Self-Supervised Learning (SSL)
AI 835
Aug-Dec 2023

Introduction

1st Aug 2023

Dr. V. Ramasubramanian
Professor
International Institute of Information Technology - Bangalore (IIIT-B)
Bangalore, India

Google-Drive Link – View-permission only

Main-folder: SSL-Course-Aug-Dec-2023/

https://drive.google.com/drive/folders/1-DDvA6HilokLH40Kh40zpiIH1AuJbjGK?usp=sharing

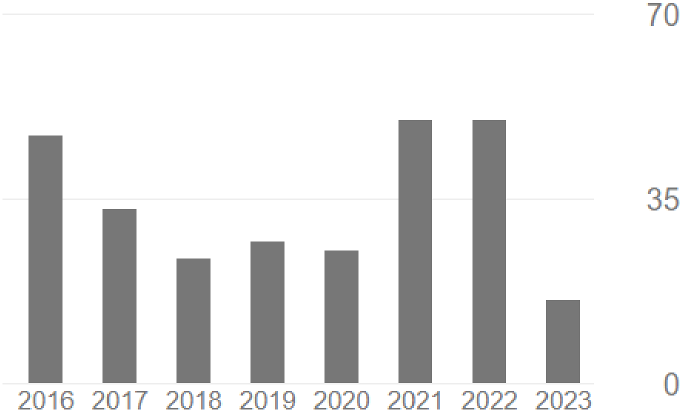
- Class-Slides-Notes
- Reference-Papers
- Reference-Slides
- Review-Survey-Papers

V. Ramasubramanian – Profile & Research @ IIIT B

- PhD [TIFR, Bombay, 1993]
- TIFR, Bombay
- Univ. Valencia, Spain
- ATR, Kyoto, Japan
- IISc, Bangalore
- Siemens Corporate Research
- PES University
- IIIT Bangalore

- Research students @ IIIT B
  - 6 PhD and 3 MS students
  - 1 PhD (Graduated)
  - 3 MS (Graduated)
- Publications
  - 80+ papers, 1 Book

	All	Since 2018
Citations	1068	192
h-index	16	7
i10-index	25	4



- ❑ Automatic speech recognition
- ❑ Machine learning, deep learning
- ❑ Few-shot learning
- ❑ Associative memory formulations

V. Ramasubramanian – Teaching @ IIIT B

Linear Algebra **[1 Term]** (GEN 504) Aug-Dec 2017

Machine Learning **[1 Term]** (CS/DS 612) Jan-Apr 2018

Maths for Machine Learning **[5 Terms]** (GEN 511) Jan-Apr 2018, (GEN 512) Aug-Dec 2018, (GEN 512) Aug-Dec 2019, (AI 512) Mar-Apr 2020, Aug-Dec 2020

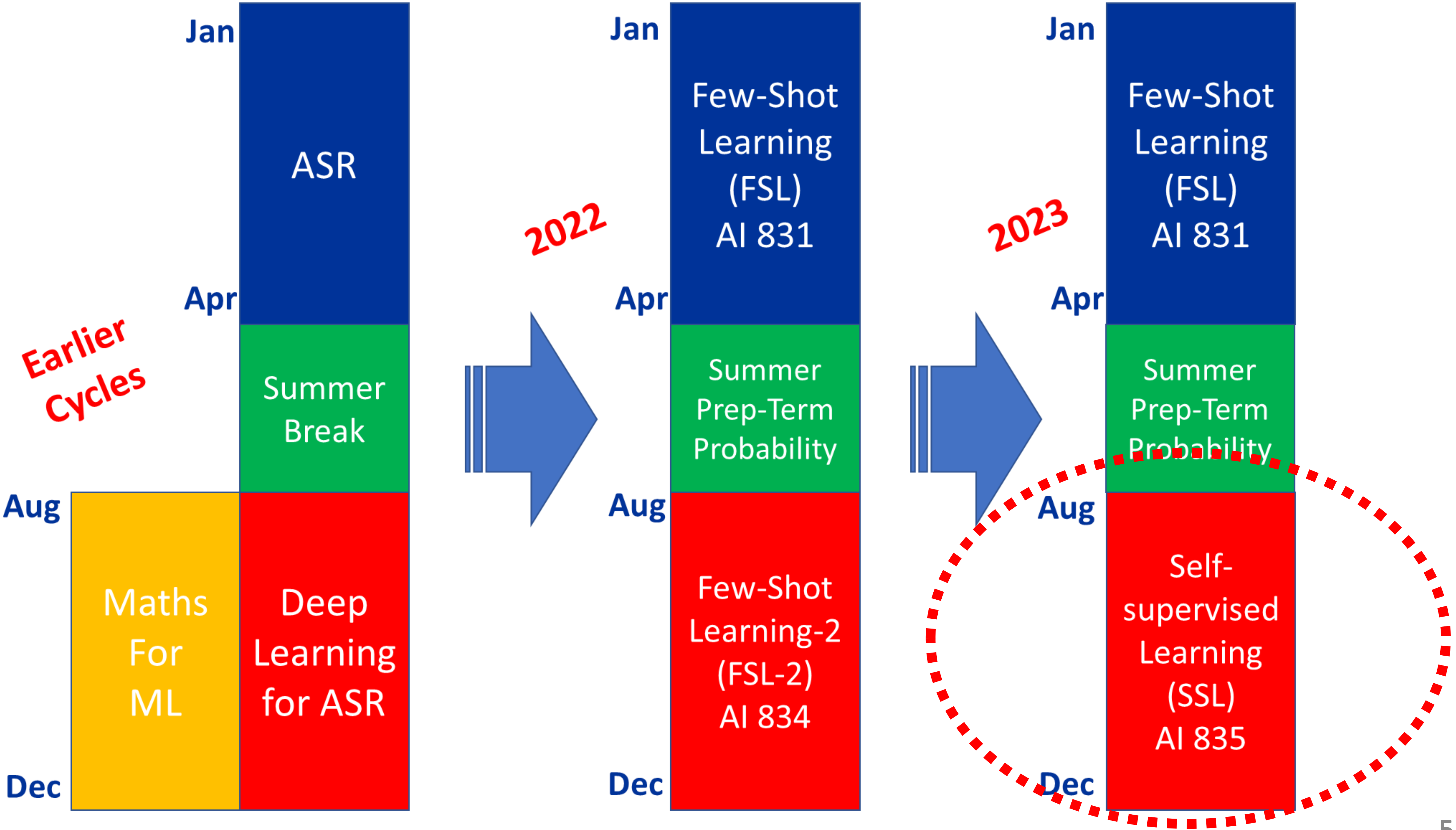
Prep-term Probability Theory [4] Aug-Sep 2020, Aug 2021, July 2022, July 2023

