

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.
 - exhaustive: very expensive
 - greedy/evolutionary solutions: StOMP, GOMP based on L0-norm, but very effective
- pre-training
 - SET
 - randomly initial init → evolutionary
- architectural: 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 compression
- low rank factorization (\neq 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/transfomers, 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 unwanted)
 - 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 \Rightarrow graph structure undergoes state transition change
 - naive: uniform structure, not really chemically sensible - conditional probabilities \Rightarrow 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 Estimate 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. discriminator)-; 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																																				
<p>PD-VOST: PARKINSON'S DISEASE VOICE SPECTROGRAM TRANSFORMER Ilias Tougui, Mehdi Zakroum, Ouassim Karakchou, Mounir Ghogho International University of Rabat, Morocco April 06 - 11, 2023 HICC (Hyderabad International Convention Center) Hyderabad, India</p> <p>Background • Parkinson's disease (PD) affects over 10 million people worldwide [1] • PD patients often experience voice and speech changes such as [2]:</p> <ul style="list-style-type: none"> - End-to-end processing: >10 seconds waveform recorded for entire duration. - Convolutional layers to 128-dimensional Mel-spectrograms (25ms Hamming window). - Final spectrogram size: 128 x 1007. - Data augmentation: Time and frequency warping. <p>Motivation • Experts achieve an accuracy of 84% after clinical diagnosis [3]. • Access to specialists is limited.</p> <p>Limitations • Previous ML studies rely on hand-crafted features.</p> <p>Methodology • Data obtained from mPower Study: 1963 PD patients, 1478 healthy controls. • Standardized recordings of voice (n = 10s)</p> <p>Methodology • Record-wise split: Recordings from the same person may appear in both sets.</p> <p>Results • Model-wise split: Recordings from different speakers.</p> <p>Prediction Aggregation (per subject) • AUC: Measures model's ability to distinguish between classes across all recordings.</p> <p>Performance Metrics • AUC: Measures model's ability to distinguish between classes across all recordings.</p> <p>Results • AUC: Measures model's ability to distinguish between classes across all recordings.</p> <p>Conclusion • Project Summary: - Baseline model: Transformer (A) uses speaker-specific pretraining. - Compared to VIT [7] as a baseline. - Transformer-based model designed for voice-based PD diagnosis.</p> <p>Fig. 1: Training and Validation Losses of ViT and AST Model. The AST Model has better performance than ViT.</p> <p>Fig. 2: Confusion Matrix of The Model Using Different Testing Strategies. The model performs well.</p> <p>Fig. 3: Performance Metrics of The Model Using Different Testing Strategies. The model performs well.</p> <p>Fig. 4: Comparison of model performance according to different testing strategies. The model performs well.</p>	<p>PD-VOST: Parkinson's Disease Voice Spectrogram Transformer <i>Ilias Tougui, Mehdi Zakroum, Ouassim Karakchou, Mounir Ghogho</i> https://ieeexplore.ieee.org/abstract/document/10889820/</p>																																				
<p>What Does an Audio Deepfake Detector Focus on? A Study in the Time Domain Gaurav Rathi, Surya, Gauravia, Roatty, Reality Defender, EPFL</p> <p>Background • Deepfakes have become a major concern in recent years due to their ability to manipulate audio and video content.</p> <p>Methodology • Deepfakes are generated by applying a transformation to the source audio or video.</p> <p>Results • Deepfakes are generated by applying a transformation to the source audio or video.</p> <p>Conclusion • Deepfakes are generated by applying a transformation to the source audio or video.</p>	<p>The EPFL combinational benchmark suite <i>pee/ginbergaefp.ch, facia, surya, gauravia roatty reality defender al</i> https://infoscience.epfl.ch/entities/publication/309aea67{-}b5a1{-}4532{-}8a6f{-}0a141d8f1ab3/full</p>																																				
