Dynamic SUPERB

DYNAMIC-SUPERB: TOWARDS A DYNAMIC, COLLABORATIVE, AND COMPREHENSIVE INSTRUCTION-TUNING BENCHMARK FOR SPEECH

Chien-yu Huang[1], *Ke-Han Lu*[*1], *Shih-Heng Wang*[*1], *Chi-Yuan Hsiao*[†1], *Chun-Yi Kuan*[†1], *Haibin Wu*[†1]
Siddhant Arora[§2], *Kai-Wei Chang*[§1], *Jiatong Shi*[2], *Yifan Peng*[2], *Roshan Sharma*[2], *Shinji Watanabe*[2]
Bhiksha Ramakrishnan[2,3], *Shady Shehata*[3], *Hung-yi Lee*[1]

[1]National Taiwan University, Taiwan, [2]Carnegie Mellon University, USA
[3]Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates

https://arxiv.org/abs/2309.09510

Project page: https://github.com/dynamic-superb/dynamic-superb

Outline

Hung-yi Lee

- Motivation
- Setting
- Baselines

Chien-yu Huang

- Provide more details

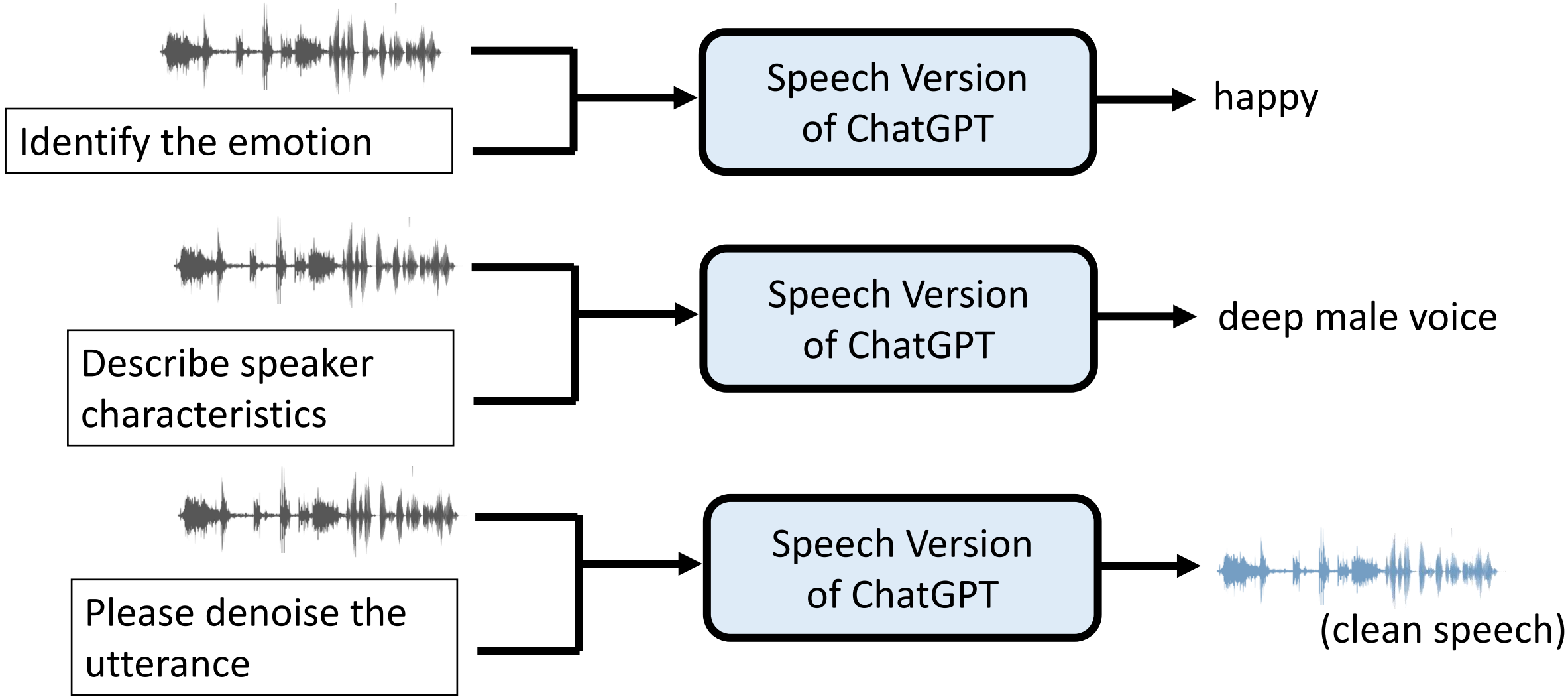
Chien-yu Huang

Motivation

Motivation

Evaluating **universal speech models** using **instruction tuning** to perform multiple tasks in a **zero-shot** fashion.

…… Speech Version of ChatGPT



(This page is not a complete survey.)

What is still missing?

