

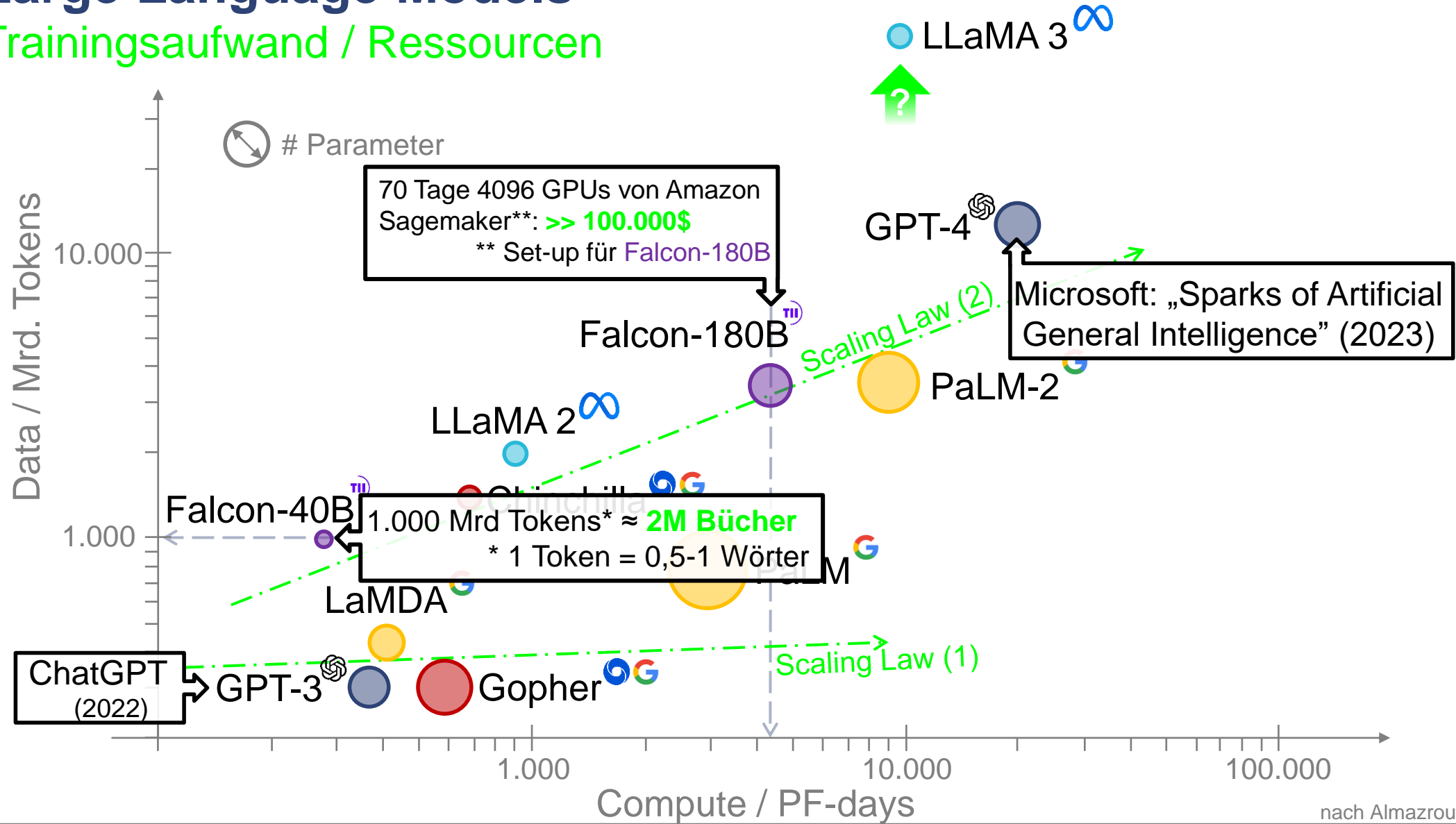
# Warum ChatGPT & Co. nicht in die Produktion gehören

