Foundation Models for Chest Radiography:

Knowledge Distillation and Impacts on Performance and Bias Propagation

Motivation

Fadi Zahar

26/06/2024

Plan

Motivation

1. Deep Learning in Medical Imaging
2. Foundation Models
3. Associated Challenges
4. Need for Knowledge Distillation

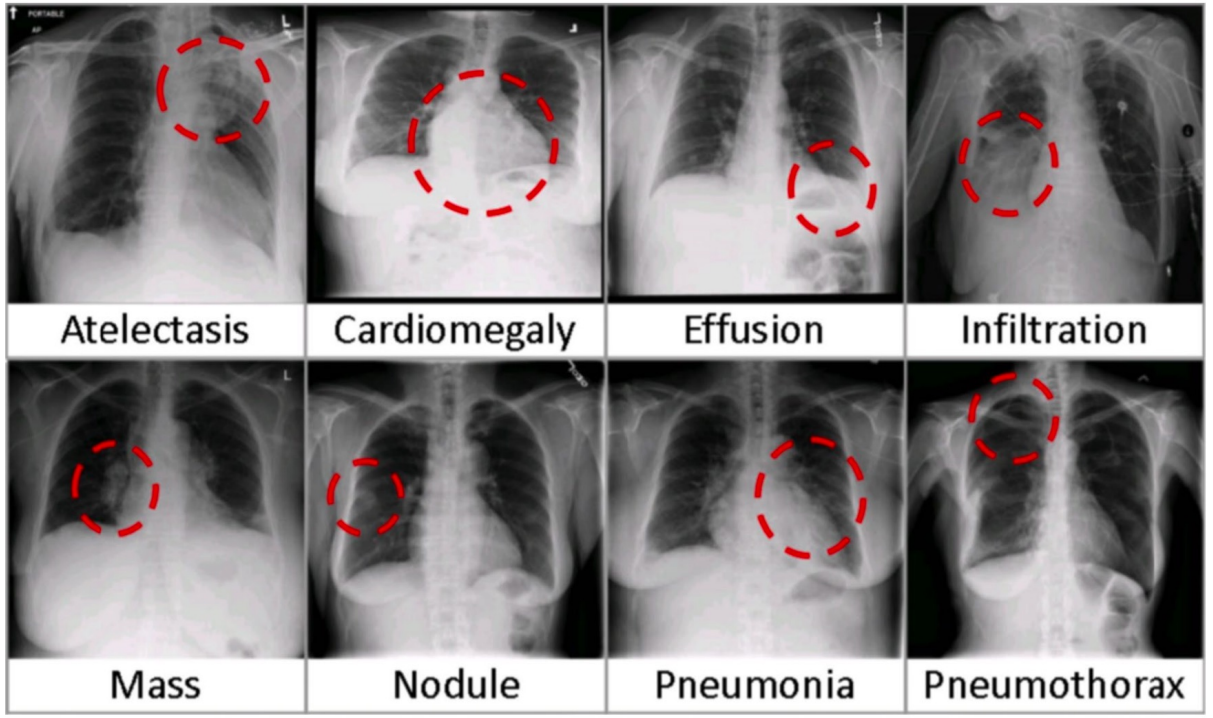
Motivation

1. Deep Learning in Medical Imaging

Deep Learning(DL)-based models have shown significant success in medical imaging applications:

- **Accelerate** medical image analysis (automation)
- Match or **exceed human expert** performance

e.g., thoracic disease detection using chest X-rays (CXRs)



Common thoracic diseases observed in Chest X-rays [1]

1. Deep Learning in Medical Imaging

Deep Learning(DL)-based models have shown significant success in medical imaging applications:

- **Accelerate** medical image analysis (automation)
- Match or **exceed human expert** performance

e.g., thoracic disease detection using chest X-rays (CXRs)

However, they need very large amounts of representative data for good generalisation