Speech

SUPERB series: 17 tasks

https://arxiv.org/abs/2105.01051
https://arxiv.org/abs/2203.06849
https://arxiv.org/abs/2210.08634
https://arxiv.org/abs/2210.07185
https://arxiv.org/abs/2110.06280

We need something bigger.

with instruction ☺

NLP

General Language Understanding Evaluation (GLUE): 9 tasks

https://arxiv.org/abs/1804.07461

Super GLUE: 8 tasks

https://arxiv.org/abs/1905.00537

FLAN: 62 tasks

https://arxiv.org/abs/2109.01652

CrossFit: 160 tasks

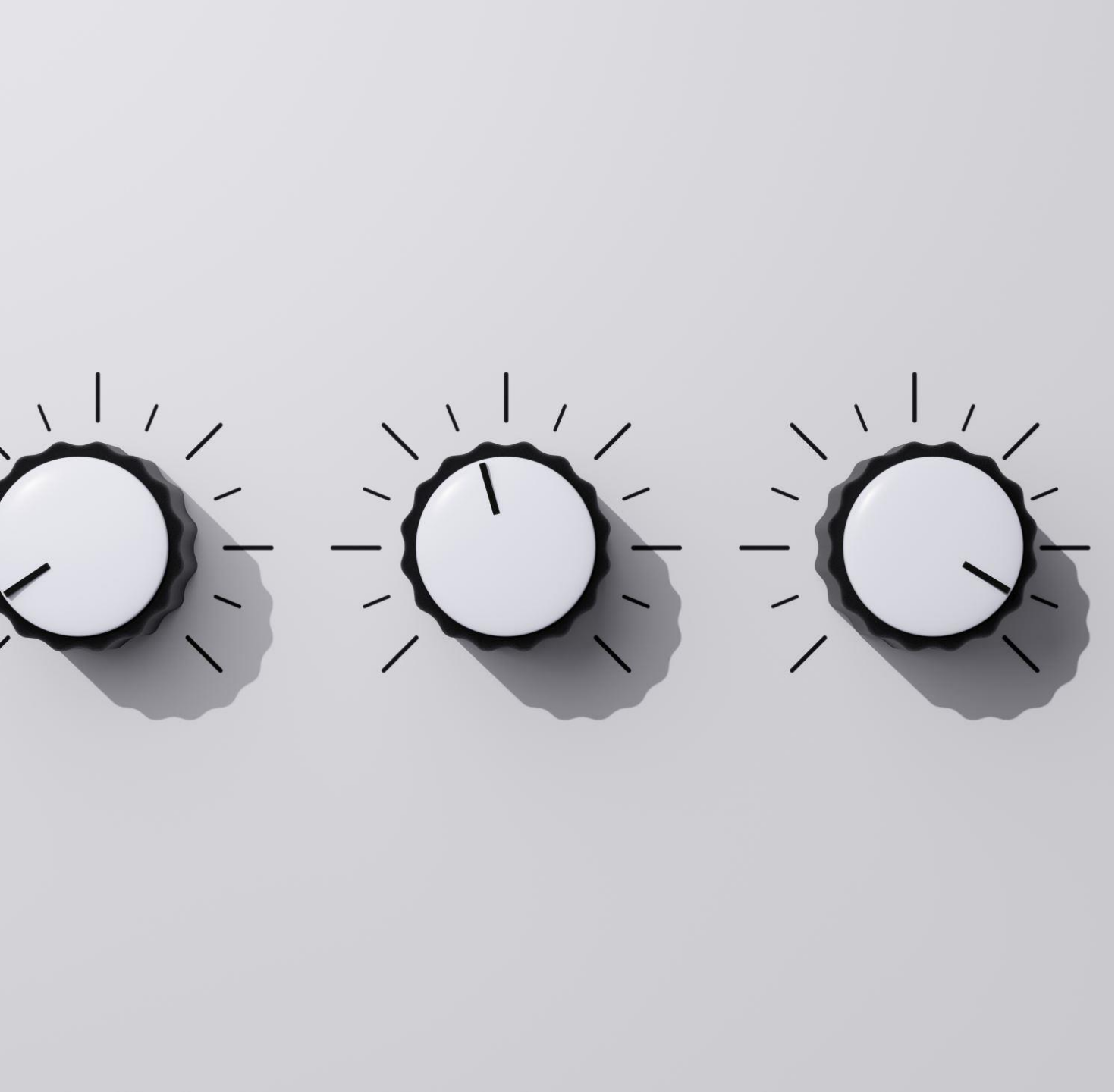
https://arxiv.org/abs/2104.08835

BIG-bench: 204 tasks

https://arxiv.org/abs/2206.04615

natural-instructions: 1616 tasks

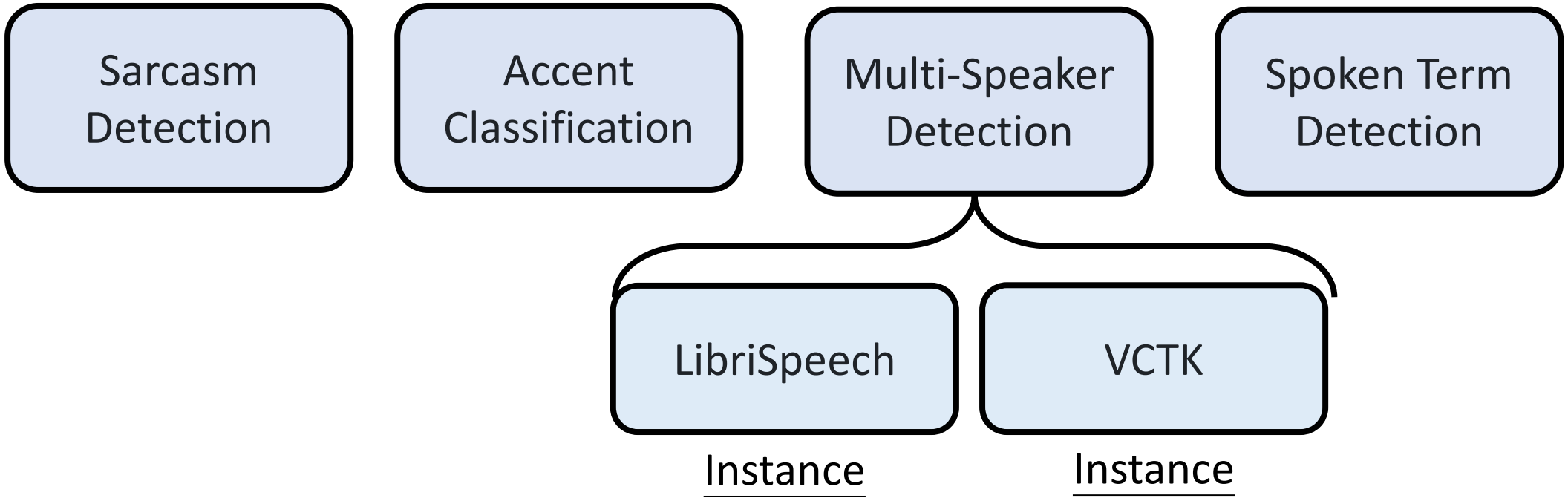
https://arxiv.org/abs/2204.07705



Setting

Terminology

Task: a specific type of processing or operation to be carried out



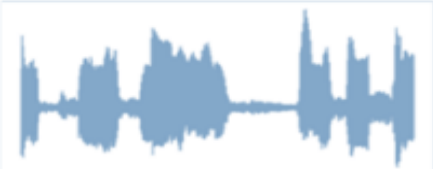
Instance refers to the specific combination of a task and a dataset

Examples in an Instance

Instance

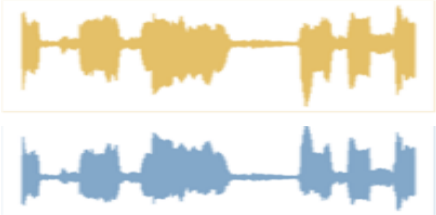
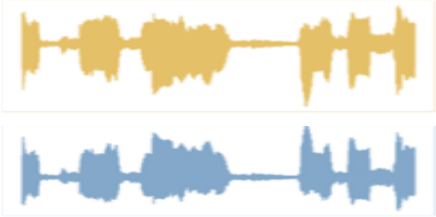
Speaker Count

LibriTTS

Text Instruction	Audio Input	Output
Identify the total number of speakers in the audio. The answer could be one, two, three, four, or five.		one