KI in der Produktion

Webinar | Max Schwenger | 2024-07-03 | macils | öffentlich

# Large Language Models

## Trainingsaufwand / Ressourcen



nach Almazrouei et al. 2023

# Large Language Models

## Limitierungen

### Begrenztes Wissen

Das Wissen von *Foundation-Modellen* ist bereits bei ihrer Veröffentlichung veraltet

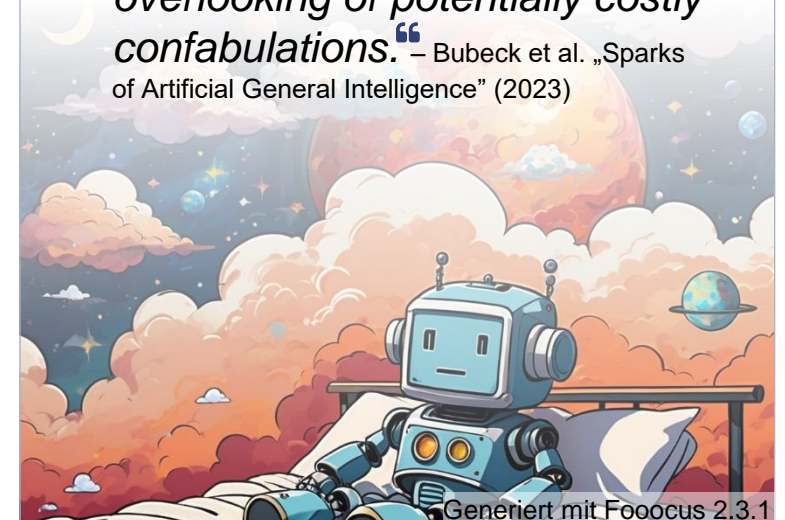
„GPT-4 generally lacks knowledge of events that have occurred after the vast majority of its pre-training data cuts off in September 2021, and does not learn from its experience.“ – OpenAI: GPT-4 Technical Report 2023

(Trainingsdauer **Falcon-180B**: 70 Tage)

### Halluzinieren

Wortwahrscheinlichkeiten sind keine Wahrheit

„Over-reliance on generations can lead to a missing or overlooking of potentially costly confabulations.“ – Bubeck et al. „Sparks of Artificial General Intelligence“ (2023)



Generiert mit Fooocus 2.3.1

# Large Language Models

## Einsatzfelder in der Produktion - Selbstwahrnehmung

### ChatGPT (GPT-3.5)

1. Predictive Maintenance
2. Quality Control
3. Supply Chain Management
4. Process Automation
5. Data Analysis and Reporting
6. Human-Machine Interaction
7. Product Development
8. Compliance and Risk Management
9. Customer Service and Support

### Claude 3

1. Predictive Maintenance
2. Process Optimization
3. Quality Control
4. Supply Chain Management
5. Technical Documentation
6. Customer Support
7. Safety and Compliance
8. Research and Development

### Falcon-180B

1. Predictive maintenance
2. Quality control
3. Customer service
4. Supply chain management
5. Process optimization
6. Safety
7. Product development

