RE und KI @ GI

Juristische Anforderungen an KI: Architektur und Design-Entscheidungen

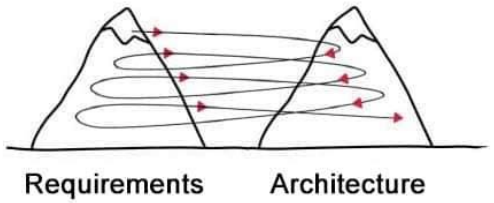
Prof. Dr. Alexander Rachmann

Fachbereich Technology
Professur für Anwendungsorientierte Informatik

E-Mail: a.rachmann@cbs.de

Disclaimer

- Die hier vorgestellten Inhalte bilden den aktuellen, nicht-finalen Stand des „AI Acts“ ab.
- Die Inhalte des AI Acts werden von aus der Sicht des Requirements Engineer (& Software Architektur) interpretiert. Denkt hierbei an das Twin Peaks-Model.
- Die Inhalte interpretiere ich nicht als Jurist – entsprechende Kompetenz maße ich mir nicht an.
- In meiner sonstigen Arbeit untersuche ich i.d.R. betriebliche Informationssysteme und behalte diesen Fokus auch hier bei. Daher werde ich wenig Augenmerk auf gesellschaftliche oder militärische Anwendungen legen.
- Für den Inhalt des AI Acts beziehe ich mich auf die deutsche Fassung unter eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206 (August 2023).
  - Noch im Juni 2023 wurden „substanzielle Änderungen“ am Text vorgenommen, die erst im Ausblick aufgenommen werden konnten.
  - Der Trilog-Prozess (Europäische Kommission, Rat der EU, Europäischas Parkament) wurde erst im Juli abgeschlossen.



Agenda

- AI Act
- Architektur eines KI-Systems nach Huyen
- Erweiterte Architektur
- Schluss & Ausblick

Struktur AI Act

Titel 1: Allgemeine Bestimmungen

- Artikel 1 Gegenstand
- Artikel 2 Anwendungsbereich
- Artikel 3 Begriffsbestimmungen
- Artikel 4 Änderungen des Anhans

Titel 2: Verbotene Praktiken im Bereich der künstlichen Intelligenz

Titel 3: Hochrisiko-KI-Systeme

- Kapitel 1: Klassifizierung von KI-Systemen als Hochrisiko-Systeme
- Kapitel 2: Anforderungen an Hochrisiko-KI-Systeme
- Kapitel 3: Pflichten der Anbieter und Nutzer von Hochrisiko-KI-Systemen und anderer Beteiligter

Ablauf Gesetzgebung EU

- Entwurfsphase
  - Europäische Kommission erstellt Entwurf
  - Einbindung von Expertengruppen
  - Folgenabschätzung
- Diskussionsphase
  - Veröffentlichung des Entwurfs
  - Überprüfung durch den Rat und das Parlament
  - Trilog-Verfahren
- Abstimmungsphase und Abschluss
  - Annahme der Verordnung
  - Veröffentlichung im Amtsblatt

Gegenstand des AI Acts

In dieser Verordnung wird Folgendes festgelegt:

a) harmonisierte Vorschriften für das Inverkehrbringen, die Inbetriebnahme und die Verwendung von Systemen der künstlichen Intelligenz (im Folgenden „KI-Systeme“) in der Union;

b) Verbote bestimmter Praktiken im Bereich der künstlichen Intelligenz;

c) besondere Anforderungen an Hochrisiko-KI-Systeme und Verpflichtungen für Betreiber solcher Systeme;

d) harmonisierte Transparenzvorschriften für KI-Systeme, die mit natürlichen Personen interagieren sollen, für KI-Systeme zur Emotionserkennung und zur biometrischen Kategorisierung sowie für KI-Systeme, die zum Erzeugen oder Manipulieren von Bild-, Ton- oder Videoinhalten verwendet werden;

e)Vorschriften für die Marktbeobachtung und Marktüberwachung.

Anwendungsbereichs des AI Acts

Diese Verordnung gilt für:

a) Anbieter, die KI-Systeme in der Union in Verkehr bringen oder in Betrieb nehmen, unabhängig davon, ob diese Anbieter in der Union oder in einem Drittland niedergelassen sind;

b) Nutzer von KI-Systemen, die sich in der Union befinden;

c) Anbieter und Nutzer von KI-Systemen, die in einem Drittland niedergelassen oder ansässig sind, wenn das vom System hervorgebrachte Ergebnis in der Union verwendet wird.

Definition von KI

„System der künstlichen Intelligenz“ (KI-System) ist eine Software,

- die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und
- im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren.

Anhang I:

- a) Konzepte des maschinellen Lernens, mit beaufsichtigtem, unbeaufsichtigtem und bestärkendem Lernen unter Verwendung einer breiten Palette von Methoden, einschließlich des tiefen Lernens (Deep Learning);
- b) Logik- und wissensgestützte Konzepte, einschließlich Wissensrepräsentation, induktiver (logischer) Programmierung, Wissensgrundlagen, Inferenz- und Deduktionsmaschinen, (symbolischer) Schlussfolgerungs- und Expertensysteme;
- c) Statistische Ansätze, Bayessche Schätz-, Such- und Optimierungsmethoden.

Verbotene Praktiken

- das Inverkehrbringen, die Inbetriebnahme oder die Verwendung eines KI-Systems, das Techniken der unterschwelligen Beeinflussung außerhalb des Bewusstseins einer Person einsetzt, um das Verhalten einer Person in einer Weise wesentlich zu beeinflussen, die dieser Person oder einer anderen Person einen physischen oder psychischen Schaden zufügt oder zufügen kann;
- […] das eine Schwäche oder Schutzbedürftigkeit einer bestimmten Gruppe von Personen aufgrund ihres Alters oder ihrer körperlichen oder geistigen Behinderung ausnutzt […]
- […] KI-Systemen durch Behörden oder in deren Auftrag zur Bewertung oder Klassifizierung der Vertrauenswürdigkeit natürlicher Personen über einen bestimmten Zeitraum auf der Grundlage ihres sozialen Verhaltens oder bekannter oder vorhergesagter persönlicher Eigenschaften oder Persönlichkeitsmerkmale, […]
- die Verwendung biometrischer Echtzeit-Fernidentifizierungssysteme in öffentlich zugänglichen Räumen zu Strafverfolgungszwecken,

Hochrisikosysteme

(1) Ungeachtet dessen, ob ein KI-System unabhängig von den unter den Buchstaben a und b genannten Produkten in Verkehr gebracht oder in Betrieb genommen wird, gilt es als Hochrisiko-KI-System, wenn die beiden folgenden Bedingungen erfüllt sind:

- a) das KI-System soll als Sicherheitskomponente eines unter die in Anhang II aufgeführten Harmonisierungsrechtsvorschriften der Union fallenden Produkts verwendet werden oder ist selbst ein solches Produkt;
- b) das Produkt, dessen Sicherheitskomponente das KI-System ist, oder das KI-System selbst als Produkt muss einer Konformitätsbewertung durch Dritte im Hinblick auf das Inverkehrbringen oder die Inbetriebnahme dieses Produkts gemäß den in Anhang II aufgeführten Harmonisierungsrechtsvorschriften der Union unterzogen werden.

(2) Zusätzlich zu den in Absatz 1 genannten Hochrisiko-KI-Systemen gelten die in Anhang III genannten KI-Systeme ebenfalls als hochriskant.

