ICASSP '25 notes and interesting posters

Benedikt Kantz

April 22, 2025

1 Montag Vormittag

1.1 Tutorial: Generative AI and Model Optimization

Problem: (compute) cost, current foundation models not sustainable Solutions:

1.1.1 Sparsity

→ scalability, less overfitting, interpretability, adaptive ways to introduce sparsity

- post training: optimal brain damage (OBD)/ optimal brain surgery (OBS)
  - dropout by contribution to error, scale by Hessian $\mathcal{H}$ contribution
- training:
  - L1-loss: Convex optim.; no free lunch: initial model very large!, more eqs.
  - exaustive: very expensive
  - greedy/evolutionary solutions: StOMP, GOMP based on L0-norm, but very effective
- pre-training
  - SET
  - randomly initial init → evolutionary
- architecutral: grow and shrink networks...

Problem: doesn't really work with LMs (empirical study), but well for other networks (esp. low-weight dropout)

1.1.2 Compression

- filter: storage compresion
- low rank factorization (≠ LoRA), during train time not fine-tuning
- knowledge distillation

2 Dienstag Nachmittag

2.1 Talk: Underwater Communications

- Problem: very slow comm underwater, ≈10 kHz range
- Towards moving target, Doppler correction using active SP correction, very manual work

Comment: interesting manual process, tedious work to sample

3 Mittwoch Nachmittag

3.1 Talk: AI+SP

Comment: some basics on diffusion/transformers, a little bit of SP in NNs

4 Donnerstag Vormittag

4.1 Talk: Multiomics

- Genomics: DNA understanding
- Transcriptomics: DNA-¿RNA understanding
- Proteomics: RNA-¿Protein structures
- Knowledge graphs: how do these systems influnce each other
- Flow:
  - identify DNA mutation that triggers illness
  - find possible RNA mechanism
  - find good fitting small ring structure

  → check for side effects in knowledge graph! (certain protein effects unwantend)

  → then test → animal tests, reduce through ML!
- Graph diffusion for drug discovery: noise schedule for diffusion essential, i.e. cosine-square schedule
  - diffuse graphs from atoms & edges as adjacency matrix
  - what is noise: discrete noise: each atom is discrete state $\implies$ graph structure undergoes state transition change
  - naive: uniform structure, not really chemically sensible - conditional probabibilites $\implies$ not uniform but marginal distribution of molecules in training (just logical!), same for edge (with deletion!)
  - one step further: consider carbon rings, restriction based on maximum bonds of atom (freie radikale)
  - SMILE-file, QED: Quantitative Esitmate of Drug likeness (from RDKit)
  - Existing methods: Time-consuming, progress slow, very few good molecules
  - Their work: jointly perturb rings+nodes
  - other approaches: motives as super-node with rings, difficulty: ring attachments - only $\approx$1 % improvement!
  - novelty however high, one molecule of them even patented!
- Knowledge graphs:
  - GNN link prediction
  - none of the existing benchmarks include features!
  - maybe talk to author!

Comment: focused on drug discovery using diffusion, not much on multiomics...

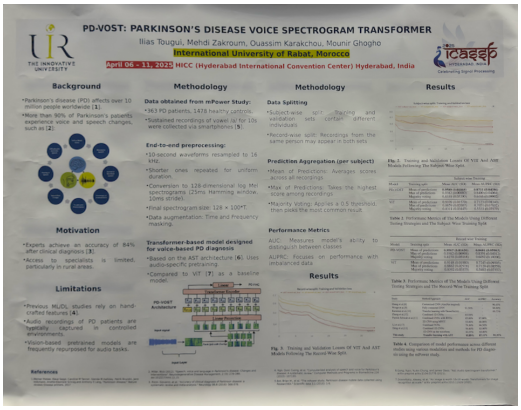
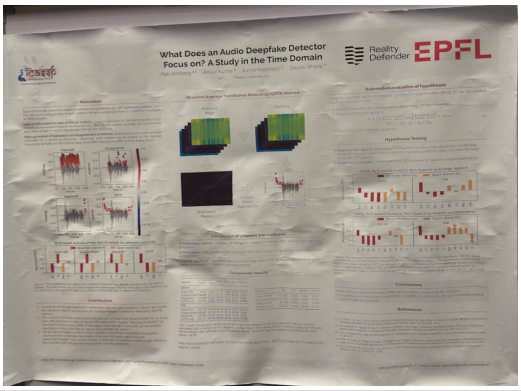
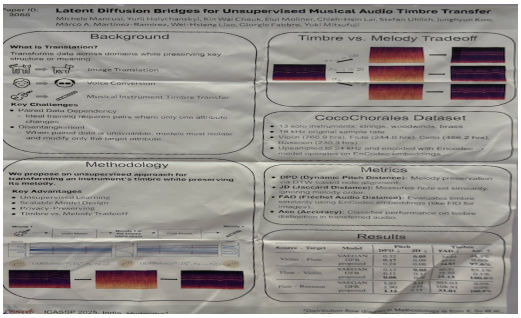
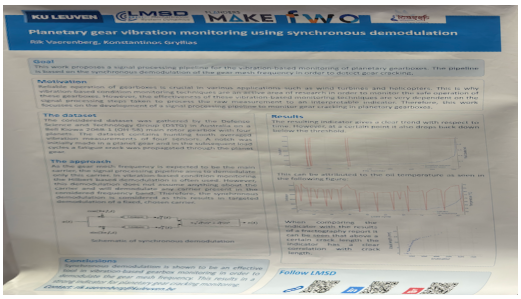
5 Lectures/Orals