Speech Processing **[1 Term]** (DS / NC 822) Jan-Apr 2018

Automatic Speech Recognition (ASR) **[4 Terms]** (DS / NE 821) Jan-Apr 2017, (DS / NC 824) Aug-Dec 2018, (DS / SP 823) Jan-Apr 2019, Jan-Apr 2020

Deep Learning for Automatic Speech Recognition (DL-ASR) **[4 terms]** (DS / NC 871) Aug-Dec 2017, (DS / SP 826) Aug-Dec 2019, (AI 826) Aug-Dec 2020

Few-shot Learning (FSL) **(AI 831) Jan – May 2022**
Few-shot Learning - 2 (FSL-2) **(AI 832) Aug – Dec 2022**
Few-shot Learning (FSL) **(AI 831) Jan – May 2023**
Self-supervised Learning (SSL) **(AI 835) Aug – Dec 2023**


Earlier Cycles
Jan
Apr
Aug
Dec
ASR
Summer Break
Maths For ML
Deep Learning for ASR
2022
Jan
Apr
Aug
Dec
Few-Shot Learning (FSL) AI 831
Summer Prep-Term Probability
Few-Shot Learning-2 (FSL-2) AI 834
2023
Jan
Apr
Aug
Dec
Few-Shot Learning (FSL) AI 831
Summer Prep-Term Probability
Self-supervised Learning (SSL) AI 835

Course Code / Course Name	AI 835 / Self-Supervised Learning (SSL)	
Course Instructor Name(s)	Prof. V. Ramasubramanian	
Credits (L:T:P) **(Lecture : Tutorial : Practical)**	**Hours**	**Component**
	3	Lecture (3 hrs)
	0	Tutorial (0 hrs)
	2	Practical (2 hrs) – Assignments / Homework
	L:T:P = 3:0:2	**Total Credits = 4**

- Area of Specialization – AI/ML
- Course Category – General
- iMTech and Mtech
- ECE and CSE

Course Pre-requisites: Maths for ML and ML

AI 835 Self-supervised Learning (SSL) V Ramasubramanian
TUE, THU 2-3:30pm @ R203

TAs

- **Dhanya Eledath** – PhD student (2018-) – Few-shot Learning for E2E ASR
- **Tirthankar Banerjee** – PhD student (2023-) – Cross-domain FSL for ASR