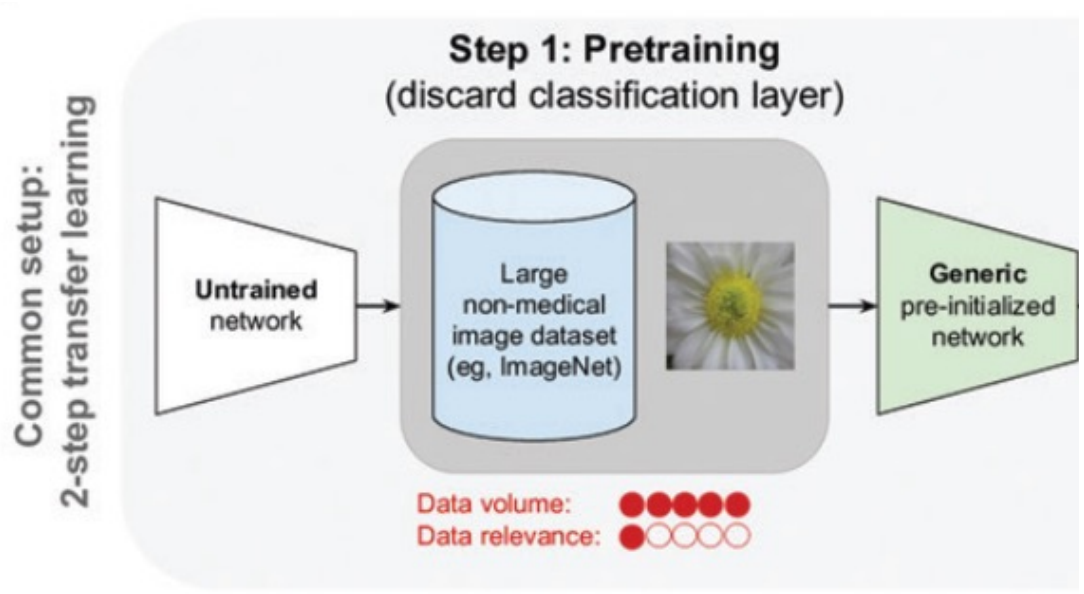
Medical images: Expensive labelling & data privacy concerns

Otherwise, they risk to **fail** due to data distribution shifts in different clinical settings:

- Population shifts *(demographics)*
- Acquisition shifts *(imaging modalities)*

2. Foundation Models

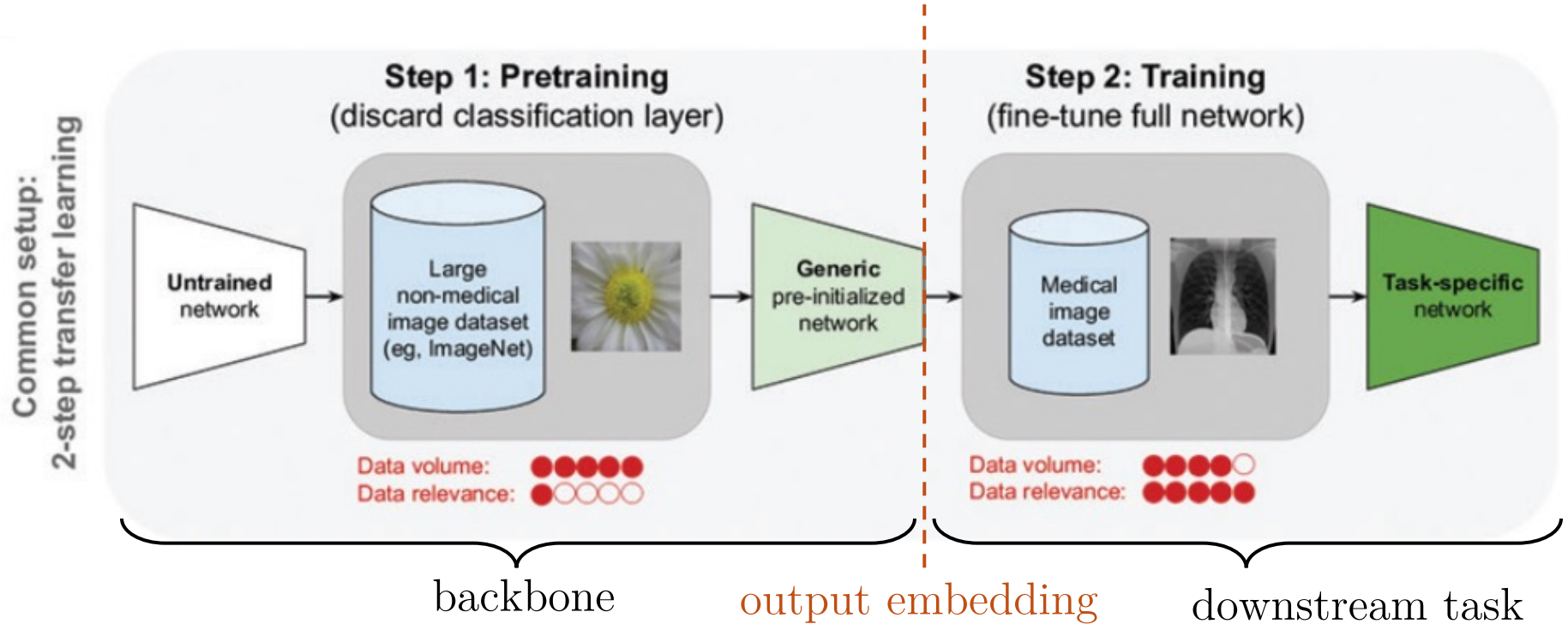
Solution: ***Foundation Models*** → pretrained on heterogeneous, diverse, large-scale datasets;
often via self-supervised or semi-supervised learning