Table 1:

Lecture	URL	Notes
Diversity-Seeking Techniques for Red-Teaming LLMs	https://ieeexplore.ieee.org/document/10890844/	Add RL-Loss to train similar to GAN by backpropagating if the model returns very similar output (i.e. discrimimnator)-¿ Very fragile learning; Limited further studies
FDR Control for Complex-Valued Data	https://ieeexplore.ieee.org/document/10889705	similar to LASSO; sparsifying system under certain guarantees
SpectralCam: High-Resolution Low-Cost Spectral Imaging Using DSLR Cameras	https://ieeexplore.ieee.org/document/10887725	Interesting concept of applying photo filter to DSLR sensor, Bayesian pattern restoration "learned" using diffusion & attenuation mtx
Fusing Multimodality of Large Language Models and Satellite Imagery	https://ieeexplore.ieee.org/document/10889624	Could be interesting in combination with HEREDITARY geospatial data once we have access
Controllable Forgetting Mechanism for Few-Shot Class-Incremental Learning	https://arxiv.org/pdf/2501.15998	Using embedding space to classify, add new classifier based on distance, seems rather hyperparameter-sensitive

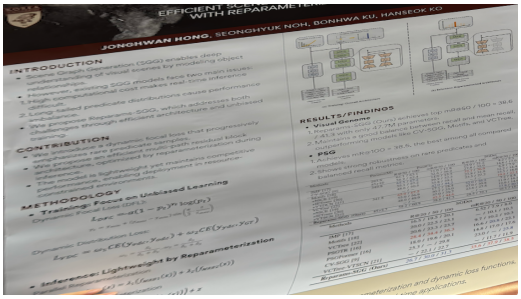
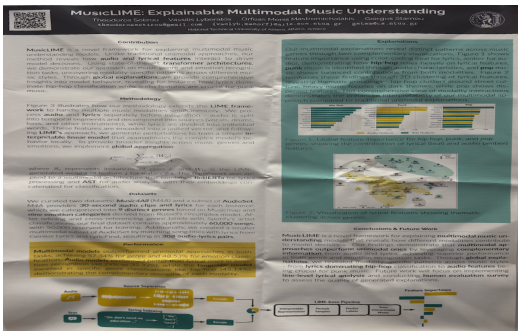
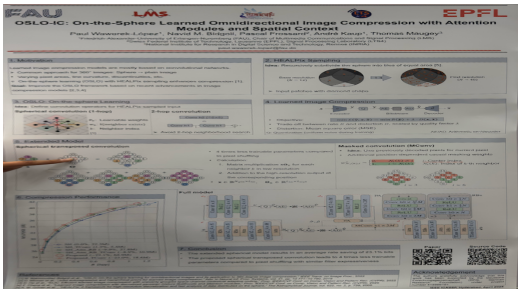
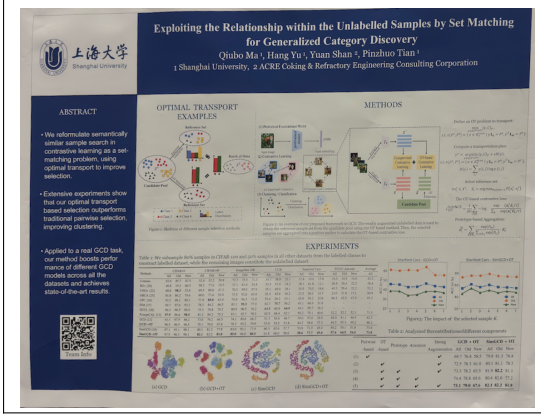
6 Posters

Table 2:

Poster	Information
	**PD-VOST: Parkinson's Disease Voice Spectrogram Transformer** *Ilias Tougui, Mehdi Zakroum, Ouassim Karakchou, Mounir Ghogho* https://ieeexplore.ieee.org/abstract/document/10889820/
	**The EPFL combinational benchmark suite** *pee/ginbergaepf.ch, facia, surya, gauravia roatty defender al* https://infoscience.epfl.ch/entities/publication/309aea67{-}b5a1{-}4532{-}8a6f{-}0a141d8f1ab3/full
	**Latent Diffusion Bridges for Unsupervised Musical Audio Timbre Transfer** *Michele Mancusi, Yurii Halychanskyi, Kin Wai Cheuk, Eloi Moliner, Chieh-Hsin Lai, Stefan Uhlich, Junghyun Koo.* https://ieeexplore.ieee.org/abstract/document/10890708/
	**do a crow endus ported ten band mentorine at pear enoses. The pipeline** *do a crow endus ported ten band mentorine at pear enoses. The pipeline*