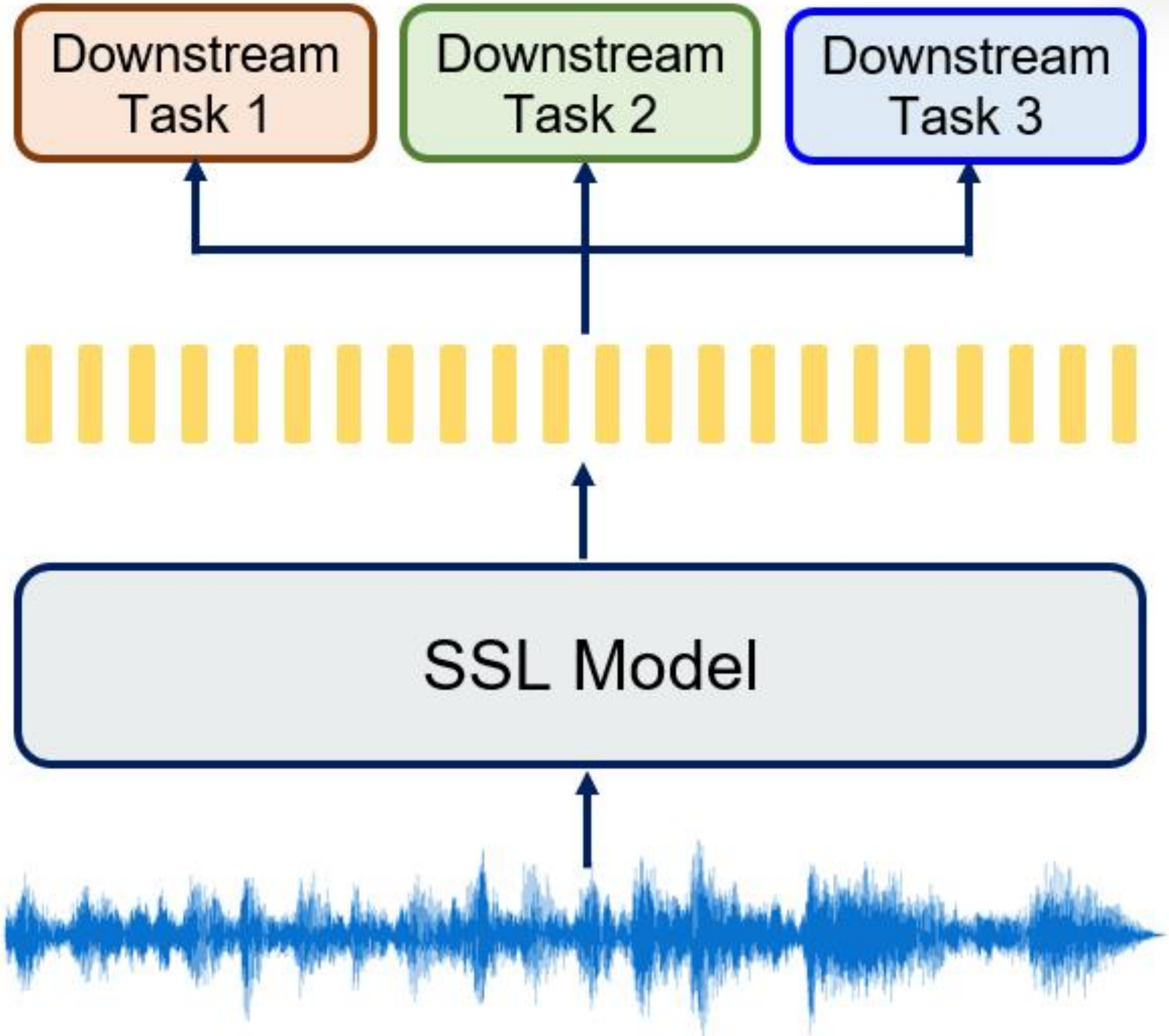
Hands-on / Tutorials

Module – 4 – By TAs

Domain-specific Foundation Models
(e.g. wav2vec and vq-wav2vec and wav2vec 2.0)
in detail.

Applications

Operating Systems



Let’s welcome the era of self-supervised Learning.

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

Self-Supervised Learning (SSL)

- What makes Foundation Models (FMs) tick?
- What is “under-the-hood” of FMs?

Foundation Models

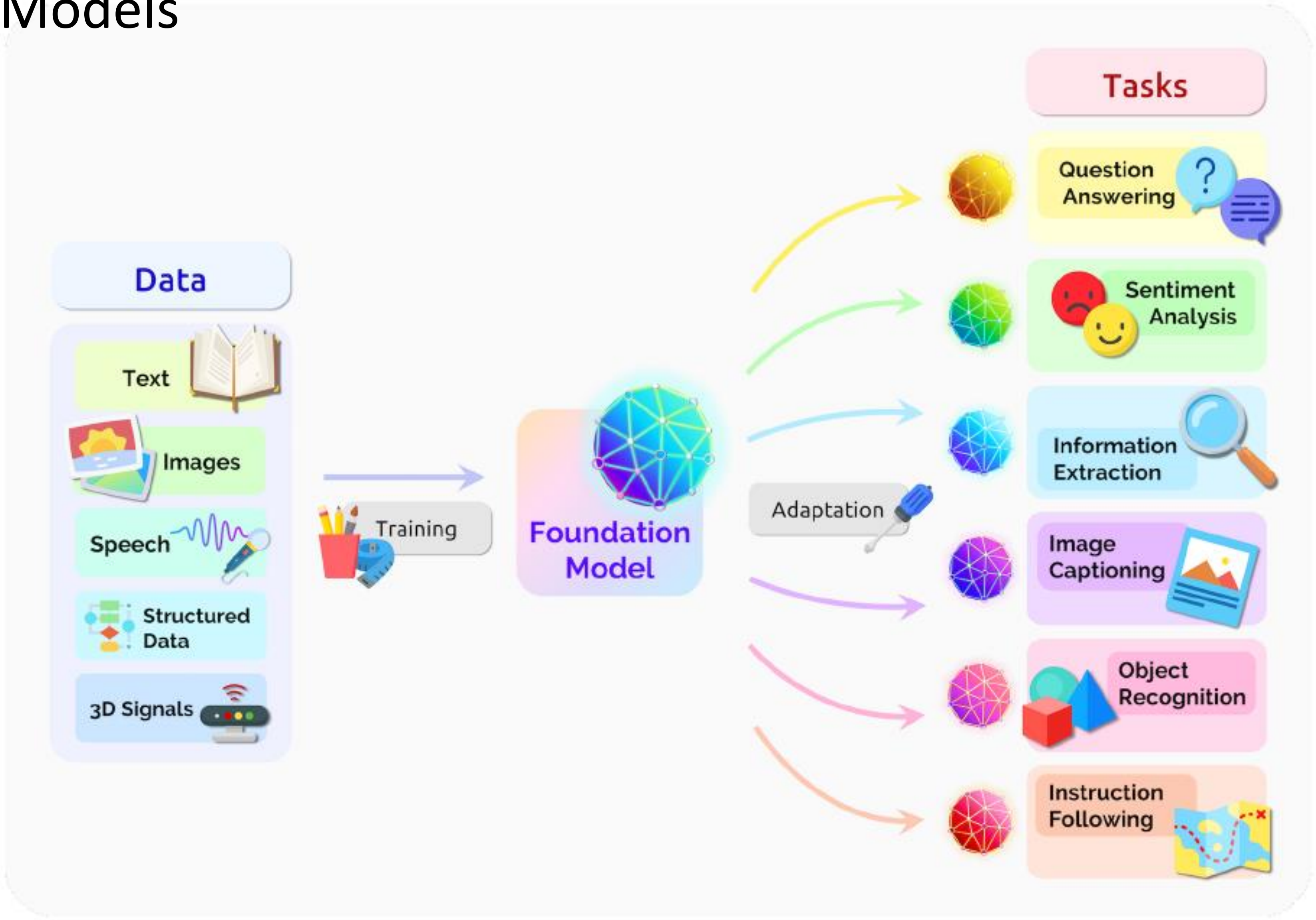


Fig. 2. A foundation model can centralize the information from all the data from various modalities. This one model can then be adapted to a wide range of downstream tasks.

