Ε.Μ.Π. - Σχολή Ηλεκτρολόγων Μηχ/κών & Μηχ/κών Υπολογιστών

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ[1] 2026-2027

Καθηγητής Πέτρος Μαραγκός, maragos@cs.ntua.gr, και Συνεργάτες/ιδες
Εργαστήρια: CVSP: http://cvsp.cs.ntua.gr, και IRAL: http://robotics.ntua.gr

Οι ενδιαφερόμενοι/ες παρακαλούνται να αποστείλουν ηλεκτρονικά σε ένα Zip με το όνομά τους τα εξής: i) την αναλυτική τους βαθμολογία από ΣΗΜΜΥ-ΕΜΠ με μια εκτίμηση του τρέχοντος μέσου όρου, ii) το βιογραφικό τους, και iii) τον αριθμό των οφειλομένων μαθημάτων αν βρίσκονται στο 4ο ή ανώτερο έτος σπουδών. Η αποστολή της πληροφορίας αυτής να γίνεται στην Γραμματέα του Εργαστηρίου κα. Βίκυ Πλατίτσα (email: <vickyplatitsa@gmail.com>).

Προαπαιτούμενα μαθήματα (Ροής Σ) με επιτυχή και συνεπή παρακολούθηση και πολύ καλή απόδοση:
- Για θέματα Ν: Σήματα & Συστήματα-ΣΣ (ή και ΨΕΣ), και Μηχανική Μάθηση-ΜΜ ή Αναγνώριση Προτύπων-ΑΠ (*).
- Για θέματα V: Οραση Υπολογιστών-ΟΥ, ή πιθανώς και ΜΜ ή ΑΠ (*).
- Για θέματα R που συνδυάζουν Ρομποτική με CV/SP/ML: ΟΥ, Ρομποτική και ΜΜ/ΑΠ (*).
- Για θέματα AVM: ΟΥ (ή και ΨΕΣ), και ΜΜ/ΑΠ (*)
- Για θέματα L: ΟΥ και ΜΜ ή ΑΠ (*).
- Για θέματα S: ΣΣ και κάποιο μαθημα ΣΑΕ απο Ροη Σ, ή πιθανώς και ΜΜ ή ΑΠ (*).
- Για θέματα Τ: ΟΥ και ΜΜ/ΑΠ (*).

* = μπορεί να παρακολουθείται ταυτόχρονα με διπλωματική.

- Γενικά, για να υπάρχει ποικιλία ερευνητικών θεμάτων, δίνονται περισσότερα θέματα από τις θέσεις που μπορούν να επιβλεφθούν εντός ενός έτους. **Τα ενεργά θέματα είναι αυτά με το σήμα (*) στον τίτλο τους.**
- Μετά την εκδήλωση ενδιαφέροντος οι ενδιαφερόμενοι φοιτητές θα συναντηθούν μαζί με τον υπεύθυνο καθ. Π. Μαραγκό και Επιστημονικούς/ές Συνεργάτες/ιδες καθώς και Υποψήφιους/ες Διδάκτορες της ερευνητικής ομάδας και θα γίνει προσπάθεια να αντιστοιχισθούν θέματα και φοιτητές με όσο το δυνατόν καλύτερο ταίριασμα ενδιαφερόντων και ικανοτήτων.
- Η τελική αποδοχή αίτησης για εκπόνηση διπλωματικής θα εξαρτηθεί από την επίδοση στα σχετικά μαθήματα του εργαστηρίου, τον αριθμό των οφειλομένων μαθημάτων, και την χρονική διαθεσιμότητα του θέματος.
- Η 1η φάση της Διπλωματικής είναι η μελέτη της σχετικής Βιβλιογραφίας. Εν συνεχεία (ή και παράλληλα με την ενασχόληση της 1ης φάσης) ακολουθεί η 2η (κύρια) φάση της έρευνας.
- Ενδιάμεση Πρόοδος: Με συχνότητα περίπου μια φορά ανά τρίμηνο θα προσκαλούνται οι ενεργές Διπλωματικές για σύντομη παρουσίαση της προόδου τους σε όλα τα μέλη του Εργαστηρίου.

Internships: Για φοιτητές/ριες που έχουν προχωρήσει τις Διπλωματικές τους σε ώριμο στάδιο, μπορεί να υπάρξουν προοπτικές για Internships στο Ευρωπαϊκό Κέντρο Αριστείας HERON.

Physiological Data Processing and Learning from Wearable Sensors for Healthcare Applications:

Η τεχνολογική πρόοδος που έχει σημειωθεί τα τελευταία χρόνια στις φορητές συσκευές (κινητά τηλέφωνα, smart-watch, κτλ.) έχει πλέον επιτρέψει την αξιόπιστη καταγραφή πολλών και διαφορετικών βιομετρικών δεικτών, καθώς επίσης και ακουστικών σημάτων, τα οποία πλέον χρησιμοποιούνται και στην ιατρική.

(*) N1. Foundation models για την ανάλυση ακατέργαστων (raw) σημάτων από έξυπνα ρολόγια με σκοπό τον εντοπισμό και την πρόβλεψη υποτροπών σε ασθενείς με ψυχωτικές διαταραχές

Η μακροχρόνια καταγραφή σημάτων από έξυπνα ρολόγια (επιταχυνσιόμετρο, γυροσκόπιο, καρδιακός ρυθμός) κατά την καθημερινότητα ασθενών με ψυχωτικές διαταραχές επιτρέπει την εξαγωγή του ψηφιακού τους φαινοτύπου (digital phenotyping) και τον εντοπισμό αποκλίσεων που ενδέχεται να προαναγγέλλουν μια υποτροπή [1]. Οι μέχρι σήμερα προσεγγίσεις βασίζονται κυρίως σε handcrafted features (π.χ. δείκτες HRV, δραστηριότητας και ύπνου) και σε εξατομικευμένα μοντέλα ανίχνευσης ανωμαλιών [1, 2], τα οποία εκπαιδεύονται σε περιορισμένο αριθμό ασθενών λόγω της σπανιότητας των υποτροπών και της έλλειψης

[1] Μερικά από τα ανωτέρω θέματα έχουν προοπτική για Διδακτορικό με οικονομική υποστήριξη από ερευνητικά προγράμματα.

