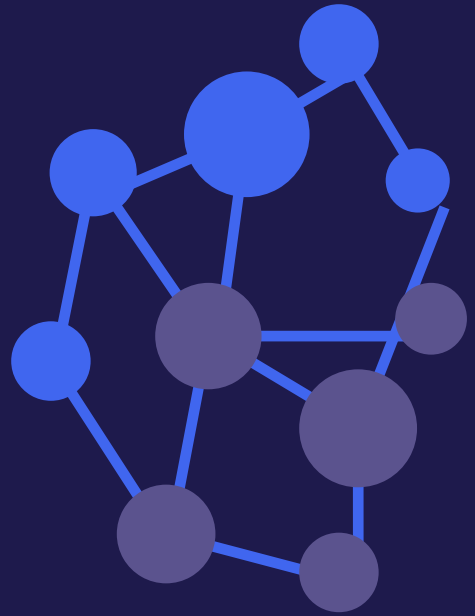
Foundation Models in der Radiologie:

Hype oder Paradigmenwechsel?



Lisa C. Adams

Institut für Diagnostische und Interventionelle Radiologie · Klinikum rechts der Isar · TUM

April 2026

SEIT DEM JAHR 2000

Warum jetzt? Die radiologische Realität

+2–3 %

mehr Radiolog:innen / Jahr

+5–7 %

mehr Untersuchungen / Jahr

× 5–6

mehr Bilddaten pro Untersuchung

FDA-ZUGELASSENE KI/ML-MEDIZINPRODUKTE · STAND ENDE 2025

1.104

Radiologische KI-Anwendungen

Platz 2: Kardiologie · ~130

Radiologie führt weiterhin mit Faktor ~8.

Quelle: FDA-Tracker (theimagingwire.com, 2026)

> 80 %

der zugelassenen KI-Software adressiert medizinische Bilder

DEFINITION

Was ist ein Foundation Model?

> *Ein Modell, das mittels Selbstüberwachung auf großen, breiten Datenmengen trainiert und an viele nachgelagerte Aufgaben angepasst werden kann.*
>
> nach Bommasani et al., Stanford CRFM, 2021 (paraphrasiert)

GRÖßE

Sehr viele Parameter, sehr viele Trainingsdaten.

BREITE

Vortraining ist nicht auf eine Aufgabe zugeschnitten.

ANPASSBARKEIT

Eine Basis – viele Anwendungen.