Anhang II: 19 Richtlinien und Verordnungen, z.B. Richtlinien für Maschinen, Sicherheit von Spielzeug, Funkanlagen, Druckgeräten, Medizinprodukte, Zivilluftfahrt, Kraftfahrzeugen

Anhang III: siehe folgende Seite

Hochrisikosysteme

- Biometrische Identifizierung und Kategorisierung natürlicher Personen
- Verwaltung und Betrieb kritischer Infrastrukturen
- Allgemeine und berufliche Bildung
- Beschäftigung, Personalmanagement und Zugang zur Selbstständigkeit
- Zugänglichkeit und Inanspruchnahme grundlegender privater und öffentlicher Dienste und Leistungen
- Strafverfolgung
- Migration, Asyl und Grenzkontrolle
- Rechtspflege und demokratische Prozesse

Hochrisikosysteme: Studie zur Risikoklassifizierung

Unternehmensfunktion	Hochrisiko	Unklar	Niedrigrisiko	Verboten	Summe
Buchhaltung und Finanzen	3	7			10
Einkauf		2	6		8
Forschung und Entwicklung		5	4		9
IT und Sicherheit	2	6	2	1	11
Kundenservice	4	4	6		14
Logistik und Lieferketten		6	5		11
Marketing und Vertrieb		1	13		14
Personalwesen	9	1			10
Produktion und Herstellung		5	4		9
Rechtswesen	1	5	4		10
Summe	**19**	**42**	**44**	**1**	**106**

https://www.appliedai.de/hub/ai-act-risikoklassifizierung-von-ki-systemen-aus-einer-praktischen-perspektive

Anforderungen an Hochrisiko-KI-Systeme

- Artikel 9: Risikomanagementsystem
- Artikel 10: Daten und Daten-Governance
- Artikel 11: Technische Dokumentation
- Artikel 12: Aufzeichnungspflichten
- Artikel 13: Transparenz und Bereitstellung von Informationen für die Nutzer
- Artikel 14: Menschliche Aufsicht
- Artikel 15: Genauigkeit, Robustheit und Cybersicherheit

Fachliche Pflichten

- Artikel 17: Qualitätsmanagementsystem
- Artikel 18: Pflicht zur Erstellung der technischen Dokumentation
- Artikel 19: Konformitätsbewertung
- Artikel 20: Automatisch erzeugte Protokolle
- Artikel 21: Korrekturmaßnahmen
- Artikel 22: Informationspflicht
- Artikel 23: Zusammenarbeit mit den zuständigen Behörden

Akteurspezifische Pflichten

- Artikel 24: Pflichten der Produkthersteller
- Artikel 25: Bevollmächtigte
- Artikel 26: Pflichten der Einführer
- Artikel 27: Pflichten der Händler
- Artikel 28: Pflichten der Händler, Einführer, Nutzer oder sonstiger Dritter
- Artikel 29: Pflichten der Nutzer von Hochrisiko-KI-Systemen

Artikel 28

Pflichten der Händler, Einführer, Nutzer oder sonstiger Dritter

(1) In den folgenden Fällen gelten Händler, Einführer, Nutzer oder sonstige Dritte als Anbieter für die Zwecke dieser Verordnung und unterliegen den Anbieterpflichten gemäß Artikel 16:

a) wenn sie ein Hochrisiko-KI-System unter ihrem Namen oder ihrer Marke in Verkehr bringen oder in Betrieb nehmen;

b) wenn sie die Zweckbestimmung eines bereits im Verkehr befindlichen oder in Betrieb genommenen Hochrisiko-KI-Systems verändern;

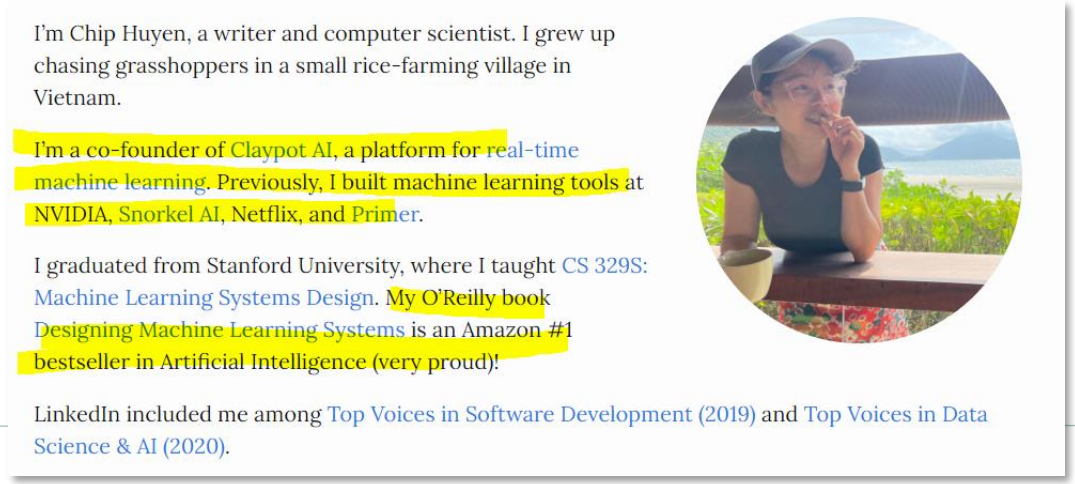
c) wenn sie eine wesentliche Änderung an dem Hochrisiko-KI-System vornehmen.

(2) Unter den in Absatz 1 Buchstabe b oder c genannten Umständen gilt der Anbieter, der das Hochrisiko-KI-System ursprünglich in Verkehr gebracht oder in Betrieb genommen hatte, nicht mehr als Anbieter für die Zwecke dieser Verordnung.

Architektur eines KI-Systems nach Huyen

„Designing Machine Learning Systems“

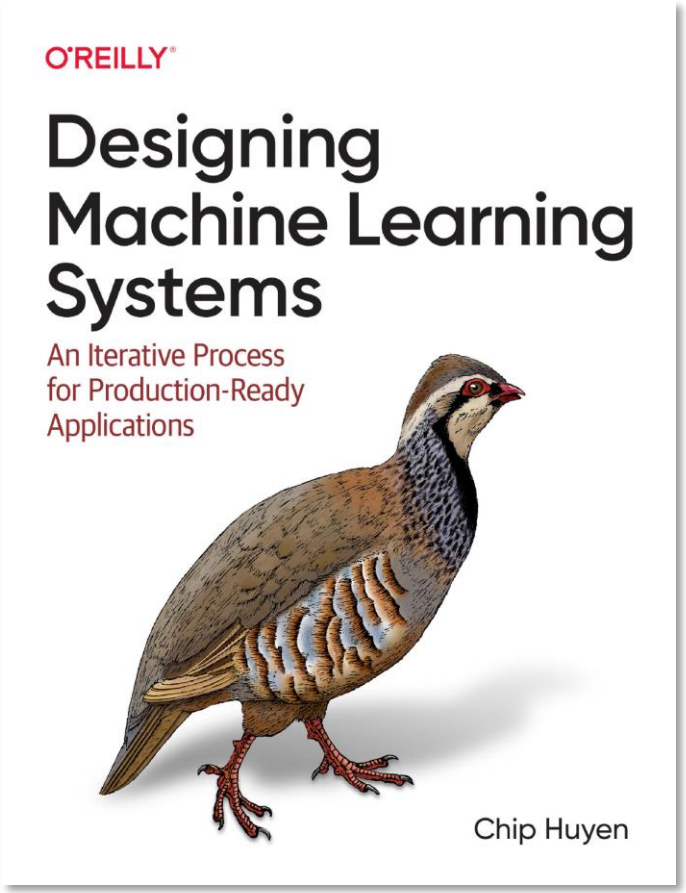
- „This book is for anyone who wants to leverage ML to solve real-world problems. ML in this book refers to both deep learning and classical algorithms, with a leaning toward ML systems at scale, such as those seen at medium to large enterprises and fast-growing startups. Systems at a smaller scale tend to be less complex and might benefit less from the comprehensive approach laid out in this book.
- Because my background is engineering, the language of this book is geared toward engineers, including ML engineers, data scientists, data engineers, ML platform engineers, and engineering managers.”