Foundation Models

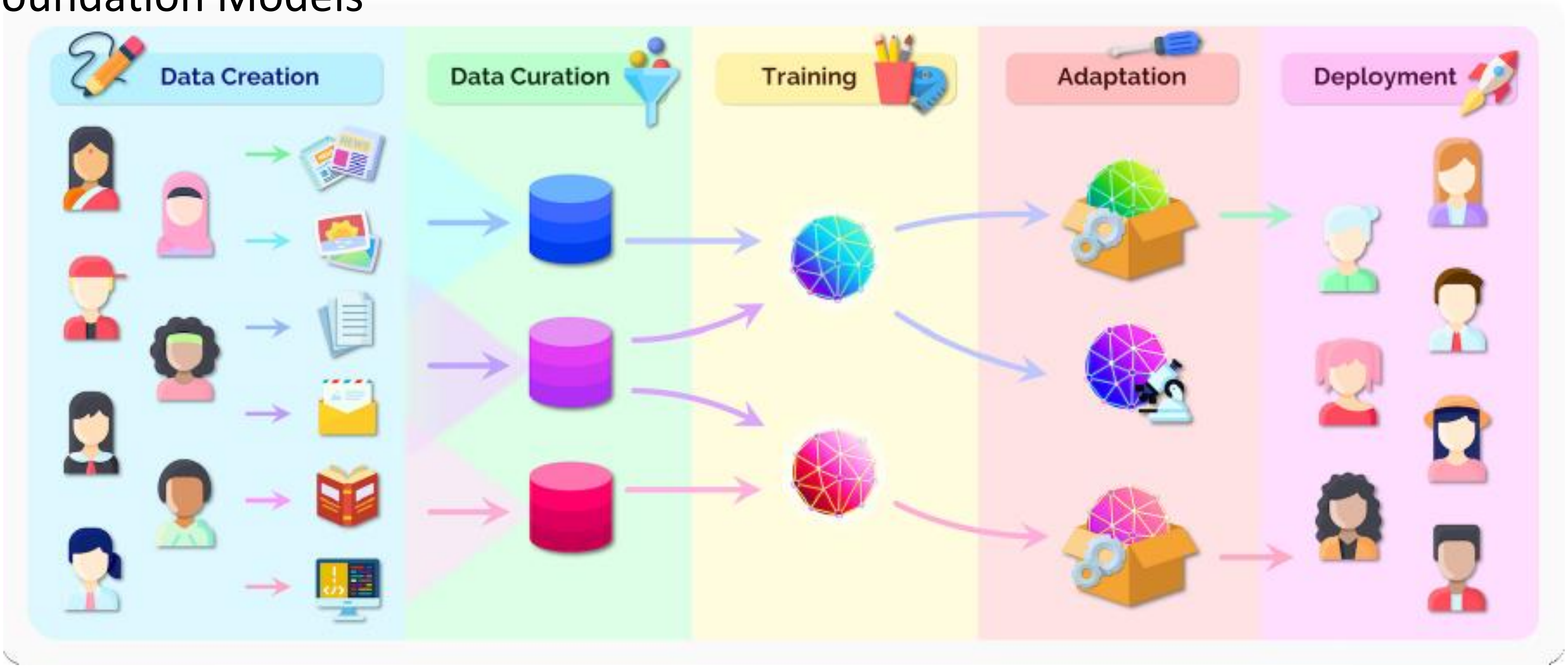


Fig. 3. Before reasoning about the social impact of foundation models, it is important to understand that they are part of a broader ecosystem that stretches from data creation to deployment. At both ends, we highlight the role of people as the ultimate source of data into training of a foundation model, but also as the downstream recipients of any benefits and harms. Thoughtful data curation and adaptation should be part of the responsible development of any AI system. Finally, note that the deployment of adapted foundation models is a decision separate from their construction, which could be for research.

Module – 1

Basic learning paradigms (supervised, unsupervised, semi-supervised, self-supervised) definitions and settings.

Origins of Self-Supervised Learning (SSL) – early unsupervised learning paradigms (RBM, AE, Word2vec, AR, etc.).