# Large Language Models

## Einsatzfelder in der Produktion - Selbstwahrnehmung

### ChatGPT (GPT-3.5)

1. Predictive Maintenance

2. Quality Control

3. Supply Chain Management

4. Process Automation

5. Data Analysis and Reporting

6. Human-Machine Interaction

7. Product Development

8. Compliance and Risk Management

9. Customer Service and Support

### Claude 3

1. Predictive Maintenance

2. Process Optimization

3. Quality Control

4. Supply Chain Management

5. Technical Documentation

6. Customer Support

7. Safety and Compliance

8. Research and Development

### Falcon-180B

1. Predictive maintenance

2. Quality control

3. Customer service

4. Supply chain management

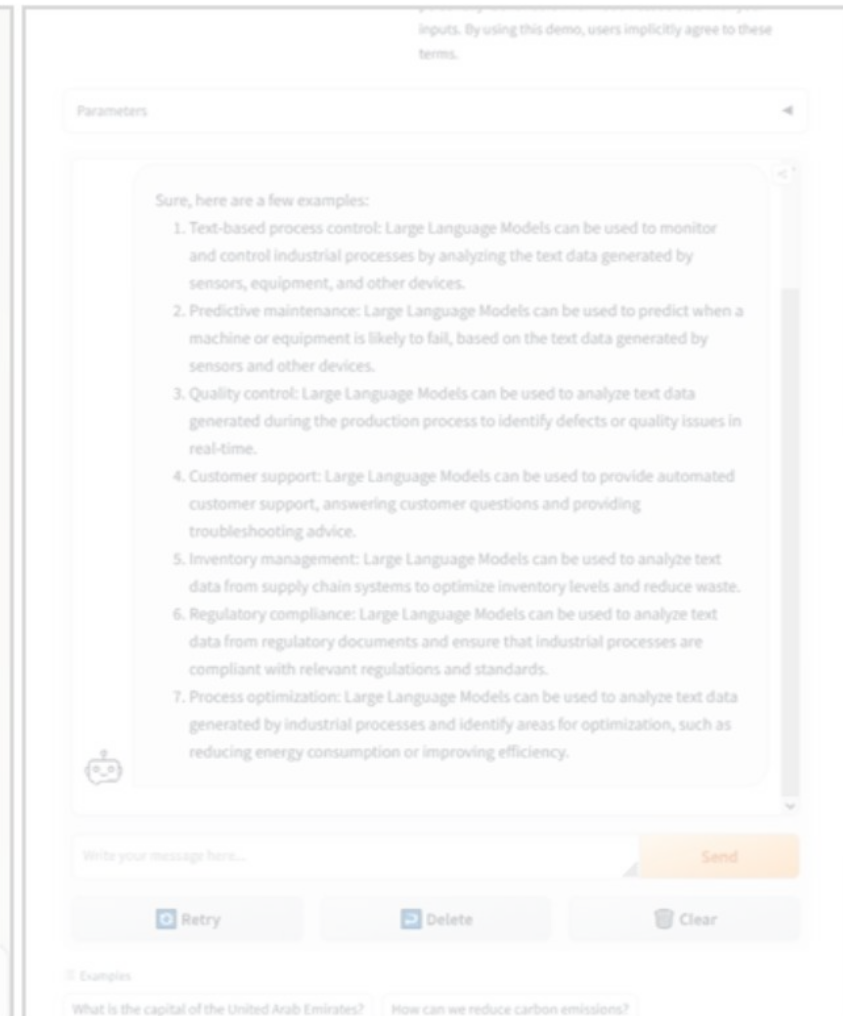
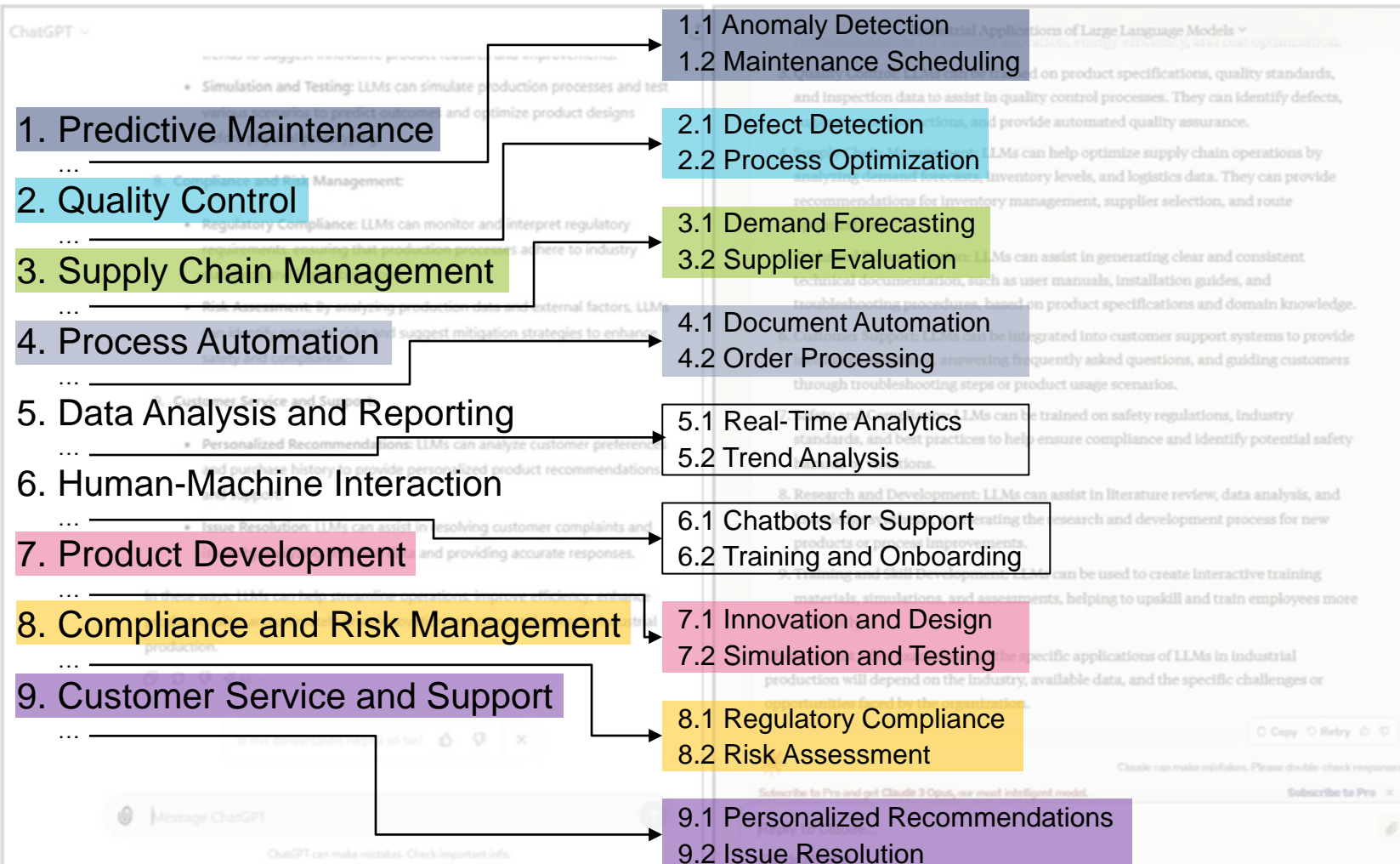
5. Process optimization

