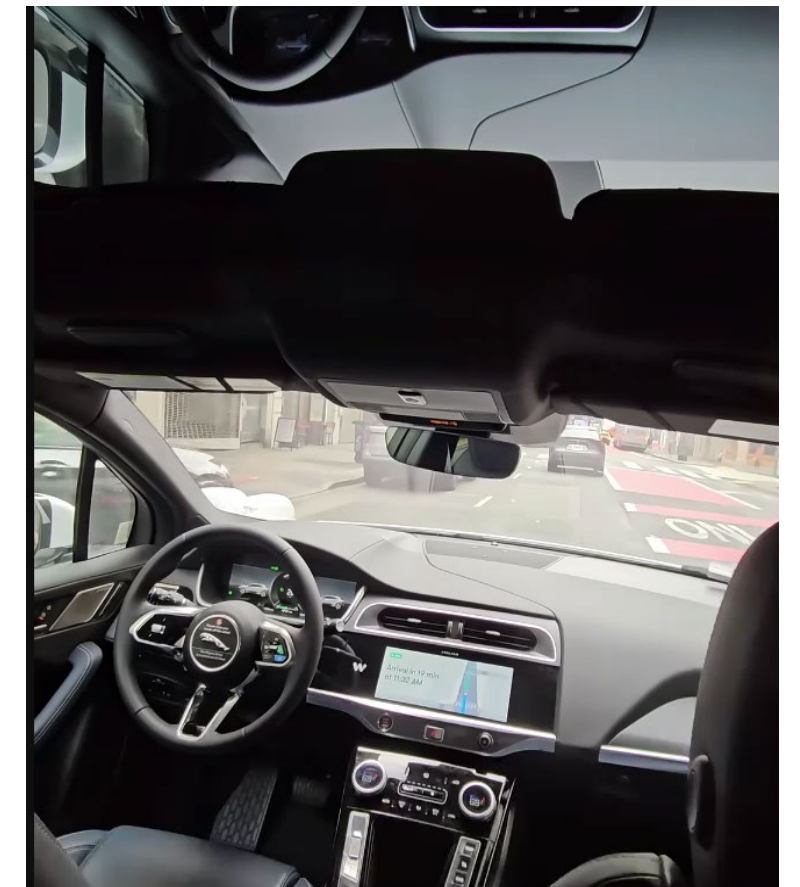
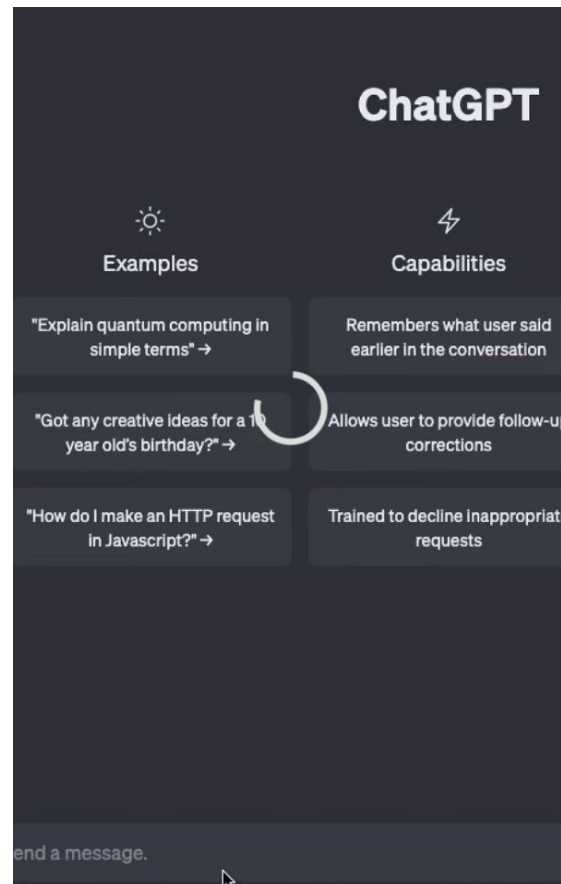
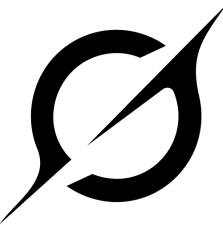