επισημειωμένων δεδομένων. Παράλληλα, έχουν εμφανιστεί foundation models για σήματα wearable συσκευών, τα οποία προεκπαιδεύονται με αυτο-επιβλεπόμενη μάθηση σε τεράστιους όγκους μη επισημειωμένων δεδομένων [3, 4], καθώς και γενικά foundation models για χρονοσειρές [5]. Τα μοντέλα αυτά δεν έχουν ακόμη εφαρμοστεί στην ανίχνευση ψυχωτικών υποτροπών. Στόχος της διπλωματικής είναι η αξιοποίηση δημόσια διαθέσιμων foundation models απευθείας πάνω στα ακατέργαστα σήματα του ρολογιού, χωρίς ενδιάμεσο στάδιο χειροποίητων χαρακτηριστικών, για την εξαγωγή αναπαραστάσεων του καθημερινού προφίλ κάθε ασθενούς. Η αξιολόγηση θα γίνει στο σύνολο δεδομένων e-Prevention [1], με σύγκριση έναντι των state-of-the-art μεθόδων που βασίζονται σε χαρακτηριστικά [2].

(Συνεργάτες: Δρ. Νίκη Ευθυμίου και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

References:

[1] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, "E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures," Sensors, vol. 22, no. 19, Oct. 2022.
[2] Tsalkitzis, N., Maragos, P., & Efthymiou, N. "Beyond Personalization: Cluster-Aware Shared Representation Learning for Wearable-Based Psychotic Relapse Detection" . In *IEEE-EMBS Int. Conference on Body Sensor Networks 2026*.
[3] M. A. Xu, J. Narain, G. Darnell, H. Hallgrimsson, H. Jeong, D. Forde, R. Fineman, K. Raghuram, J. M. Rehg, and S. Ren, "RelCon: Relative Contrastive Learning for a Motion Foundation Model for Wearable Data," in Proc. International Conference on Learning Representations (ICLR), 2025.
[4] H. Yuan, S. Chan, A. P. Creagh, C. Tong, A. Acquah, D. A. Clifton, and A. Doherty, "Self-supervised learning for human activity recognition using 700,000 person-days of wearable data," npj Digital Medicine, vol. 7, 91, 2024.
[5] M. Goswami, K. Szafer, A. Choudhry, Y. Cai, S. Li, and A. Dubrawski, "MOMENT: A Family of Open Time-series Foundation Models," in Proc. International Conference on Machine Learning (ICML), 2024..

(*) N2. Ανάλυση ηχητικών σημάτων για υποστήριξη ασθενών σε θέματα ψυχικής υγείας.

Στόχος αυτής της διπλωματικής εργασίας είναι ο εντοπισμός, ή και η πρόβλεψη, υποτροπιαζουσών καταστάσεων σε ασθενείς στο διπολικό ή στο σχιζοειδές φάσμα βασιζόμενοι σε δεδομένα ομιλίας των ασθενών [1-5]. Ενδεικτικές κατευθύνσεις αποτελούν 1) η χρήση παράλληλων τροπικοτήτων σε συνδυασμό με την ομιλία (πχ καταγραφές κειμένου των συνομιλιών) [6], 2) εκμετάλλευση προεκπαιδευμένων γλωσσικών/ακουστικών μοντέλων ή μελέτη σε συνδυασμό με συγγενικά προβλήματα, όπως η αναγνώριση συναισθηματικής κατάστασης [7,8], 3) διερεύνηση γλωσσικών δεικτών (πχ συναισθηματικά φορτισμένο λεξιλόγιο, αυτοαναφορική γλώσσα) χρησιμοποιώντας μεθόδους επεξεργασίας φυσικής γλώσσας και καταγραφές κειμένου των κλινικών συνομιλιών [11-13] ή/και άλλων βάσεων δεδομένων με έμφαση στην ερμηνευσιμότητα [9, 10].

(Συνεργάτες: Δρ. Νάνσυ Ζλατίντση, ΥΔ Χρήστος Γαρούφης ΕΜΠ, και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

[1] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, "E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures," Sensors, vol. 22, no. 19, Oct. 2022.
[2] Garoufis, C., Zlatintsi, A., Filntisis, P. P., Efthymiou, N., Kalisperakis, E., Karantinos, T., ... & Maragos, P. Towards unsupervised subject-independent speech-based relapse detection in patients with psychosis using variational autoencoders. In Proc. EUSIPCO 2022
[3] J. Gideon, K. Matton, S. Anderau, M.G. McInnis, and E.M. Provost,"When to Intervene: Detecting Abnormal Mood using Everyday Smart-phone Conversations,"arXiv preprint arXiv:1909.11248, 2019.
[4] K.-Y. Huang, C.-H. Wu, and M.-H. Su, "Attention-Based Convolutional Neural Network and Long Short-Term Memory for Short-Term Detection of Mood Disorders Based on Elicited Speech Responses," Pattern Recognition, 2019.
[5] Liu, S., A. Mallol-Ragolta, B. W. Schuller. "COVID-19 Detection from Speech in Noisy Conditions." ICASSP 2023.
[6] S. Yoon et al. Attentive modality hopping mechanism for speech emotion recognition. *ICASSP 2020.*
[7] Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., ... & Wei, F. Wavlm: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1505-1518.

[8] Morais, E., Hoory, R., Zhu, W., Gat, I., Damasceno, M., & Aronowitz, H. Speech emotion recognition using self-supervised features. In Proc. ICASSP 2022
[9] Ringeval, F. et al., "AVEC 2017: Real-life depression, and affect recognition workshop and challenge." In *Proceedings of the 7th annual workshop on audio/visual emotion challenge*
[10] Wang, K. et al., "MEAD: A large-scale audio-visual dataset for emotional talking-face generation." Proc. ECCV 2020
[11] V. Lyberatos, E. G. Dervakos, E. Adamidi, A. Voulodimos, G Stamou, Exploration of Perceptual Speech Features for Clinical Decision-Support in Mental Health Care, ACL 2026 Workshop
[12] J. N. de Boer, A. E. Voppel, S. G. Brederoo, H. G. Schnack, K. P. Truong, F. N. K. Wijnen, I. E. C. Sommer, Acoustic speech markers for schizophrenia-spectrum disorders: a diagnostic and symptom-recognition tool, Psychol. Med. 2023.
[13] Y. Zhang, A. Carrió, J. Sevilla-Llewellyn-Jones, E. Gutiérrez, A. Calvo, J.-B. Navarro, and A. Barajas, "Linguistic markers in at-risk mental states using natural language processing: A systematic review," Healthcare, Apr. 2026.