I'm Chip Huyen, a writer and computer scientist. I grew up chasing grasshoppers in a small rice-farming village in Vietnam.

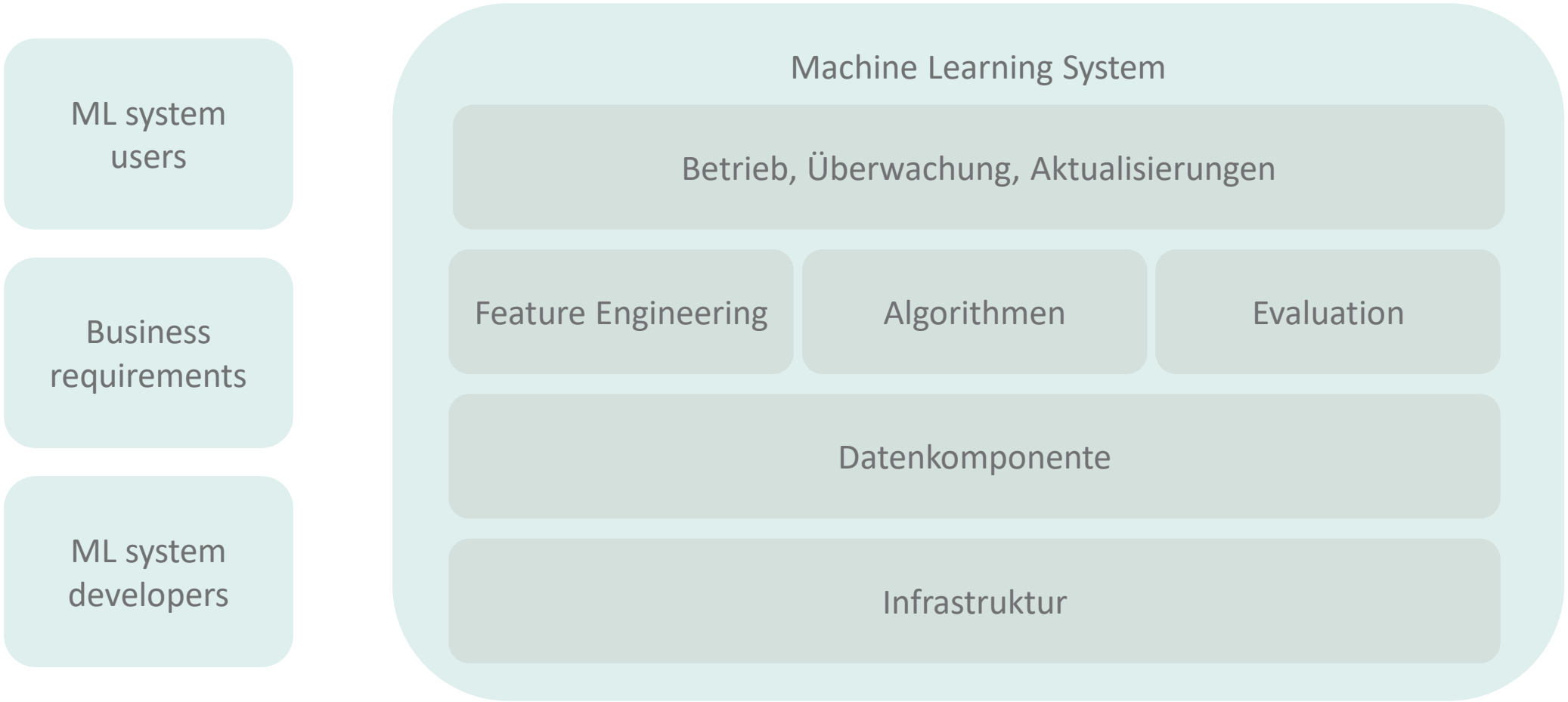
I'm a co-founder of Claypot AI, a platform for real-time machine learning. Previously, I built machine learning tools at NVIDIA, Snorkel AI, Netflix, and Primer.

I graduated from Stanford University, where I taught CS 329S: Machine Learning Systems Design. My O'Reilly book Designing Machine Learning Systems is an Amazon #1 bestseller in Artificial Intelligence (very proud)!

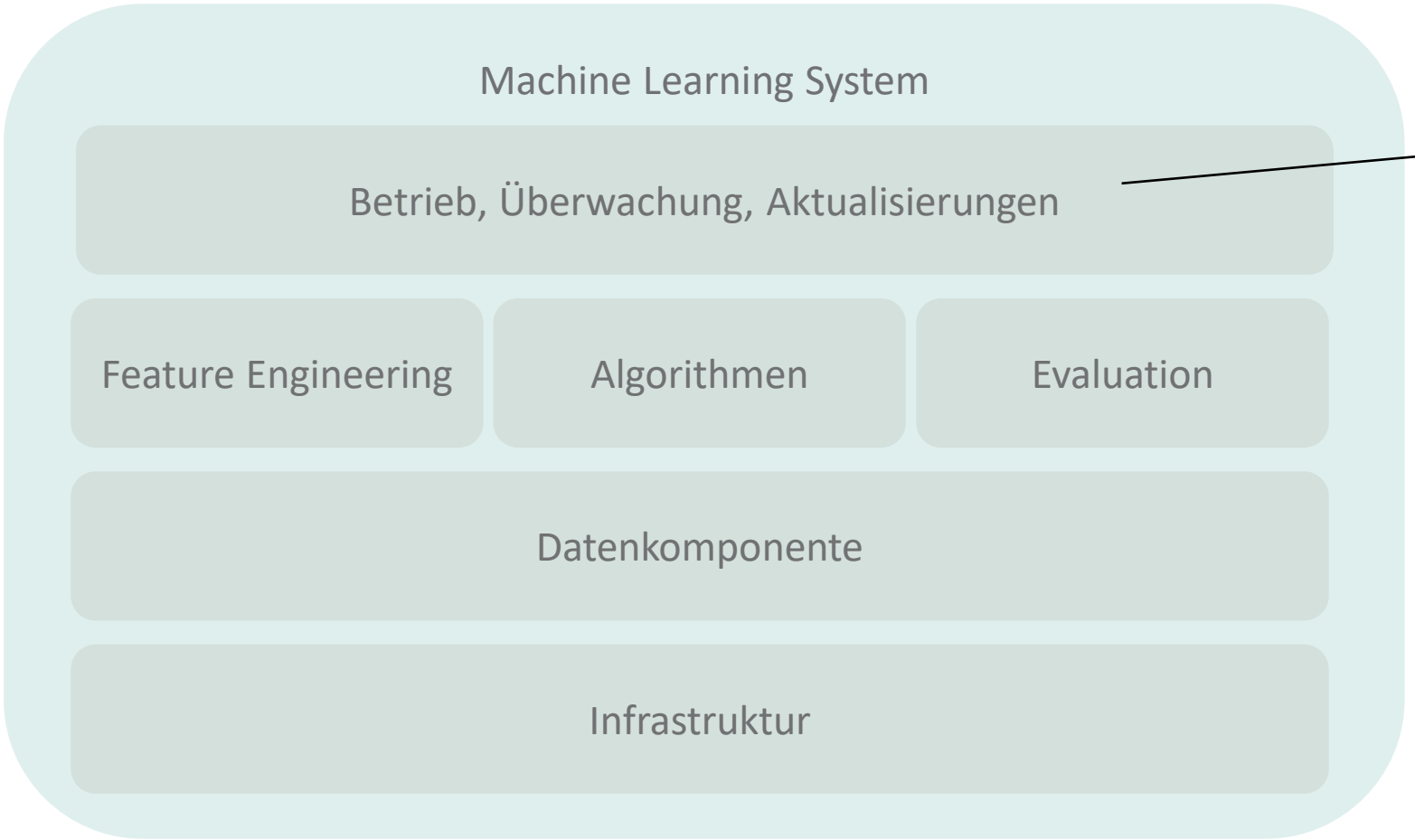
LinkedIn included me among Top Voices in Software Development (2019) and Top Voices in Data Science & AI (2020).