Determine the number of speakers detected in the audio recording. The answer could be one, two, three, four, or five.		two
Count the distinct voices present in the audio recording. The answer could be one, two, three, four, or five.		three

Different Example Formats

Speaker Verification

Text Instruction	Audio Input	Output
Listen carefully to both audio recordings and judge if they are spoken by the same individual. The answer could be yes or no.		Yes
Do the speech patterns in the two audio recordings come from the same speaker? The answer could be yes or no.		No

Emotion Recognition

Instruction	Input	Output
Please identify the emotion in the audio. The answer could be ……		“Happy”

ASR

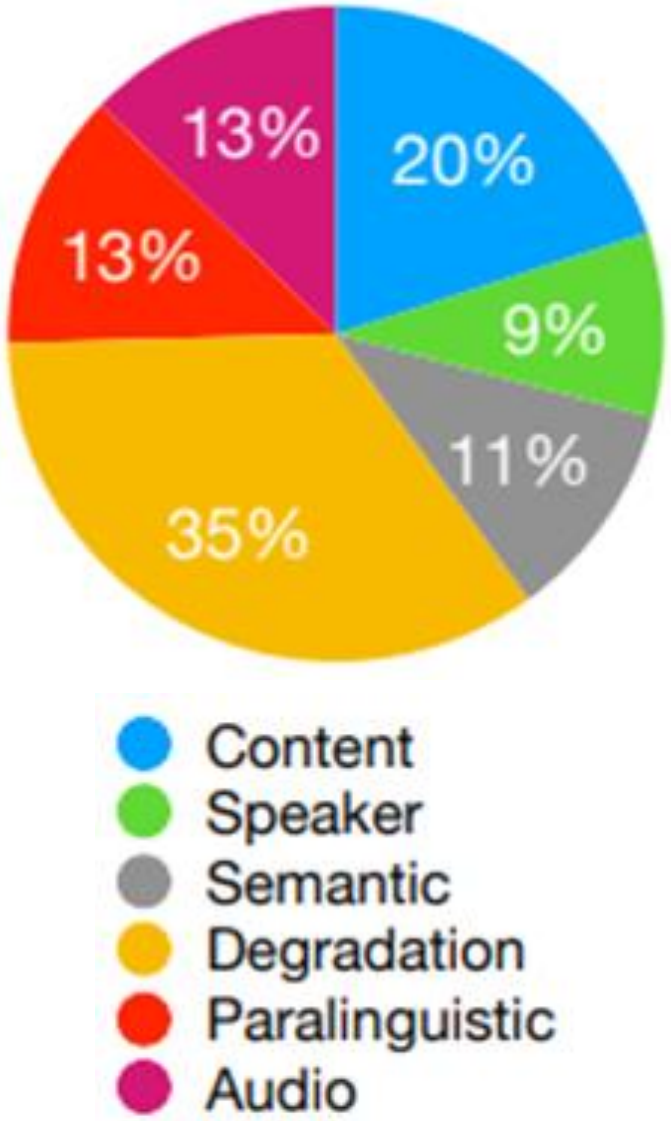
Instruction	Input	Output
Please transcribe the utterance.		“How are you”

Speech Enhancement

Instruction	Input	Output
Please denoise the utterance.		

Current Status

- 55 instances
- Covering 6 dimensions
  - Content: speech command recognition
  - Speaker: speaker verification
  - Semantics: sarcasm detection
  - Degradation: noise SNR prediction
  - Paralinguistic: emotion recognition
  - Audio: environmental sound classification



55 is not a big number …
They are all classification tasks ...

Let’s work together!

