# From General to Clinical: Adapting Foundation Models for Medical Images

**Kick-off Presentation**
July 10th, 2025

# Who are we?

**Computational Imaging and AI in Medicine**

Prof. Dr. Julia Schnabel

# Who are we?

**Sameer Ambekar**

Research interest:
- Distribution shifts
- Test-time adaptation
- Foundation models

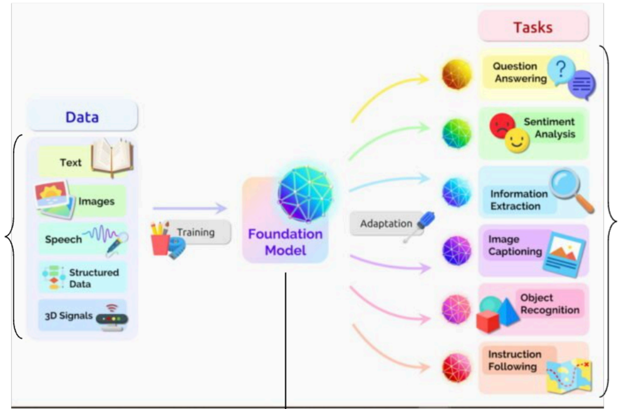
**Dr. Laura Daza**

Research interest:
- Multi-modal Learning
- Foundation models
- Segmentation

# Towards Foundation Models

**Massive** amounts of data

Semi-, **self-**, unsupervised training




https://blogs.nvidia.com/wp-content/uploads/2023/03/Transformer-apps.jpg

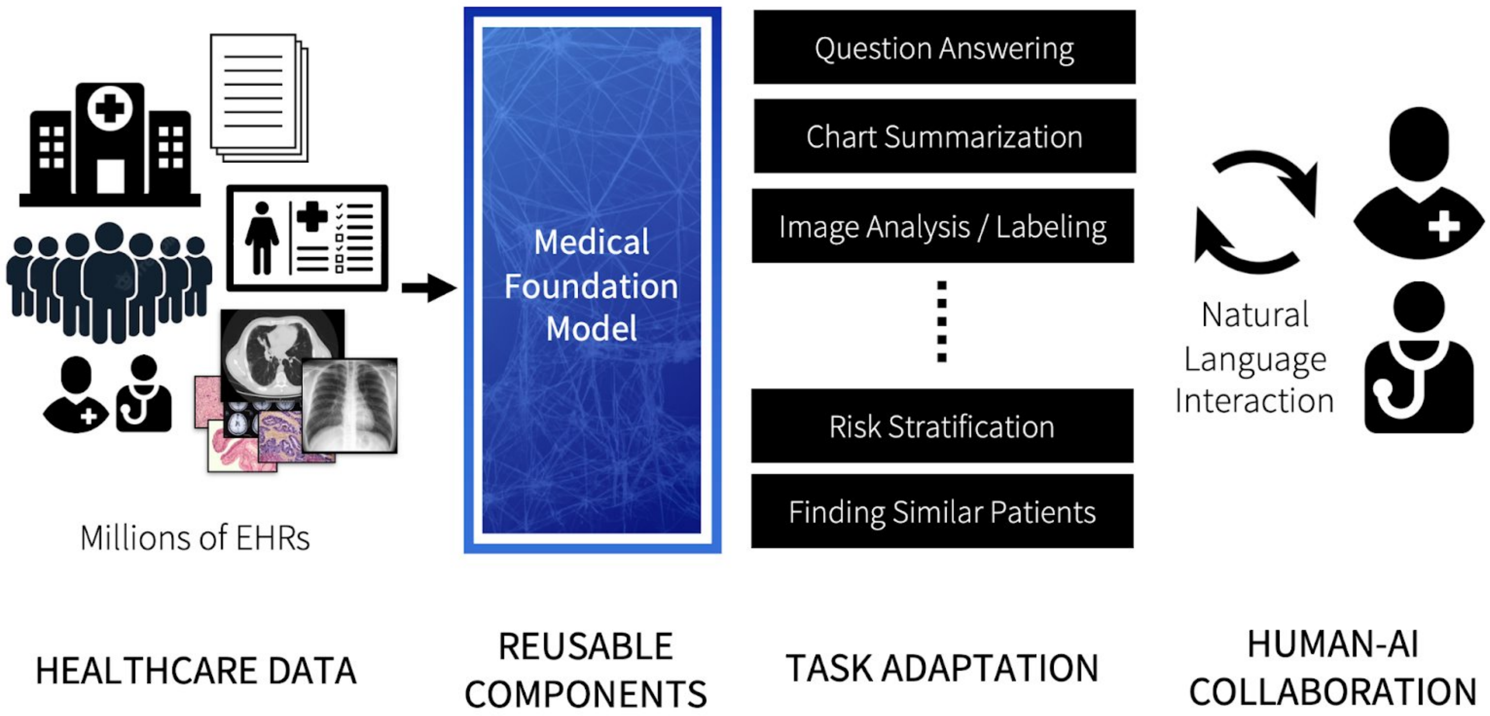
Can be adapted to **many** downstream tasks