Architektur eines ML-Systems

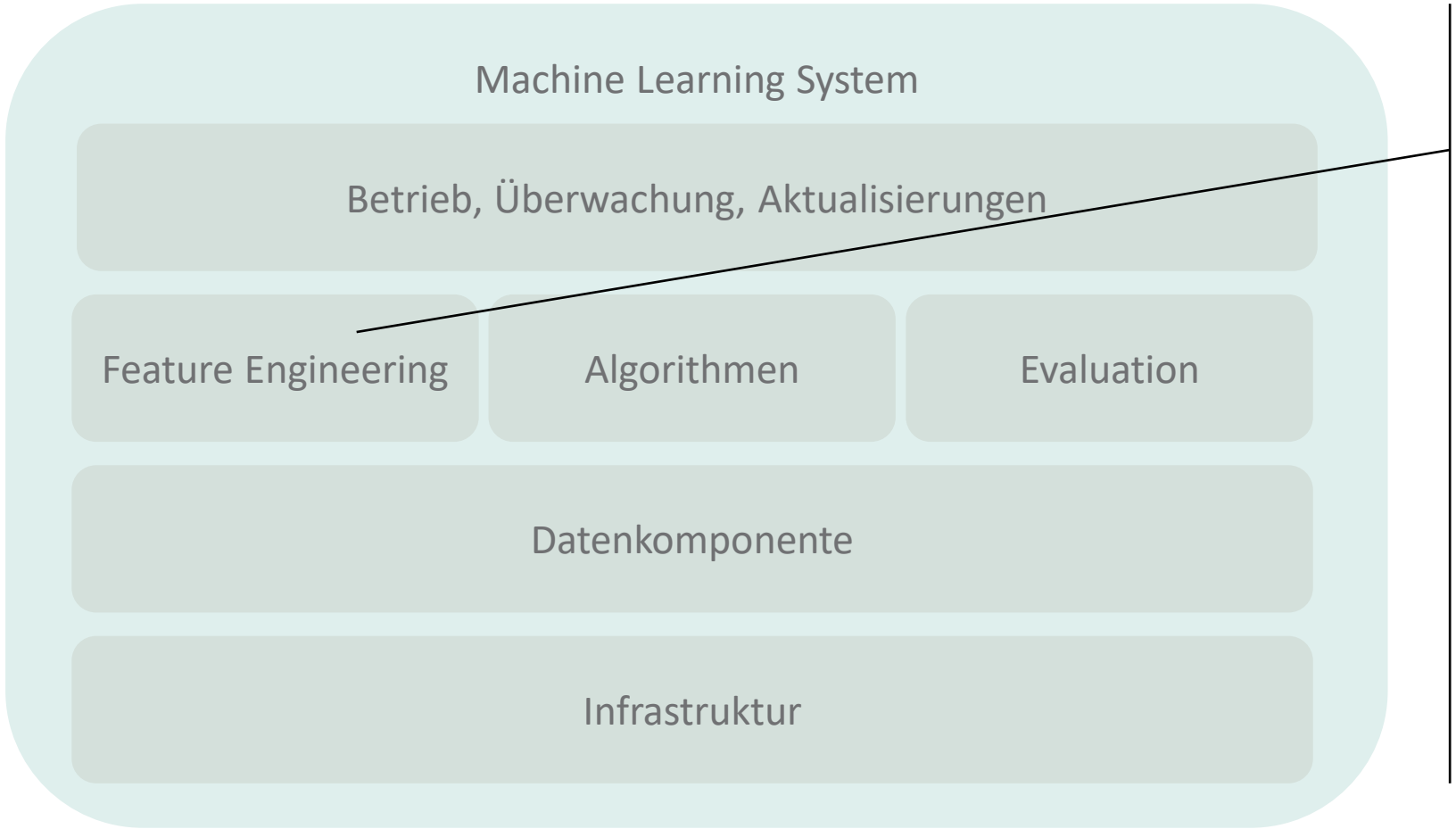


Betrieb, Überwachung, Aktualisierungen



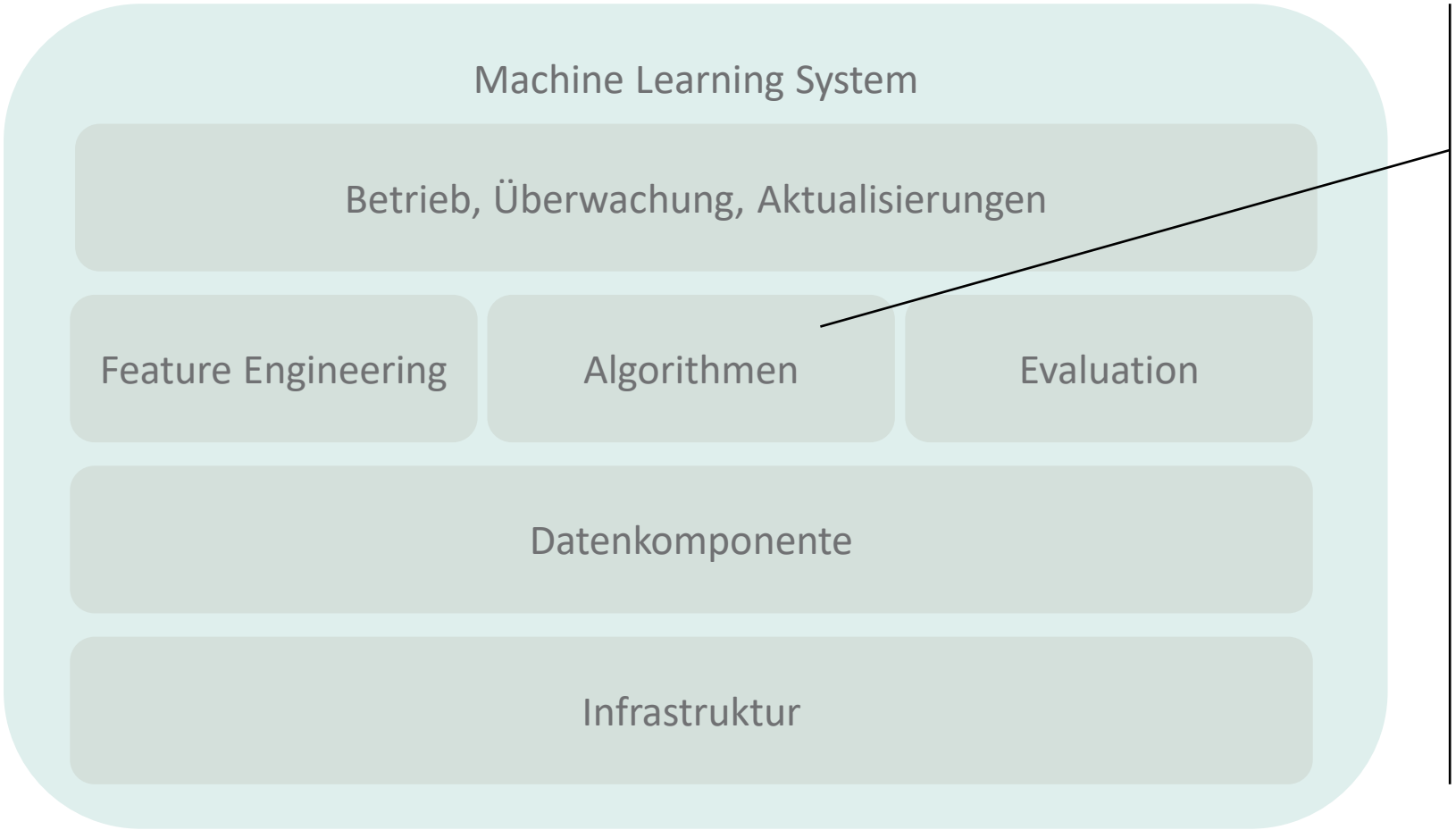
- different ways to deploy a model, comparing online prediction with batch prediction, and ML on the edge with ML on the cloud
- production data differing from training data, edge cases, and degenerate feedback loops
- how to continually update your models in production to adapt them to changing data distributions; trade-offs between model iteration and data iteration

