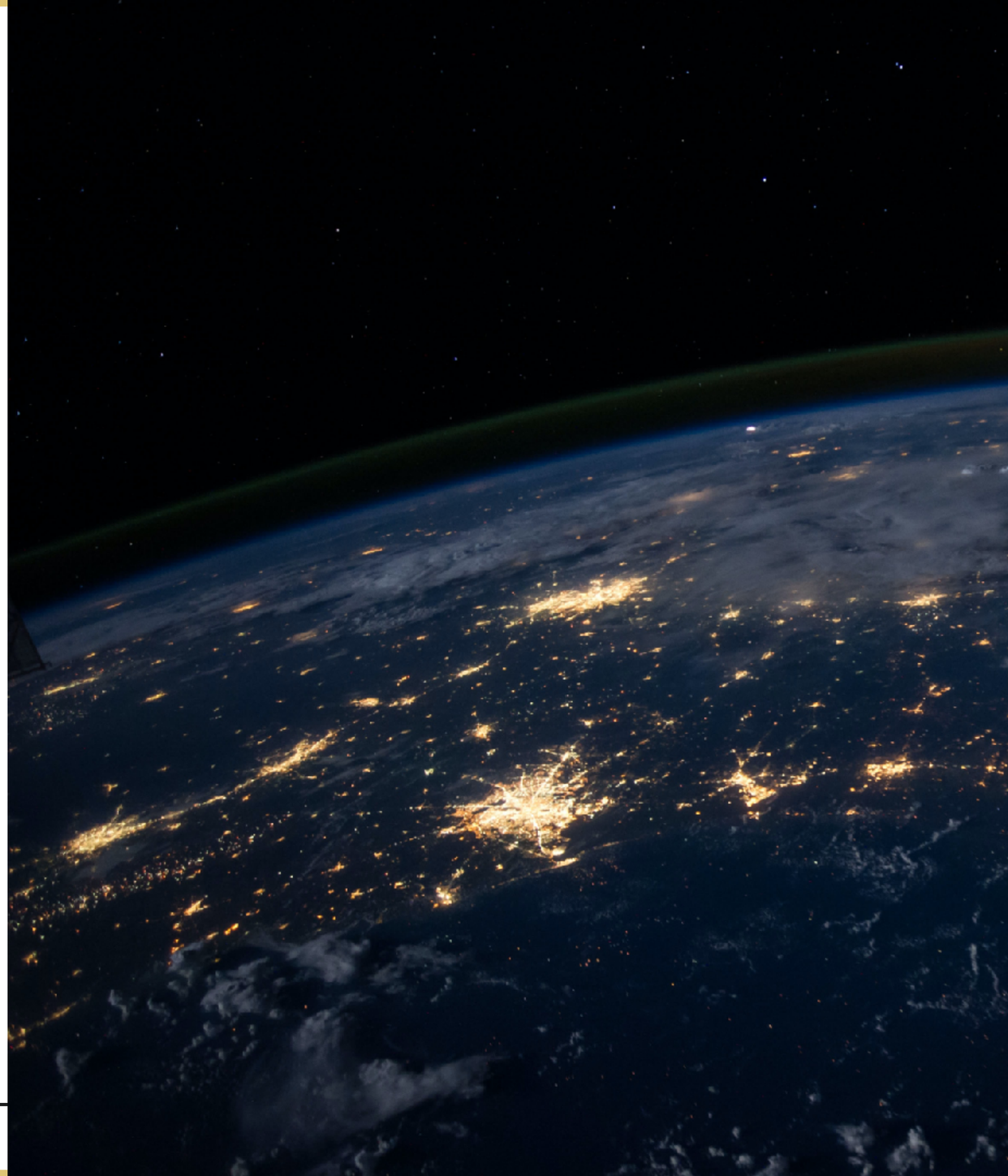
Living on the Edge

Edge-Analytics in der Praxis @ TamedAI GmbH

29.10.2020

Ole Meyer
ole.meyer@tamed.ai



Beyond Clouds - TamedAI

- Die TamedAI GmbH ist eine Forschungsausgründung der Universität Duisburg-Essen
- Aufgabengebiet:
  - Entwicklung von KI-basierten Speziallösungen und/oder kritischen Systemen
- Vorgehen:
  - Produktbasiert durch Verwendung von vorgebauten Komponenten
  - Hands-On: Weniger Consultancy, mehr Taten
- Ziel
  - Anpassbarkeit: Vollständige Kontrolle über alle KI-Module
  - Sicherheit: Keine externen Services, die Zielplattform ist frei wählbar
  - Forschungsnah: Lösungen auf dem aktuellen Stand der Forschung

Edge-Analytics - Entscheidungsgründe in der Praxis

Tamed!AI

Geschwindigkeit

- Zielgröße:
  - Antwortzeit im Bereich weniger Millisekunden
- Beispiel:
  - AR im Operationssaal
- Herausforderungen
  - Insbesondere State-of-the-Art-Modelle häufig nicht Echtzeitfähig ohne besondere Hardware
    - Beispiel GPT2: ca. 1.5 Milliarde Parameter, Ausführungszeit ohne Hochleistungs-GPUs/TPUs > 1 Minute
  - Hardware steht auf Edge-Devices oft nicht zur Verfügung oder ist ein wesentlicher Kostentreiber
- Lösungen
  - Knowledge Distillation
  - Wahl passender Netzwerkarchitekturen (z.B. MobileNet vs NASNet in der Bildverarbeitung)