Challenge nell'AI per Immagini Mediche

Esperienze e Risultati

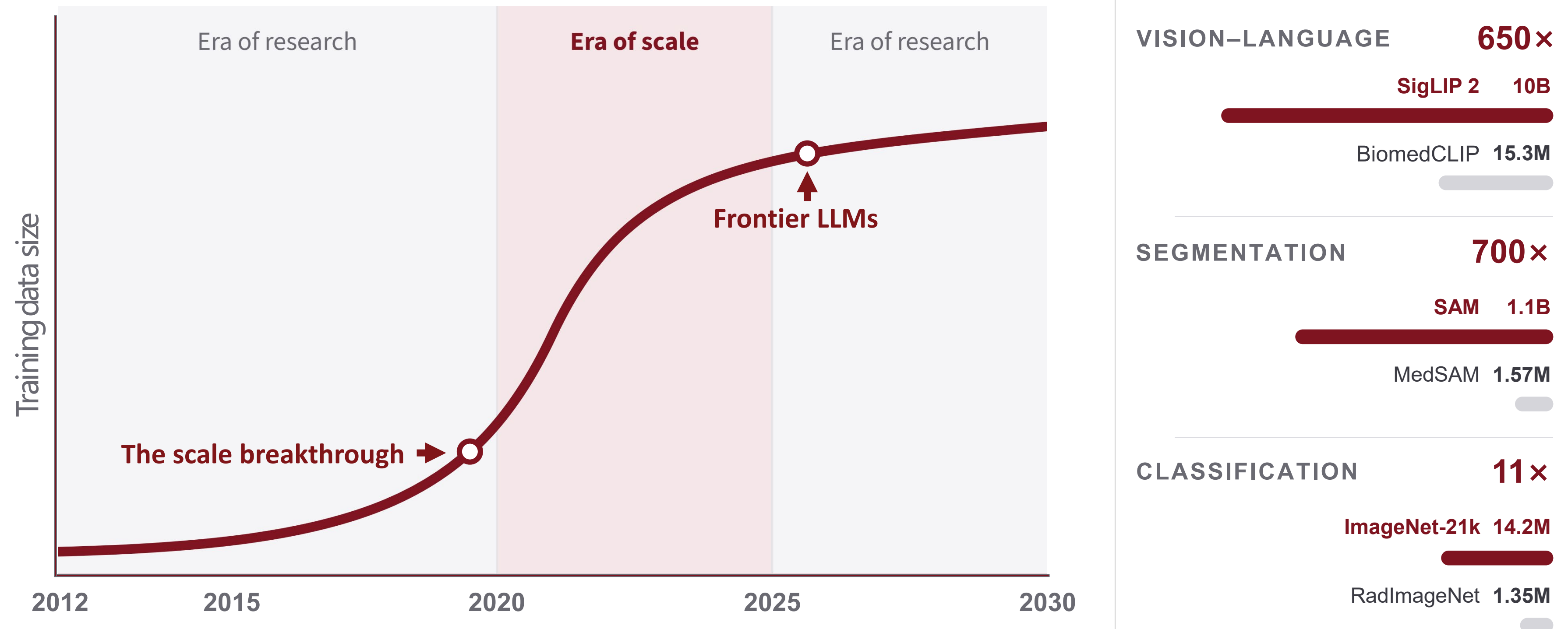
Federico Bolelli

AlmageLab — University of Modena and Reggio Emilia · federico.bolelli@unimore.it



Why Medical 3D Vision Has No ImageNet Moment (Yet)?

General models crossed into the **era of scale**.