(*) N3. Ανίχνευση stress από φορετές συσκευές καρπού για την έγκαιρη αναγνώριση κρίσιμων συμβάντων

Τα κρίσιμα συμβάντα, όπως τα επεισόδια stress, οξείας δυσφορίας ή συμπεριφορικών εκρήξεων, συνοδεύονται από μεταβολές στον καρδιακό ρυθμό [1], στη μεταβλητότητά του (HRV) και στην ηλεκτροδερμική δραστηριότητα (EDA), τις οποίες μπορούν να καταγράφουν συνεχώς οι φορετές (wearable) συσκευές. Η ανάλυση αυτών των σημάτων με μηχανική μάθηση επιτρέπει την ανίχνευση του stress σε πραγματικό χρόνο, κυρίως μέσω εξατομικευμένων μοντέλων [2], ακόμη και την πρόβλεψη επιθετικής συμπεριφοράς πριν από την εκδήλωσή της [3]. Ωστόσο, τα περισσότερα συστήματα έχουν αξιολογηθεί σε ελεγχόμενες εργαστηριακές συνθήκες [5], η σωματική δραστηριότητα προκαλεί φυσιολογική απόκριση παρόμοια με αυτή του στρες [2], και οι επισημειώσεις στην καθημερινή ζωή είναι σπάνιες, γεγονός που έχει οδηγήσει στη συλλογή νέων συνόλων δεδομένων εκτός εργαστηρίου [6], [7]. Στόχος της διπλωματικής είναι η ανάπτυξη μεθόδων για τη συνεχή ανίχνευση στρες από τα σήματα ενός ρολογιού καρπού ως έγκαιρη ένδειξη κρίσιμων συμβάντων. Θα διερευνηθούν προσεγγίσεις όπως η ανίχνευση ανωμαλιών, όπου ένα εξατομικευμένο μοντέλο (π.χ. autoencoder) μαθαίνει τη φυσιολογική «βάση» κάθε ατόμου και εντοπίζει τις αποκλίσεις από αυτήν αντί για απόλυτα κατώφλια [4], καθώς και η πολυτροπική σύμμειξη της πληροφοριας με ξεχωριστούς κωδικοποιητές ανά σήμα και μηχανισμό cross-attention, ώστε η πληροφορία του επιταχυνσιομέτρου να διακρίνει το στρες από τη σωματική δραστηριότητα [8] και το μοντέλο να λειτουργεί ακόμη και όταν κάποιο σήμα λείπει ή είναι θορυβώδες.

(Συνεργάτες: Δρ. Νίκη Ευθυμίου)

References:
[1] E. Kildal, K. Stadskleiv, E. S. Boysen, T. Øderud, I.-L. Dahl, T. M. Seeberg, S. Guldal, F. Strisland, C. Morland, and B. Hassel, "Increased heart rate functions as a signal of acute distress in non-communicating persons with intellectual disability," Scientific Reports, vol. 11, 6479, 2021.
[2] S. de Vries, F. van Oost, H. Smaling, N. de Knegt, P. Cluitmans, R. Smits, and E. Meinders, "Real-time stress detection based on artificial intelligence for people with an intellectual disability," Assistive Technology, 2023. DOI: 10.1080/10400435.2023.2261045
[3] T. Imbiriba, A. Demirkaya, A. Singh, D. Erdogmus, and M. S. Goodwin, "Wearable Biosensing to Predict Imminent Aggressive Behavior in Psychiatric Inpatient Youths With Autism," JAMA Network Open, vol. 6, no. 12, e2348898, 2023.
[4] N. Efthymiou, G. Retsinas, P. P. Filntisis, and P. Maragos, "Augmenting Transformer Autoencoders with Phenotype Classification for Robust Detection of Psychotic Relapses," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024.
[5] T. Iqbal, A. J. Simpkin, D. Roshan, et al., "Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset," Sensors, vol. 22, no. 21, 8135, 2022.
[6] S. Hosseini et al., "A multimodal sensor dataset for continuous stress detection of nurses in a hospital," Scientific Data, vol. 9, 255, 2022.
[7] S. Kang, W. Choi, C. Y. Park, N. Cha, A. Kim, A. H. Khandoker, L. Hadjileontiadis, H. Kim, Y. Jeong, and U. Lee, "K-EmoPhone: A Mobile and Wearable Dataset with In-Situ Emotion, Stress, and Attention Labels," Scientific Data, vol. 10, 351, 2023.
[8] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks" (PPG-DaLiA dataset), Sensors, vol. 19, no. 14, 3079, 2019.

(*) N4. Απόσταξη γνώσης (knowledge distillation) μοντέλων αναγνώρισης συμβάντων από σήματα wearable συσκευων για εκτέλεση σε πραγματικό χρόνο σε έξυπνα ρολόγια

Τα μεγάλα μοντέλα βαθιάς μάθησης και τα foundation models για σήματα wearable συσκευών [4] επιτυγχάνουν υψηλή ακρίβεια σε εργασίες όπως η αναγνώριση δράσεων [6], η ταξινόμηση σταδίων ύπνου [7] και η ανίχνευση υποτροπών [5], όμως οι απαιτήσεις τους σε μνήμη και ενέργεια δεν επιτρέπουν την εκτέλεσή τους στο ίδιο το ρολόι. Η απόσταξη γνώσης (knowledge distillation) μεταφέρει τη γνώση ενός μεγάλου μοντέλου-«δασκάλου» σε ένα συμπαγές μοντέλο-«μαθητή», ακόμη και όταν ο μαθητής έχει πρόσβαση σε λιγότερες ή διαφορετικές τροπικότητες από τον δάσκαλο, π.χ. μόνο στο επιταχυνσιόμετρο του ρολογιού αντί για βίντεο [1] ή για αισθητήρες υψηλής ποιότητας [2]. Στόχος της διπλωματικής είναι η ανάπτυξη συμπαγών μοντέλων που εκτελούνται σε πραγματικό χρόνο στο ρολόι, μέσω απόσταξης γνώσης σε συνδυασμό με τεχνικές όπως το quantization [3]. Θα διερευνηθεί επίσης αν οι ίδιες τεχνικές γενικεύουν σε εργασίες με πολύ διαφορετική χρονική κλίμακα και αν ένας κοινός πολυεργασιακός (multi-task) μαθητής μπορεί να εξυπηρετήσει περισσότερες από μία εργασίες.