6. Safety

7. Product development

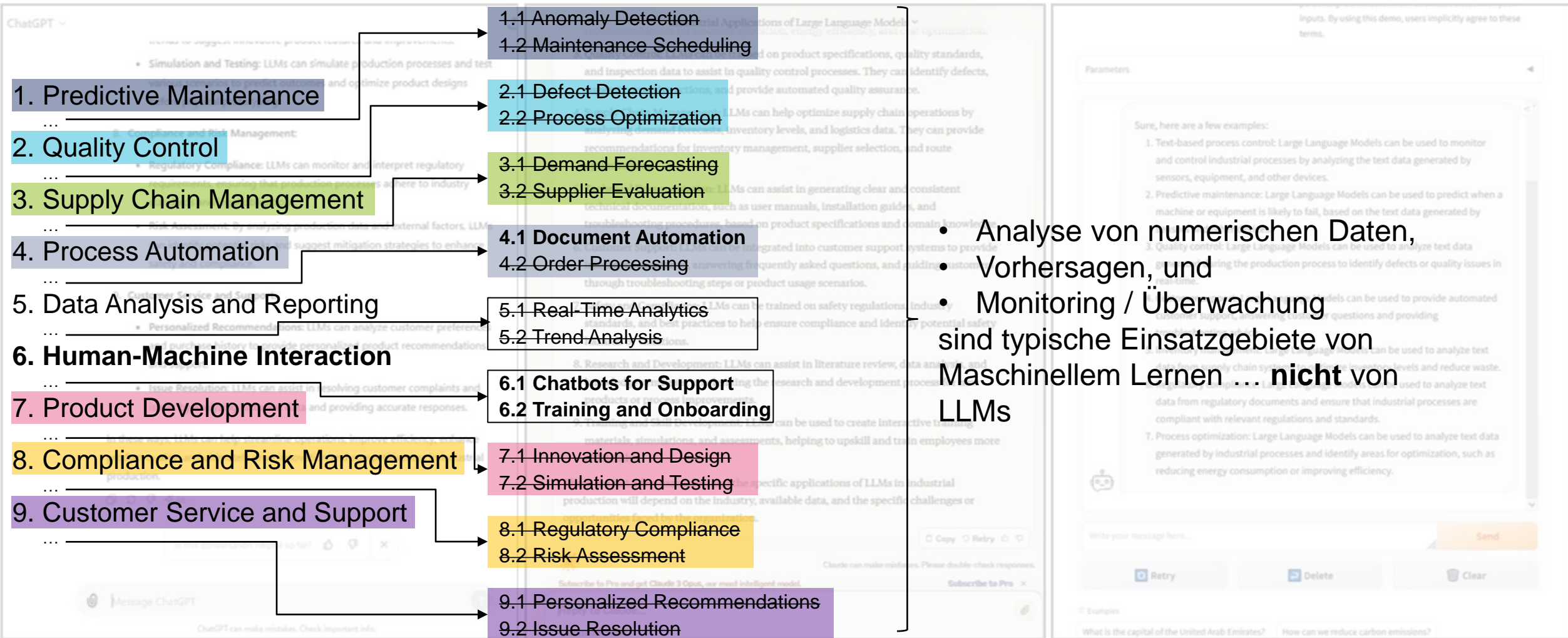
# Large Language Models

## Einsatzfelder in der Produktion - Selbstwahrnehmung



# Large Language Models

## Einsatzfelder in der Produktion - Selbstwahrnehmung



# Large Language Models

## Einsatzfelder in der Produktion

- **Fehlersuche / Instandhaltung:** als (bessere?) Suchmaschine in der Maschinendokumentation  
⇒ Kontextlänge < 60 Seiten
- **Textanalyse von Instandhaltungsmeldungen:** Identifikation wiederkehrender Fehlerbilder und übersehener Bottlenecks  
⇒ einmalige Optimierung; **haben Sie die Daten?**
- **Co-Programmieren:** Unterstützung beim (SPS-)Programmieren neuer Anlagen
- **Dikturfunktion zu Dokumentation:** Das ist kein Sprachmodell sondern eine Transkription
- **Mensch-Datenbank-Interface bei der Datenanalyse:** das LLM übersetzt Fragestellungen in Code für eine Datenbankabfrage und visualisiert das Ergebnis ggf.  
⇒ **Text-to-SQL**

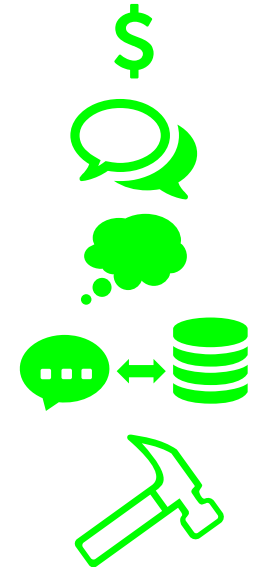




# Fazit

## Warum LLMs nichts in der Produktion verloren haben

- **RESSOURCEN:** RAM, FLOPs, Daten
- **DATENTYP:** Sprachmodelle in einer Welt voller Zahlen
- **HALLUZINIEREN:** Wollen Sie (ausgedachte) Fehler tolerieren?
- **STÄRKEN NUTZEN:** informelles Brainstorming, Coding, als Schnittstelle
- **KNOW YOUR TOOLS:** Nur weil LLMs nun das erste Mal sind, dass auch das Management wesentlich mit KI in Berührung kommt, ist es nicht die Antwort auf alles



Kontakt:

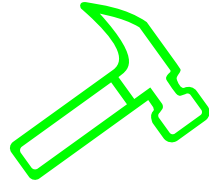
Max Schwenzer

Production Data Scientist

Tel. 0173 1527023

max.schwenzer@voith.com

*IF THE ONLY TOOL YOU HAVE IS A HAMMER,  
IT IS TEMPTING TO TREAT EVERYTHING AS  
IF IT WERE A NAIL – Abraham Maslow*



THE  
END

# Quellen

Almazrouei, Ebtesam, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, u. a. „The Falcon Series of Open Language Models“. arXiv, 29. November 2023. <http://arxiv.org/abs/2311.16867>.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, u. a. „Sparks of Artificial General Intelligence: Early experiments with GPT-4“. arXiv, 13. April 2023. <http://arxiv.org/abs/2303.12712>.