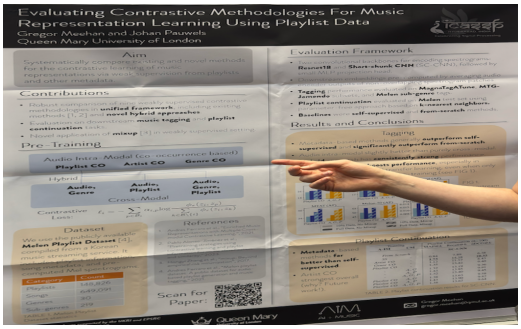
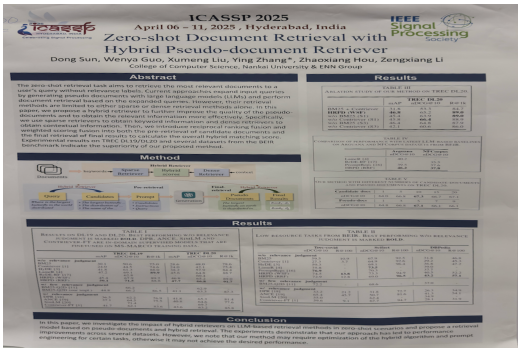
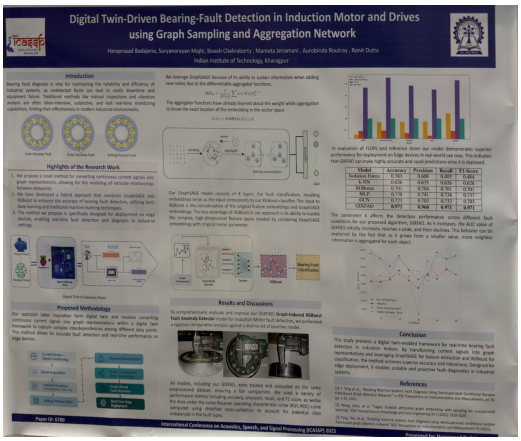
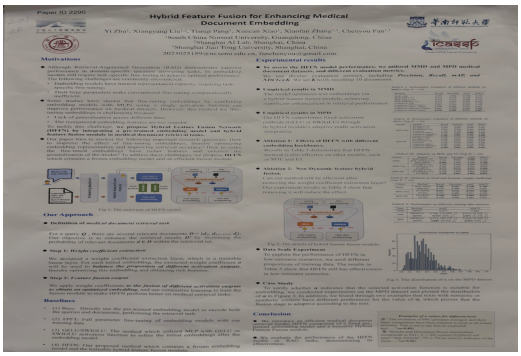
Continued on next page

Table 2: (Continued)

	**EFFICETH REPARAMNE** *outperforming models like CV-SOG, Motifs, and VCTree.*
	**Musiclime: Explainable multimodal music understanding** *tasks, achieving 57.34% for genre and 48.53% for emotion Classi-* https://ieeexplore.ieee.org/abstract/document/10889771/
	**OSLO-IC: On-the-Sphere Learned Omnidirectional Image Compression with Attention Modules and Spatial Context** *Bidgoll, Pascal Frossard?, André Kaup", Thomas Maugey?* https://ieeexplore.ieee.org/abstract/document/10889131/
	**Exploiting the Relationship within the Unlabelled Samples by Set Matching for Generalized Category Discovery** *Qiubo Ma', Hang Yu%, Yuan Shan 3, Pinzhuo Tian 1* https://ieeexplore.ieee.org/abstract/document/10889522/

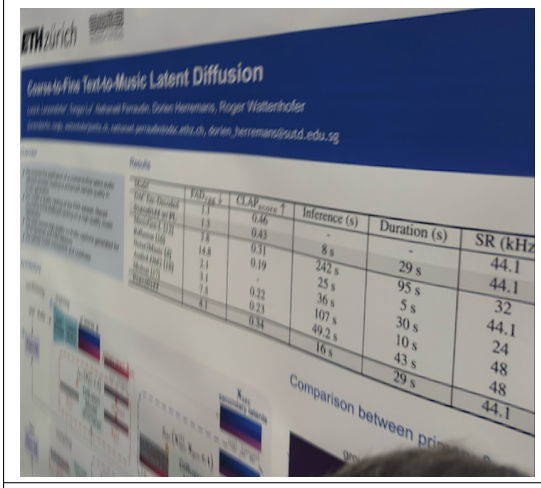
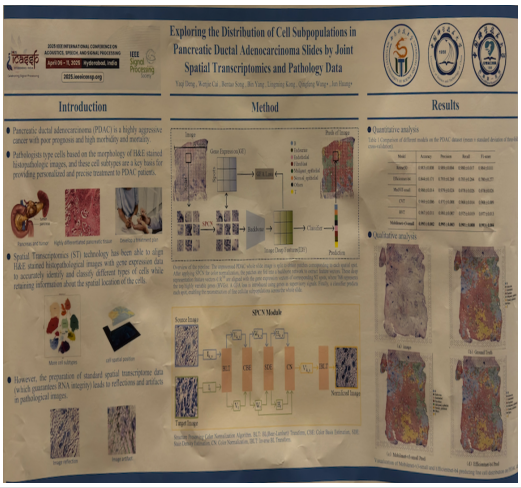
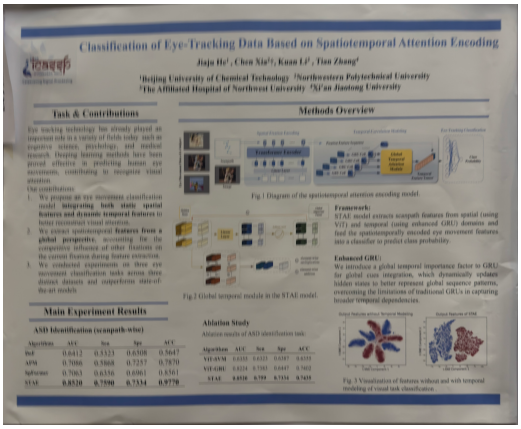
Continued on next page

Table 2: (Continued)