(Συνεργάτες: Δρ. Νίκη Ευθυμίου)

References:
[1] B. Chen, W. Wongso, Z. Li, Y. Khaokaew, H. Xue, and F. Salim, "COMODO: Cross-Modal Video-to-IMU Distillation for Efficient Egocentric Human Activity Recognition," Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT/UbiComp), 2026.
[2] Z. Zhao, M. Mortazavi, and N. Yan, "PULSE: Privileged Knowledge Transfer from Electrodermal Activity to Low-Cost Sensors for Stress Monitoring," in Proc. Machine Learning for Health (ML4H), 2025.
[3] J. Lin, L. Zhu, W.-M. Chen, W.-C. Wang, C. Gan, and S. Han, "On-Device Training Under 256KB Memory," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2022.
[4] A. Pillai, D. Spathis, F. Kawsar, and M. Malekzadeh, "PaPaGei: Open Foundation Models for Optical Physiological Signals," in Proc. International Conference on Learning Representations (ICLR), 2025.
[5] Tsalkitzis, N., Maragos, P., & Efthymiou, N. "Beyond Personalization: Cluster-Aware Shared Representation Learning for Wearable-Based Psychotic Relapse Detection" . In *IEEE-EMBS Int. Conference on Body Sensor Networks 2026*.
[6] S. Chan, H. Yuan, C. Tong, A. Acquah, A. Schonfeldt, J. Gershuny, and A. Doherty, "CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition," Scientific Data, vol. 11, 1135, 2024.
[7] K. Wang, J. Yang, A. Shetty, and J. Dunn, "DREAMT: Dataset for Real-time sleep stage EstimAtion using Multisensor wearable Technology," PhysioNet, 2024.

Επικοινωνία Ανθρώπου-Ρομπότ και Αυτόνομη Πλοήγηση: Πολυτροπικές Ευφυείς Διεπαφές, Οραση, Ανάλυση Οπτικο-Ακουστικών Σημάτων, Ανίχνευση, Αναγνώριση, Πλοήγηση:

(*) R1. Αναγνώριση ανθρωπίνων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ, με εφαρμογές σε assistive robotics:

Μια από τις προκλήσεις για την ανάπτυξη ευφυών ρομπότ που αλληλεπιδρούν με φυσικό τρόπο με το περιβάλλον τους, είναι η ικανότητά τους να αναγνωρίζουν τις δράσεις και τις χειρονομίες που εκτελούν οι άνθρωποι ανά πάσα στιγμή. Στόχος της διπλωματικής αυτής είναι ο αυτόματος χρονικός *εντοπισμός* δράσεων/χειρονομιών σε πραγματικό χρόνο, η μοντελοποίηση και η *αναγνώρισή* τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων, όπως RGB-D ή/και ανθρώπινη πόζα. Οι μέθοδοι αυτές μπορούν να χρησιμοποιηθούν σε μεγάλες βάσεις δεδομένων ή/και σε ρομποτικές διατάξεις υποβοήθησης ηλικιωμένων και ατόμων με κινητικά προβλήματα.

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης)

References:
[1] I. Rodomagoulakis, N. Kardaris, V. Pitsikalis, E. Mavroudi, A. Katsamanis, A. Tsiami and P. Maragos, "Multimodal Human Action Recognition in Assistive Human-Robot Interaction", Proc. ICASSP-2016, Shanghai, China, Mar. 2016.

[2] A. Zlatintsi, I. Rodomagoulakis, V. Pitsikalis, P. Koutras, N. Kardaris, X. Papageorgiou, C. Tzafestas and P. Maragos, "Social Human-Robot Interaction for the Elderly: Two Real-life Use Cases", *Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI-2017)*, Vienna, Austria, March 2017.
[4] J. Carreira, & A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset". *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017)*, Honolulu, Hawaii, July 2017
[5] M. Xu, M. Gao, Y. Chen, L. S. Davis, D. J. Crandall, "Temporal recurrent networks for online action detection", *Proc. IEEE Int'l Conference on Computer Vision (ICCV-2019)*, Seoul, Korea, Oct.-Nov. 2019.
[6] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, Ahmed A. A. Osman, D. Tzionas, and M. J. Black, "Expressive Body Capture: 3D Hands, Face, and Body from a Single Image", *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2019)*, Long Beach, CA, United States, June 2019.
[7] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding", in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2019
[8] G. Chalvatzaki, P. Koutras, A. Tsiami, C. S. Tzafestas and P. Maragos, "i-Walk Intelligent Assessment System: Activity, Mobility, Intention, Communication", *Proc. 16th European Computer Vision Conference Workshops (ECCVW) – 8th Int'l Workshop on Assistive Computer Vision and Robotics (ACVR-2020)*, Aug. 2020.
[9] V. Vasileiou, N. Kardaris, P. Maragos, Exploring Temporal Context and Human Movement Dynamics for Online Action Detection in Videos, in *Proc. 29th European Signal Processing Conference (EUSIPCO 2021)*, Dublin, Ireland, 2021.
[10] G. Moustris, N. Kardaris, A. Tsiami, G. Chalvatzaki, P. Koutras, A. Dometios, P. Oikonomou, C. Tzafestas, P. Maragos, E. Efthimiou, X. Papageorgiou, S.-E. Fotinea, Y. Koumpouros, A. Vacalopoulou, E. Papageorgiou, A. Karavasili, F. Koureta, D. Dimou, A. Nikolakakis, K. Karaiskos and P. Mavridis, "The i-Walk Lightweight Assistive Rollator: First Evaluation Study", in *Frontiers in Robotics and AI*, vol. 8, pp. 272, Sep. 2021.
[11] K. Spathis, N. Kardaris, P. Maragos, "Multi-task Learning for Joint Action and Gesture Recognition", in 9th Workshop and Competition on Affective & Behavior Analysis in-the-wild, in conjunction with the International Conference on Computer Vision (ICCV), 2025.