"surrogate" task

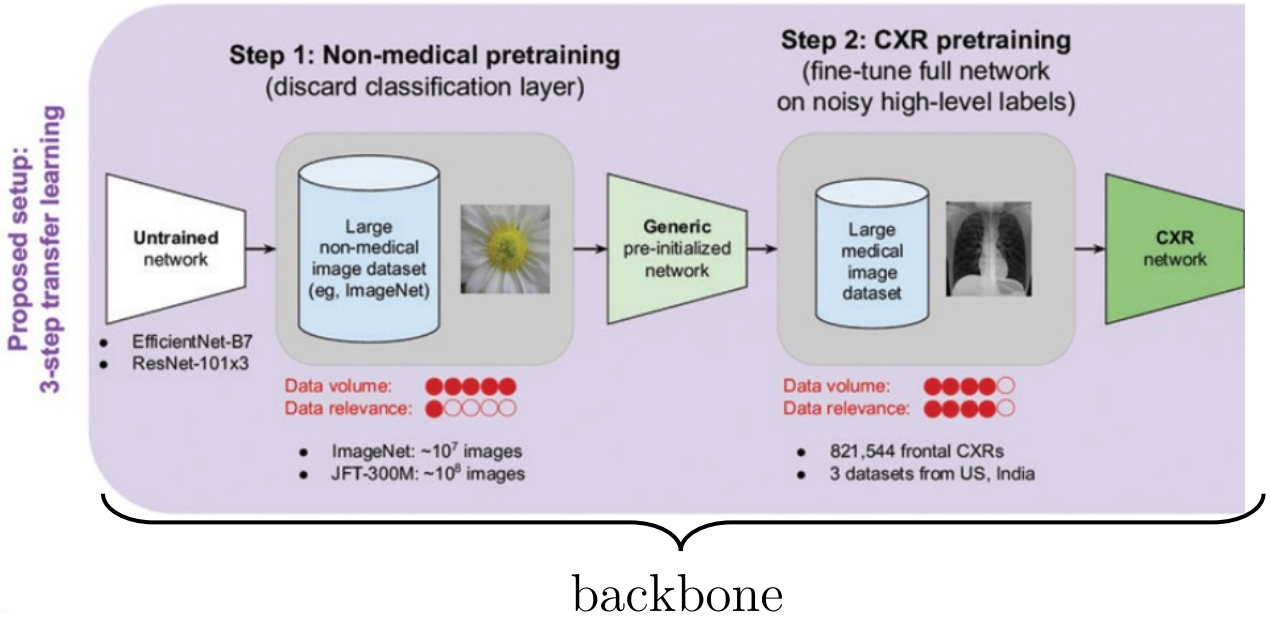
2. Foundation Models

Solution: ***Foundation Models*** → pretrained on heterogeneous, diverse, large-scale datasets;
often via self-supervised or semi-supervised learning
→ generated feature representations (embeddings) then used in
downstream task adaptation with much less training data