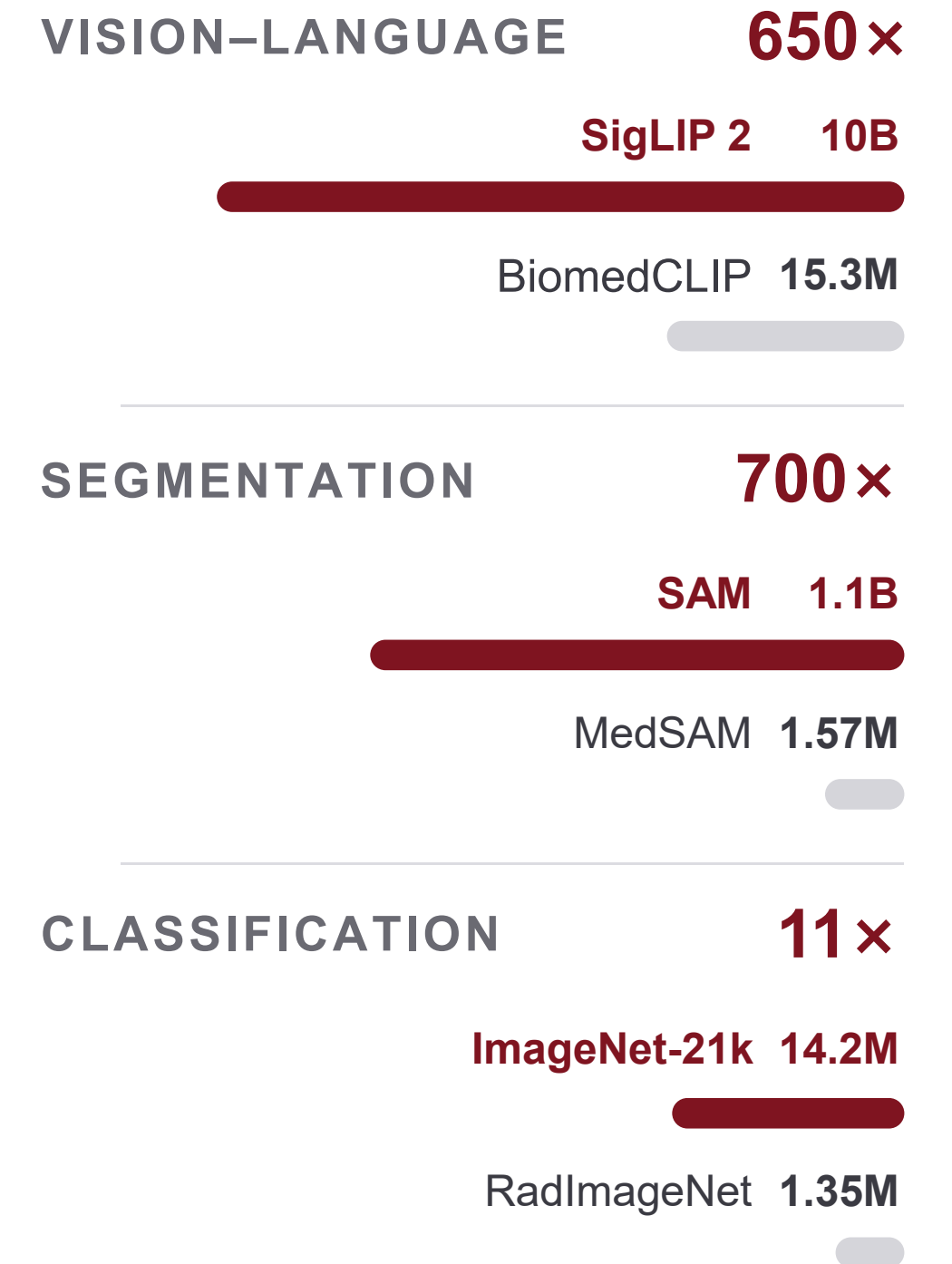
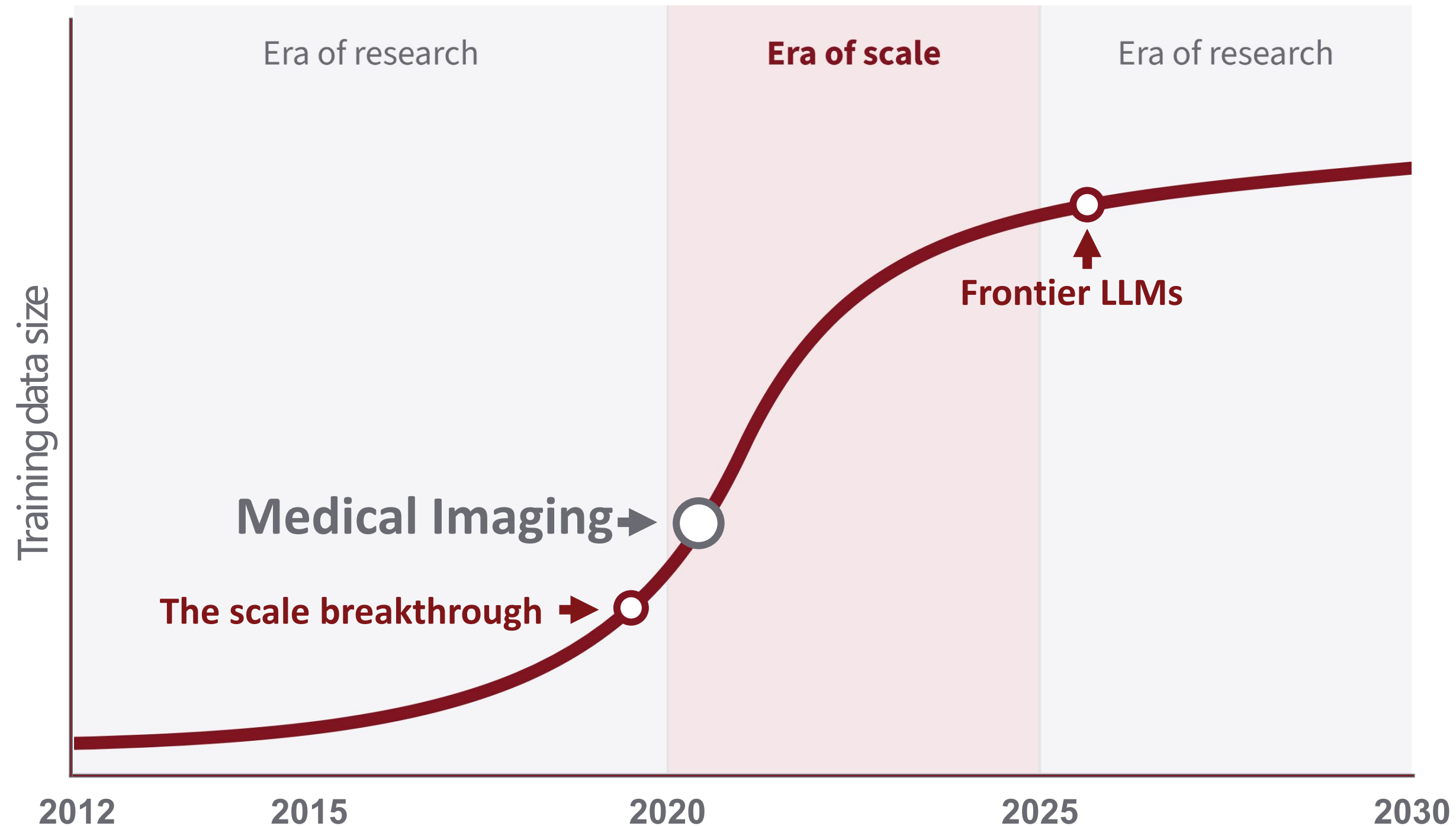


Counts as reported by the authors: WebLI (SigLIP 2), PMC-15M (BiomedCLIP), SA-1B (SAM), MedSAM, ImageNet-21k, RadImageNet.