(*) R2. Visual SLAM systems with geometric and/or semantic enhancements

Τα συστήματα vSLAM [1,2] επιτυγχάνουν ταυτόχρονα την 3D ανακατασκευή του άγνωστου χώρου στον οποίο κινούνται, καθώς και την εκτίμηση της τροχιάς που διαγράφουν. Το vSLAM αποτελεί τη real-time εκδοχή του κλασικού προβλήματος της ΟΥ, Structure from Motion (SfM), ενώ έχει επίσης μελετηθεί παλιότερα υπό τη μορφή Time-of-Flight (ToF) μεθόδων (lasers ή sonars) και στο άλλο κλασικό πρόβλημα της Ρομποτικής, το SLAM. Πλέον, τα παραπάνω ενοποιούνται σε ένα σύστημα, με χρήση mono, stereo ή RGBD καμερών, καθώς και είναι δυνατό να συνδυαστούν με πληροφορία από LIDARs ή IMUs [7] οδηγώντας σε παραλλαγές συστημάτων vSLAM.
Έμφαση δίνεται:

- στην αξιοποίηση συνθετότερων γεωμετρικών δομών της εικόνας (ευθείες, επίπεδα, κωνικές) [3,4,6],
- στην εξαγωγή και αξιοποίηση σημασιολογικής πληροφορίας [4,5],
- στην ενσωμάτωση δυναμικών άκαμπτων αντικειμένων με ταυτόχρονο υπολογισμό της πόζας τους [8],
- στη χρήση 3D Gaussian Splats για την αύξηση της απεικονιστικής δυνατότητας των χαρτών [9]
- στη χρήση σημείων φυγής για αποσύζευξη μεταφορικής και περιστροφικής συνιστώσας της πόζας [10].

Τα ανωτέρω υλοποιούνται με σκοπό την αύξηση της ευρωστίας, της ακρίβειας, του εύρους των περιβαλλόντων εφαρμογής, ή τη μείωση της υπολογιστικής πολυπλοκότητας. Ευρέως χρησιμοποιούμενα σύνολα δεδομένων σε εξωτερικούς χώρους είναι το [11] και σε εσωτερικούς τα [12,13].

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμιγκας)

References:

[1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System", *IEEE Transactions on Robotics*, 2015.
[2] D. Schlegel, M. Colosi and G. Grisetti, "ProSLAM: Graph SLAM from a Programmer's Perspective", *IEEE International Conference on Robotics and Automation*, 2018.
[3] F. Nardi, B. D. Corte and G. Grisetti, "Unified Representation and Registration of Heterogeneous Sets of Geometric Primitives", *IEEE Robotics and Automation Letters*, 2019.
[4] L. Nicholson, M. Milford and N. Sünderhauf, "QuadricSLAM: Dual Quadrics From Object Detections as Landmarks in Object-Oriented SLAM", *IEEE Robotics and Automation Letters*, 2019.
[5] I. Asmanis, P. Mermigkas, G. Chalvatzaki, J. Peters and P. Maragos, "A Semantic Enhancement of Unified Geometric Representations for Improving Indoor Visual SLAM", *19th International Conference on Ubiquitous Robots (UR)*, 2022.
[6] M. Hosseinzadeh, Y. Latif, T. Pham, N. Sünderhauf and I. Reid, "Structure Aware SLAM Using Quadrics and Planes", *ACCV*, 2018.

[7] A. Rosinol, M. Abate, Y. Chang and L. Carlone, “Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping”, *IEEE International Conference on Robotics and Automation*, 2020.
[8] A. Manetas, P. Mermigkas and P. Maragos, “SDPL-SLAM: Introducing Lines in Dynamic Visual SLAM and Multi-Object Tracking”, *Proc. of IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2024.
[9] P. Mermigkas, A. Manetas, and P. Maragos, “GLAM-SLAM: Real-time Gaussian LArge-scale Mapping via Flow Densification and Spatial Decomposition”, *Proc. of IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2026.
[10] J. Liu and Z. Meng, "Visual SLAM With Drift-Free Rotation Estimation in Manhattan World", *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6512-6519, Oct. 2020.
[11] KIT, Visual Odometry/SLAM Evaluation, [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry.php
[12] TUM, RGB-D SLAM Dataset and Benchmark, [Online]. Available: https://vision.in.tum.de/data/datasets/rgbd-dataset
[13] ICL-NUIM, Dataset for benchmarking VO and SLAM algorithms, [Online]. Available: https://www.doc.ic.ac.uk/~ahanda/VaFRIC/iclnuim.html

(*) R3. Visual localization with Deep Learning techniques using RGB images for robotic inspection in power substations

Στα πλαίσια της διπλωματικής επιθυμούμε την απόκτηση ενός πυκνού νέφους σημείων για την αναπαράσταση του χώρου ενός Κέντρου Υψηλής Τάσης (ΚΥΤ) (π.χ. pointcloud που έχει παραχθεί με μεθόδους LiDAR SLAM ή visual SLAM) και κατόπιν την επίλυση του προβλήματος της αυτόματης χωροθέτησης μιας άγνωστης εικόνας προερχόμενης από το ίδιο ΚΥΤ. Η παραπάνω μεθοδολογία που συναντάται με το όνομα visual localization βρίσκει εφαρμογές σε πεδία όπως η αυτόνομη πλοήγηση, η επαυξημένη πραγματικότητα [1], κ.ά. Έρευνες έχουν εστιάσει στη χρήση μικρού συνόλου εικόνων (few-shot learning) [2], στην ταυτόχρονη χρήση τρισδιάστατης και σημασιολογικής πληροφορίας [3], καθώς και πρόσφατα στην ενσωμάτωση 3D Gaussian Splatting [1]. Το θέμα αυτό μπορεί να εφαρμοστεί πάνω στην αυτόματη επισκόπηση ηλεκτρολογικού εξοπλισμού σε κέντρα υψηλής τάσης με τη βοήθεια αυτόνομου ρομπότ.