<p>Latent Diffusion Bridges for Unsupervised Musical Audio Timbre Transfer Michele Manucusi, Yurii Halychanskyi, Kin Wai Cheuk, Eloi Moliner, Chieh-Hsin Lai, Stefan Uhlich, Junghyun Koo</p> <p>Background • What is Transferring? • How to transfer musical audio from one domain while preserving key structure or meaning?</p> <p>Methodology • We propose an unsupervised approach for latent diffusion bridges for timbre transfer.</p> <p>Results • Metrics</p> <table border="1"> <thead> <tr> <th>Acoustic Target</th> <th>Model</th> <th>DPD</th> <th>JD</th> <th>FAO</th> <th>Auc</th> </tr> </thead> <tbody> <tr> <td>Violin</td> <td>WAVGAN</td> <td>0.61</td> <td>0.60</td> <td>0.60</td> <td>0.61</td> </tr> <tr> <td>Flute</td> <td>WAVGAN</td> <td>0.61</td> <td>0.60</td> <td>0.60</td> <td>0.61</td> </tr> <tr> <td>Drum</td> <td>WAVGAN</td> <td>0.17</td> <td>0.16</td> <td>0.16</td> <td>0.17</td> </tr> <tr> <td>Piano</td> <td>WAVGAN</td> <td>1.00</td> <td>0.99</td> <td>0.99</td> <td>1.00</td> </tr> <tr> <td>Bassoon</td> <td>WAVGAN</td> <td>1.00</td> <td>0.99</td> <td>0.99</td> <td>1.00</td> </tr> </tbody> </table>	Acoustic Target	Model	DPD	JD	FAO	Auc	Violin	WAVGAN	0.61	0.60	0.60	0.61	Flute	WAVGAN	0.61	0.60	0.60	0.61	Drum	WAVGAN	0.17	0.16	0.16	0.17	Piano	WAVGAN	1.00	0.99	0.99	1.00	Bassoon	WAVGAN	1.00	0.99	0.99	1.00	<p>Latent Diffusion Bridges for Unsupervised Musical Audio Timbre Transfer <i>Michele Manucusi, Yurii Halychanskyi, Kin Wai Cheuk, Eloi Moliner, Chieh-Hsin Lai, Stefan Uhlich, Junghyun Koo</i> https://ieeexplore.ieee.org/abstract/document/10890708/</p>
Acoustic Target	Model	DPD	JD	FAO	Auc																																
Violin	WAVGAN	0.61	0.60	0.60	0.61																																
Flute	WAVGAN	0.61	0.60	0.60	0.61																																
Drum	WAVGAN	0.17	0.16	0.16	0.17																																
Piano	WAVGAN	1.00	0.99	0.99	1.00																																
Bassoon	WAVGAN	1.00	0.99	0.99	1.00																																
<p>Planetary gear vibration monitoring using synchronous demodulation KU Leuven, CLMSD, MAKE IWC, RIK Visserberg, Konstantinos Gryffas</p> <p>Background • This work proposes a signal processing pipeline for the vibration-based monitoring of planetary gearboxes. The pipeline is able to detect the presence of gearboxes in various applications such as wind turbines and helicopters. This is achieved by the use of a deep learning model that is trained on a dataset of planetary gearbox vibration signals.</p> <p>Methodology • The dataset consists of vibration signals from various sources, including wind turbines and helicopters. The signals are processed using a deep learning model to identify the presence of planetary gearboxes.</p> <p>Results • The resulting indicator gives a clear trend with respect to the ambient temperature. This trend is also shown in other deep learning models.</p> <p>Conclusion • The proposed pipeline is shown to be an effective method for monitoring planetary gearboxes. The use of a deep learning model allows for the detection of planetary gearboxes in various applications, such as wind turbines and helicopters. The use of a deep learning model also allows for the detection of planetary gearboxes in various applications, such as wind turbines and helicopters.</p>	<p>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</p>																																				