Role Model: BIG-bench

204 tasks

444 authors across 132 institutions

https://github.com/google/BIG-bench

BEYOND THE IMITATION GAME: QUANTIFYING AND EXTRAPOLATING THE CAPABILITIES OF LANGUAGE MODELS

Alphabetic author list:*

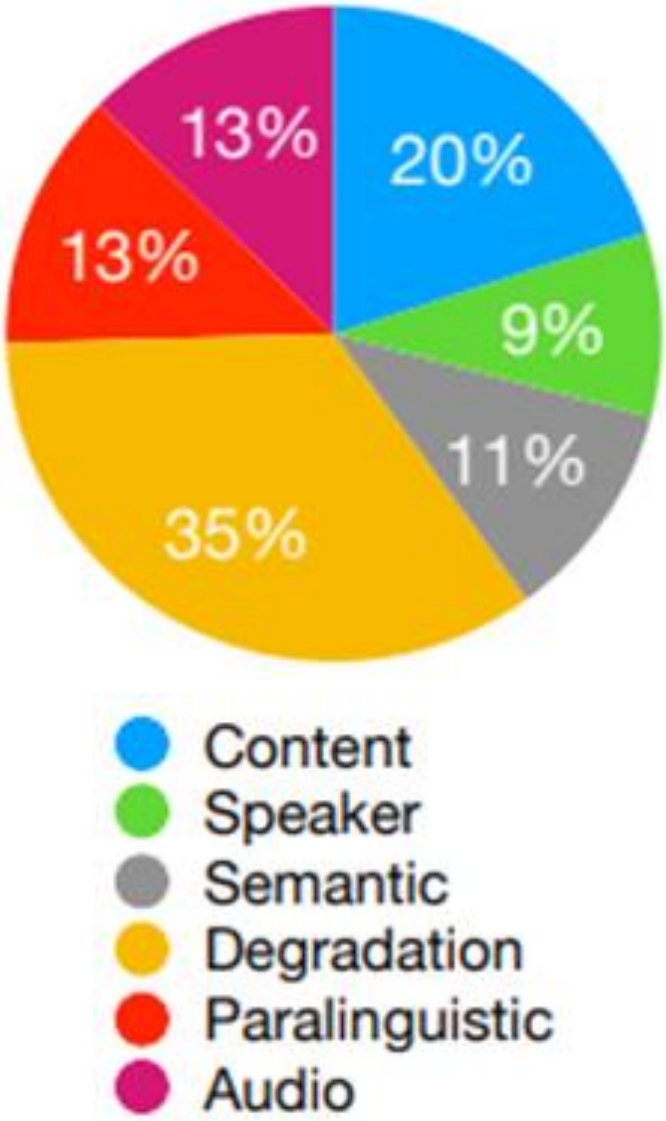
Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W. Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda Askell, Amanda Dsouza, Ambrose Slone, Ameet Rahane, Anantharaman S. Iyer, Anders Andreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew Dai, Andrew La, Andrew Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum, Arul Menezes, Arun Kirubarajan, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut Erdem, Ayla Karakas, B. Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej Bojanowski, Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno Stein, Berk Ekmekci, Bill Yuchen Lin, Blake Howald, Cameron Diao, Cameron Dour, Catherine Stinson, Cedrick Argueta, César Ferri Ramírez, Chandan Singh, Charles Rathkopf, Chenlin Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Chris Waites, Christian Voigt, Christopher D. Manning, Christopher Potts, Cindy Ramirez, Clara E. Rivera, Clemencia Siro, Colin Raffel, Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman, Dan Roth, Daniel Freeman, Daniel Khashabi, Daniel Levy, Daniel Moseguí González, Danielle Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Deniz Yuret, Derek Chen, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho Mollo, Diyi Yang, Dong-Ho Lee, Ekaterina Shutova, Ekin Dogus Cubuk, Elad Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele Rodola, Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan A. Chi, Ethan Dyer, Ethan Jerzak, Ethan Kim, Eunice Engefu Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar, Fernando Martínez-Plumed, Francesca Happé, François Chollet, Frieda Rong, Gaurav Mishra, Genta Indra Winata, Gerard de Melo, Germán Kruszewski, Giambattista Parascandolo, Giorgio Mariani, Gloria Wang, Gonzalo Jaimovitch-López, Gregor Betz, Gury Gur-Ari, Hana Galijasevic, Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar, Henry Shevlin, Hinrich Schütze, Hiromu Yakura, Hongming Zhang, Hugh Mee Wong, Ian Ng, Isaac Noble, Jaap Jumelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon Lee, Jaime Fernández Fisac, James B. Simon, James Koppel, James Zheng, James Zou, Jan Kocoń, Jana Thompson, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason Phang, Jason Wei, Jason Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy Kim, Jeroen Taal, Jesse Engel, Jesujoba Alabi, Jiaching Xu, Jiaming Song, Jillian Tang, Joan Waweru, John Burden, John Miller, John U. Balis, Jonathan Berant, Jörg Frohberg, Jos Rozen, Jose Hernandez-Orallo, Joseph Boudeman, Joseph Jones, Joshua B. Tenenbaum, Joshua S. Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth, Karthik Gopalakrishnan, Katerina Ignatyeva, Katja Markert, Kaustubh D. Dhole, Kevin Gimpel, Kevin Omondi, Kory Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar, Kyle McDonell, Kyle Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin, Lidia Contreras-Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble, Ludwig Schmidt, Luheng He, Luis Oliveros Colón, Luke Metz, Lütfi Kerem Şenel, Maarten Bosma, Maarten Sap, Maartje ter Hoeve, Maheen Farooqi, Manaal Faruqui, Mantas Mazeika, Marco Baturan, Marco Marelli, Marco Maru, Maria Jose Ramírez Quintana, Marie Tolkiehn, Mario Giulianelli, Martha Lewis, Martin Potthast, Matthew L. Leavitt, Matthias Hagen, Mátyás Schubert, Medina Orduna Baitemirova, Melody Arnaud, Melvin McElrath, Michael A. Yee, Michael Cohen, Michael Gu, Michael Ivanitskiy, Michael Starritt, Michael Strube, Michał Swędrowski, Michele Bevilacqua, Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimee Xu, Mirac Suzgun, Mo Tiwari, Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhdeh Gheini, Mukund Varma T, Nanyun Peng, Nathan Chi, Nayeon Lee, Neta Gur-Ari Krakover, Nicholas Cameron, Nicholas Roberts, Nick Doiron, Nikita Nangia, Niklas Deckers, Niklas Muennighoff, Nitish Shirish Keskar, Niveditha S. Iyer, Noah Constant, Noah Fiedel, Nuan Wen, Oliver Zhang, Omar Agha, Omar Elbaghdadi, Omer Levy, Owain Evans, Pablo Antonio Moreno Casares, Parth Doshi, Pascale Fung, Paul Pu Liang, Paul Vicol, Pegah Alipoormolabadi, Peiyuan Liao, Percy Liang, Peter Chang, Peter Eckersley, Phu Mon Htut, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour, Priti Oli, Qiaozhu Mei, Qing Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer Gabriel, Rahel Habacker, Ramón Risco Delgado, Raphael Millière, Rhythm Garg, Richard Barnes, Rif A. Saurous, Riku Arakawa, Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman Sitelew, Ronan LeBras, Rosanne Liu, Rowan Jacobs, Rui Zhang, Ruslan Salakhutdinov, Ryan Chi, Ryan Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif M. Mohammad, Sajant Anand, Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel R. Bowman, Samuel S. Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah A. Rous, Sarik Ghazarian, Sayan Ghosh, Sean Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi, Shadi Hamdan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi, Shixiang Shane Gu, Shubh Pachchigar, Shubham Toshniwal, Shyam Upadhyay, Shyamolima (Shammie) Debnath, Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini, Soo-Hwan Lee, Spencer Torene, Sriharsha Hatwar, Stanislas Dehaene, Stefan Divic, Stefano Ermon, Stella Biderman, Stephanie Lin, Stephen Prasad, Steven T. Piantadosi, Stuart M. Shieber, Summer Misherghi, Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li, Tao Yu, Tariq Ali, Tatsu Hashimoto, Te-Lin Wu, Theo Desbordes, Theodore Rothchild, Thomas Phan, Tianle Wang, Tiberius Nkinyili, Timo Schick, Timofei Kornev, Timothy Telleen-Lawton, Tina Tunkasiri, Tobias Gerstenberg, Trenton Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Ramasesh, Vinay Uday Prabhu, Vishakh Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Yiding Hao, Yifu Chen, Yonatan Belinkov, Ya Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J. Wang, Zirui Wang, Ziyi Wu