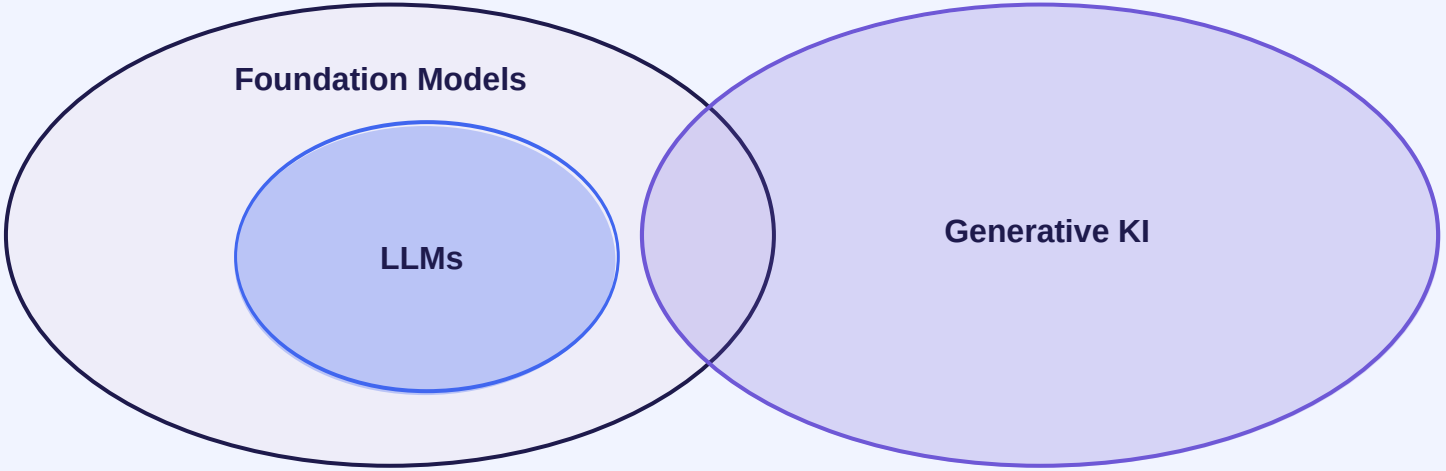
ABGRENZUNG

Foundation Model vs. klassische KI

Traditionelle KI	Foundation Modell
Aufgabenspezifisch	Aufgaben-agnostisch
Von Grund auf mit Labels trainiert	Auf Rohdaten oder schwach gelabelten Daten vortrainiert
Ein Modell pro Aufgabe	Ein Backbone (Grundgerüst), viele Heads (Aufsätze)
Typischerweise ~10.000 Labels	Millionen bis Milliarden von Elementen
Spitzenleistung in der eng gefassten Aufgabe	Breitere Kompetenz, oft geringere Spitze pro Einzelaufgabe

Altes Paradigma: 10 aufgabenspezifische Modelle. **Neues Paradigma:** 1 Basismodell, auf multiple Anwendungen anpassbar

Foundation Model, LLM, Generative KI



Foundation Model *beschreibt,* ***wie*** *ein Modell trainiert wurde.* ***Generative KI*** *beschreibt,* ***was*** *es tut.*

FM, kein LLM, keine GenKI:

RAD-DINO, MedSigLIP

LLMs (FM + GenKI):

GPT, Llama, Gemma, Claude

GenKI, kein FM:

ältere GANs

Eine Analogie

KLASSISCHE KI

Hat das Medizinstudium übersprungen

Keine medizinische Ausbildung. Bekommt einen Stapel Thorax-Röntgen mit dem Hinweis „das ist Pneumothorax, das nicht". Nach einigen tausend Beispielen erkennt sie Pneumothorax sehr gut. Kann kein CT lesen. Erkennt keinen anderen Befund im selben Röntgenbild. Keine Anatomie, keine Physiologie, kein Kontext.

FOUNDATION MODEL

Frisch approbiert

Sechs Jahre Medizinstudium, noch keine praktische Erfahrung. Kennt Sprache, Anatomie, Muster. Am ersten Klinik-Tag noch nicht produktiv. Mit kurzer Rotation lernt sie neue Aufgaben schnell. Weniger anfällig, wenn sich die Eingabedaten ändern.

Warum Foundation Models in der Radiologie?

1

Vorhandene Supervision (Beschriftungen)

PACS und RIS enthalten bereits Millionen Bild-Befund-Paare. FMs nutzen diese vorhandenen Beschriftungen – keine neuen Pixel-Annotationen nötig.

2

Modalitäten-Breite

Dieselbe Architektur, mit anderen Daten trainiert, ergibt Modelle für Röntgen, CT, MRT, PET. Kein Neustart mehr pro Modalität.

3

Nachgelagerte Flexibilität

Ein Basismodell für Klassifikation, Segmentierung, Suche, Befundentwurf. Ein neues Infrastruktur-Muster für klinische KI.

FMs lernen nicht ohne menschliche Beschriftungen. Sie nutzen die Beschriftungen, die wir im Klinikalltag ohnehin produzieren.