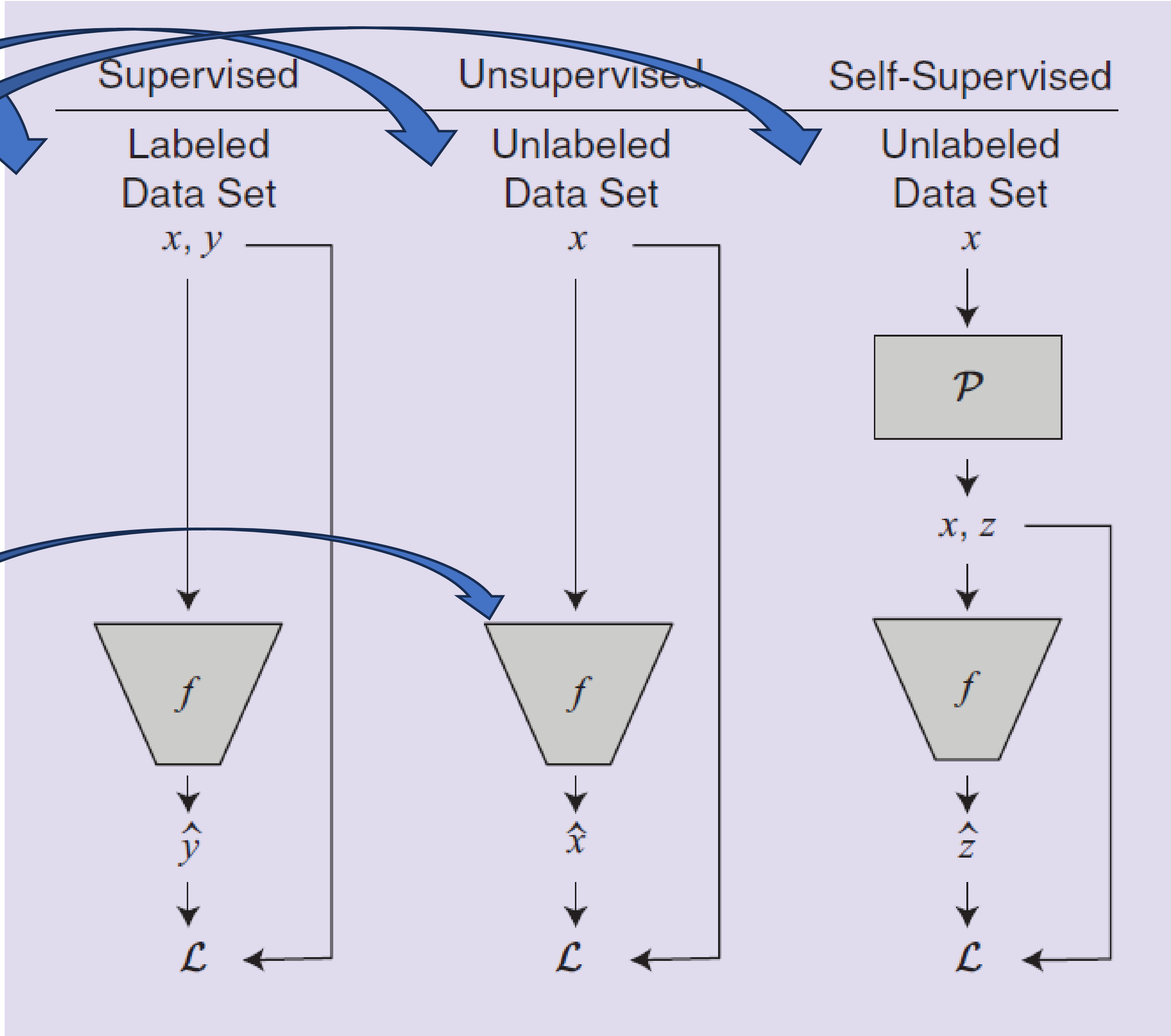


FIGURE 1. Contrasting supervised, unsupervised and self-supervised learning paradigms for training a model f using raw data x, labels y, and loss function $\mathcal{L}$. Self-supervision methods introduce pretext tasks $\mathcal{P}$ that generate pseudolabels z for discriminative training of f.

Module – 2

Basics of Foundation Models, SSL formalisms, definitions and examples of

- **Pretext tasks**
- **Losses**
- **Downstream adaptations**

in a domain-agnostic setting

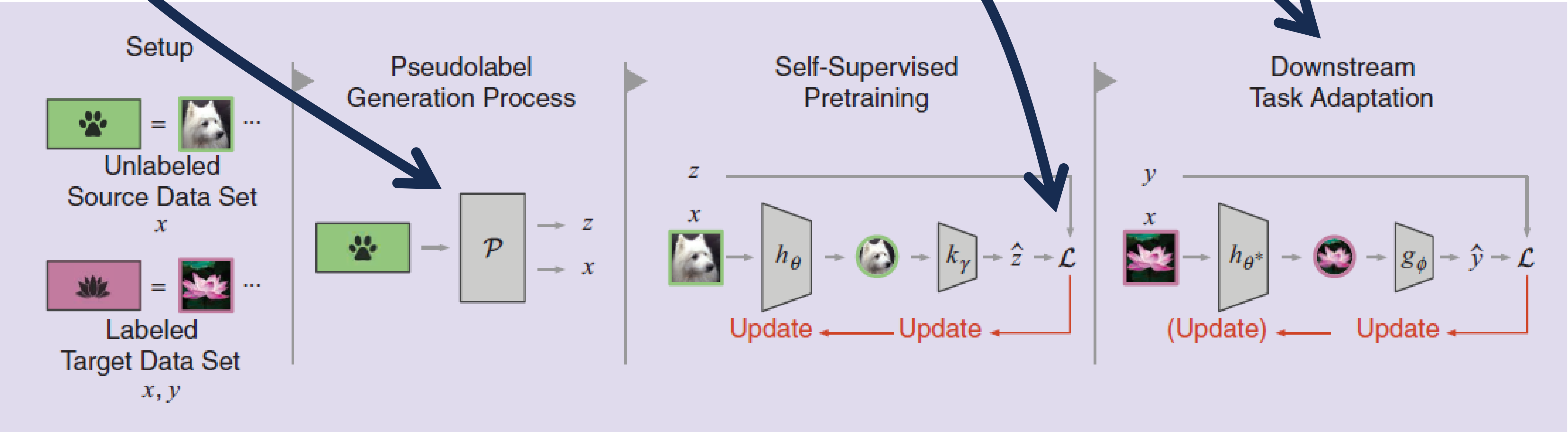


FIGURE 2. The self-supervised workflow starts with an unlabeled source data set and a labeled target data set. As defined by the pretext task, pseudolabels are programmatically generated from the unlabeled set. The resulting inputs, x and pseudolabels z, are used to pretrain the model $k_\gamma(h_\theta(\cdot))$—composed of feature extractor h_θ and output k_γ modules—to solve the pretext task. After pretraining is complete, the learned weights θ^* of the feature extractor h_{θ^*} are transferred and used together with a new output module g_ϕ to solve the downstream target task.