	**Evaluating Contrastive Methodologies for Music Representation Learning Using Playlist Data** *methods [1, 2] and novel hybrid approaches* `https://ieeexplore.ieee.org/abstract/document/10888157/`
	**Fine-tuning and prompt optimization: Two great steps that work better together** *Dong Sun, Wenya Guo, Xumeng Liu, Ying Zhang*, Zhaoxiang Hou, Zengxiang Li* `https://arxiv.org/abs/2407.10930`
	**Digital Twin-Driven Bearing-Fault Detection in Induction Motor and Drives using Graph Sampling and Aggregation Network** *Haraprasad Badajena, Suryanarayan Majhi, Bivash Chakraborty, Mamata Jenamani, Aurobinda Routray, Ronit Dutta* `https://ieeexplore.ieee.org/abstract/document/10889484/`
	**Yi Zhu', Xiangyang Liu!?, Tianqi Pang', Xuncan Xiao!, Xiaofan Zhang33, Chenyou Fan!.*** *Yi Zhu', Xiangyang Liu!?, Tianqi Pang', Xuncan Xiao!, Xiaofan Zhang33, Chenyou Fan!.**

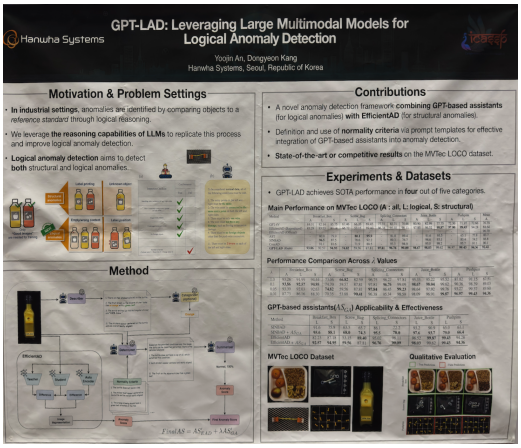
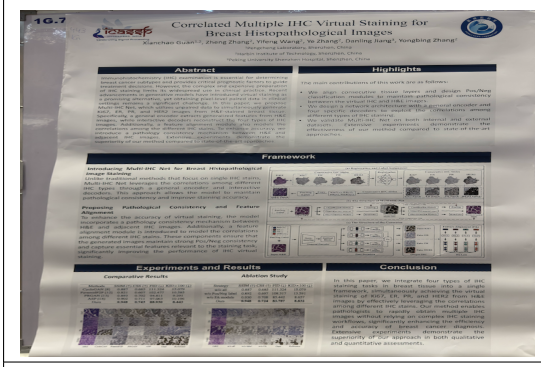
Continued on next page

Table 2: (Continued)

	**Text to music audio generation using latent diffusion model: A re-engineering of audioldm model** *1nh tonne Wegner Neteal Ponesin, Deren Hertemans, Rogger Wattenhofer* `https://www.diva{-}portal.org/smash/record.jsf?pid=diva2:1845150`
	**Exploring the Distribution of Cell Subpopulations in Pancreatic Ductal Adenocarcinoma Slides by Joint Spatial Transcriptomics and Pathology Data** *Yagi Deng, Wenjie Cai, Bentao Song, Bin Yang, Lingming Kong, Qingfeng Wang✉, Jun Huang* `https://ieeexplore.ieee.org/abstract/document/10890640/`
	**Classification of Eye-Tracking Data Based on Spatiotemporal Attention Encoding** *Maju He, Chen Xia't, Kuan L, Tan Zhangt* `https://ieeexplore.ieee.org/abstract/document/10889246/`

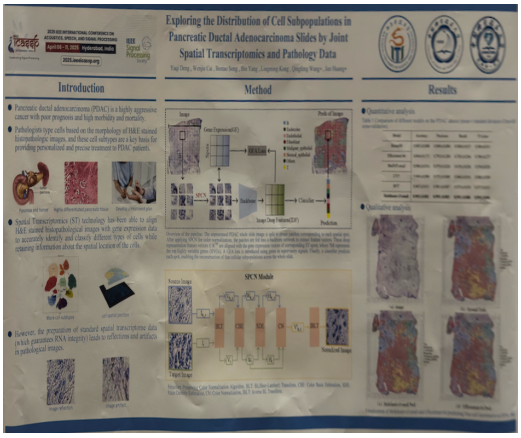
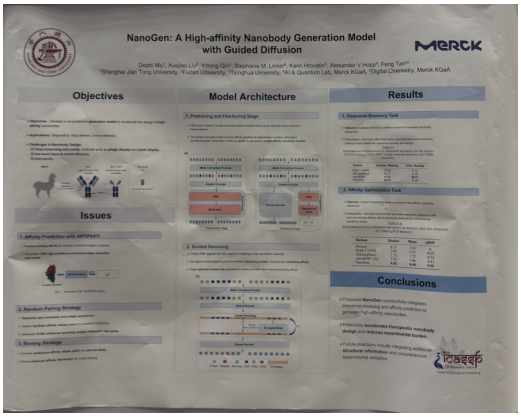
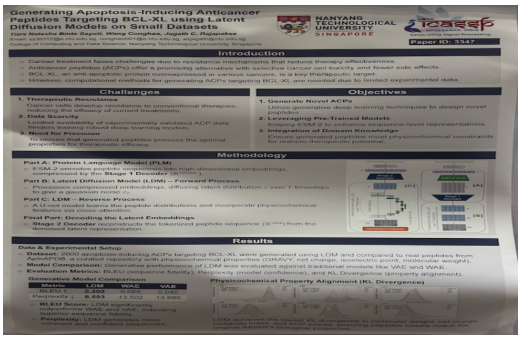
Continued on next page

Table 2: (Continued)