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμιγκας, Νίκος Κάρδαρης)

[1] Zhai, Hongjia et al, “SplatLoc: 3D Gaussian Splatting-based Visual Localization for Augmented Reality”, 10.48550/arXiv.2409.14067.
[2] Dong, Siyan, et al, "Visual localization via few-shot scene region classification", *Proc. International Conference on 3D Vision (3DV)*, 2022.
[3] Schönberger, Johannes L. et al, "Semantic visual localization", *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
[4] N. Kardaris, P. Mermigkas, G. Moustris, C. Tzafestas, P. Maragos, “Automated Thermal Fault Detection in Ultra-High Voltage Substation Equipment”, *Proc. IFAC Workshop on Smart Energy Systems for Efficient and Sustainable Smart Grids and Smart Cities (SENSYS 2025)*, Bari, Italy, 2025.

(*) R4. Joint 6D Pose and Deformation Estimation of Objects with High Shape Variability for Robotic Grasping

Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη μεθόδων όρασης υπολογιστών για την από κοινού εκτίμηση της 6D πόζας και της 3D παραμόρφωσης αντικειμένων με μεγάλη ενδοκατηγορική μεταβλητότητα σχήματος, όπως φρούτα και λαχανικά. Καθώς οι κλασικές μέθοδοι εκτίμησης πόζας [3] προϋποθέτουν ακριβές CAD μοντέλο για κάθε στιγμιότυπο, κάτι πρακτικά ανέφικτο για τέτοια αντικείμενα, πρόσφατες προσεγγίσεις παραμορφώνουν ένα πρότυπο μοντέλο κατηγορίας μέσω πλέγματος (lattice deformation) [5, 6]. Ξεκινώντας από αυτές τις μεθόδους και το benchmark PEAR [6], η εργασία θα διερευνήσει πιο εκφραστικές αναπαραστάσεις παραμόρφωσης, όπως μαθημένους χώρους σχήματος χαμηλής διάστασης, νευρωνικά πεδία παραμόρφωσης ή μοντέλα που γενικεύουν σε πολλές και άγνωστες κατηγορίες, ενώ όπου απαιτείται το PEAR θα επεκταθεί με νέες κατηγορίες και πιο απαιτητικά σενάρια (clutter, επικαλύψεις). Ως τελικό στάδιο, οι μέθοδοι θα ενσωματωθούν σε ρομποτικό σύστημα χειρισμού με τους βραχίονες Franka Emika Panda [1] και SO-101 [2], όπου η εκτιμώμενη γεωμετρία θα τροφοδοτεί αλγορίθμους παραγωγής λαβών [4], ώστε να εξεταστεί πώς η ακριβέστερη μοντελοποίηση του σχήματος βελτιώνει την επιτυχία λαβής σε πραγματικά πειράματα.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, Δρ. Κατερίνα Παπαδημητρίου)

References:
[1] Franka Robotics, Franka Emika Panda. https://franka.de/
[2] TheRobotStudio & Hugging Face, SO-101 Robot Arm (LeRobot). https://github.com/TheRobotStudio/SO-ARM100

[3] Labbé, Y., et al. (2022). MegaPose: 6D Pose Estimation of Novel Objects via Render & Compare. CoRL.
[4] Mousavian, A., Eppner, C., Fox, D. (2019). 6-DOF GraspNet: Variational Grasp Generation for Object Manipulation. ICCV.
[5] Glytsos, M., Filntisis, P. P., Retsinas, G., Maragos, P. (2025). Category-Level 6D Object Pose Estimation in Agricultural Settings Using a Lattice-Deformation Framework and Diffusion-Augmented Synthetic Data (PLANTPose). arXiv:2505.24636.
[6] Chatzis, N., Tsinouka, A., et al., Filntisis, P. P. (2026). Mind the Shape Gap: A Benchmark and Baseline for Deformation-Aware 6D Pose Estimation of Agricultural Produce (PEAR/SEED). IROS 2026.

(*) R5. Audio-visual speaker localization, distance estimation, and diarization in multimodal, multichannel environments (Οπτικο-ακουστικός εντοπισμός, εκτίμηση απόστασης και καταλογοποίηση ομιλητών σε πολυτροπικά και πολυκαναλικά περιβάλλοντα)

Μια από τις προκλήσεις σε περιβάλλοντα πολλαπλών ομιλητών είναι ο εντοπισμός τους (localization) στο χώρο [1,2], συμπεριλαμβανομένης της εκτίμησης της κατεύθυνσης και της απόστασής τους από το σύστημα καταγραφής, όπως και ο προσδιορισμός του ποιος μιλάει πότε (diarization) [3]. Τέτοια περιβάλλοντα / σενάρια μπορεί να αφορούν αλληλεπίδραση ενός ρομπότ με ομάδα πολλαπλών χρηστών, μία συνάντηση / συζήτηση σε αίθουσα συνεδριάσεων γύρω από ένα τραπέζι συνεδριάσεων, ή πολλαπλά «παράθυρα» ομιλητών σε τηλεοπτικό ενημερωτικό πρόγραμμα. Η επίλυση των προβλημάτων αυτών είναι σημαντική για μια σειρά από downstream tasks, όπως αυτόματος διαχωρισμός, βελτίωση και αναγνώριση φωνής των διαφόρων ομιλητών, εξαγωγή μεταδεδομένων της όλης αλληλεπίδρασης, καθώς και ενεργή ρομποτική αλληλεπίδραση, όπου ένα ρομπότ μπορεί να εντοπίζει τον ενεργό ομιλητή και να προσεγγίζει αυτόνομα τη θέση του.

Στη διπλωματική αυτή, για την αποτελεσματική αντιμετώπιση αυτών των προβλημάτων, θα χρησιμοποιηθούν τεχνικές βαθιάς μηχανικής μάθησης βασιζόμενες σε δεδομένα πολλαπλών πολυτροπικών αισθητήρων, οπτικών (μίας ή περισσοτέρων καμερών) και ακουστικών (μικροφωνικών συστοιχιών). Ιδιαίτερη έμφαση θα δοθεί στην οπτικο-ακουστική εκτίμηση της χωρικής θέσης των ομιλητών, τόσο ως προς την κατεύθυνση όσο και ως προς την απόστασή τους, με στόχο την απόκτηση τρισδιάστατης πληροφορίας για τη θέση της ηχητικής πηγής. Θα χρησιμοποιηθούν διαθέσιμες βάσεις δεδομένων από τη βιβλιογραφία [4,5], ενώ θα δοκιμαστεί και πρακτική αλγοριθμική υλοποίηση στη ρομποτική πλατφόρμα TIAGo [6] που είναι διαθέσιμη στο εργαστήριο. Στο ρομποτικό σενάριο θα διερευνηθεί η αξιοποίηση της εκτιμώμενης θέσης του ενεργού ομιλητή για την αυτόνομη πλοήγηση και προσέγγιση του ρομπότ προς αυτόν.