Feature Engineering



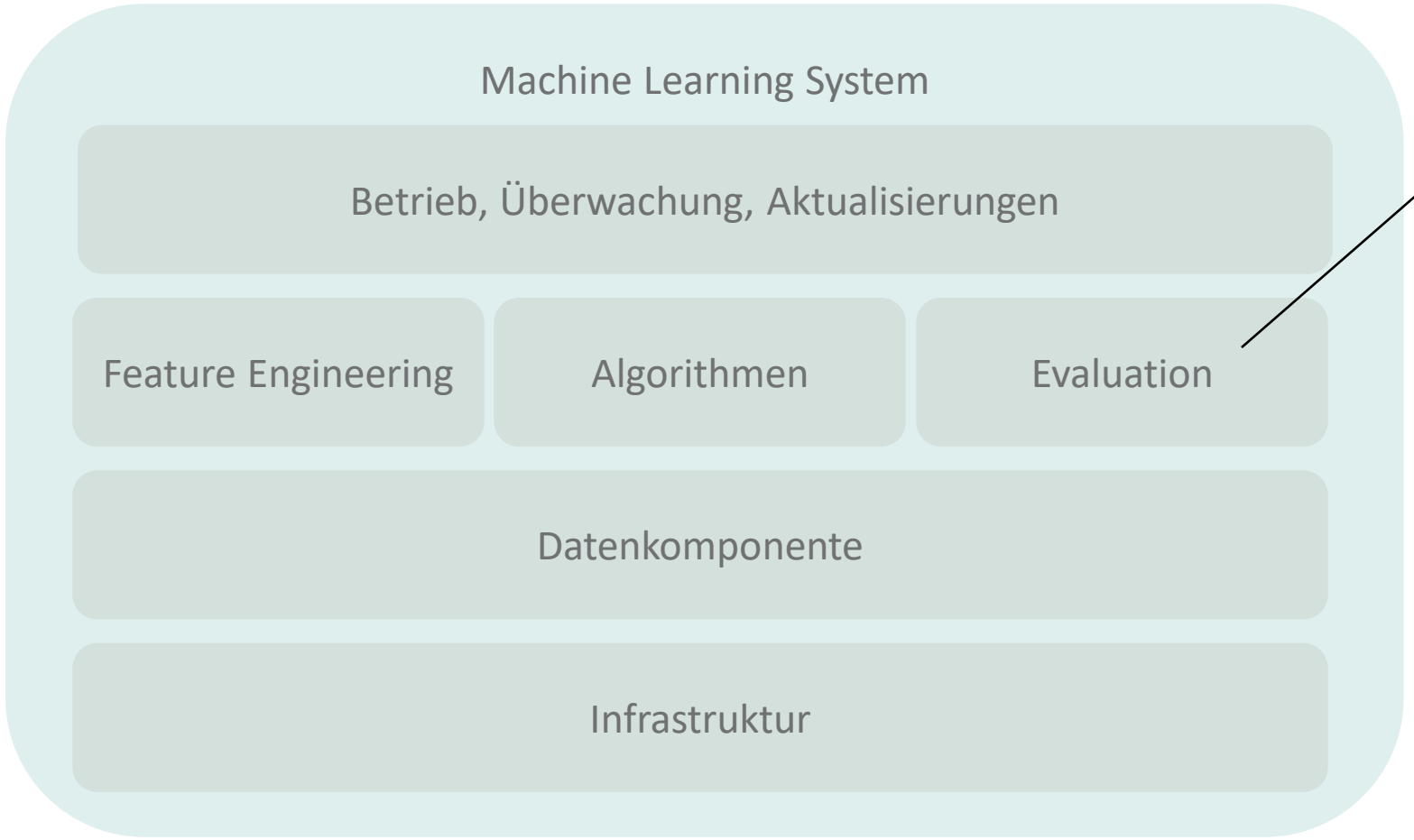
- How to engineer good features is a complex question with no foolproof answers. The best way to learn is through experience: trying out different features and observing how they affect your models' performance.
- "a feature is an individual measurable property or characteristic of a phenomenon" (Wikipedia)