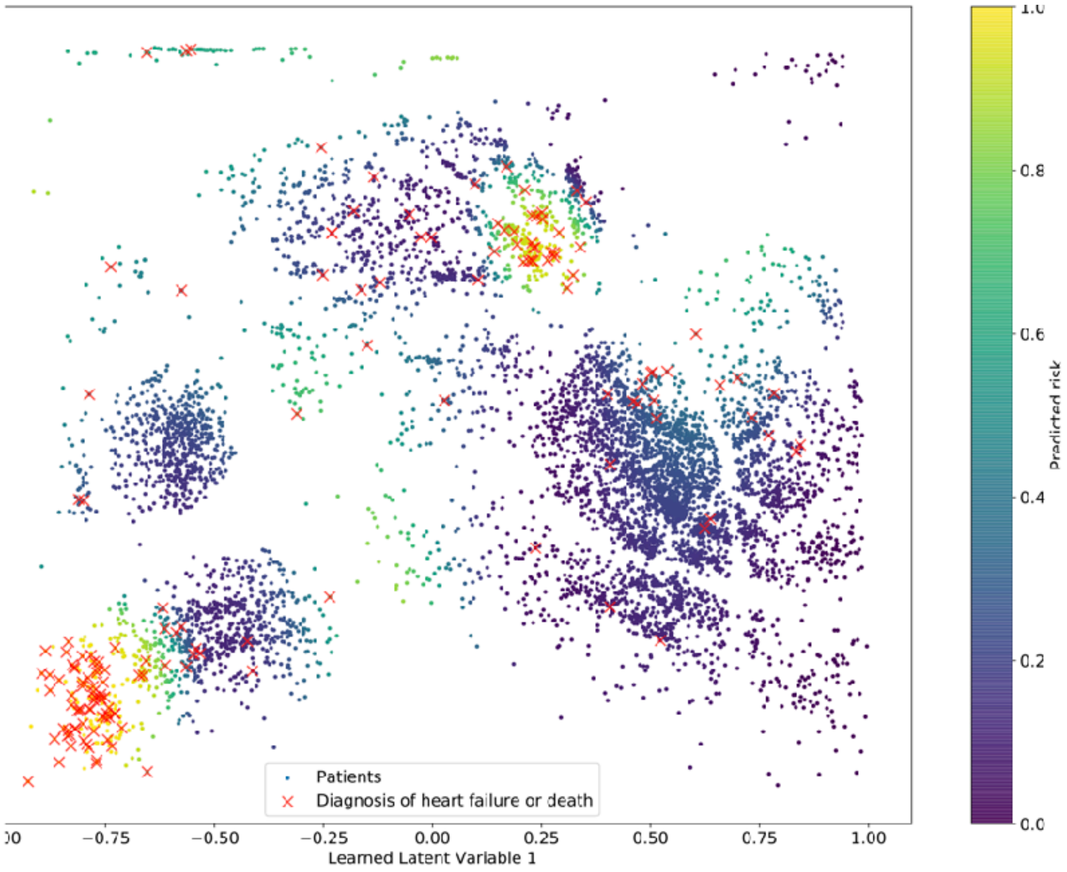
Ausfallsicherheit

- Zielgröße:
  - Ausfalltoleranz <10 Minuten
- Beispiel:
  - Kraftwerksteuerung, Steuerung von Windkraftanlagen
- Herausforderungen
  - Viele Algorithmen sind stochastisch.
  - Abgleich bei redundanten Systemen ist deutlich komplexer
  - Korrupte Daten sind schwerer Erkennbar
- Lösungen
  - Integration von „Safety-Frames“: KI-basierte Steuerung wird auf einen sicheren Operationsbereich eingeschränkt
  - Integration von Anomaly Detection (bestimmt Verfahren wie AutoEncoder lassen dies direkt auf Modellebene zu)

Privacy

- Beispiel:
  - Schutz von medizinischen Daten
- Herausforderungen
  - Zusammenführung unterschiedlicher Daten ohne gegenseitiges Vertrauen
  - Gemixte Verwendung von Cloud- und Edge-Ressourcen
- Lösungen
  - Encoding in einen Latenten-Space auf den Edge-Devices
  - Verwendung von Federated-Learning bei mehreren Teilnehmern

Kodierung von Patientendaten



5G - Eingrenzung der möglichen Anwendungen

- Hochsicherheitsanwendungen
  - z.B. Unmittelbare Steuersysteme für autonome Fahrzeuge
  - Verwendung von Funktechnologie stellt oftmals keine Alternative dar
- Hochkapazitive Algorithmen mit hoher Ausführungszeit
  - z.B. State-of-the-Art NLP / Image Processing mit GPT-2/NASNet
  - Ausführung möglich, häufig keine entscheidenden Geschwindigkeitsvorteile durch 5G (relativ)
- Mittlere Modellgrößen mit Echtzeitanforderungen ohne hohe Sicherheitsanforderungen
  - z.B. AR Anwendungen oder Live-Übersetzungen
  - Einsatz von 5G ist optimal
    - Schnelle Modellausführung macht Geschwindigkeitsunterschiede sichtbar
    - Einsatz von Funktechnologien passt zur Sicherheitsanforderung

We build AI-based systems.

Tamed for your business.

Wild enough for a revolution.

www.tamed.ai | ole.meyer@tamed.ai

Tamed!AI