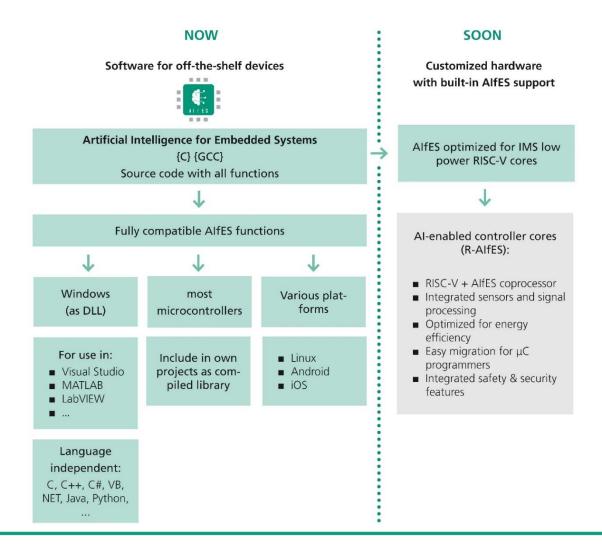






# **Embedded AI - Roadmap**





# Neue Möglichkeiten und Kompatibilität



- Rekonfiguration des KNN zur Laufzeit (ein Gerät mit verschiedenen Aufgaben)
  - Ändern Sie die KNN-Struktur und laden Sie neue Gewichte
- Trainieren auf dem Mikrocontroller
- Dezentralisierte KI ohne Cloud
- Kompatibilität zu anderen Frameworks
  - Importiere ein KNN aus TensorFlow / Keras
- Ein Python-Wrapper extrahiert die Modellparameter (Lizenzversion)
  - Manuell auch in der Open Source Version möglich
- Das Modell kann ausgeführt und sogar weiter trainiert werden







### **Funktionsübersicht**

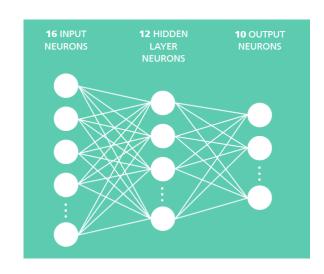


#### Feedforward neural network inference

- Frei konfigurierbar (inputs, hidden layer, outputs)
- Viele Aktivierungsfunktionen
  - Sigmoid, softsign, linear, RELU, Leaky RELU, softmax, tanh, ELU

### Feedforward neural network Training

- Vollständiger Backpropagation und ADAM Algorithmus
- Trainingsarten
  - Online, Batch, Minibatch
- Verschiedene loss Funktionen
  - Sum of squares, softmax, mean squared error (MSE)



### Was kommt als nächstes?



- Fixpoint Kalkulation mit Quantisierung der Gewichte
  - 32 Bit und 8 Bit Integer
- AlfES Express Funktionen
  - Inferenz und Training mit einem Funktionsaufruf
- CNN / ConvNets
  - Inferenz und Training
- Reinforcement Learning



# **AlfES – Demonstrator: Handschrifterkennung**



- Handschrifterkennung Ziffern von 0-9
  - Verwendet ein kapazitives Standard-PS/2-Touchpad
  - Eine spezielle Merkmalsextraktion wurde entwickelt
  - Nur 16 Eingänge
  - Sehr kleines KNN mit nur 12 Neuronen in einer versteckten Schicht
  - Die Erkennung benötigt etwa 25ms (16 MHz Taktfrequenz)
  - Vorab auf PC trainiert
  - 10 Personen wurden trainiert