	**GPT-LAD: Leveraging Large Multimodal Models for Logical Anomaly Detection** *Main Performance on MVTec LOCO (A : all, L: logical, S: structural)* `https://ieeexplore.ieee.org/abstract/document/10888757/`
	**Virtual multiplex immunohistochemistry: application on cell block of effusion and aspiration cytology** *Xianchao Guan12, Zheng Zhang?, Yifeng Wang?, Ye Zhang?, Danling Jiang", Yongbing Zhang?* `https://onlinelibrary.wiley.com/doi/abs/10.1002/dc.24344`
	**Vie sin an Aude Spectresun Taedome 981 on the Speechcommanda** *Vie sin an Aude Spectresun Taedome 981 on the Speechcommanda*

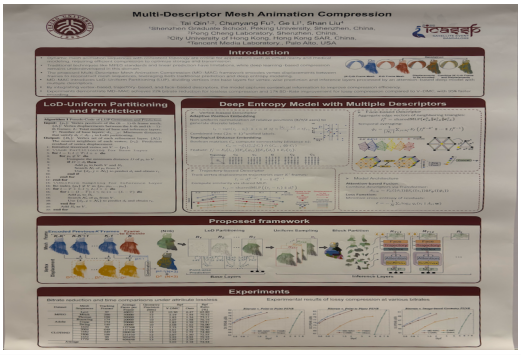
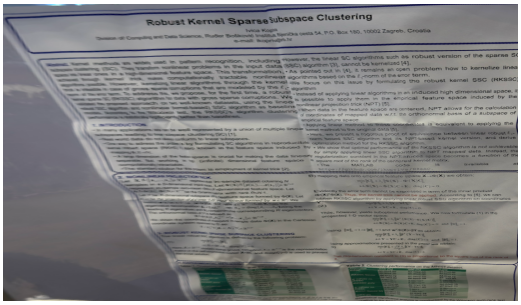
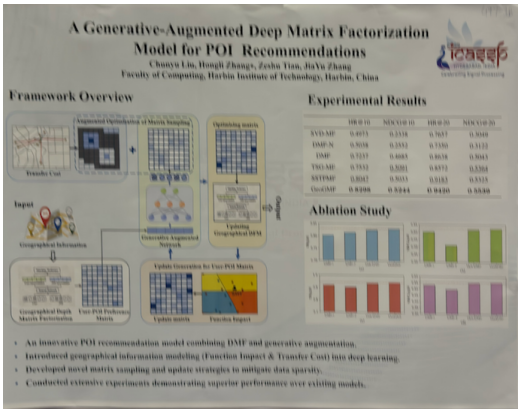
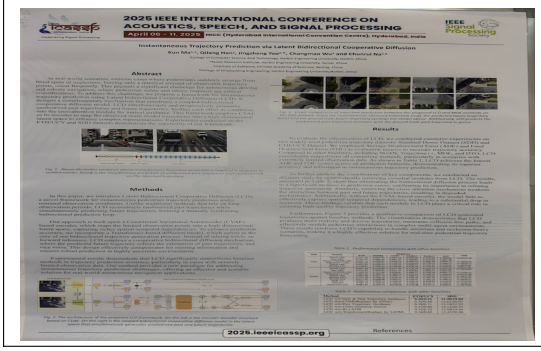
Continued on next page

Table 2: (Continued)

	**Exploring the Distribution of Cell Subpopulations in Pancreatic Ductal Adenocarcinoma Slides by Joint Spatial Transcriptomics and Pathology Data** *Yagi Deng, M cnje Cai, Benao Song, Bin Yang, Limphing Kung , Qedeng Pung» , Jam H* https://ieeexplore.ieee.org/abstract/document/10890640/
	**NanoGen: A High-affinity Nanobody Generation Model with Guided Diffusion** *Dezhi Wu', Xuejiao Liu?, Yiming Qin", Stephanie M. Linker', Karin Hrovatin", Alexander V.Hopp*, Feng Tan*** https://ieeexplore.ieee.org/abstract/document/10888039/
	**ApinAPDE, a curated repository with physicochemical proporties (GRAVY, net charge, isoelectric point, molecular weight).** *Emas: 2230112006.M.00u.59. cong na00012e.ntu.edu.sg. as.293th@ntu.edu.5g*
	**Toward robust early detection of alzheimer's disease via an integrated multimodal learning approach** *Yifei Chen, Shenghao Zhu. Zhaojie Fang. Chang Liu, Binfeng Zou, Linwei Qiu, Yuhe Wang. Shuo Chang. Fan Jia, Felwel Qin". Jin Fang. Yong Peng, Changmiao Wang* https://ieeexplore.ieee.org/abstract/document/10888363/