Current Status

- 55 tasks created from 33 datasets
- Covering 6 dimensions
  - Content: speech command recognition
  - Speaker: speaker verification
  - Semantics: sarcasm detection
  - Degradation: noise SNR prediction
  - Paralinguistic: emotion recognition
  - Audio: environmental sound classification

Everyone can add new tasks! ➡ **Dynamic**

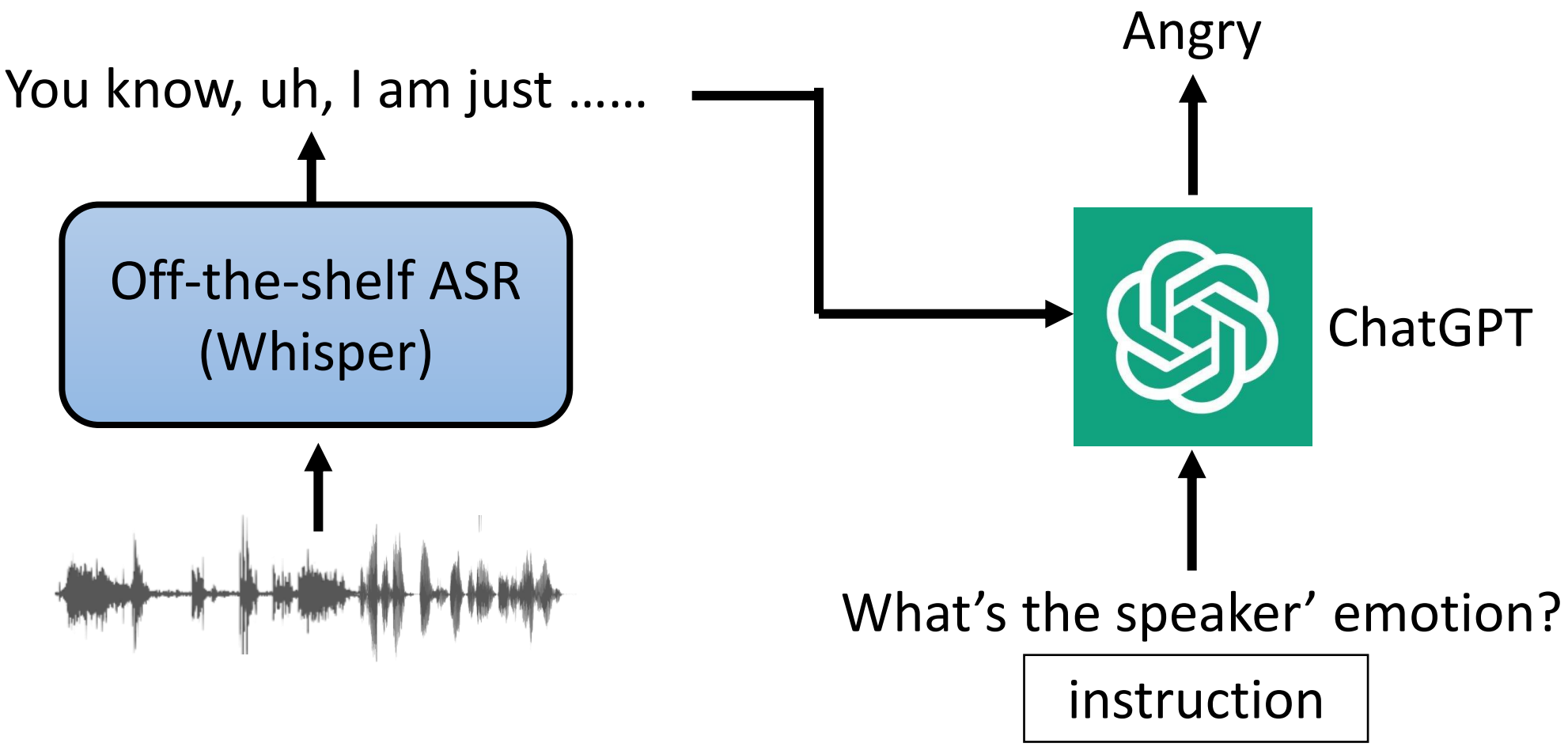
(We can write a big paper together like BIG-bench in the future.)