Continued on next page

Table 2: (Continued)

	<p>EFFICETH REPARAMNE outperforming models like CV-SOG, Motifs, and VCTree.</p>
	<p>MusicLiME: Explainable multimodal music understanding tasks, achieving 57.34% for genre and 48.53% for emotion Classification https://ieeexplore.ieee.org/abstract/document/10889771/</p>
	<p>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/</p>
	<p>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/</p>

Continued on next page

Table 2: (Continued)

	<p>Evaluating Contrastive Methodologies for Music Representation Learning Using Playlist Data methods [1, 2] and novel hybrid approaches https://ieeexplore.ieee.org/abstract/document/10888157/</p>
	<p>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</p>
	<p>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/</p>
	<p>Yi Zhu', Xiangyang Liu!?, Tianqi Pang', Xuncan Xiao!, Xiaofan Zhang33, Chenyou Fan!.* Yi Zhu', Xiangyang Liu!?, Tianqi Pang', Xuncan Xiao!, Xiaofan Zhang33, Chenyou Fan!.*</p>

Continued on next page

Table 2: (Continued)

	<p>Text to music audio generation using latent diffusion model: A re-engineering of audiodlm model 1nh tonne Wegner Neteal Ponesin, Deren Hertemans, Rogger Wattenhofer https://www.diva{-}portal.org/smash/record.jsf?pid=diva2:1845150</p>																																																																		
	<p>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/</p>																																																																		
	<p>Classification of Eye-Tracking Data Based on Spatiotemporal Attention Encoding Maju Hei, Chen Xia't, Kuan L, Tan Zhangt Beijing University of Chemical Technology / Northwest Polytechnical University *The Affiliated Hospital of Northwest University / **Xian Jiaotong University</p> <p>Task & Contributions Eye tracking has already played an important role in a variety of fields today, such as user interaction, game development, and medical research. Deep learning methods have been proved effective in predicting human eye movement, contributing to disease visual attention.</p> <ol style="list-style-type: none"> We propose an eye movement classification framework based on spatial features and dynamic temporal features to better reconstruct visual attention. We introduce features from a global perspective, accounting for the competitive influence of other features in the classification process. We conducted experiments on three eye tracking datasets and drew three distinct conclusions and improvements about our work model. <p>Main Experiment Results</p> <table border="1"> <thead> <tr> <th colspan="6">AID Identification (w/epoch=0)</th> </tr> <tr> <th></th> <th>AUC</th> <th>F1</th> <th>Sp</th> <th>NPC</th> <th>AUC</th> </tr> </thead> <tbody> <tr> <td>Web</td> <td>0.6412</td> <td>0.5323</td> <td>0.6310</td> <td>0.5647</td> <td>0.6412</td> </tr> <tr> <td>APM</td> <td>0.7086</td> <td>0.5868</td> <td>0.7237</td> <td>0.7870</td> <td>0.7086</td> </tr> <tr> <td>Sphere</td> <td>0.7063</td> <td>0.6556</td> <td>0.6961</td> <td>0.7462</td> <td>0.7063</td> </tr> <tr> <td>SMC</td> <td>0.8350</td> <td>0.7590</td> <td>0.7324</td> <td>0.7479</td> <td>0.8350</td> </tr> </tbody> </table> <p>Methods