[1] Curtis P. Langlotz - Toward a Radiology Foundation Model

Why Medical 3D Vision Has No ImageNet Moment (Yet)?

General models crossed into the **era of scale**.



Counts as reported by the authors: WebLI (SigLIP 2), PMC-15M (BiomedCLIP), SA-1B (SAM), MedSAM, ImageNet-21k, RadImageNet.

[1] Curtis P. Langlotz - Toward a Radiology Foundation Model.

Three layers now make methods more **comparable**.

01

Shared data and task

Same cohort, image inputs, reference labels and train/test split.

BraTS · CAMELYON · Decathlon

02

Controlled evaluation

Hidden test labels, central scoring and a fixed submission protocol.

MICCAI challenge ecosystem

03

Standardized evidence

Task-aware metric selection, ranking rules and transparent reporting.

BIAS · Metrics Reloaded

[1] L. Maier-Hein, et al. “Why rankings of biomedical image analysis competitions should be interpreted with care.” *Nature Communications* 2018.

[2] L. Maier-Hein, et al. “Metrics reloaded: recommendations for image analysis validation.” *Nature Methods* 2024.

A **benchmark dataset** defines what the task actually is.

What is a benchmark dataset?

A shared reference dataset used to measure and fairly compare methods on a specific task.



What is fixed?

A fixed test set, predefined task and data split, standardized acquisition protocol.



Why it matters?

Enables reproducible, consistent, and meaningful comparison across methods.

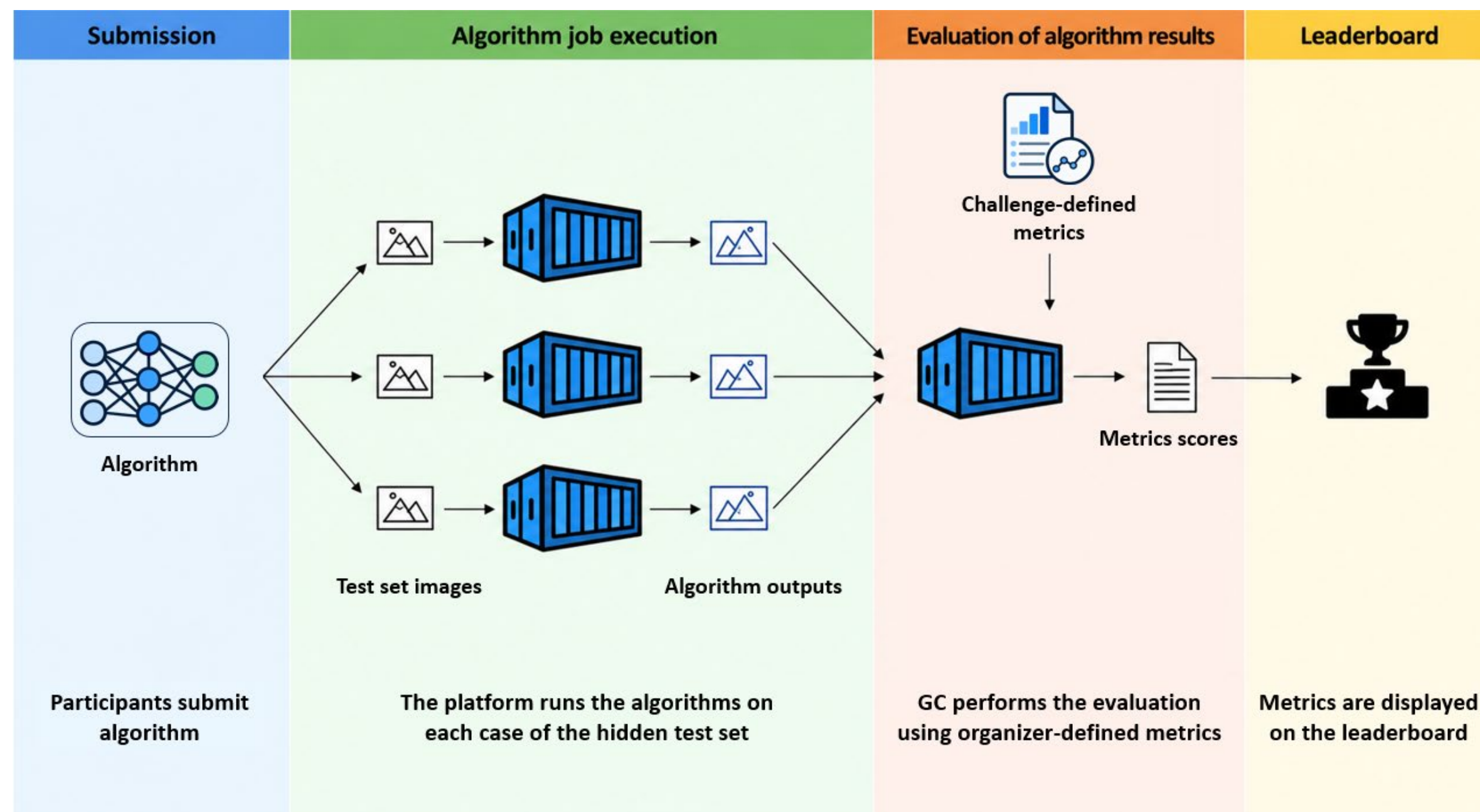


[1] B. H. Menze, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS)." IEEE TMI 2015.