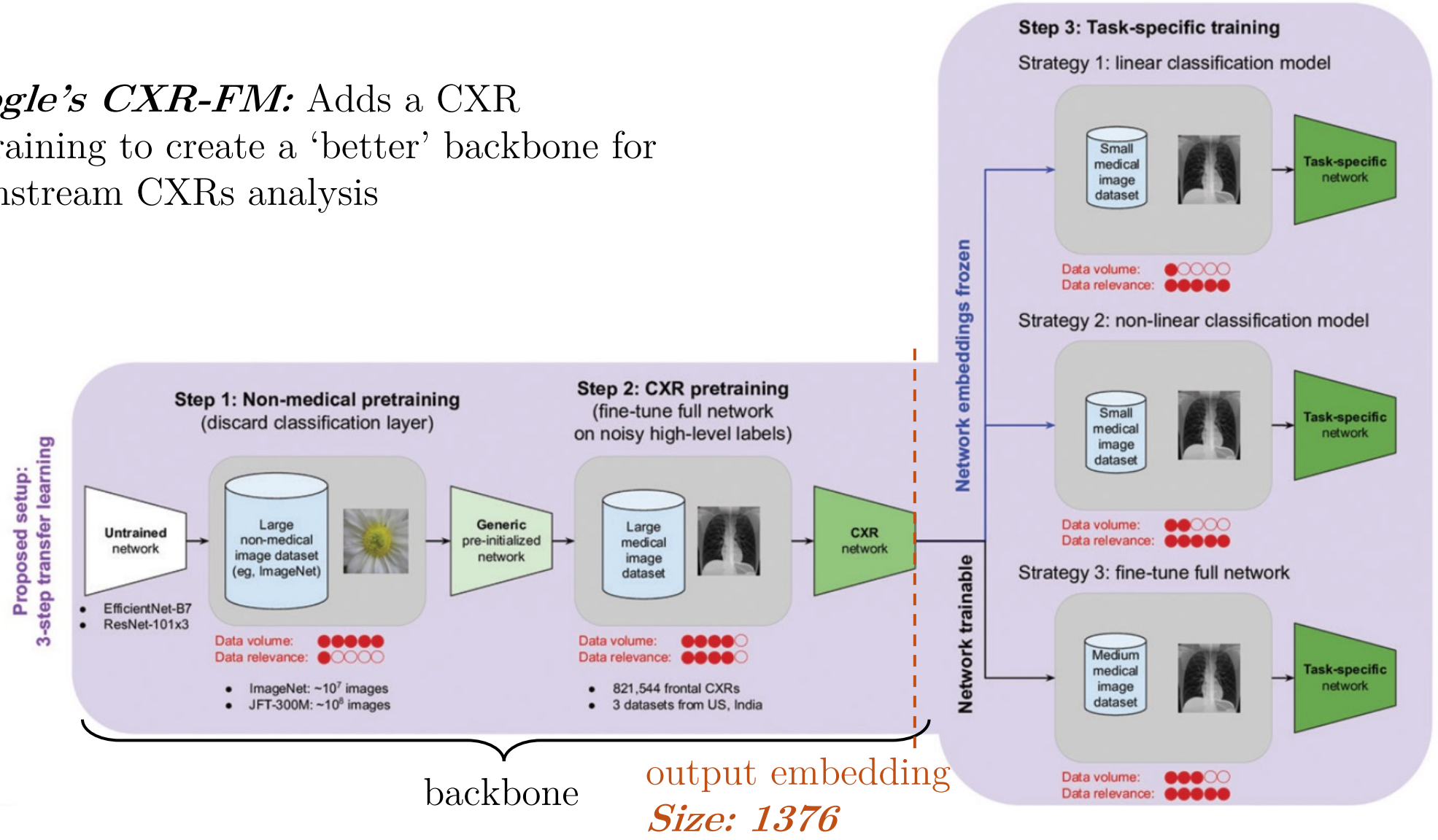
2. Foundation Models

Google's CXR-FM: Adds a CXR pretraining to create a 'better' backbone for downstream CXRs analysis