Overview</p> <p>Fig 1 Diagram of the spatiotemporal attention encoding model</p> <p>Framework This framework extracts spatiotemporal features from spatial (using ViT) and temporal (using enhanced GRU) features and then sequentially fuses eye movement features into a classifier to predict class probability.</p> <p>Enhanced GRU We introduce a global temporal modulating factor to GRU for global temporal information, which draws more global hidden states to better represent global sequence patterns, overcoming the limitations of traditional GRU in capturing long-term dependencies.</p> <p>Fig 2 Global temporal module in the STAE model</p> <p>Ablation Study Ablation analysis of AID identification task</p> <table border="1"> <thead> <tr> <th></th> <th>AUC</th> <th>F1</th> <th>Sp</th> <th>NPC</th> <th>AUC</th> </tr> </thead> <tbody> <tr> <td>Web</td> <td>0.6412</td> <td>0.5323</td> <td>0.6310</td> <td>0.5647</td> <td>0.6412</td> </tr> <tr> <td>APM</td> <td>0.7086</td> <td>0.5868</td> <td>0.7237</td> <td>0.7870</td> <td>0.7086</td> </tr> <tr> <td>Sphere</td> <td>0.7063</td> <td>0.6556</td> <td>0.6961</td> <td>0.7462</td> <td>0.7063</td> </tr> <tr> <td>SMC</td> <td>0.8350</td> <td>0.7590</td> <td>0.7324</td> <td>0.7479</td> <td>0.8350</td> </tr> </tbody> </table> <p>Fig 3 Visualization of feature visual and with temporal modeling of visual task classification</p>	AID Identification (w/epoch=0)							AUC	F1	Sp	NPC	AUC	Web	0.6412	0.5323	0.6310	0.5647	0.6412	APM	0.7086	0.5868	0.7237	0.7870	0.7086	Sphere	0.7063	0.6556	0.6961	0.7462	0.7063	SMC	0.8350	0.7590	0.7324	0.7479	0.8350		AUC	F1	Sp	NPC	AUC	Web	0.6412	0.5323	0.6310	0.5647	0.6412	APM	0.7086	0.5868	0.7237	0.7870	0.7086	Sphere	0.7063	0.6556	0.6961	0.7462	0.7063	SMC	0.8350	0.7590	0.7324	0.7479	0.8350
AID Identification (w/epoch=0)																																																																			
	AUC	F1	Sp	NPC	AUC																																																														
Web	0.6412	0.5323	0.6310	0.5647	0.6412																																																														
APM	0.7086	0.5868	0.7237	0.7870	0.7086																																																														
Sphere	0.7063	0.6556	0.6961	0.7462	0.7063																																																														
SMC	0.8350	0.7590	0.7324	0.7479	0.8350																																																														
	AUC	F1	Sp	NPC	AUC																																																														
Web	0.6412	0.5323	0.6310	0.5647	0.6412																																																														
APM	0.7086	0.5868	0.7237	0.7870	0.7086																																																														
Sphere	0.7063	0.6556	0.6961	0.7462	0.7063																																																														
SMC	0.8350	0.7590	0.7324	0.7479	0.8350																																																														