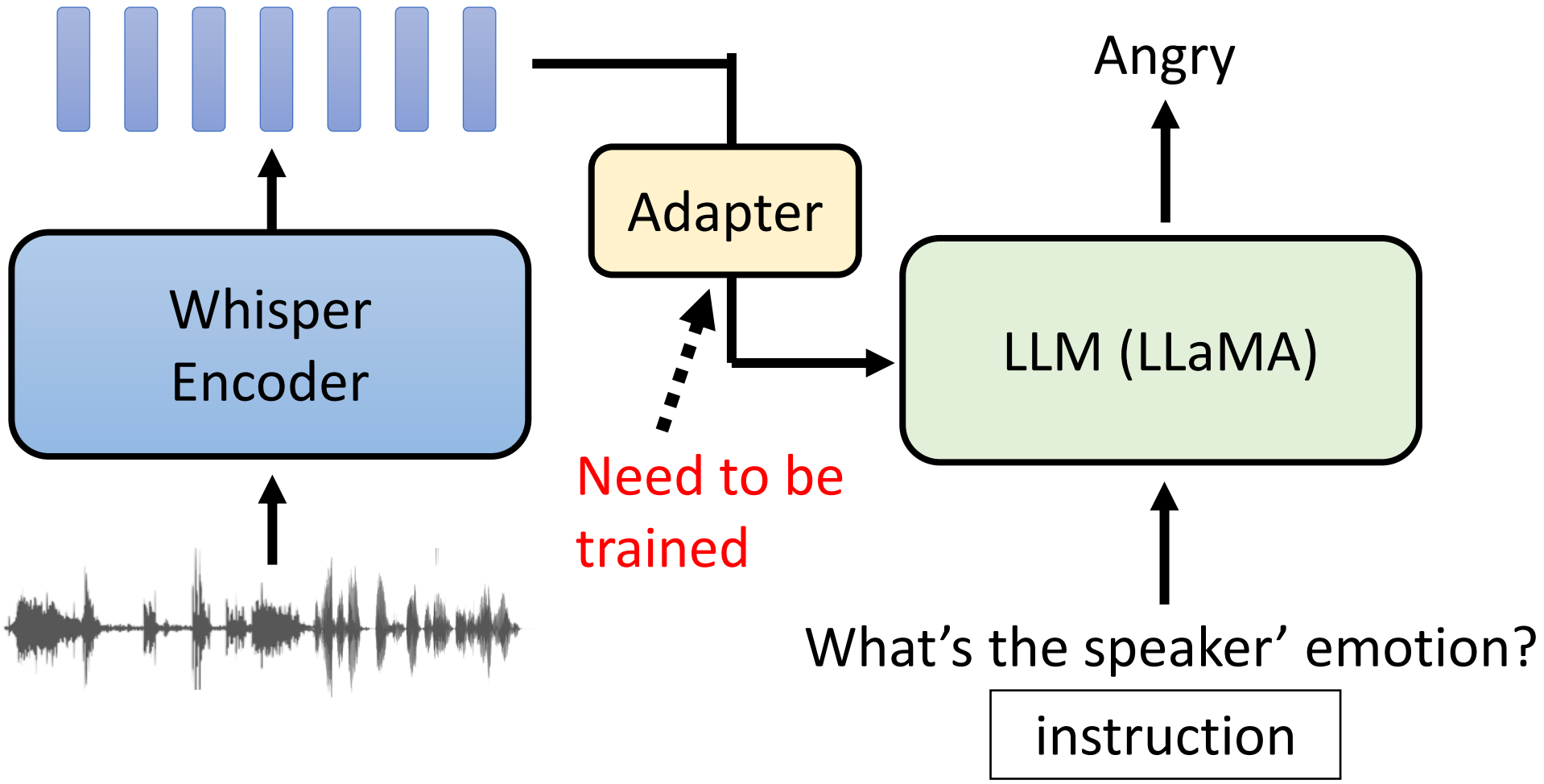
Let's work together!