Module – 3

- **Domain-agnostic** pretext task formulations and corresponding **domain-specific** pretext formulations along with associated losses:
  - the generic principles and definitions of **pretext tasks** underlying a wide variety of SSL frameworks in different domains such as WaveNet, WaveRNN, VAE, VQ-VAE, CPC, Wav2vec, Wav2vec2.0, BERT, HuBERT, GPT, XLNet, PixelCNN, PixelRNN, iGPT, SimCLR, Barlow Twins, data2vec etc. and
  - a range of **associated losses** used in the pretext-task learning such as Contrastive loss, Triplet loss, Lifted structured loss, Multi-class n-pair loss, Noise contrastive estimation (NCE), InfoNCE etc.

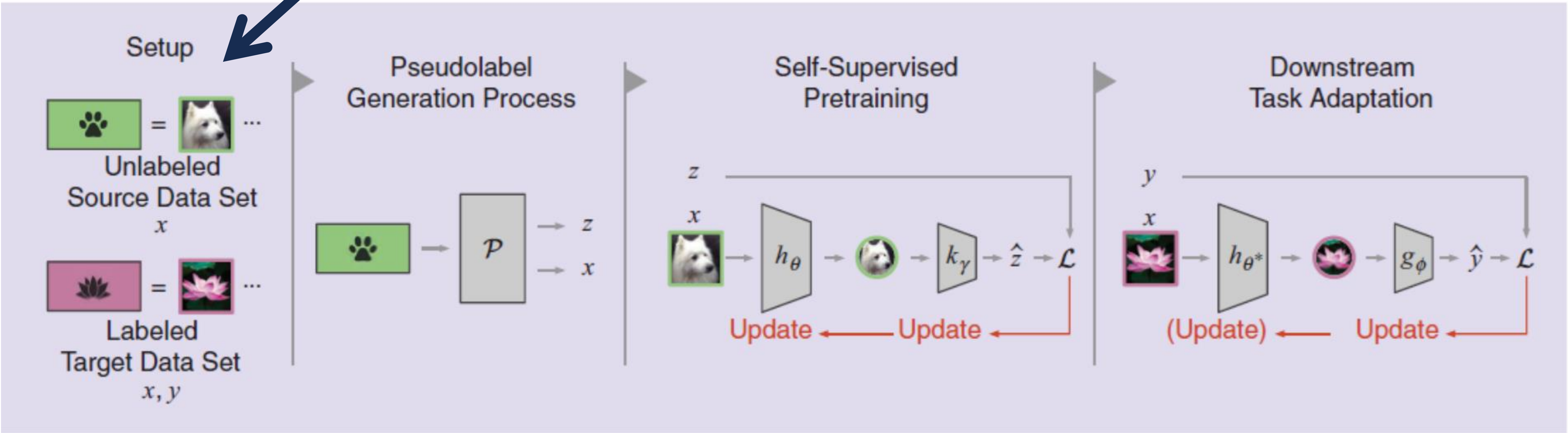


FIGURE 2. The self-supervised workflow starts with an unlabeled source data set and a labeled target data set. As defined by the pretext task, pseudolabels are programmatically generated from the unlabeled set. The resulting inputs, x and pseudolabels z, are used to pretrain the model $k_\gamma(h_\theta(\cdot))$—composed of feature extractor h_θ and output k_γ modules—to solve the pretext task. After pretraining is complete, the learned weights θ^* of the feature extractor h_{θ^*} are transferred and used together with a new output module g_ϕ to solve the downstream target task.