Continued on next page

Table 2: (Continued)

Continued on next page

Table 2: (Continued)

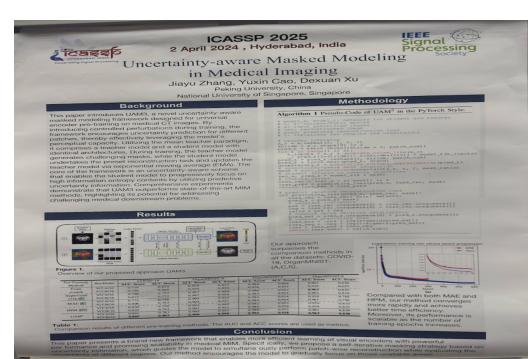
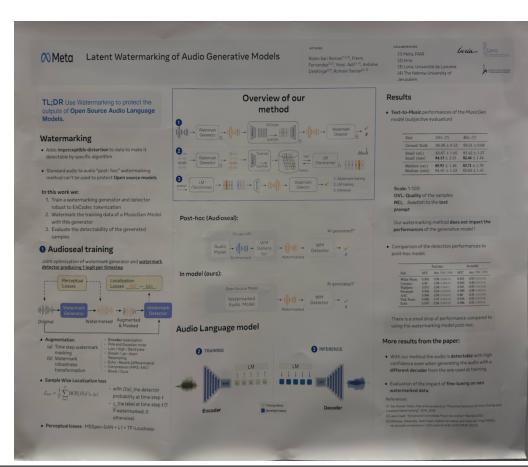
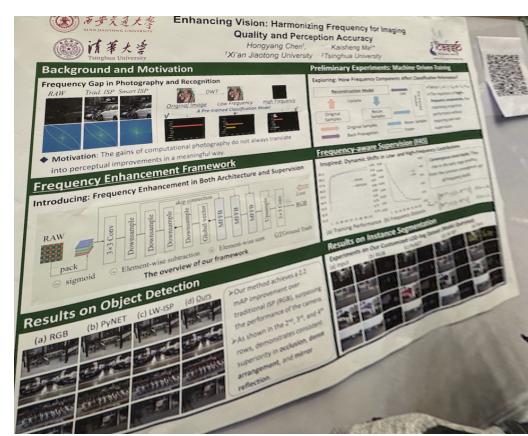
	<p>Exploring the Distribution of Cell Subpopulations in Pancreatic Ductal Adenocarcinoma Slides by Joint Spatial Transcriptomics and Pathology Data Yagi Deng, M cnje Cai, Benao Song, Bin Yang, Limphing Kung , Qedeng Pung*, Jam H https://ieeexplore.ieee.org/abstract/document/10890640/</p>
	<p>NanoGen: A High-affinity Nanobody Generation Model with Guided Diffusion Dezhij Wu*, Xuejiao Liu*, Yiming Qin*, Stephanie M. Linker*, Karin Hrovatin*, Alexander V.Hopp*, Feng Tan** https://ieeexplore.ieee.org/abstract/document/10888039/</p>
	<p>ApinAPDE, a curated repository with physicochemical properties (GRAVY, net charge, isoelectric point, molecular weight). Emas: 2230112006.M.00.u.59. cong na00012e.ntu.edu.sg. as.293th@ntu.edu.5g</p>
	<p>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/</p>

Continued on next page

Table 2: (Continued)

Continued on next page

Table 2: (Continued)

 <p>This paper introduces UAMA, a novel uncertainty-aware masked modeling framework for medical image segmentation. It addresses the challenge of learning from incomplete and noisy training data by introducing context-aware uncertainty-aware masked patches. The framework consists of three main components: a teacher network, a student network, and a uncertainty-aware masked patch generation module. The teacher network generates uncertainty maps, while the student model generates corresponding masks. The uncertainty-aware masked patch generation module then generates challenging medical image patches for training. The results show that UAMA achieves state-of-the-art performance on various medical image segmentation tasks.</p>	<p>Masked image modeling advances 3d medical image analysis <i>Der formace and promising scalability in medical MIM. Spect ically, we propose a tel-literative masking stratcoy based on</i> https://openaccess.thecvf.com/content/WACV2023/html/Chen_Masked_Image_Modeling_Advances_3D_Medical_Image_Analysis_WACV_2023_paper.html</p>
 <p>This poster presents a method for protecting the outputs of open-source audio language models. It uses watermarking to detect forged samples and ensure the integrity of generated samples. The method involves adding imperceptible noise to the data for training, generating a watermark, and then extracting it from the generated samples. The results show that the method can effectively protect open-source audio language models.</p>	<p>Latent watermarking of audio generative models <i>Losses*Inlt..</i> https://ieeexplore.ieee.org/abstract/document/10889782/</p>
 <p>This poster introduces HYMAN, a hybrid memory and attention network for unsupervised anomaly detection. It uses a dual-path architecture with a shared feature space to handle complex anomalies. The network consists of an encoder, a multi-head self-attention layer, and a skip connection. The results show that HYMAN outperforms existing methods on various datasets.</p>	<p>HYMAN: Hybrid Memory and Attention Network for Unsupervised Anomaly Detection <i>Jiahao Li, Yiqiang Chen, Yunbing Xing, Yang Gu, Xiangyu Lan</i> https://ieeexplore.ieee.org/abstract/document/10890028/</p>
 <p>This poster presents a frequency enhancement framework for improving image quality and perception accuracy. It introduces frequency enhancement in both architecture and supervision. The results show that the proposed method achieves better performance than traditional methods on various datasets.</p>	<p>A general framework for object detection <i>>As shown in the 29, 3P%, and 4°</i> https://ieeexplore.ieee.org/abstract/document/710772/</p>