Huge models
(hundreds of billions of parameters)

Pretraining stage is extremely expensive

Fine tuning is way cheaper than training for every task

# Towards Medical Foudation Models



# How to use shots for Foundation models?

## Zero-Shot

```
Translate English to French:     ← task description
cheese =>                        ← prompt
```

Model predicts the answer with a natural language task description.

No Gradient updates

## One-Shot

```
Translate English to French:     ← task description
sea otter => loutre de mer       ← example
cheese =>                        ← prompt
```

In addition to task description, the model sees a single example of the task.

No Gradient updates

## Few-Shot

```
Translate English to French:     ← task description
sea otter => loutre de mer       ← examples
peppermint => menthe poivrée
plush girafe => girafe peluche
cheese =>                        ← prompt
```

In addition to task description, the model sees a few examples of the task.

No Gradient updates

# Goals of the Seminar

**Understand:**

**(i) What to adapt:**
Identify which model parameters to adapt, and explore common strategies for adaptation

**(ii) How to adapt:**
The various strategies by which the models can be adapted to the target data

**(iii) Evaluation for clinical context:**
Discuss methods for handling distributional shifts that mirror real-world clinical scenarios

**Learn:**

How to read and present a scientific paper
How to design and present a scientific poster

**Know better:**

A wide range of foundation models and how to adapt and align them to medical imaging applications

International guest speaker talks on the topic

# Papers for Teaching and Presentation (1/3)

[1] R. Bommasani and et al. On the opportunities and risks of foundation models. arXiv, 2021.

[2] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PmLR, 2020.

[3] A. Radford and et al. Learning transferable visual models from natural language supervision. ICML, 2021

[4] Caron and et al. Emerging properties in self-supervised vision transformers. ICCV, 2021

[5] Dosovitskiy, Alexey, et al. "An image is worth 16×16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

[6] Z. Liu and et al. Swin transformer: Hierarchical vision transformer using shifted windows. In ICCV, 2021

[7] Kenton, Jacob Devlin Ming-Wei Chang, and Lee Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *Proceedings of naacL-HLT*. Vol. 1. No. 2. 2019.

# Papers for Teaching and Presentation (2/3)


[8] K. He and et al. Masked autoencoders are scalable vision learners. CVPR, 2022.

[9] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models. arXiv 2021." ICLR 2022

[10] Ma and et al. Medsam: Segment anything model for medical images. arXiv, 2023

[11] Ge and et al. Domain adaptation via prompt learning. IEEE TNNLS, 2023

[12] Y. Zhang and et al. Biomedclip: Open biomedical contrastive language-image pretraining. arXiv, 2023

[13] Hoopes, Andrew, et al. "Voxelprompt: A vision-language agent for grounded medical image analysis." *arXiv preprint arXiv:2410.08397* (2024).

[14] Zhao, Theodore, et al. "Biomedparse: a biomedical foundation model for image parsing of everything everywhere all at once." *arXiv preprint arXiv:2405.12971* (2024).

[15] Wald, Tassilo, et al. "An OpenMind for 3D medical vision self-supervised learning." *arXiv preprint arXiv:2412.17041* (2024).

# Papers for Teaching and Presentation (3/3)

[16] Remy, François, Kris Demuynck, and Thomas Demeester. "Biolord: Learning ontological representations from definitions (for biomedical concepts and their textual descriptions)." *arXiv preprint arXiv:2210.11892* (2022).

[17] Yuan, Zheng, et al. "Improving biomedical pretrained language models with knowledge." *arXiv preprint arXiv:2104.10344* (2021).

[18] Huang, Ziyan, et al. "Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training." *arXiv preprint arXiv:2304.06716* (2023).