Baselines

Example Baseline: ASR + ChatGPT

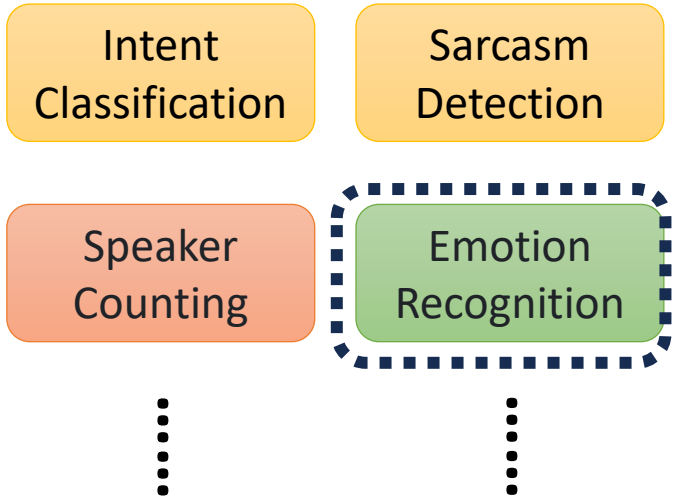


Example Baseline: Whisper Encoder + LLM



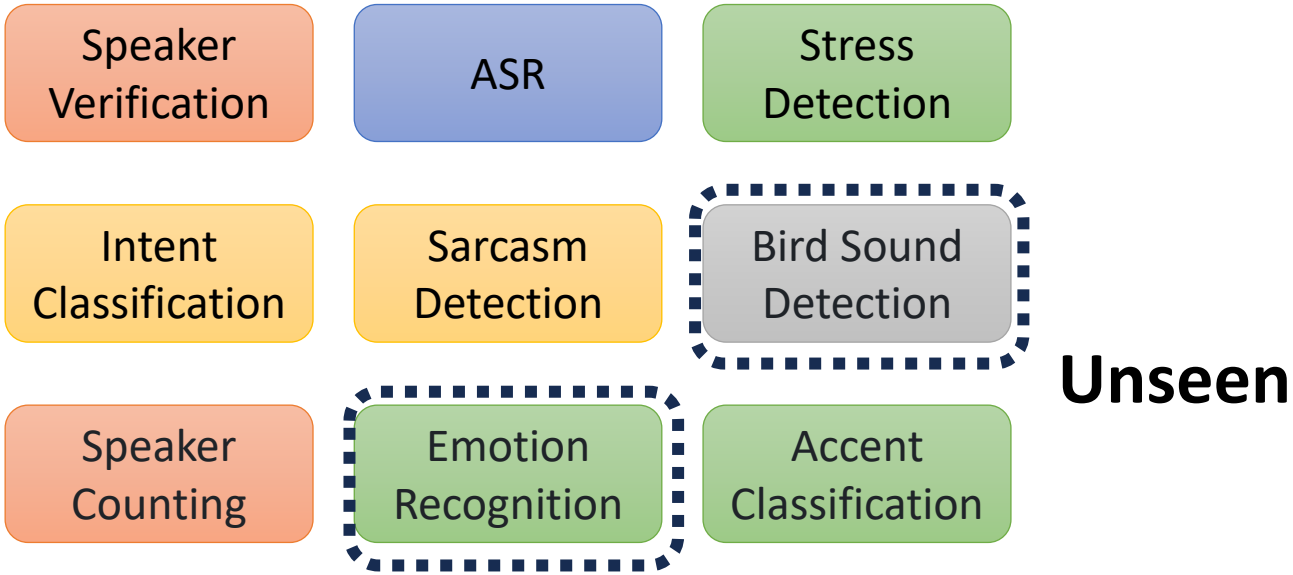
Training Data vs. Testing Data

Training Data (23 instances)