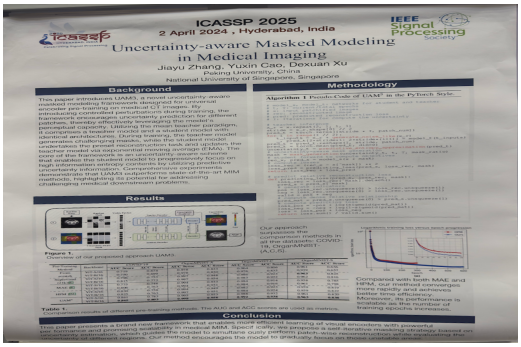
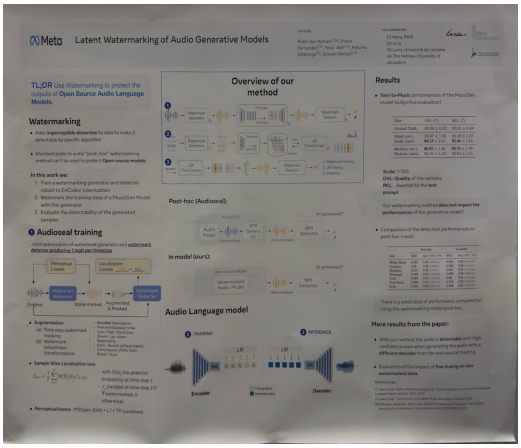
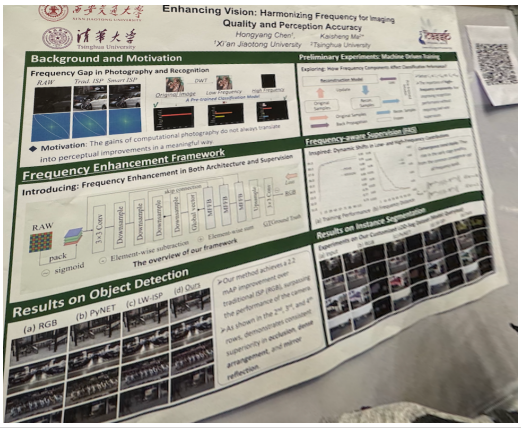
Continued on next page

Table 2: (Continued)

	**Experiments demonstrato MD-MAC achieves 22% bitrate reduction for lossless compression and 17% BD-Rate improvement for lossy compression compared to V-DMC, with 95% faster** *1Shenzhen Graduate School, Peking University, Shenzhen, China,*
	**Robust Kernel Sparse Subspace Clustering** *da cesta 54, P.O. Box 180, 10002 Zagrob, Croatia* `https://ieeexplore.ieee.org/abstract/document/10888170/`
	**A Generative-Augmented Deep Matrix Factorization Model for POI Recommendations** *Chonse Lio, Hongl Zhange, Aesta Man, Masa Zhang* `https://ieeexplore.ieee.org/abstract/document/10890082/`
	**2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)** *Kun Ma12, Qilong Han', Jingzheng Yao?.*, Changmao Wu' and Chunrul NazA* `https://scholar.archive.org/work/utteultwdberxnhwxuaxknu6be/access/wayback/http://ieeexplore.ieee.org/ielx5/79/34703/01657809.pdf?tp=&arnumber=1657809&isnumber=34703`

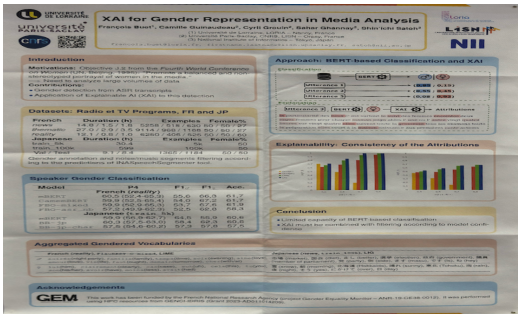
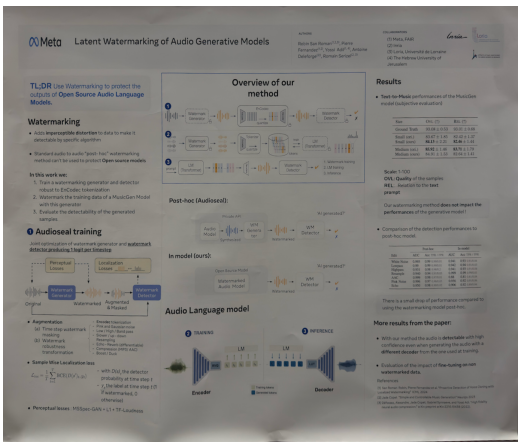
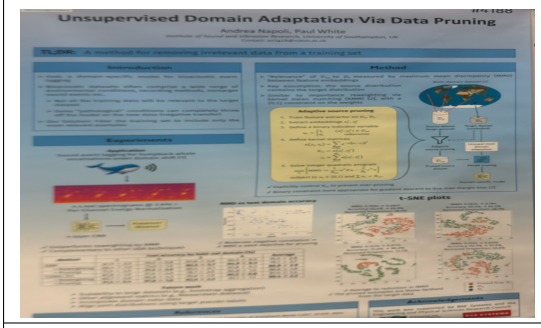
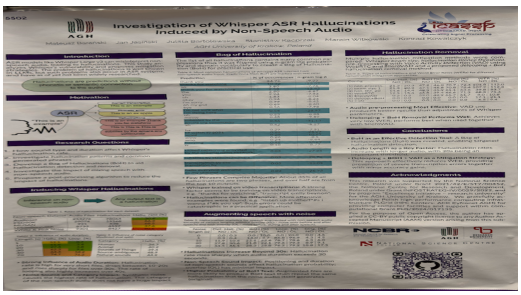
Continued on next page

Table 2: (Continued)

	**Masked image modeling advances 3d medical image analysis** *Der formance and promising scalability in medical MIM. Spect ically, we propose a telliterative masking stratcoy based on* https://openaccess.thecvf.com/content/WACV2023/html/Chen_Masked_Image_Modeling_Advances_3D_Medical_Image_Analysis_WACV_2023_paper.html
	**Latent watermarking of audio generative models** *Losses .....*Inlt..* https://ieeexplore.ieee.org/abstract/document/10889782/
	**HYMAN: Hybrid Memory and Attention Network for Unsupervised Anomaly Detection** *Jiahao Li, Yigiang Chen, Yunbing Xing, Yang Gu, Xiangyuan Lan* https://ieeexplore.ieee.org/abstract/document/10890028/
	**A general framework for object detection** *›As shown in the 29, 3P%, and 4˚* https://ieeexplore.ieee.org/abstract/document/710772/