[2] A. L. Simpson, M. Antonelli, et al. "A large annotated medical image dataset for the development and evaluation of segmentation algorithms." 2019.

[3] P. Bandi, et al. "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge." IEEE TMI (2018).

Generalization can only be measured on a **hidden test set**. Grand Challenge makes this evaluation **controlled**, **reproducible**, and **comparable**.



[1] L. Maier-Hein, et al. "Why rankings of biomedical image analysis competitions should be interpreted with care." *Nature Communications* 2018.

[2] grand-challenge.org — platform for centralized, reproducible evaluation of biomedical image analysis algorithms.

Reporting and **metrics** determine what the numbers actually mean.

BIAS statement

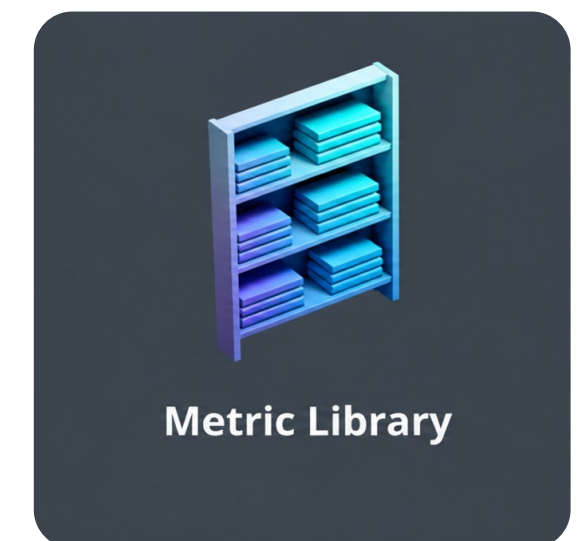
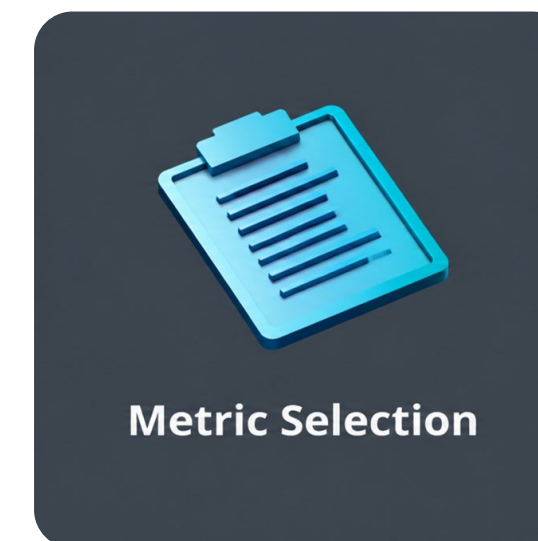
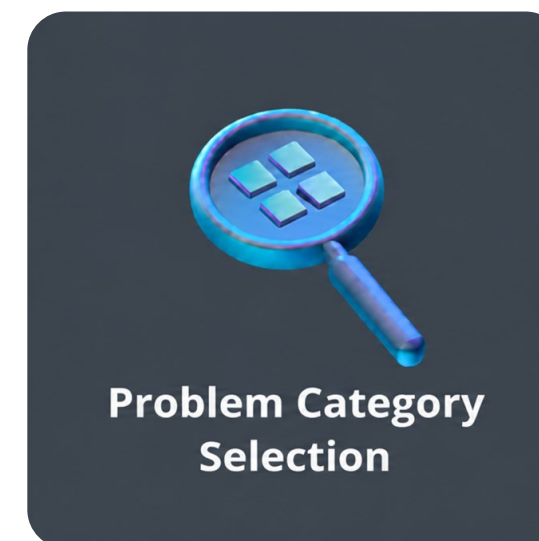
Biomedical Image Analysis ChallengeS is a structured reporting checklist covering the dataset/cohort, annotations, evaluation metrics, and ranking scheme.



Task-aware metrics

Choose metrics based on the task and its failure modes. Overlapping metrics alone (e.g., Dice score) can hide boundary, small-structure, and clinically relevant errors.

Metrics Reloaded Framework

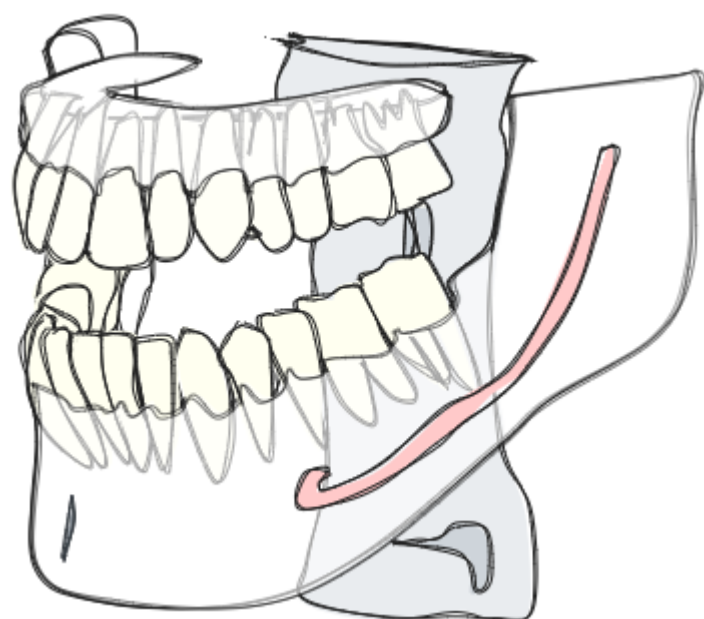


[1] L. Maier-Hein, A. Reinke, M. Kozubek, et al. "BIAS: Transparent reporting of biomedical image analysis challenges." *Medical Image Analysis* 2020.