(Συνεργάτες: Δρ. Κατερίνα Παπαδημητρίου, Αν. Καθ. Γεράσιμος Ποταμιάνος, Τμήμα ΗΜΜΥ, Παν/μιο Θεσσαλίας)

References:
[1] Fu et al., “AuralNet: Hierarchical attention-based 3D binaural localization of overlapping speakers,” *Proc. Interspeech*, 2025.
[2] Shimada et al., *Stereo sound event localization and detection in regular video content*, Available at: https://dcase.community/challenge2025/index#task3 (2025)
[3] N. Makishima et al., “Unified audio-visual modeling for recognizing which face spoke when and what in multi-talker overlapped speech and video,” *Proc. Interspeech*, 2025.
[4] *CHiME-9 Task 1 – McoRec*. Available at: https://www.chimechallenge.org/current/task1/index (2025)
[5] *The AMI corpus*. Available at: https://groups.inf.ed.ac.uk/ami/download/
[6] *PAL Robotics, TIAGo Robot - A Mobile Manipulator for Research and Industry*. Available at: https://pal-robotics.com/robot/tiago/

(*) R6. Learning to Climb: A Robot in the Playground

Οι δομές αναρρίχησης στις παιδικές χαρές αποτελούν απαιτητικά περιβάλλοντα, στα οποία τα παιδιά εξασκούν δεξιότητες όπως η ισορροπία, ο συντονισμός ολόκληρου του σώματος και ο σχεδιασμός κινήσεων. Η παρούσα διπλωματική εργασία διερευνά αν αντίστοιχες δεξιότητες μπορούν να αποκτηθούν από προσομοιωμένα ρομπότ μέσω αλληλεπίδρασης με ένα σύνθετο περιβάλλον τύπου “space net”. Θα αναπτυχθεί περιβάλλον προσομοίωσης σε MuJoCo, Isaac Gym ή παρόμοια πλατφόρμα, με έναν αρθρωτό ή ανθρωποειδή πράκτορα και ένα τρισδιάστατο άκαμπτο δίκτυο σχοινιών. Με χρήση μεθόδων ενισχυτικής μάθησης, ο πράκτορας θα εκπαιδευτεί σε εργασίες όπως η αναρρίχηση, η προσέγγιση συγκεκριμένων σημείων και η ασφαλής μετακίνηση στο δίκτυο. Αρχικά μπορούν να χρησιμοποιηθούν πλήρεις πληροφορίες για την κατάσταση του ρομπότ και του περιβάλλοντος. Ιδιαίτερο ενδιαφέρον παρουσιάζει η δυνατότητα μεταφοράς δεξιοτήτων μεταξύ διαφορετικών σωμάτων και η χρήση curriculum learning. Εναλλακτικά, μπορεί να μελετηθεί ιεραρχικός έλεγχος, με διαχωρισμό του

σχεδιασμού επαφών από τον χαμηλού επιπέδου έλεγχο κίνησης. Στόχος είναι η μελέτη της μάθησης επαναχρησιμοποιήσιμων κινητικών δεξιοτήτων.

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, Επ. Καθ. Παναγιώτης Βαρθολομαίος, ΥΔ Στέλιος Ζαρίφης)

References:

[1] C. Pawlowski, et al. "The role of playgrounds in the development of children's fundamental movement skills: A scoping review." PLoS one 18.12 (2023): e0294296.
[2] S. Zhao, et al. "LadderMan: Learning Humanoid Perceptive Ladder Climbing." arXiv preprint arXiv:2606.05873 (2026).
[3] X. Sun, et al. "Planning and control of stable ladder climbing motion for the four-limbed robot "WAREC-1"." 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.
[4] E. Todorov, T. Erez, and Y. Tassa. "MuJoCo: A physics engine for model-based control." 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2012, pp. 5026–5033.

(*) R7. Ενσωμάτωση της πρόβλεψης ανθρώπινων δράσεων σε μοντέλα Vision-Language-Action σε συνεργατικές διαδικασίες

Τα μοντέλα Vision-Language-Action (VLA) [1, 2] μετατρέπουν εικόνα και γλωσσικές εντολές σε ενέργειες ρομπότ, αλλά δρουν μόνο όταν ο άνθρωπος διατυπώνει ρητά τι χρειάζεται. Στην πραγματική συνεργασία, όμως, ένας καλός συνεργάτης προβλέπει τις ανάγκες του άλλου και ενεργεί πριν του ζητηθεί. Η προληπτική ρομποτική βοήθεια έχει μελετηθεί κυρίως με ρητά μοντέλα της εργασίας και των προτιμήσεων του χρήστη, π.χ. σε διαδικασίες συναρμολόγησης [3], ενώ η πρόβλεψη ανθρώπινων δράσεων από βίντεο έχει μελετηθεί εκτενώς [4], χωρίς όμως να έχει ενσωματωθεί στην ίδια την πολιτική του ρομπότ. Επιπλέον, τα περισσότερα VLA αποφασίζουν μόνο από την τρέχουσα παρατήρηση, ενώ σε διαδικασίες πολλών βημάτων η γνώση του τι έχει ήδη γίνει είναι απαραίτητη [5]. Στόχος της διπλωματικής είναι η μελέτη του κατά πόσο τα σημερινά VLA μπορούν να υποστηρίξουν προληπτικά έναν άνθρωπο σε μια συνεργατική διαδικασία, όπως η συναρμολόγηση, και η πρόταση τροποποιήσεων που τα καθιστούν ικανά γι' αυτό. Η μελέτη θα γίνει σε έναν απλό βραχίονα της πλατφόρμας LeRobot, όπου VLA ανοιχτού κώδικα, όπως το SmolVLA [1] και το π0 [2], θα εκπαιδευτούν περαιτέρω με επιδείξεις τηλεχειρισμού και θα συγκριθούν τόσο με ρητές εντολές όσο και με μια γενική οδηγία, όπου το ρομπότ πρέπει να καταλάβει από την κίνηση του ανθρώπου τι θα χρειαστεί στη συνέχεια. Θα διερευνηθούν τροποποιήσεις όπως η εισαγωγή tokens για την πόζα των χεριών και του σώματος του ανθρώπου, ένας μηχανισμός μνήμης δύο επιπέδων, όπου η βραχυπρόθεσμη μνήμη αποτυπώνει την πρόσφατη κίνηση του ανθρώπου και η μακροπρόθεσμη την πρόοδο της διαδικασίας με τη βοήθεια μιας βοηθητικής κεφαλής που προβλέπει το επόμενο βήμα [5], και μια ρητή ενέργεια αναμονής, ώστε το μοντέλο να μαθαίνει όχι μόνο τι να κάνει αλλά και πότε να παρέμβει.