RAD-DINO

Microsoft, Pérez-García et al., Nat Mach Intell 2025 · Reiner Bild-Codierer für Thorax-Röntgen

- DINOv2-Selbstüberwachung, ausschließlich Bilder
- Fünf öffentliche Datensätze: MIMIC-CXR, CheXpert, NIH-CXR, PadChest, BRAX
- ViT-B/14 als Basismodell, nur Vektor-Ausgaben (kein Sprachkopf)
- Wird als Bild-Codierer in MAIRA-2 verwendet
- Lizenz: MSRLA, zugangsbeschränkt, nur für Forschung

880 Tsd.

Trainingsbilder

87 Mio.

Parameter

<8 GB

Grafikspeicher für Anwendung

Microsofts eigene Skalierungs-Studie 2025 (mit Mayo Clinic) zeigt, dass textüberwachte Ansätze ab einigen Millionen Bildern besser skalieren. Die These „Bild allein reicht" gilt nicht immer.

MAIRA-2: Anspruch vs. Realität

Microsoft, Bannur et al., 2024 · Vorab-Veröffentlichung, nicht begutachtet

DAS PAPIER

- RAD-DINO-Codierer + Vicuna 7B
- ~7 Mrd. Parameter insgesamt
- Erzeugt Thorax-Röntgen-Befunde mit eingezeichneten Markierungen
- Eingaben: frontal + optional lateral, Voraufnahme, Indikation
- Stand der Technik bei MIMIC-CXR-Befunden zum Erscheinungszeitpunkt
- Verbesserte Werte bei RadGraph-F1, RadFact, CheXbert-F1

DIE REALITÄT

Lim et al., Seoul National University Hospital, Dez 2025 (arXiv 2512.00271)
478 Notaufnahme-Patient:innen, 1.434 Reports, drei Thorax-Befunder, taggleiches CT als Goldstandard

17,4 % Halluzinationsrate (Radiolog:innen: 0,1 %)

24,5 % RADPEER 3b (gravierender Lesefehler)

0 % Sensitivität für seltene Befunde

Hervorragende Benchmark-Werte übersetzen sich nicht in klinische Reife bei unbekannten Populationen. Die Lücke vom Labor zum Patientenbett – in Zahlen.

CT-CLIP und CT-RATE

Hamamci et al., Nat Biomed Eng 2026 · 3D-Thorax-CT, vergleichendes Bild-Text-Lernen

- Erster öffentlicher 3D-Thorax-CT-Datensatz mit gepaarten Befunden
- Ein Standort, Istanbul. Befunde maschinell aus dem Türkischen ins Englische übersetzt; Stichprobe nachträglich von bilingualen Medizinstudent:innen geprüft
- 3D-ViT + CXR-BERT, vergleichendes Vortraining
- Schlägt das vollüberwachte CT-Net auf externen Datensätzen aus Duke und der UPMC
- FAU Erlangen Hochleistungsrechner + DFG-Förderung (deutscher Bezug)

25.692

Thorax-CT-Volumen

21.304

Patient:innen (1 Standort)

0,73 / 0,90+

AUROC: zero-shot intern / fine-tuned

Lizenz CC-BY-NC-SA 4.0: nicht-kommerziell und Share-Alike. Schließt jede umsatzgenerierende klinische Nutzung aus. Eine deutschsprachige Version existiert nicht (eine japanische, CT-RATE-JPN, gibt es).

Merlin

Stanford, Blankemeier et al., Nature 2026 · 3D-Abdomen-CT, Bild-Sprach-Modell

- Trainiert auf CT + ICD-Codes + Befundtexten
- I3D-ResNet-152 + Clinical Longformer, ~300 Mio. Parameter
- Auf einer einzigen Grafikkarte A6000 (48 GB) trainiert
- Evaluiert auf 752 Aufgaben in 6 Kategorien
- Schlägt CT-CLIP auf Thorax-CT – obwohl nur auf Abdomen-CT trainiert

15.331

Trainings-Scans

44.098

Scans, externe Validierung

MIT

Lizenz

Peer-reviewed in Nature, offene MIT-Lizenz, externe Validierung, läuft auf einer einzelnen GPU. Offene Punkte: Interessenkonflikte (Gründer-Rollen, Google-Co-Autor:innen), bislang keine unabhängige Befunder-Studie.