[2] A. Reinke, L. Maier-Hein, et al. "Metrics reloaded: recommendations for image analysis validation." *Nature Methods* 2024.



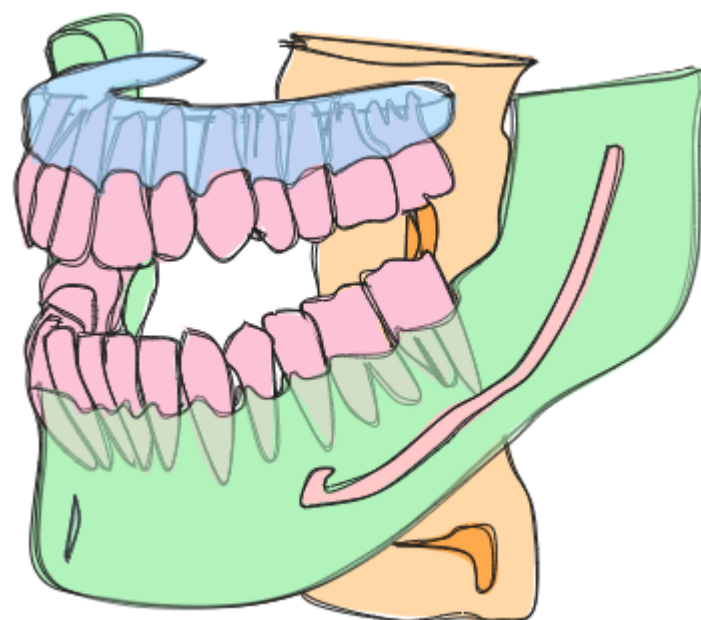
1 class



2023



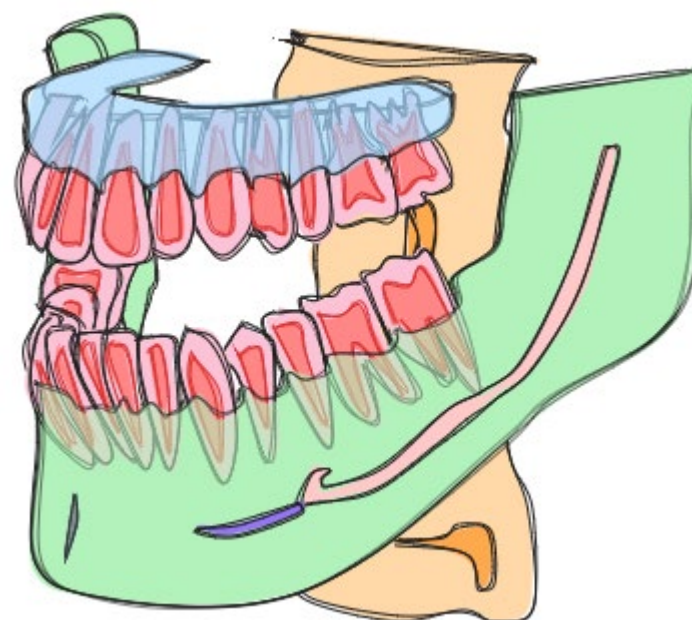
42 classes



2024



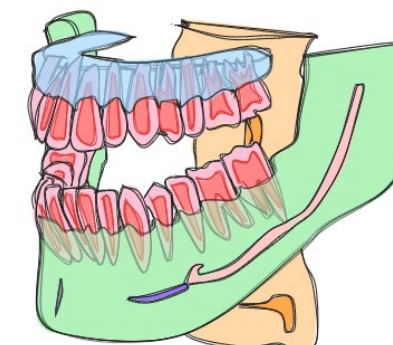
77 classes



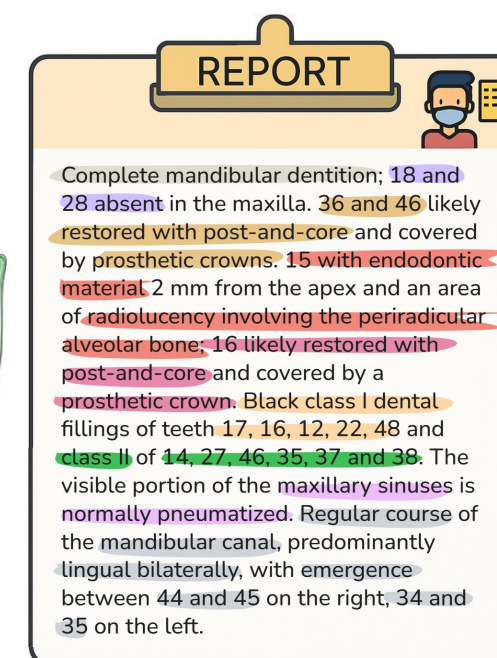
2025



77 classes + Reports



2026



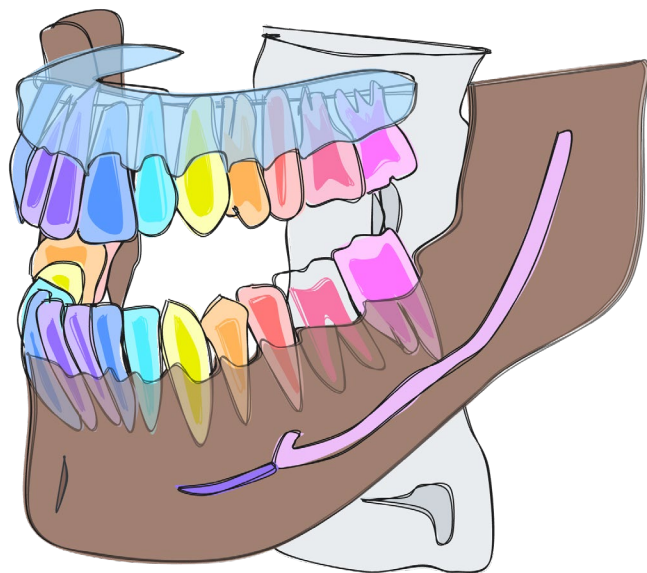
[1] Bolelli F., et al. "Segmenting the Inferior Alveolar Canal in CBCT Volumes: The ToothFairy Challenge" IEEE Transactions on Medical Imaging 2024.

[2] Bolelli F., et al. "Multi-structure segmentation in CBCT volumes: The ToothFairy2 challenge." Medical Image Analysis 2026.