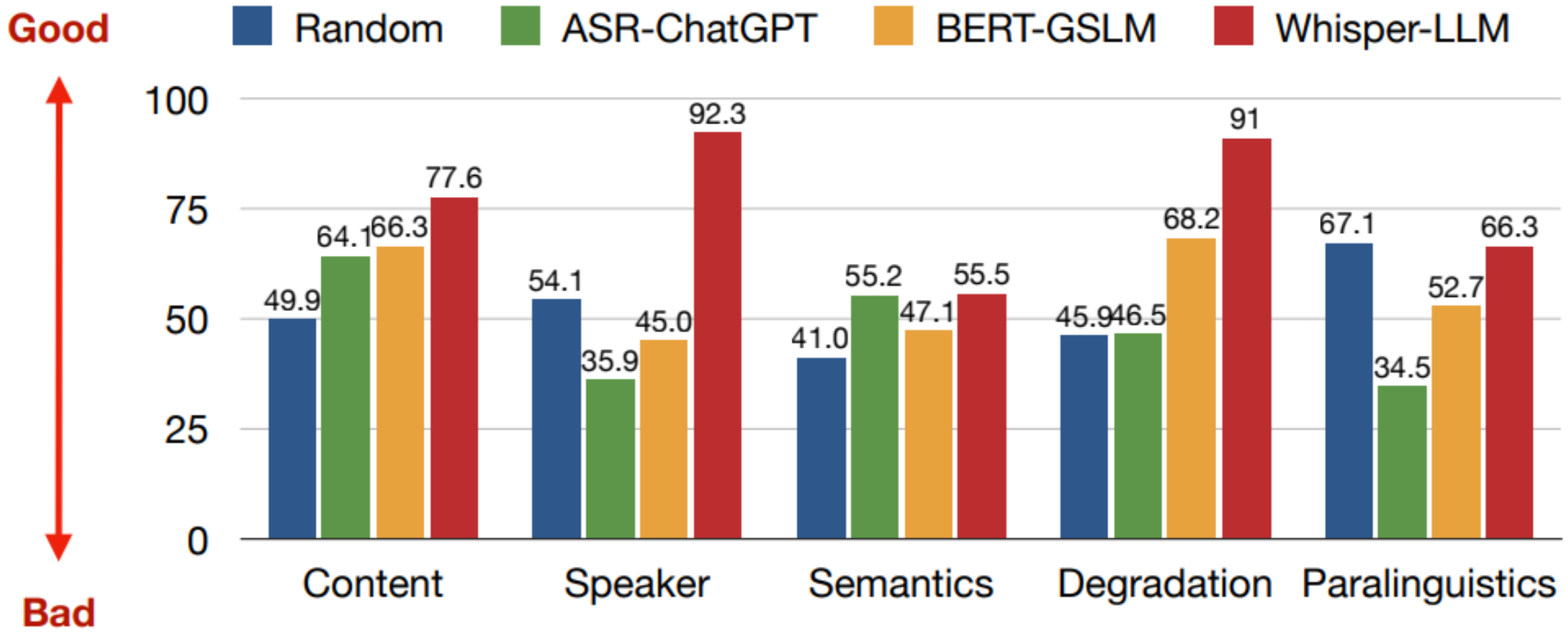
Training model parameters

Dynamic-SUPERB (55 instances)

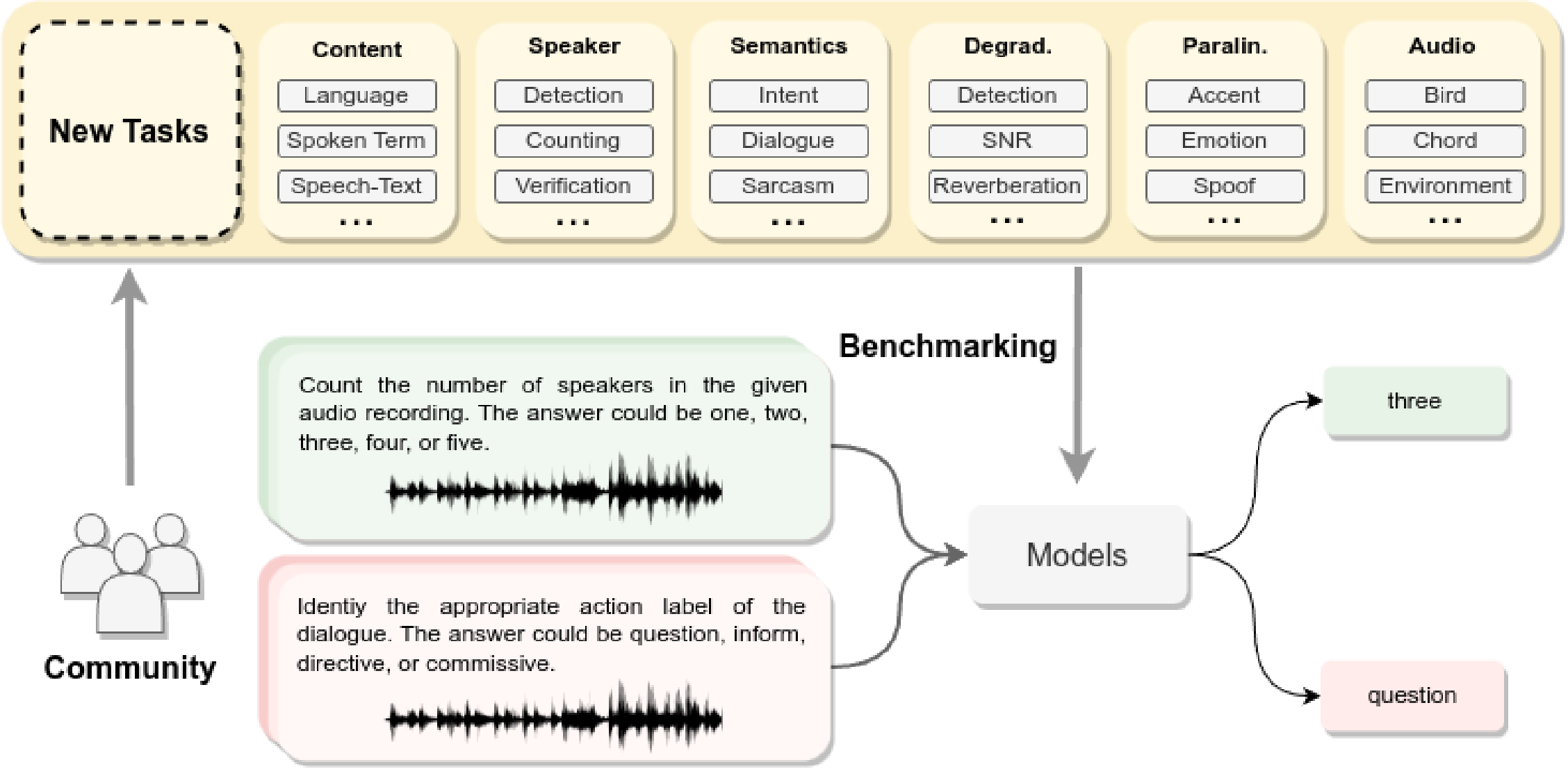


Seen (not the same data)

Overall Results



Dynamic-SUPERB



Action Item

- ASRU 2023 satellite workshop - **Sp**eech foundation models and their performance benchm**arks** (SPARKS)
- Submit a paper to Dynamic-SUPERB session
  - Short paper: 2-3 pages, detailing your methodology, findings, in-depth analysis or new tasks with the Dynamic-SUPERB benchmark.
  - Deadline: **Nov. 24, 2023** (not the workshop deadline)
  - Present at the Dynamic-SUPERB session

https://sites.google.com/g.ntu.edu.tw/sparks