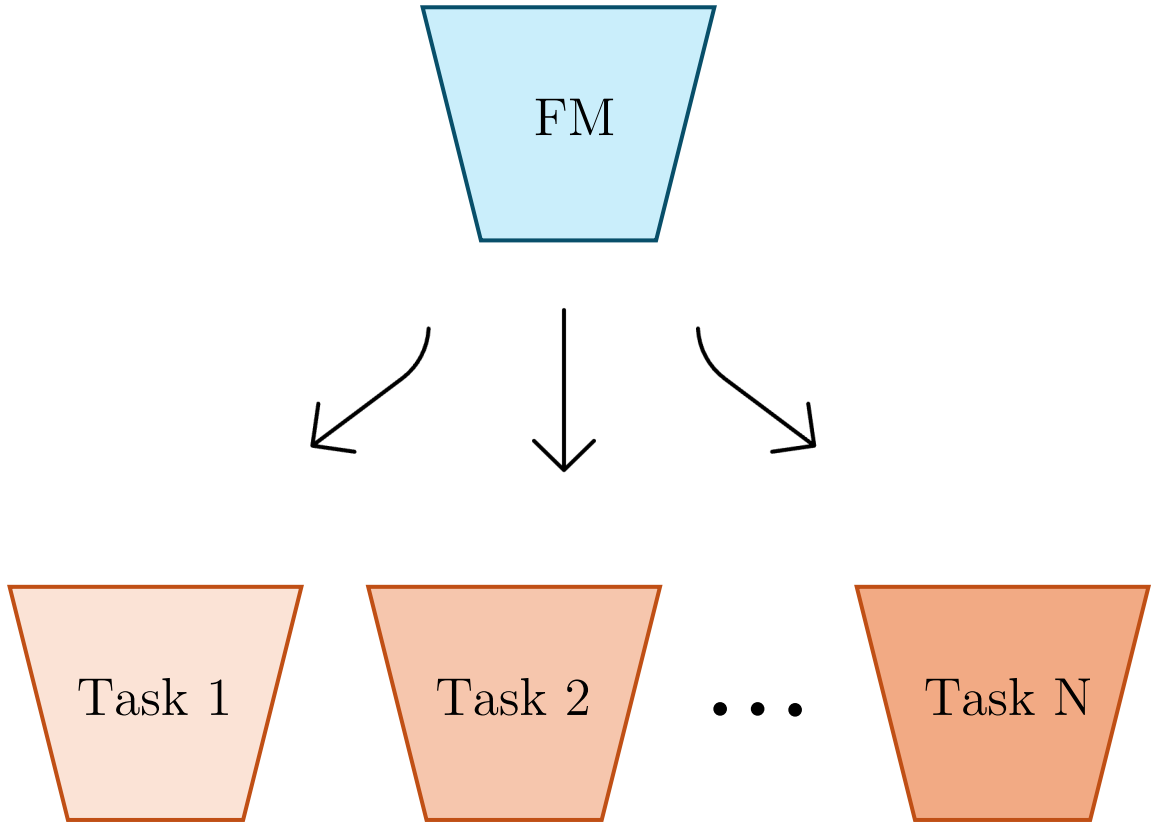
2. Foundation Models

Google's CXR-FM: Adds a CXR pretraining to create a 'better' backbone for downstream CXRs analysis