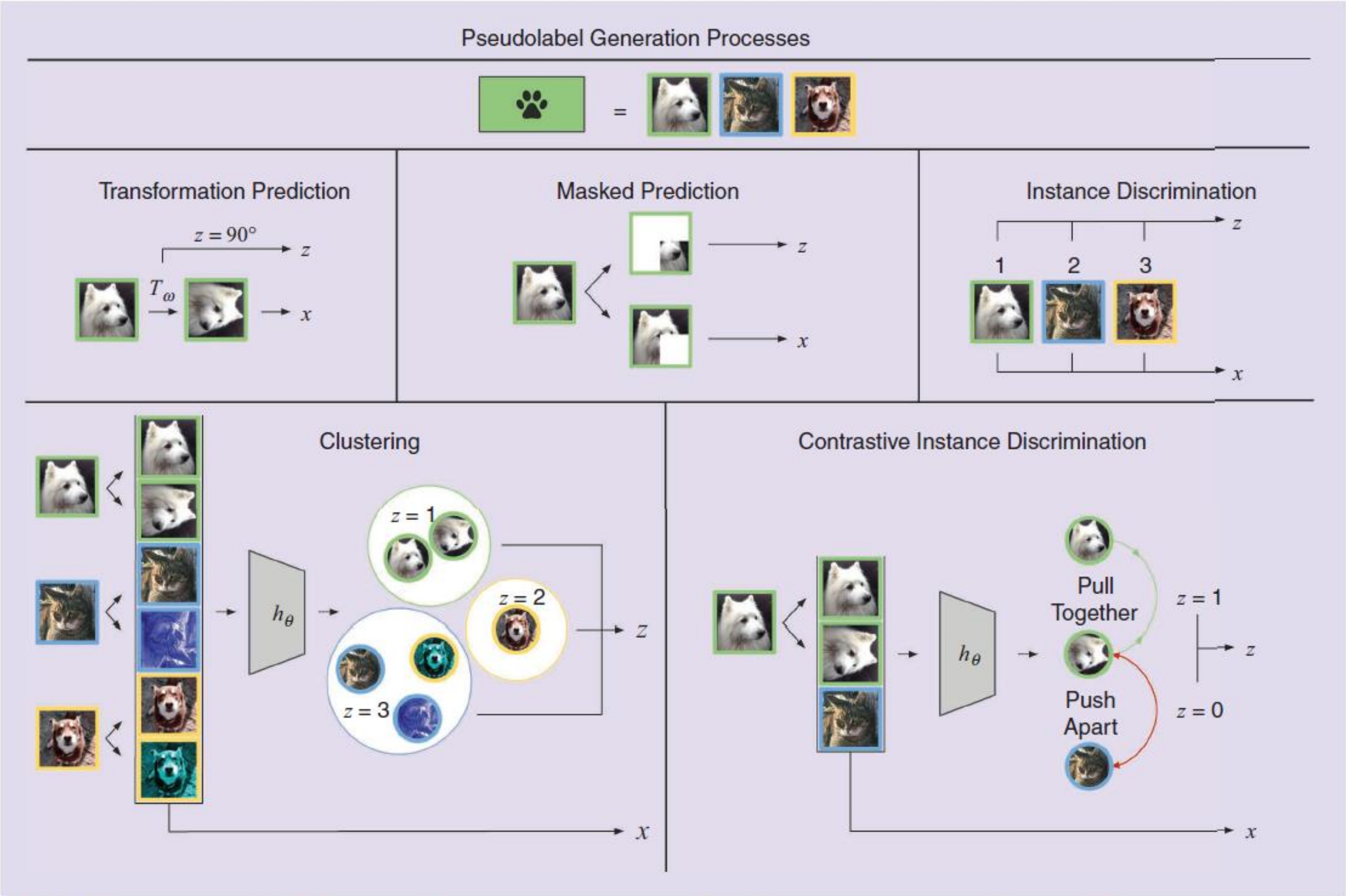


FIGURE 3. Illustrative examples of the way pseudolabels are generated in the four families of pretext tasks of our taxonomy: TP, masked prediction, instance discrimination, and clustering. An additional depiction is included of the popular version of instance discrimination using contrastive losses. The squares represent inputs x, while circles portray the feature vectors of those inputs, $h_\theta(x)$.