Algorithmen



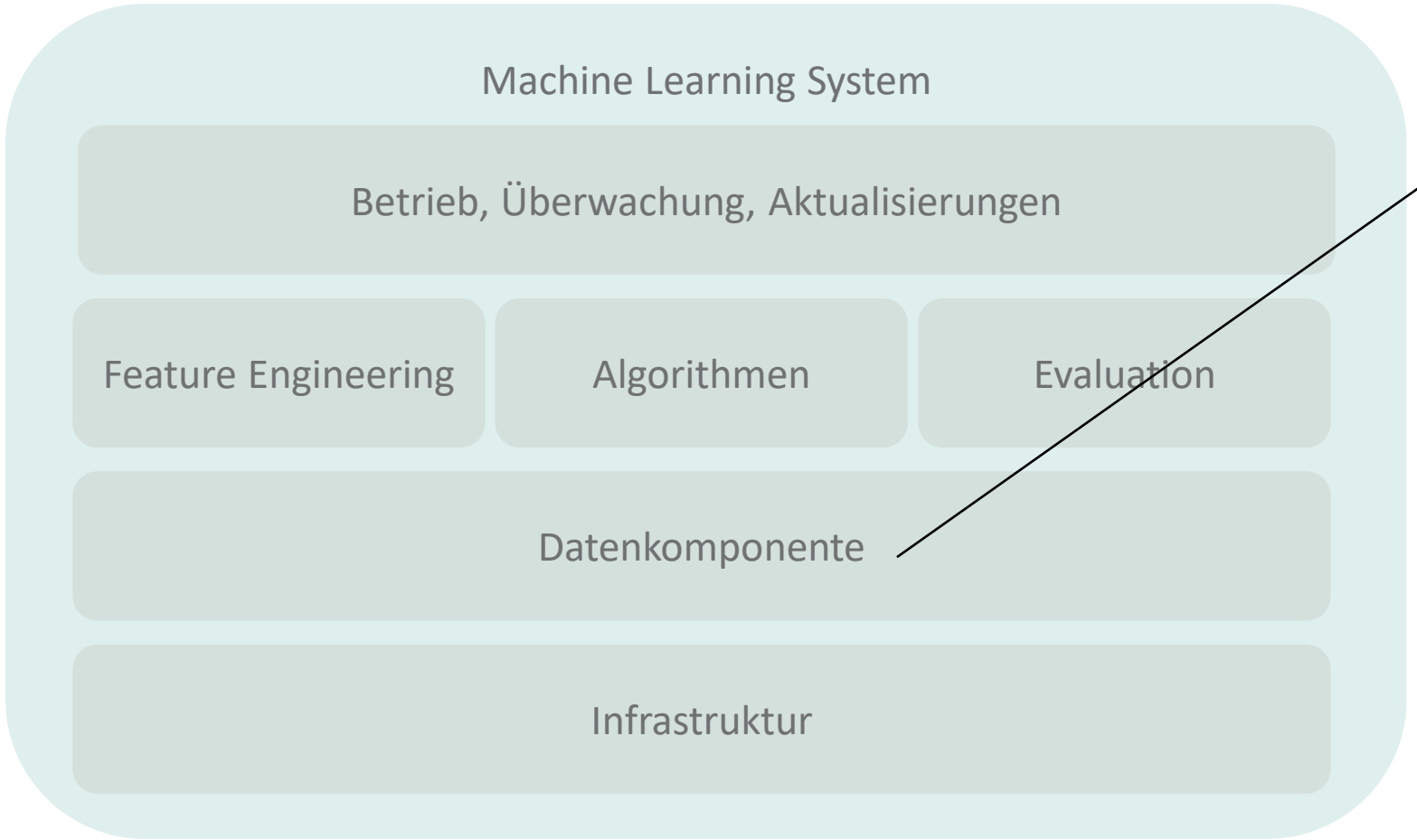
- “part of ML systems, which many ML practitioners consider to be the most fun part of an ML project lifecycle.”

Evaluation



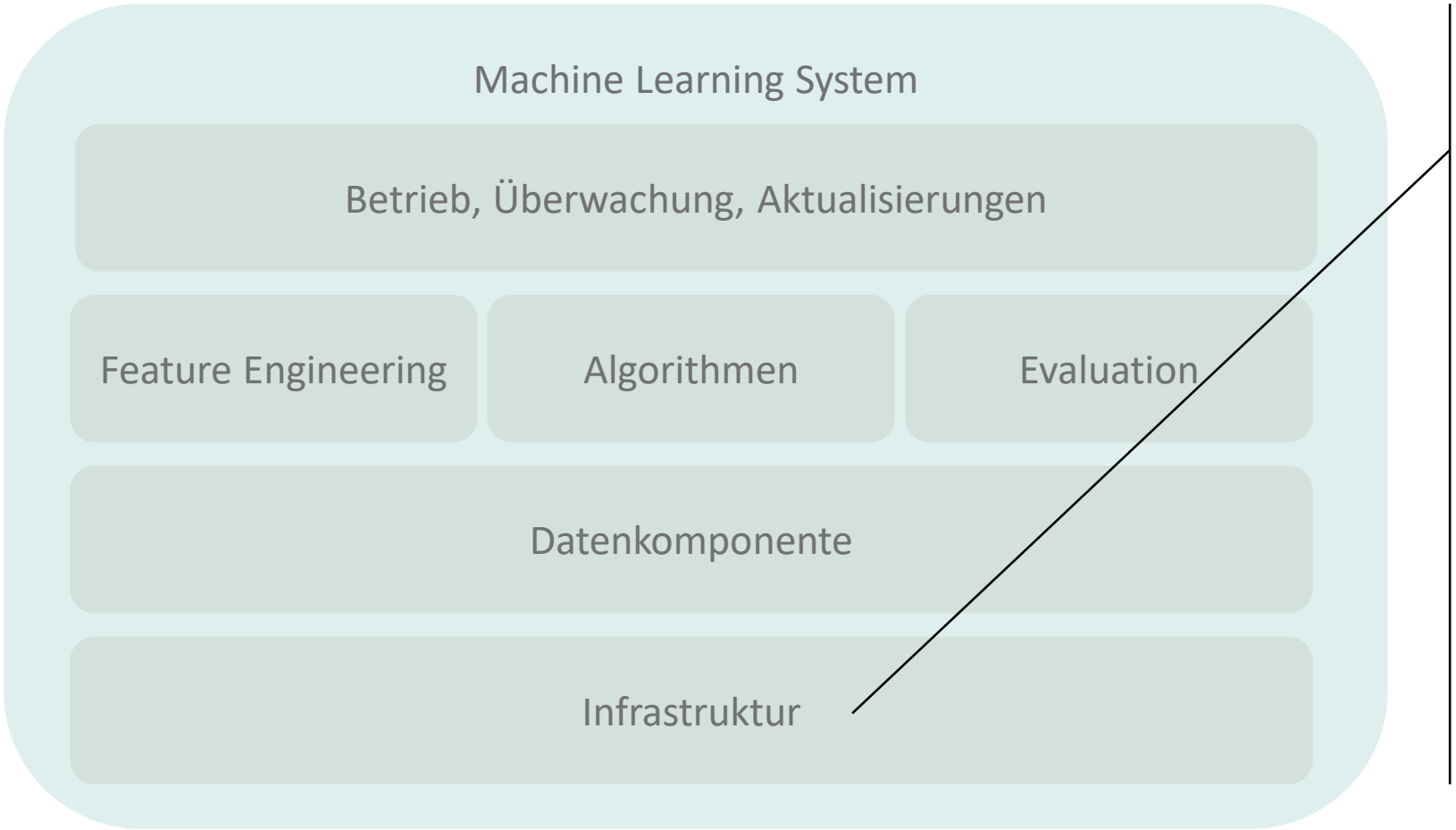
- “how to evaluate your models to pick the best one to deploy.
- Evaluation metrics don’t mean much unless you have a baseline to compare them to, and we covered different types of baselines you might want to consider for evaluation. “

Datenkomponente



- “data models: relational, document, and graph”
- “data storage engines and processing”
- “real-time transport like Apache Kafka and RabbitMQ”
- “No matter how clever your algorithms might be, if your training data is bad, your algorithms won’t be able to perform well.”