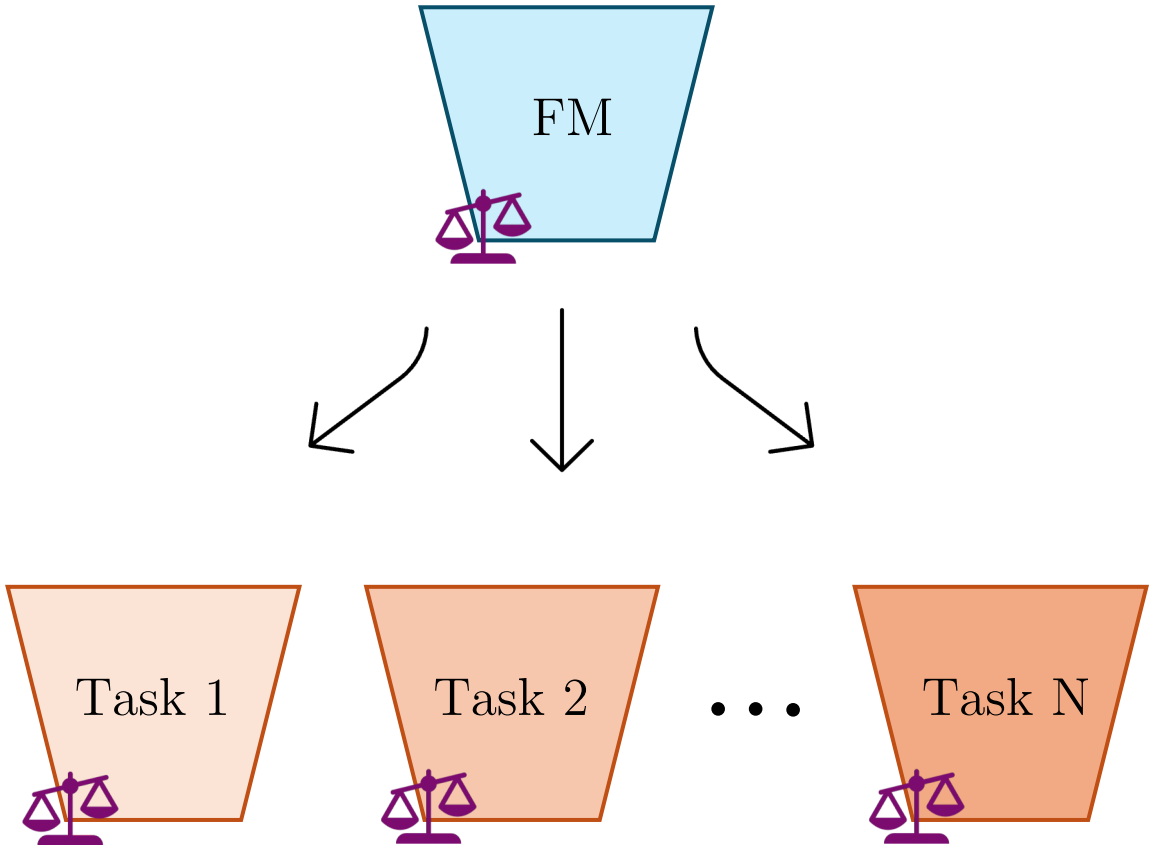
3. Associated Challenges

- General lack of profound understanding of Foundation Models (FMs)' mechanisms and impacts



3. Associated Challenges

- General lack of profound understanding of Foundation Models (FMs)' mechanisms and impacts
- Can propagate encoded biases trough downstream tasks: *any defect is likely 'inherited' by all adapted models.*



3. Associated Challenges

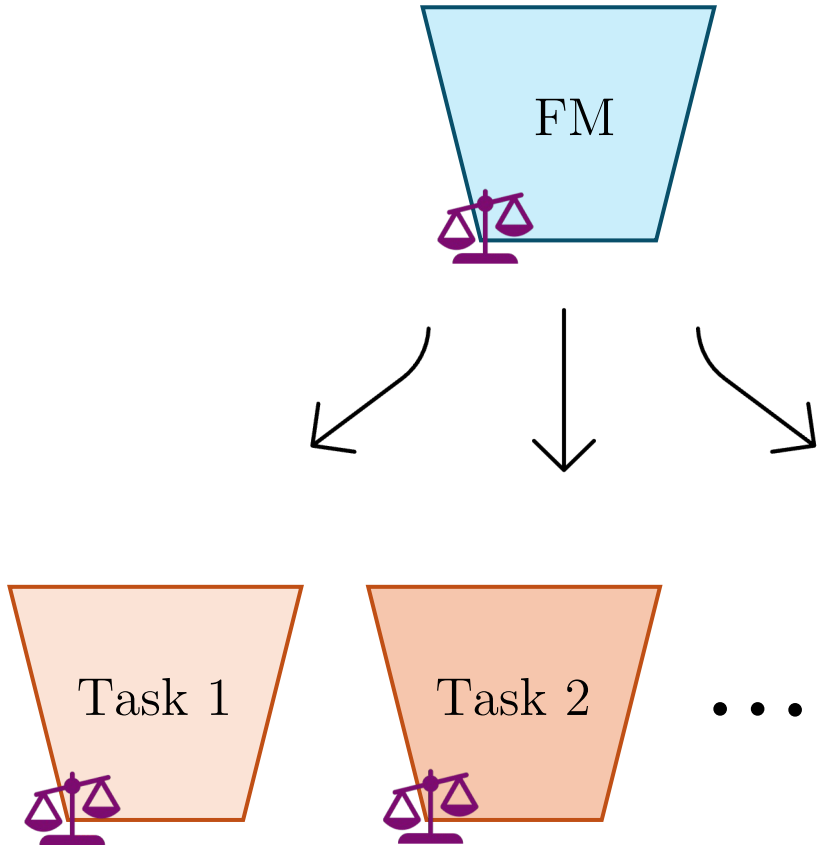
- General lack of profound understanding of Foundation Models (FMs)' mechanisms and impacts
- Can propagate encoded biases trough downstream tasks: *any defect is likely 'inherited' by all adapted models*
- Indeed, [3] showed that the CXR-FM exhibited biases related to race and biological sex: *making it unsafe for clinical deployment*


Radiology: Artificial Intelligence

ORIGINAL RESEARCH

Risk of Bias in Chest Radiography Deep Learning Foundation Models

[3]