MedGemma 1.5

Google HAI-DEF, Januar 2026 · Frei verfügbares 4-Mrd.-Parameter-Modell für medizinische Daten

- Basiert auf Gemma 3 + medizinischem Bild-Codierer MedSigLIP
- Neu in 1.5: 3D-CT/MRT, Whole-Slide-Histopathologie, Verlaufsvergleiche, anatomische Lokalisation, FHIR-Patientenakten
- Stark komprimiert läuft das Modell auf einer 6-GB-Consumer-Grafikkarte
- Lizenz erlaubt kommerzielle Nutzung – aber nicht als Medizinprodukt

81 %

„radiologische Akzeptanz" (MedGemma 1, Hersteller-Eval)

Drei Einschränkungen:

- Nicht begutachtet
- n = 1 Befunder
- Unverblindet

Unabhängige Evidenz (Lim 2025, Notaufnahme-CXR): die vier offenen VLMs (Lingshu, MAIRA-2, MedGemma, MedVersa) reichen von 41,1–71,4 % klinischer Akzeptanz und 5,4–17,4 % Halluzinationsrate. Auch das beste der vier liegt deutlich unter klinischer Einsatzreife (Radiolog:innen-Halluzinationen: 0,1 %).

RadFM

Wu, C., Zhang, X., Zhang, Y. et al., Nat Comm 2025 · Universalmodell über Röntgen, CT, MRT, PET, Ultraschall

- Alle radiologischen Modalitäten in einem Modell (Ultraschall im Trainingskorpus nur marginal vertreten)
- 16 Mio. Scans (13 Mio. 2D + 615 Tsd. 3D), 14 Mrd. Parameter, MedLLaMA-13B als Basismodell
- Erstes multimodales Sprachmodell, das mehrere 3D-Bilder gleichzeitig verarbeitet

Zwei praktische Probleme:

Hardware: 14 Mrd. Parameter mit 3D-Eingaben erfordern Hochleistungs-Grafikkarten (A100 oder H100). Nicht realistisch für Klinik-Hardware.

Leistung: Die Merlin-Studie zeigt einen direkten Vorsprung bei der Befundgenerierung. Auch spezialisierte neuere Modelle schlagen RadFM bei den meisten Thorax-Röntgen-Aufgaben.

FAZIT

Zum Mitnehmen

1. **Foundation Models sind vortrainiert, breit trainiert, anpassbar.**
   Keine Magie – sondern ein anderer Punkt im Kompromissraum.

2. **Aktuell stärkste Anwendungen: Befundentwürfe, Suche, Verknüpfung von Bildern und Patientenakten.**
   Schmale, aufgabenspezifische Modelle gewinnen weiterhin bei der Spitzen-Genauigkeit pro Einzelaufgabe.

3. **Frei verfügbare medizinische Foundation Models laufen heute auf Abteilungs-Hardware.**
   Der Engpass sind Einführung und Regulierung – nicht die Modellfähigkeit.

Regulatorischer Hinweis: Keines dieser Modelle ist CE-zertifiziert. EU AI Act: Hochrisiko-Pflichten für Annex-III-Standalone-Systeme ab 2. August 2026; KI als Bestandteil eines Medizinprodukts unter MDR/IVDR erst ab 2. August 2027 (Digital Omnibus könnte verschieben). Die einsetzende Einrichtung trägt dann die Anbieterpflichten.

Vielen Dank.

Fragen?

HYPE ODER PARADIGMENWECHSEL?

Beides. Hype bei einzelnen Modellen, Paradigmenwechsel im Trainingsmuster.

lisa.adams@tum.de · Institut für Diagnostische und Interventionelle Radiologie · TUM