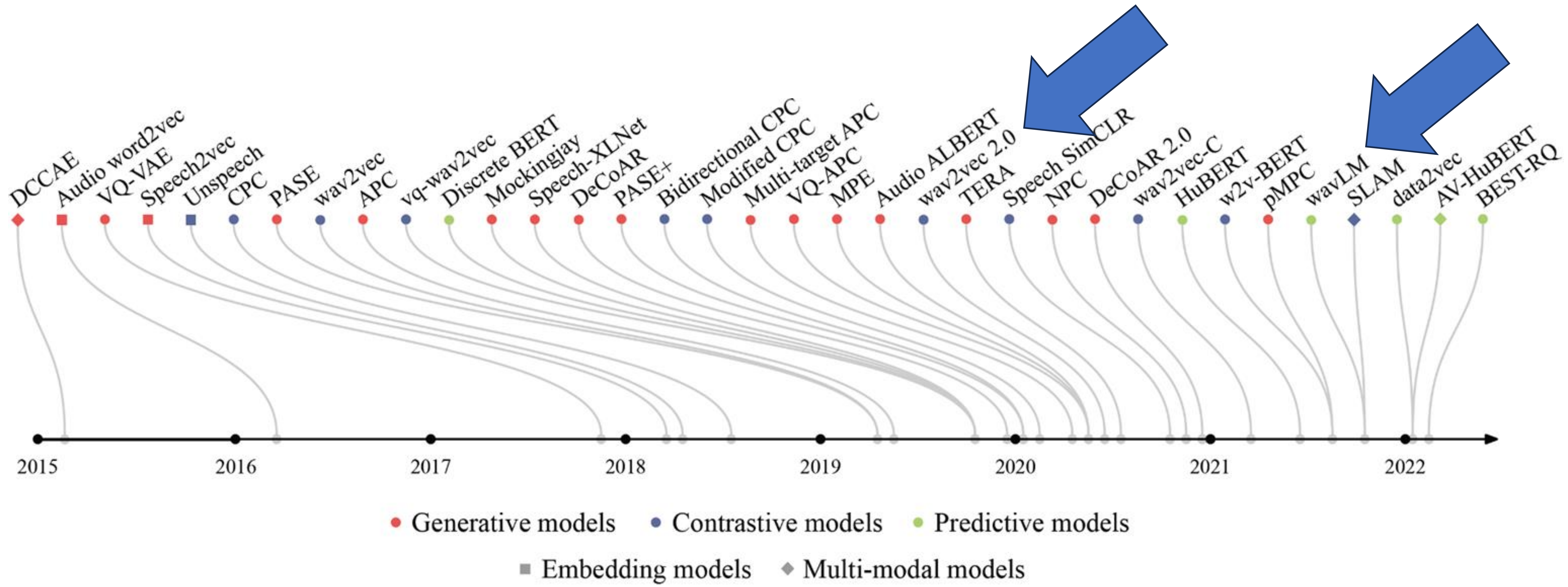
Speech representation learning methods

Contrastive approaches

Predictive approaches

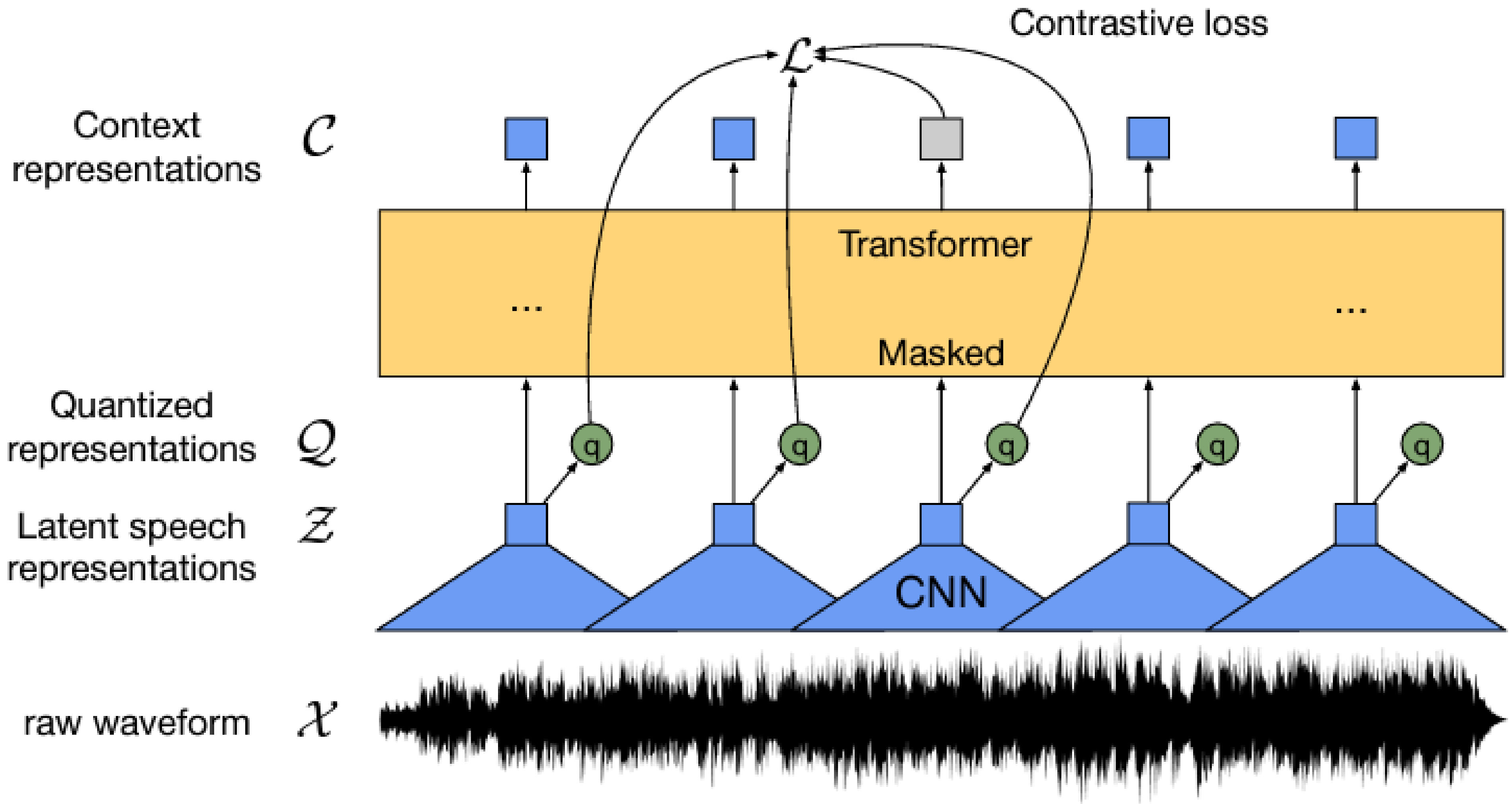
Generative approaches

Speech representation learning methods



Module – 4

Domain-specific Foundation Models (e.g. **wav2vec and vq-wav2vec and wav2vec 2.0**) in detail.



Learning Resources

1. Rishi Bommasani et al., “On the Opportunities and Risks of Foundation Models”, Technical Report, Center for Research on Foundation Models (CRFM), Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University, 2021/2022.
2. Randall Balestriero et al., “A Cookbook of Self-Supervised Learning”, arXiv:2304.12210v1 [cs.LG] 24 Apr 2023.
3. Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales, “Self-Supervised Representation Learning: Introduction, Advances and Challenges”, IEEE Signal Processing Magazine, 39 (3): 42–62, May 2022.
4. Jie Gui, Tuo Chen, Qiong Cao, Zhenan Sun, Hao Luo and Dacheng Tao, “A Survey of Self-Supervised Learning from Multiple Perspectives: Algorithms, Theory, Applications and Future Trends”, arXiv:2301.05712v1 [cs.LG] 13 Jan 2023.
5. Xiao Liu , Fanjin Zhang , Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang and Jie Tang, “Self-Supervised Learning: Generative or Contrastive”, IEEE Transaction on Knowledge and Data Engineering, pp. 857-876, vol. 35, no. 1, Jan 2023.
6. Madeline C. Schiappa, Yogesh S. Rawat and Mubarak Shah, “Self-Supervised Learning for Videos: A Survey”, ACM Computing Surveys, 2022.
7. Longlong Jing and Yingli Tian, “Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 43, no. 11, Nov. 2021.
8. Abdelrahman Mohamed, Hung-yi Lee , Lasse Borgholt, Jakob D. Havtorn, Joakim Edin, Christian Igel, Katrin Kirchhoff, Shang-Wen Li, Karen Livescu, Lars Maaløe, Tara N. Sainath and Shinji Watanabe, “Self-Supervised Speech Representation Learning: A Review”, IEEE Journal of Selected Topics in Signal Processing, pp. 1179 – 1210, vol. 16, no. 6, Oct. 2022.
9. Shuo Liu, Adria Mallol-Ragolta, Emilia Parada-Cabaleiro, Kun Qian, Xin Jing, Alexander Kathan, Bin Hu and Bjorn W. Schuller, “Audio Self-supervised Learning: A Survey”, arXiv:2203.01205v1 [cs.SD] 2 Mar 2022.
10. Review/survey articles in Journals/Tutorials in Conferences (some – as above)
11. Publications from NIPS, NeurIPS, ICML, ICLR, ECCV, ICCV, CVPR, IEEE Transactions (various domains).

Assessment Plan

- Assignment (40 Marks): 4 of 10 Marks Each
- Quiz (50 Marks): 2 of 25 Marks Each
- Class Participation: 10 Marks

Thank you !!