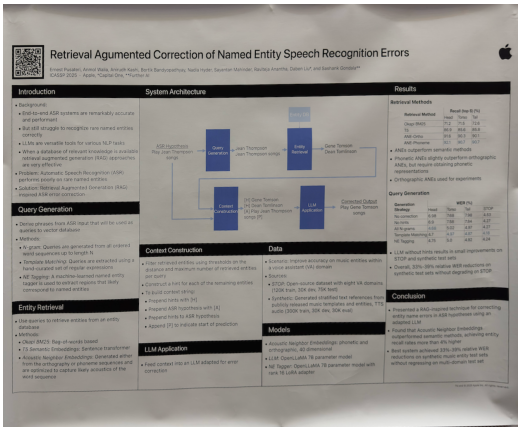
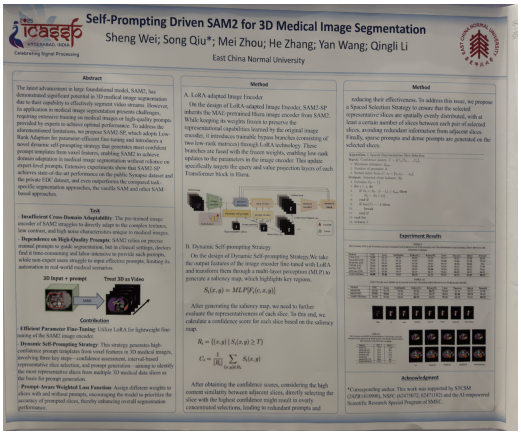
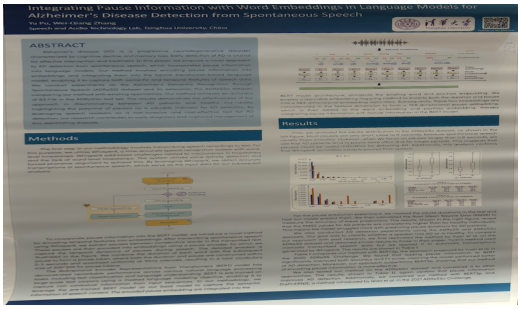
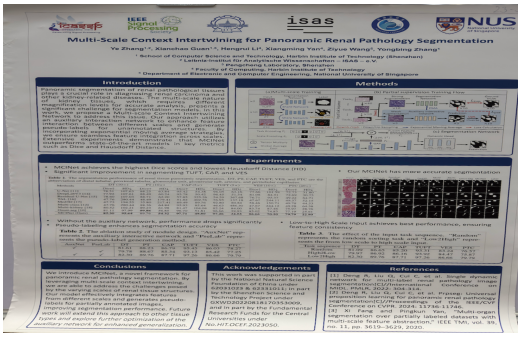
Continued on next page

Table 2: (Continued)

	**XD0h 18122089roject Gender Equalty Monitor - ANIR-19-GE30-0012). It was portormed** *Françols Buot', Camille Guinaudeau?, Cyril Grouin?, Sahar Ghannay?, Shin'ichi Satoh»*
	**Latent watermarking of audio generative models** *13) Olfonses, Alexandre, Jade Copet, Gabriel Synarve, and Yest Adt. "High fidelity* `https://ieeexplore.ieee.org/abstract/document/10889782/`
	**Unsupervised Domain Adaptation Via Data Pruning** *TWDR. & method for removing irrelevant data from a training set* `https://ieeexplore.ieee.org/abstract/document/10890190/`
	**Investigation of Whisper ASR Hallucinations Induced by Non-Speech Audio** *AGH University of Krakow. Poland* `https://arxiv.org/abs/2501.11378`

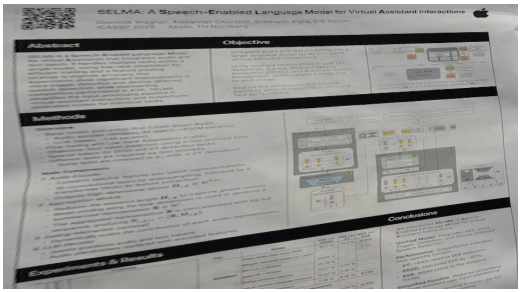
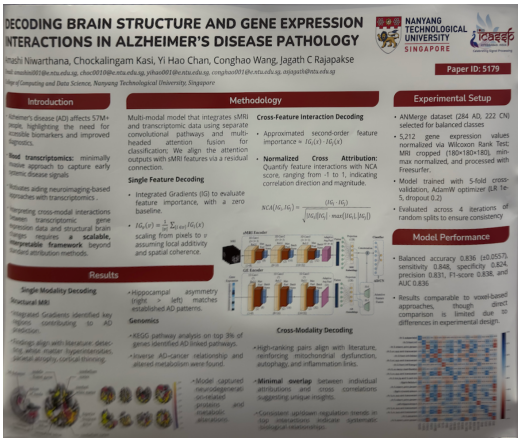
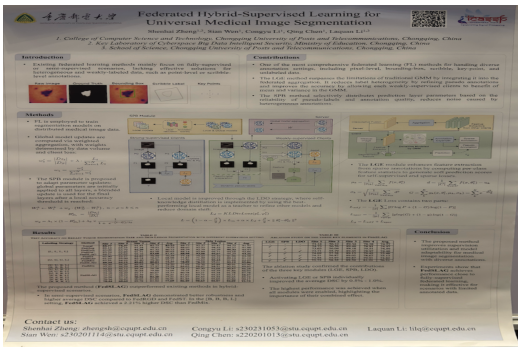
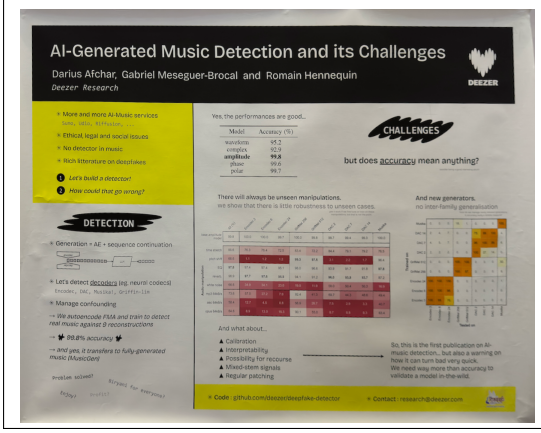
Continued on next page