4. Need for Knowledge Distillation

- For effective bias mitigation → full access to the FM is needed [3]

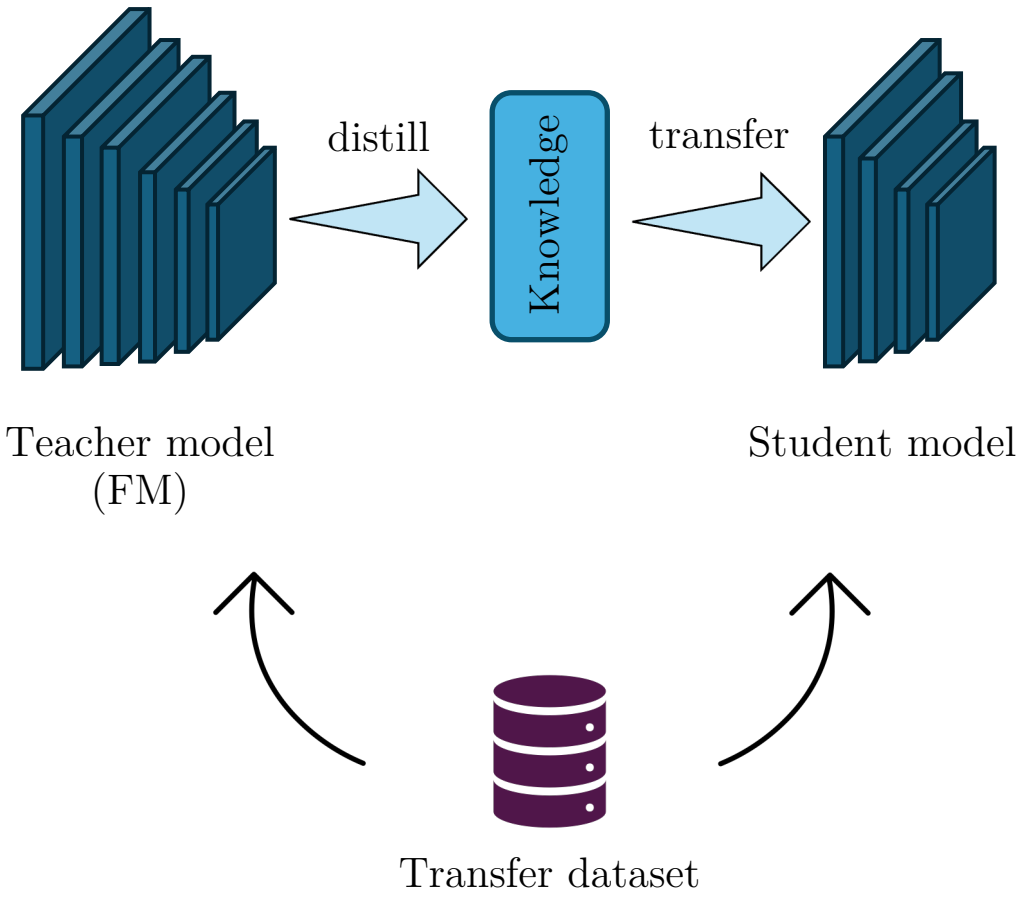
- However, CXR-FM for example is only released as an 'API call': it takes input images and generate output features BUT the parameters are not publicly available

4. Need for Knowledge Distillation

- For effective bias mitigation → full access to the FM is needed [3]

- However, CXR-FM for example is only released as an 'API call': it takes input images and generate output features BUT the parameters are not publicly available

- Interest to explore Knowledge Distillation (KD) as a means to transfer the learned knowledge from the *teacher* (e.g., CXR-FM) to a usually smaller *student* model.



References

1. Louati, H., Louati, A., Bechikh, S. et al. Topology optimization search of deep convolution neural networks for CT and X-ray image classification. BMC Med Imaging 22, 120 (2022). https://doi.org/10.1186/s12880-022-00847-w
2. Sellergren AB, Chen C, Nabulsi Z, et al (2022) Simplified Transfer Learning for Chest Radiography Models Using Less Data. Radiology 305:454–465
3. Glocker B, Jones C, Roschewitz M, Winzeck S (2023) Risk of Bias in Chest Radiography Deep Learning Foundation Models. Radiol Artif Intell 5:230060
4. Gou J, Yu B, Maybank SJ, Tao D (2020) Knowledge Distillation: A Survey. Int J Comput Vis 129:1789–1819