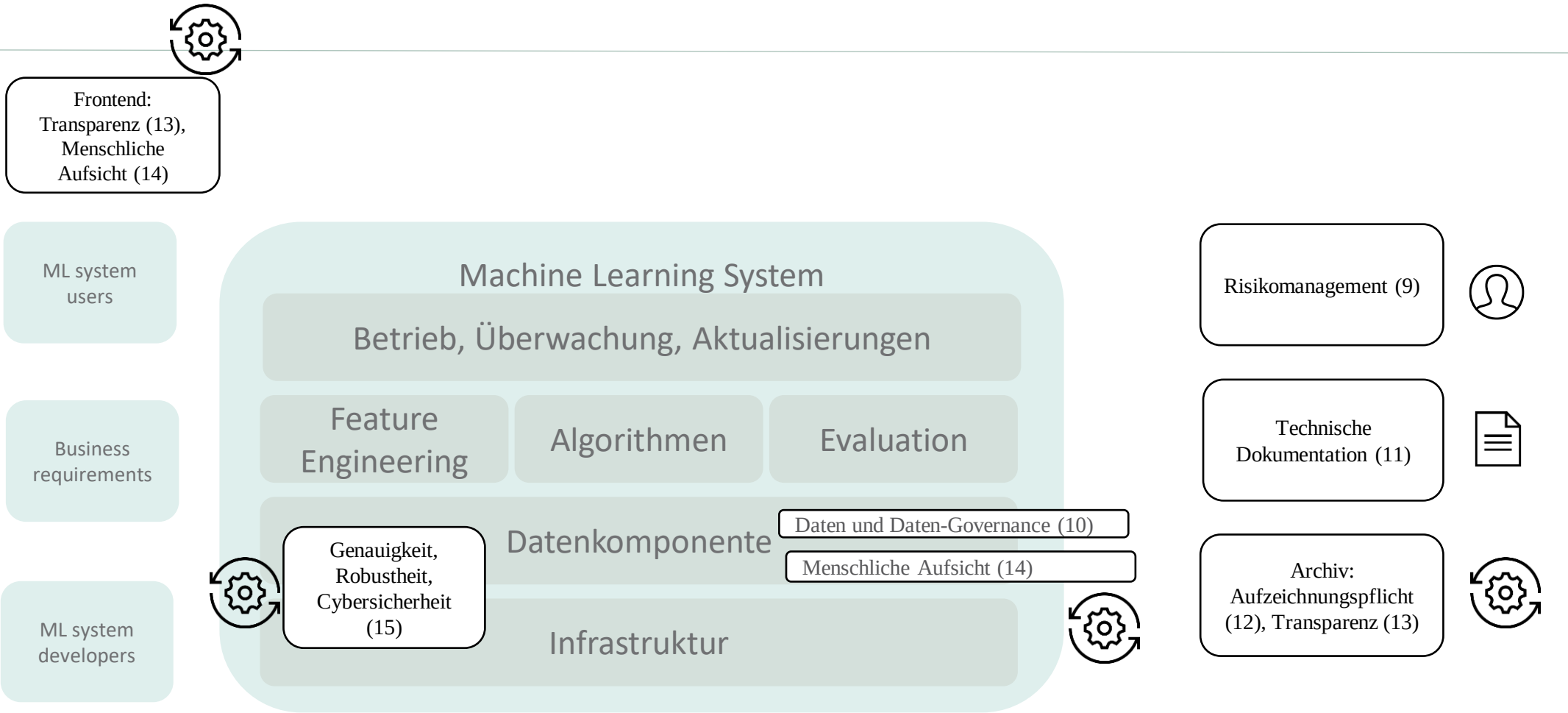
Infrastruktur



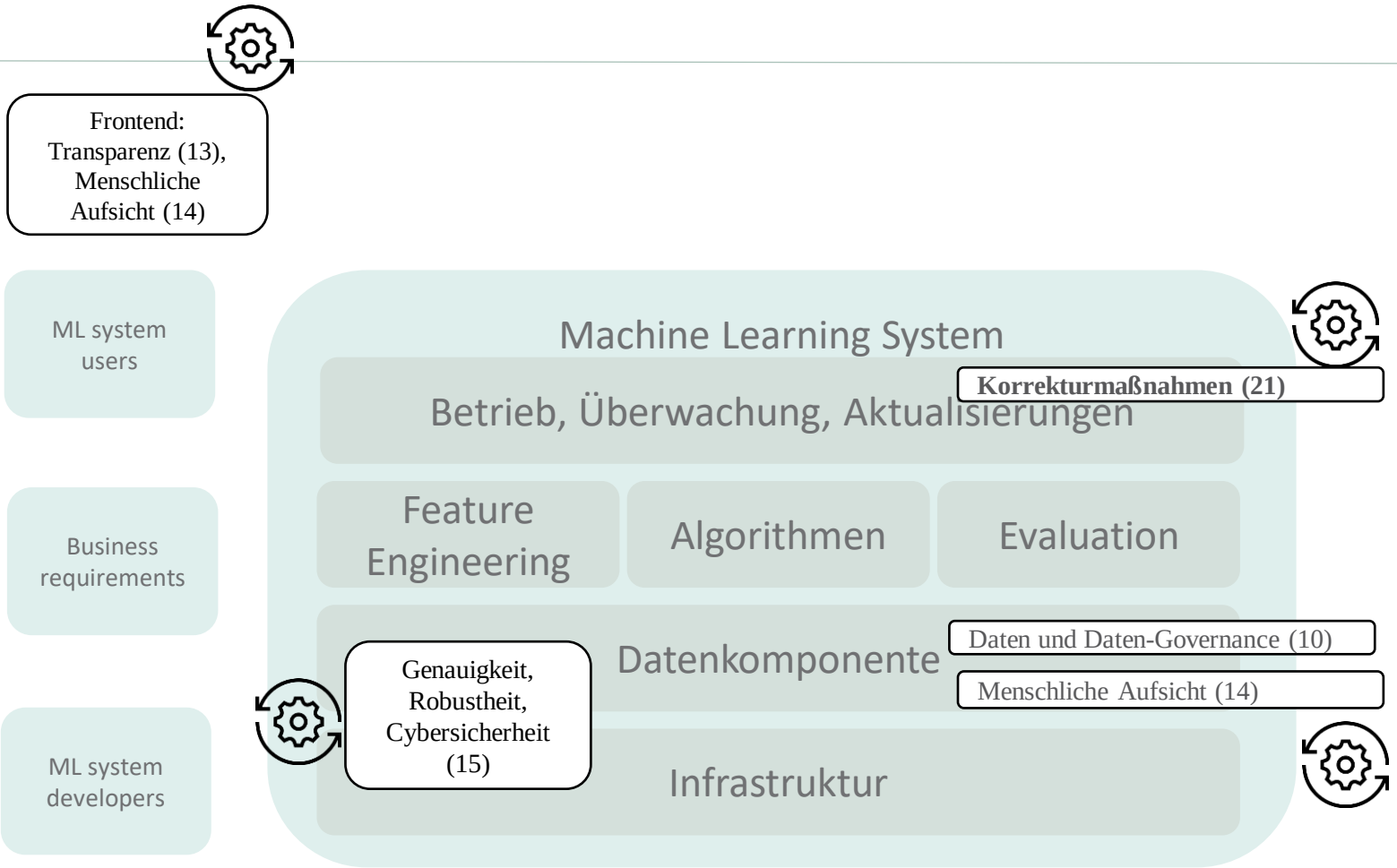
- “To enable data scientists to develop and deploy ML models, it’s crucial to have the right tools and infrastructure set up.”