### **AlfES – Demonstrator: Personalisierbare Kl**



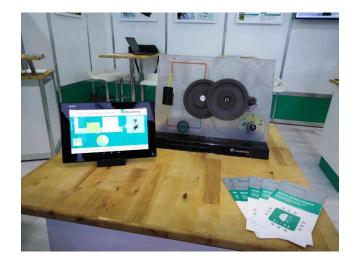
- Erkennt komplexe Gesten
  - $\blacksquare$  Erkennt in die Luft geschriebene Figuren  $\rightarrow$  spezielle Merkmalsextraktion
  - Kann einzelne Gesten direkt im System erlernen ADAM-Algorithmus
  - Es können bis zu 10 individuelle Gesten trainiert werden (nur durch den Speicher begrenzt)
  - Verwendet einen Beschleunigungssensor
  - AlfES erstellt ein KNN mit der passenden Struktur und trainiert es
  - Drei Wiederholungen pro Geste sind ausreichend
  - Das Training von drei Gesten dauert weniger als 2 Sekunden
  - Erkennung einer Geste ca. 20 100 ms (Cortex M4)

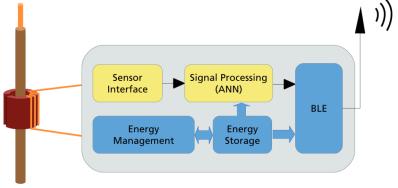


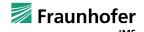
# **AlfES – Demonstrator: Intelligenter Stromsensor**



- Wireless Drahtloser Stromsensor zur Zustandsüberwachung
  - Drahtloser und energieautarker Betrieb für einfache Nachrüstung
  - Lernt den Zustand eines Gerätes anhand seines Stromverbrauchs
  - Lernalgorithmus auf dem Mikrocontroller (ATMega32U4)
  - Konfiguration über BLE
  - Sendet nur den Gerätestatus über BLE
  - Es müssen keine Messwerte übertragen werden







# Project noKat: Erkennung von Menschen mittels eingebetteter KI



- Laufendes ZIM Projekt (Zentrales Innovationsprogramm für den Mittelstand)
  - Fraunhofer IMS und van Rickelen GmbH & Co. KG
- Stromsparendes und kostengünstiges Kamerasystem (RGB)
  - Kamera bleibt stationär



- Andere Klassen möglich (Autos, Fahrräder, Tiere, etc.)
- Spezialisierte Merkmalsextraktion und ein sehr kleines KNN
  - Kein ConvNet (convolutional neural network)
- Verringerung der erforderlichen Parameter um mehr als 99 %
  - EfficientDet-D7 (77 Millionen Parameter) / noKat (1125 Parameter)
- Verarbeitungszeit auf einem Mikrocontroller (160 MHz) ca. 120ms

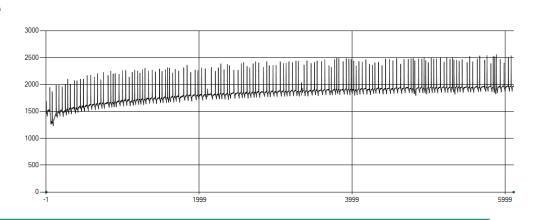




### Projekt Hinkelstein: Vorhofflimmern in EKG Daten erkennen



- Hybrid integrierte neuronale Netze zur Klassifikation von EKG-Daten mit Low-Power Schaltungstechniken
  - TU-Dortmund und Fraunhofer IMS
- BMBF Pilotinnovationswettbewerbs "Energieeffizientes KI-System"
- Offline Detektion von Vorhofflimmern in EKG Daten unter Einbindung energieeffizienter KI-Systeme
- 16.000 / 2-minütige und gelabelte Trainingsdaten von der Charité
- Hürde: Minimale Detektionsrate: 90% / Maximale Fehlalarmrate: 20%
- Ergebnis: Simulation eines ASICS in IHP 130nm Technologie
  - CNN vollständig in Hardware und CMOS kompatibel
  - 360 Nanojoule und 1,2 V Versorgungsspannung
  - 1,33 ms pro Inferenz
  - 96,5% TP / 9,8% FP (4 Sigma / Cross Validation)