Continued on next page

Table 2: (Continued)

This figure is a dashboard titled "XAI for Gender Representation in Media Analysis". It includes sections for "Introduction", "Datasets", "Speaker Gender Classification", "Aggregated Gendered Vocabularies", and "Acknowledgements". The "Speaker Gender Classification" section contains a table for French and Japanese datasets, and a "BERT-based Classification and XAI" section with a bar chart titled "Explainability: Consistency of the Attributions".

XD0h 18122089project Gender Equality Monitor - ANIR-19-GE30-0012). It was ported from François Buot[†], Camille Guinaudeau[‡], Cyril Grouin[‡], Sahar Ghannay[‡], Shin'ichi Satoh[‡]

This diagram illustrates the "Latent Watermarking of Audio Generative Models" process. It starts with "TLDI Use Watermarking to protect the outputs of Open Source Language Models". The "Overview of our method" section details the "Post-hoc (Auditioned)" and "In-model (sound)" approaches. The "Audio Language model" section shows the flow from "Encoder" to "Decoder" through "LM" and "INFERENZ". A "More from the paper:" section provides additional experimental results.

Latent watermarking of audio generative models 13) Olfenses, Alexandre, Jade Copet, Gabriel Synarve, and Yest Adt. "High fidelity watermarking of audio generative models"
<https://ieeexplore.ieee.org/abstract/document/10889782/>

This diagram details the "Unsupervised Domain Adaptation Via Data Pruning" method. It shows the "Pruning selection" and "Method" phases. The "Pruning selection" phase involves "Domain adaptation via data pruning" and "Data pruning via domain adaptation". The "Method" phase is divided into "Adaptive source pruning" and "Adaptive target pruning". A "t-SNE plots" section visualizes the data distribution.

Unsupervised Domain Adaptation Via Data Pruning TWDR. & method for removing irrelevant data from a training set
<https://ieeexplore.ieee.org/abstract/document/10890190/>

This figure is a dashboard titled "Investigation of Whisper ASR Hallucinations induced by Non-Speech Audio AGH University of Krakow. Poland". It includes sections for "Introduction", "ASR Performance Metrics", "Hallucination Induction", "Hallucination Identification", and "Conclusion". The "Hallucination Induction" section contains a table comparing "ASR Performance Metrics" for different datasets.

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)

This figure is a detailed diagram of a system architecture for named entity speech recognition error correction. It is divided into four main sections: Introduction, System Architecture, Results, and Query Generation.

- Introduction:** Discusses the motivation for the work, mentioning the need for better speech recognition for the visually impaired and the challenges of dealing with errors in ASR systems.
- System Architecture:** Shows a flow from 'ASR Inferences' (e.g., 'John Thompson') through 'Entity Generation' (e.g., 'John Thompson'), 'Query Generation' (e.g., 'John Thompson'), and finally 'Entity Resolution' (e.g., 'John Thompson').
- Results:** Compares 'Retrieval Method' and 'Phonetic ANNs' on datasets like 'WMT14', 'IWSLT14', and 'MSRA'. Phonetic ANNs show superior performance across most metrics.
- Query Generation:** Details the process of generating queries from ASR errors. It includes a 'Contact Construction' step where entities are grouped by name and a 'Data' step involving phonetic matching and distance calculations.