[3] Lumetti L., et al. "ToothFairy3: Scaling CBCT Maxillofacial Segmentation to 77 Classes with U-Mamba2." MICCAI 2026.

[4] Lumetti L., et al. "Ontology-Grounded Structured Prediction for Dental CBCT Reporting." MICCAI 2026.

A patient is not a volume



SEGMENTED VOLUME

PATH A — RAW GEOMETRY

tooth 38

1.8 cm³

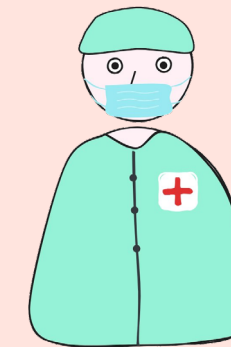
Accurate, but not yet a clinical statement



PATH B — REPORT GENERATION

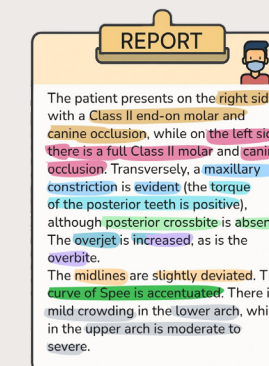
Findings stated in clinical language

Relations, severity, laterality; what the volume implies



Clinician's intervention

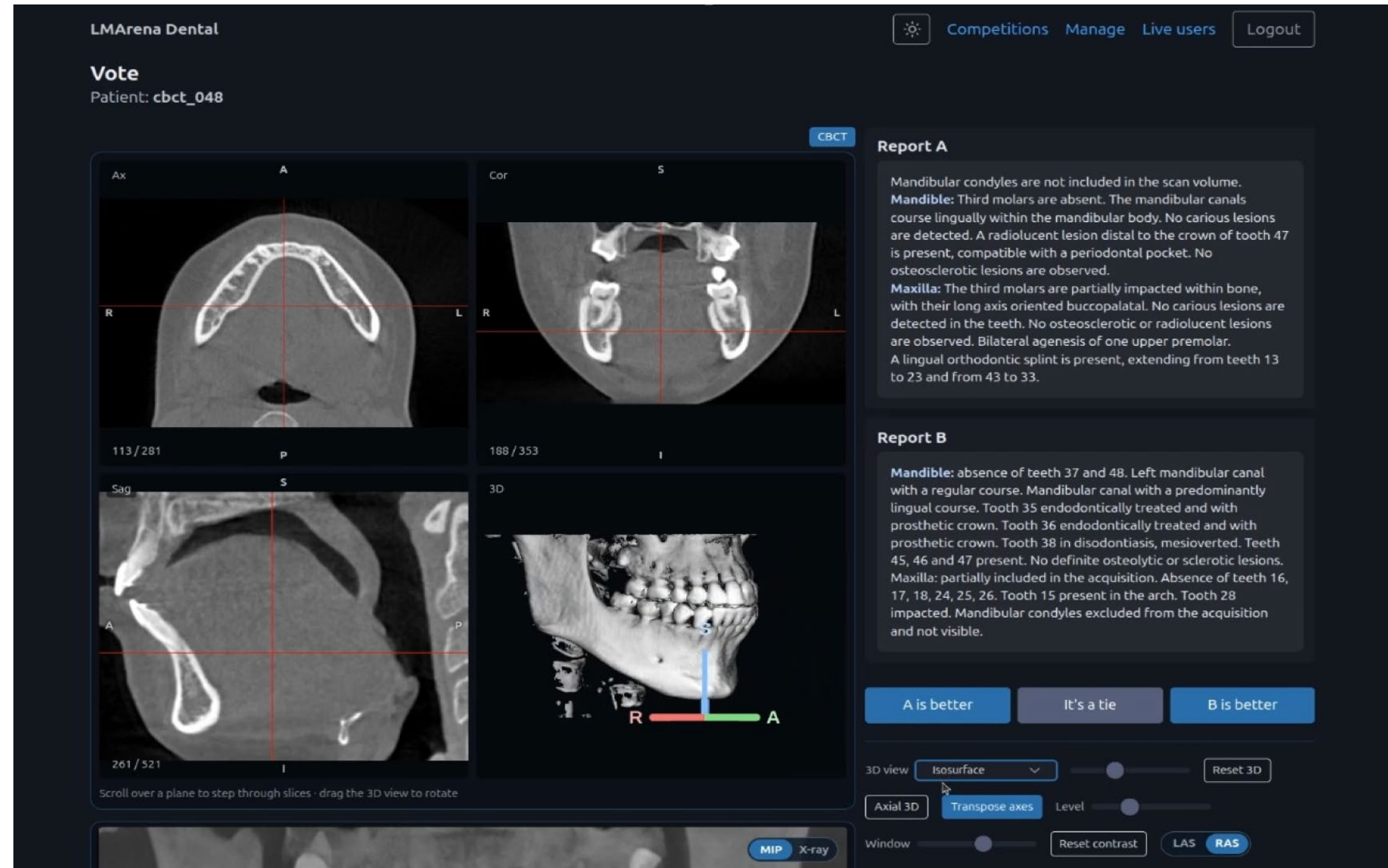
A human has to translate the numbers into findings