Erweiterte Architektur

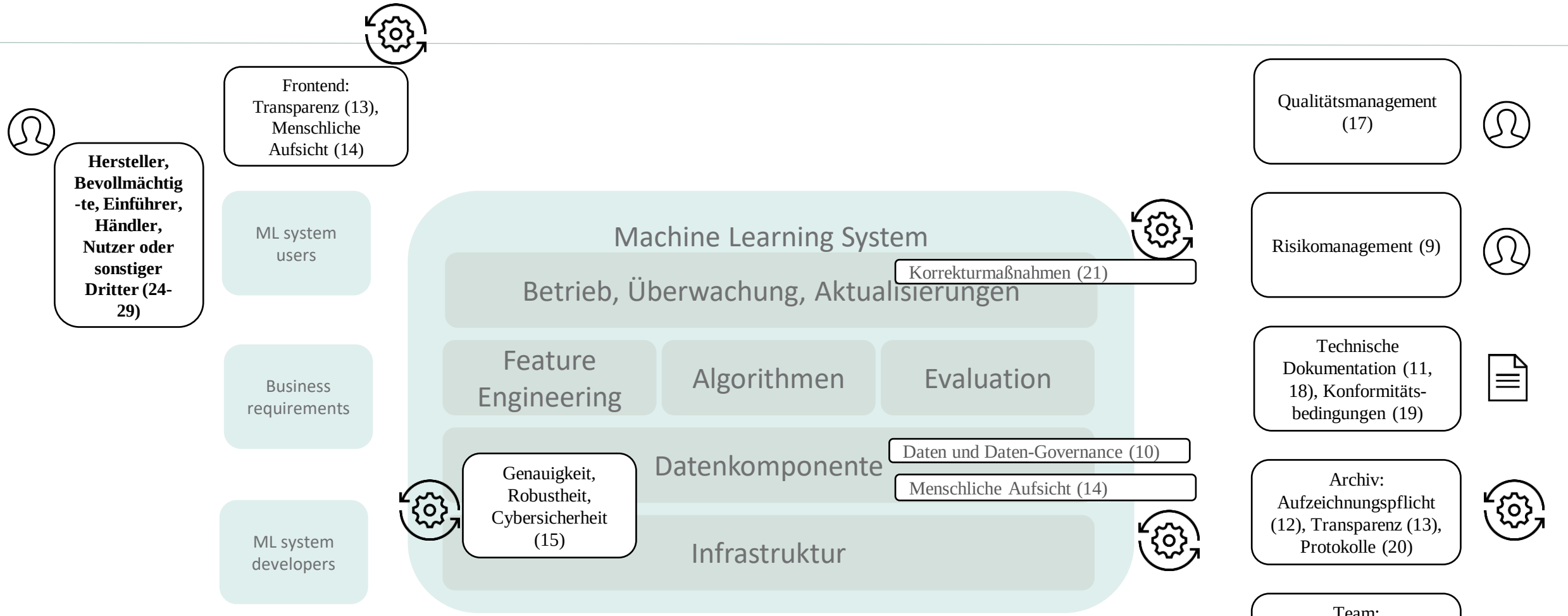
Anforderungen & Architektur



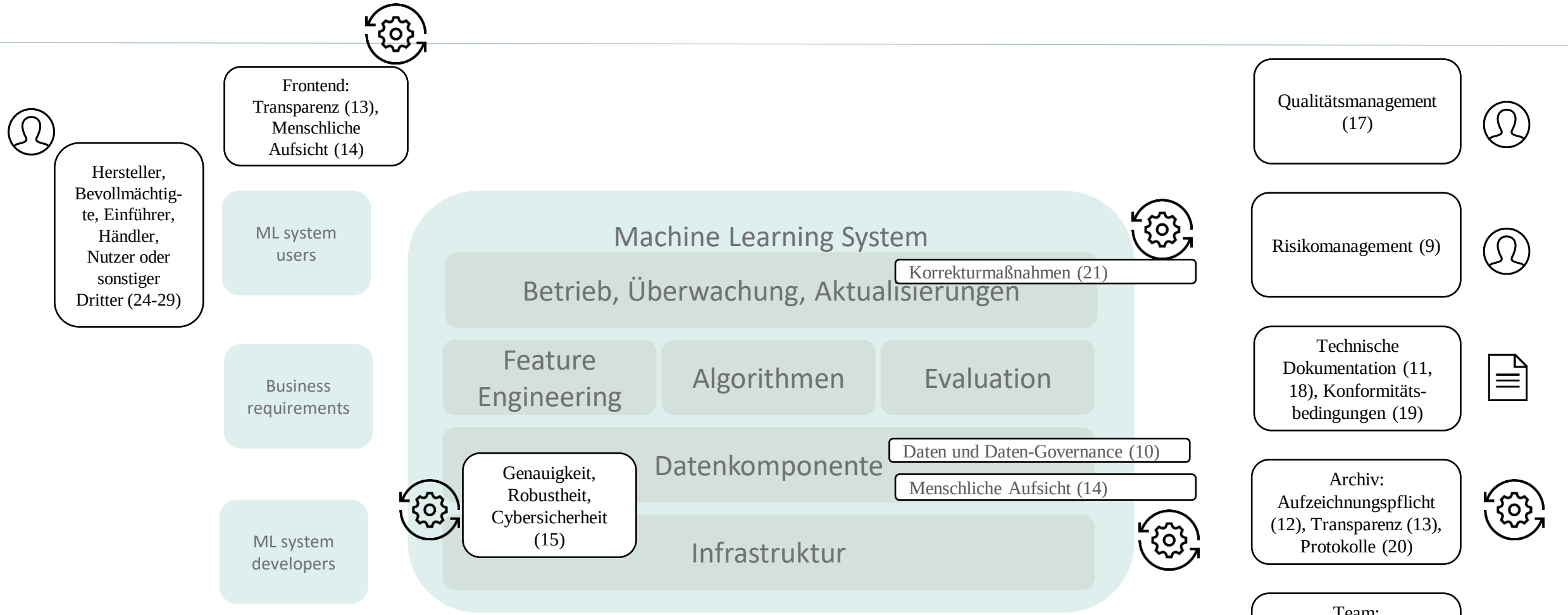
Fachliche Pflichten



Akteurspezifische Pflichten



Gesamtblick auf die erweiterte Architektur



Schluss

Schluss

- Wir haben auf einer relativ hohen Flughöhe den AI Act kennengelernt.
- Viele Details haben wir ausgelassen: Wir konnten aber erkennen, dass die Anforderungen aus dem AI Act vielfältig sind.
- Wir haben die Anforderungen aus dem AI Act zu einer möglichen Architektur zugeordnet. Dabei haben wir vereinfacht die moderne, datengetriebene KI als Maß für alle KI-Systeme verstanden.
- Es ist zu erkennen, dass bestehende Architekturen nicht die Anforderungen aus dem AI Act vollständig abbilden (zu erwarten).

Ausblick: Rückmeldung des Parlaments

- Angleichung der Definition von KI zu der OECD
- Die Liste der verbotenen KI-Systeme wurde erweitert
- Die Definition für Hochrisiko-Systeme wurde erweitert
- Es soll ein „Schichten-Ansatz“ etabliert werden für Anbieter von general-purpose-KI
- Ein KI-Büro („AI office“) soll eingerichtet werden
- Open-Source-Komponente sollen weitestgehend befreit werden vom AI Act

https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence

Ausblick: Zeitlinie

The AI Act has not yet been adopted. The European Commission proposed the regulation in April 2021, and it will need to be reviewed and approved by the European Parliament and the Council of the EU before it can become law.

On 6 December 2022, the Council of the EU published its general approach regarding the AI Act proposal. During the last votes in the Parliament, heated discussions about new, disruptive AI applications on the market – namely: ChatGPT – caused delays in the process. On 27 April 2023, the Parliament agreed on a draft of its position. The leading committees voted on the draft **on 11 May 2023**. The final vote in plenary took take place in June 2023. The final negotiations, the so-called trilogue, begins.

It is therefore possible that the AI Act will still enter into force in 2023. The majority of the provisions will then apply another 24 months later, during which time companies and organizations will have to ensure that their AI systems comply with the requirements and obligations set out in the regulation.

https://www.taylorwessing.com/en/insights-and-events/insights/artificial-intelligence-act#:~:text=It%20is%20therefore%20possible%20that,set%20out%20in%20the%20regulation.