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/>

This figure presents a novel self-prompting strategy for 3D medical image segmentation using the SAM2 model. It is organized into three main sections: Abstract, Method, and Experiment Results.

- Abstract:** States that SAM2 is a large pre-trained model, but its deployment in medical image segmentation is limited due to its inability to effectively segment video streams. The proposed method addresses this by using self-prompting to improve training and inference speed.
- Method:** Describes two main components: I. A LRaP adapted Image Encoder and II. Dynamic Self-prompting Strategy. The encoder takes a 3D input and performs feature extraction. The self-prompting strategy involves generating prompts based on the current batch of images to guide the model's learning.
- Experiment Results:** Shows quantitative results comparing the proposed method with baseline models like UNet++ and Open-UNet++ on datasets like BraTS, ISBI 2012, and 3D-Seg. The proposed method shows significant improvements in terms of both accuracy and speed.

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/>

This figure details a deep learning framework for Alzheimer's disease detection from spontaneous speech. It includes sections for Abstract, Methods, and Results.

- Abstract:** Outlines the goal of developing a deep learning model that can detect Alzheimer's disease from spontaneous speech. It highlights the use of a multi-task learning approach and the integration of pause information with word embeddings.
- Methods:** Describes the architecture, which consists of a sequence-to-sequence model with attention and a parallel language model. It also discusses the use of a sorted index for words and the integration of pause duration and word embeddings.
- Results:** Provides experimental results showing the model's performance on datasets like NIST-SRE07 and NIST-SRE14, demonstrating improved detection accuracy compared to baseline methods.

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>

This figure presents a novel multi-scale context intertwining framework for panoramic renal pathology segmentation. It includes sections for Abstract, Methods, Experiments, and Conclusions.

- Abstract:** Describes the challenge of segmenting renal pathology from panoramic images and the proposed multi-scale context intertwining framework.
- Methods:** Details the architecture, which uses a multi-scale backbone and context intertwining modules to handle complex renal structures.
- Experiments:** Compares the proposed method with baselines on datasets like T100, T200, and T300. The results show improved performance, particularly in handling complex renal structures.
- Conclusions:** Summarizes the findings and suggests future research directions.

Multi-scale Context Intertwining for Panoramic Renal Pathology Segmentation Ye Zhang'2, Xlanchao Guan13, Hengrui LI", Xiangming Yan", Ziyue Wang", Yongbing Zhang'
<https://ieeexplore.ieee.org/abstract/document/10889659/>

This figure is a detailed diagram of a system architecture for named entity speech recognition error correction. It is divided into four main sections: Introduction, System Architecture, Results, and Query Generation.

The figure includes several tables and figures, such as Table 1 (Performance comparison), Table 2 (Ablation study of auxiliary network), Table 3 (The effect of the input task sequence), and Figure 1 (Comparison of different auxiliary network structures).

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/>

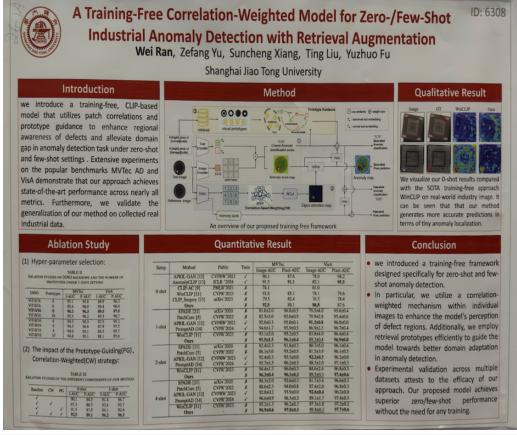
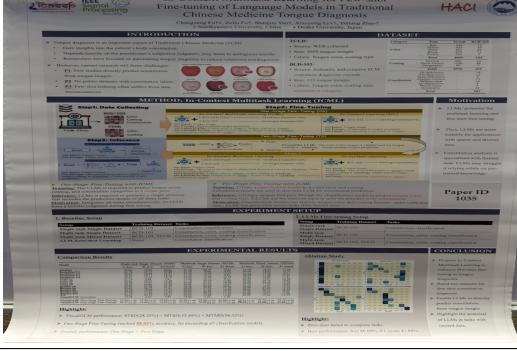
Continued on next page