Directly actionable

The output now carries the clinical meaning

Segmentation describes anatomy. Reports communicate **clinical meaning**.



Video not available in the PDF!



Video not available in the PDF!

An all in one platform: from ingestion to **annotation**.

- We moved the annotation protocol itself into the infrastructure.
- Data are organized around the patient. Data can be visualized jointly, and the annotation protocol is encoded in the platform.
- Clinicians work collaboratively, instead of producing independent datasets that we later try to reconcile.



- 01** **Benchmarks make progress measurable**

- 02** **Evaluation design matters as much as the model**

- 03** **Larger-scale datasets require better annotation infrastructure**

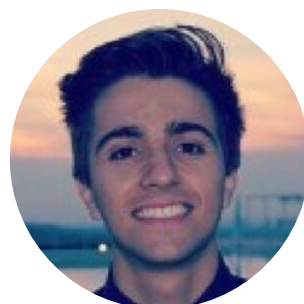
- 04** **Clinical AI must move toward multimodal, structured, meaningful outputs**

AIMAGELAB — UNIVERSITY OF MODENA AND REGGIO EMILIA

GROUP



F. Bolelli



K. Marchesini



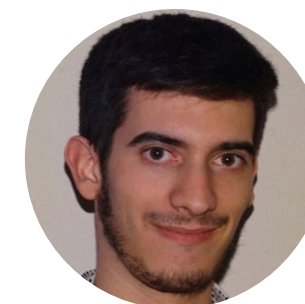
E. Candeloro



O. Carpentiero



N. Morelli



E. Vezzali



L. Borghi



G. Casari



G. Rosati



Y. Assefa



M. Lugli



L. Lumetti

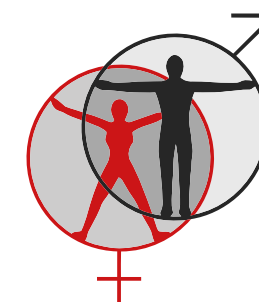


C. Grana

CLINICAL PARTNERS

D. Santi · G. Spaggiari · A. Anesi · M. Di Bartolomeo · A. Pellacani · P. Minafra · G. Pellacani · L. Bertoni · G. Bianchi · S. Puliatti · G. Besutti · R. Magistroni · F. Benuzzi · F. Lui · L. Lombardo · F. Cremonini

COLLABORATIONS



Challenge nell'AI per Immagini Mediche

Esperienze e Risultati

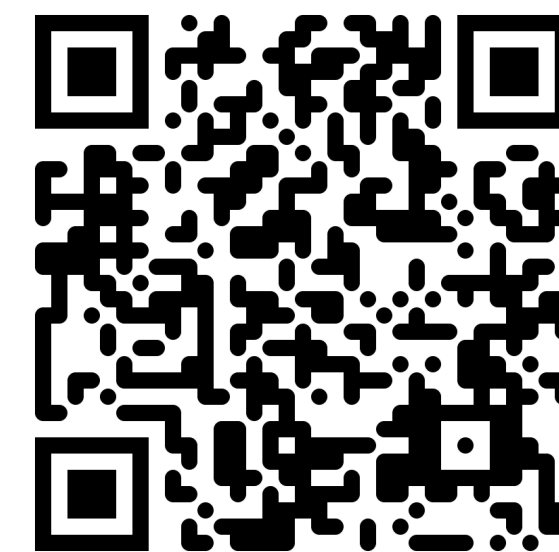
Thank You!

Federico Bolelli

AlmageLab — University of Modena and Reggio Emilia · federico.bolelli@unimore.it



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Papers, datasets, code