Table 2: (Continued)

	**Retrieval augmented correction of named entity speech recognition errors** *Ernest Pusateri, Anmol Walia, Aniruch Kashi, Bortik Bandyopadhyay, Nadia Hyder, Sayantan Mahinder, Raviteja Anantha, Daben Liu*, and Sashank Gondala*** `https://ieeexplore.ieee.org/abstract/document/10888936/`
	**Self-Prompting Driven SAM2 for 3D Medical Image Segmentation** *a Sorted index from C: s, -(6r,62..., 62...,4)* `https://ieeexplore.ieee.org/abstract/document/10889344/`
	**Integrating Pause Information with Word Embeddings in Language Models for Alzheimer's Disease Detection from Spontaneous Speech** *Speech and Audio Technology lab, Tsinghua University, China* `https://arxiv.org/abs/2501.06727`
	**Multi-scale Context Intertwining for Panoramic Renal Pathology Segmentation** *Ye Zhang'2, Xlanchao Guan13, Hengrul LI", Xiangming Yan", Ziyue Wang", Yongbing Zhang'* `https://ieeexplore.ieee.org/abstract/document/10889659/`

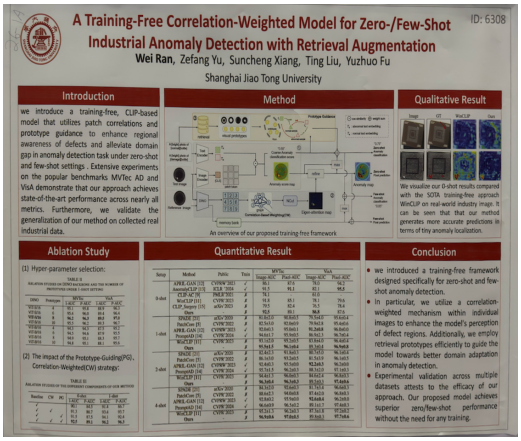
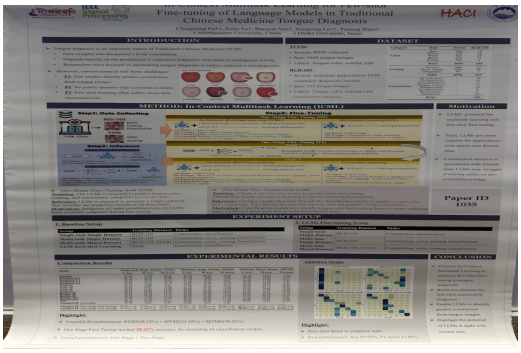
Continued on next page

Table 2: (Continued)

	**SELMA: A Speech-Enabled Language Model for Virtual Assistant Interactions** *Simplified Pigstina, Meduces complarty* https://ieeexplore.ieee.org/abstract/document/10890139/
	**Glial-neuronal interactions in Alzheimer's disease: the potential role of a 'cytokine cycle'in disease progression** *hutcmoshini001@Bentuedusg, choc0010@e.ntu.edusg,yihao001@e.ntu.edusg, conghao001@e.ntu.edusg,asjagath@ntu.edu.sg* https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1750{-}3639.1998.tb00136.x
	**Wireless Sensor Networks: 13th China Conference, CWSN 2019, Chongqing, China, October 12–14, 2019, Revised Selected Papers** *1. College of Computer Science and Technology, Chongging University of Posts and Telecommunications, Chongging. China* https://books.google.com/books?hl=en&lr=&id=o3fADwAAQBAJ&oi=fnd&pg=PR6&dq=1.+College+of+Computer+Science+and+Technology,+Chongging+University+of+Posts+and+Telecommunications,+Chongging.+China&ots=xK3vgQuzGb&sig=1RlJj8oR{-}X4PqqWdWiKUW8crgKE
	**AI-Generated Music Detection and its Challenges** *Suno, Udio, Riffusion, ...* https://arxiv.org/abs/2501.10111

Continued on next page

Table 2: (Continued)

	**A Training-Free Correlation-Weighted Model for Zero-/Few-Shot Industrial Anomaly Detection with Retrieval Augmentation** *Wei Ran, Zefang Yu, Suncheng Xiang, Ting Liu, Yuzhuo Fu* `https://ieeexplore.ieee.org/abstract/document/10890083/`
	**Parameter-efficient fine-tuning of large-scale pre-trained language models** *Changzeng Fits, Zelin Fut, Shaojun Yant. Xiaoyong Lyvt, Yuliang Zhaot* `https://www.nature.com/articles/s42256{-}023{-}00626{-}4`