Table 2: (Continued)

	<p>SELMA: A Speech-Enabled Language Model for Virtual Assistant Interactions <i>Simplified Pigstina, Meduces complarty</i> https://ieeexplore.ieee.org/abstract/document/10890139/</p>
	<p>Glial-neuronal interactions in Alzheimer's disease: the potential role of a 'cytokine cycle' in disease progression <i>hutcmoshini001@Bentuedusg, choc0010@e.ntu.edu.sg, yiha001@e.ntu.edu.sg, congao001@e.ntu.edu.sg, asjagath@ntu.edu.sg</i> https://onlinelibrary.wiley.com/doi/10.1111/j.1750{-}3639.1998.tb00136.x</p>
	<p>Wireless Sensor Networks: 13th China Conference, CWSN 2019, Chongqing, China, October 12–14, 2019, Revised Selected Papers 1. <i>College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China</i> https://books.google.com/books?hl=en&lr=&id=o3fADwAAQBAJ&oi=fnd&pg=PR6&dq=1.+College+of+Computer+Science+and+Technology,+Chongqing+University+of+Posts+and+Telecommunications,+Chongqing.+China&ots=xK3vgQuzGb&sig=1R1Jj8oR{-}X4PqqWdWiKUW8crgKE</p>
	<p>AI-Generated Music Detection and its Challenges <i>Suno, Udio, Riffusion, ...</i> https://arxiv.org/abs/2501.10111</p>

Continued on next page

Table 2: (Continued)

 <p>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 Shanghai Jiao Tong University</p> <p>Introduction we introduce a training-free, CLP-based model that utilizes prior knowledge and prototype guidance to enhance regional awareness of defects and alleviate domain gap in anomaly detection task under zero-shot or few-shot settings. Extensive experiments on three benchmarks (MoCap, MoAn and Vis4) demonstrate that our approach achieves state-of-the-art performance across nearly all metrics. Furthermore, we validate the generalization of our method on collected real industrial data.</p> <p>Method The proposed framework consists of three main components: 1. Prototype Guiding: A pre-trained CLP model is used to generate prototypes for each category. 2. Region-aware learning: The model uses a correlation-weighted mechanism to emphasize regions of interest. 3. Retrieval Augmentation: The model retrieves prototypes from a large dataset to guide the model towards better domain adaptation in anomaly detection.</p> <p>Qualitative Result We visualize our 0-shot results compared with the SOTA training-free approach. The visualizations show that our method can be seen that our method generates more accurate predictions in terms of key anomaly locations.</p>	<p>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/</p>
 <p>Parameter-efficient fine-tuning of large-scale pre-trained language models Changzeng Fits, Zelin Fut, Shaojun Yant, Xiaoyong Lyvt, Yuliang Zhaot</p> <p>INTRODUCTION Fine-tuning of Language Models in Traditional Clinical Medicine Disease Diagnosis</p> <p>Dataset HACI</p> <p>EXPERIMENTAL SETUP 1. Data Collection 2. Feature Extraction 3. Model Training 4. Evaluation</p> <p>EXPERIMENTAL RESULTS Comprehensive Results</p> <p>Conclusion Proposed framework can significantly improve the performance of pre-trained language models in clinical medicine disease diagnosis tasks.</p>	<p>Parameter-efficient fine-tuning of large-scale pre-trained language models Changzeng Fits, Zelin Fut, Shaojun Yant, Xiaoyong Lyvt, Yuliang Zhaot</p> <p>https://www.nature.com/articles/s42256{-}023{-}00626{-}4</p>