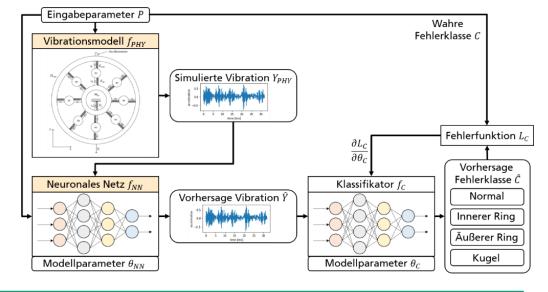


### **Projekt KI-MUSIK 4.0**



- Mikroelektronik-basierte universelle Sensor-Schnittstelle mit Künstlicher Intelligenz für Industrie 4.0
- Überwachung industrieller Anlagen und Maschinen mit neuartigen akustische Sensoren
- Physics-guided Neural Networks
  - Kombination Simulation und Daten
- Anwendungen:
  - Wälzlager
  - Werkzeugmaschinen
  - Brennstoffzelle

#### **Hybrides Modell**

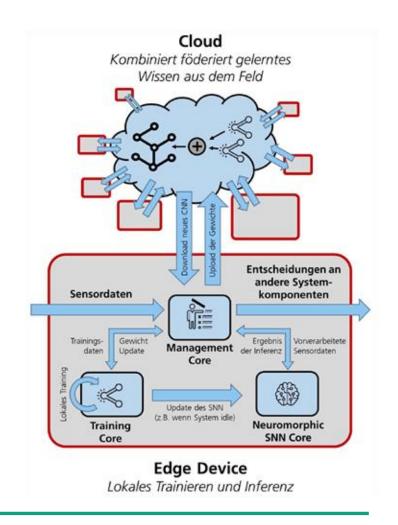




# Fraunhofer Verbundprojekt: SEC-Learn



- Projekt SEC-Learn (Sensor Edge Cloud für föderiertes Lernen)
  - Fraunhofer KI Plattform
  - 11 Fraunhofer Institute
  - Föderiertes Lernen (Federated Learning)
  - Training von neuronalen Netzen in mehreren Sensorknoten
  - Jedes Gerät erhält nur einen Teil der Trainingsdaten
  - Alle anderen Sensorknoten profitieren von dem Erlernten
  - Das fertige Modell wird abschließend auf alle Systeme verteilt
  - Spiking neural networks (SNNs) Core für low power
  - Cloudanbindung für die Kommunikation und Updates der Gewichte



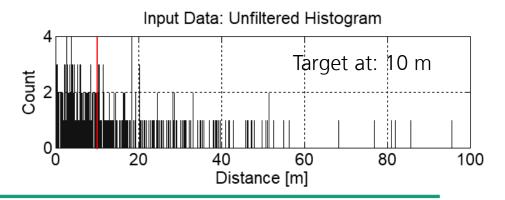


### **AIfES – Research: AI basiertes LIDAR**



- Vorverarbeitung von Sensorsignalen in Echtzeit in einer LIDAR-Kamera
- Der Einsatz von KI sollte die Leistung bei starkem Umgebungslicht verbessern
- Ein KNN wertet das Histogramm der zeitkorrelierten Einzelphotonenzahl aus
  - Besteht zu 95% aus Rauschen und nur zu 5% aus Zielinformationen.
  - Hintergrundrauschen aufgrund des Umgebungslichts stört die Messung
- Das KNN errechnet die Entfernung des Objekts und ersetzt alle Filter
- Verbesserung bei erhöhtem Umgebungslicht
  - Verbesserung der Genauigkeit um etwa 20%
  - Die KI kann die Messung bewerten







### **Contact**



#### Dr. Pierre Gembaczka

Program Manager: Industrial Al

Product Manager: AlfES

Fraunhofer IMS

Finkenstraße 61, 47057 Duisburg

Phone +49 203 3783-220

Email <u>pierre.gembaczka@ims.fraunhofer.de</u>

AlfES - Artificial Intelligence for Embedded Systems www.aifes.ai

www.aifes.de