(Συνεργάτες: Δρ. Νίκη Ευθυμίου, Δρ. Παναγιώτης Φιλντίσης)

References:

[1] M. Shukor, D. Aubakirova, F. Capuano, P. Kooijmans, S. Palma, A. Zouitine, M. Aractingi, C. Pascal, M. Russi, A. Marafioti, S. Alibert, M. Cord, T. Wolf, and R. Cadene, "SmolVLA: A Vision-Language-Action Model for Affordable and Efficient Robotics," arXiv:2506.01844, 2025.
[2] K. Black, N. Brown, D. Driess, et al., "π0: A Vision-Language-Action Flow Model for General Robot Control," in Proc. Robotics: Science and Systems (RSS), 2025.
[3] H. Nemlekar, N. Dhanaraj, A. Guan, S. K. Gupta, and S. Nikolaidis, "Transfer Learning of Human Preferences for Proactive Robot Assistance in Assembly Tasks," in Proc. ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2023.
[4] R. Girdhar and K. Grauman, "Anticipative Video Transformer," in Proc. IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
[5] H. Shi et al., "MemoryVLA: Perceptual-Cognitive Memory in Vision-Language-Action Models for Robotic Manipulation," in Proc. International Conference on Learning Representations (ICLR), 2026.

(*) R8. Ρομποτικός χειρισμός με μοντέλα όρασης–γλώσσας και οπτική ανατροφοδότηση

Τα μεγάλα μοντέλα όρασης–γλώσσας (Vision-Language Models, VLMs) και όρασης-γλώσσας-δρασης (Vision-Language-Action Models, VLAs) προσφέρουν τη δυνατότητα σύνδεσης γλωσσικών οδηγιών και οπτικών παρατηρήσεων με ενέργειες ρομποτικού χειρισμού. Σχετικές προσεγγίσεις αξιοποιούν γλωσσικά μοντέλα για τη σύνθεση προγραμμάτων ελέγχου [1], τη σύνδεση οδηγιών με χωρικές αναπαραστάσεις της σκηνής [2] και την αναθεώρηση σχεδίων βάσει ανατροφοδότησης από το περιβάλλον [3]. Στόχος της διπλωματικής είναι η ανάπτυξη

και αξιολόγηση ενός συστήματος στο οποίο ένα προεκπαιδευμένο VLM καθοδηγεί έναν πραγματικό ρομποτικό βραχίονα SO-101 μέσω της πλατφόρμας LeRobot, λαμβάνοντας εικόνες από κάμερα και οδηγίες φυσικής γλώσσας. Το μοντέλο θα επιλέγει και θα παραμετροποιεί βασικές ενέργειες, όπως μετακίνηση της αρπάγης, περιστροφή του καρπού και άνοιγμα ή κλείσιμο της λαβής, μέσω δομημένων κλήσεων εργαλείων (tool calling). Μετά την εκτέλεση κάθε ενέργειας, θα λαμβάνει νέα παρατήρηση ώστε να ελέγχει την πρόοδο και να διορθώνει την επόμενη κίνησή του. Η μελέτη θα εστιάσει σε εργασίες επιτραπέζιου χειρισμού, όπως η μεταφορά και τοποθέτηση αντικειμένων σύμφωνα με χωρικές οδηγίες. Θα διερευνηθούν η αναγνώριση αποτυχημένων ενεργειών, η διατήρηση συνοπτικής μνήμης των προηγούμενων προσπαθειών και η προσαρμογή σε αλλαγές της σκηνής. Θα συγκριθούν η εκτέλεση ενός αρχικού σχεδίου, η διαδοχική λήψη αποφάσεων με οπτική ανατροφοδότηση και η αξιοποίηση ιστορικού ενεργειών–αποτελεσμάτων. Η αξιολόγηση θα βασιστεί στην επιτυχία των εργασιών, στον αριθμό διορθωτικών κινήσεων, στον χρόνο ολοκλήρωσης και στη γενίκευση σε νέες διατάξεις αντικειμένων.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, Δρ. Νίκη Ευθυμίου, ΥΔ Παναγιωτης Ροδίτης)

References:
[1] J. Liang et al., "Code as Policies: Language Model Programs for Embodied Control," arXiv:2209.07753, 2022.
[2] W. Huang et al., "VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models," CoRL, 2023.
[3] W. Huang et al., "Inner Monologue: Embodied Reasoning through Planning with Language Models," CoRL, 2022.

(*) R9. Imitation Learning: Μάθηση συναισθηματικά διαμορφωμένων αντιδράσεων αποφυγής για ανθρωποειδή ρομπότ

Όταν ένα αντικείμενο πλησιάζει απότομα, ο άνθρωπος εκδηλώνει συντονισμένες αμυντικές αντιδράσεις, όπως σκύψιμο, απομάκρυνση του κορμού ή ανύψωση των χεριών. Η μορφή τους όμως δεν καθορίζεται μόνο από το ερέθισμα αλλά και από την εσωτερική του κατάσταση: ένας αιφνιδιασμένος άνθρωπος αντιδρά ταχύτερα και με πιο έντονη, ολόσωμη κίνηση, ενώ όταν περιμένει το ερέθισμα εκτελεί μια ελάχιστη, στοχευμένη αποφυγή. Η οπτική προσέγγιση αντικειμένων (looming) αποτελεί θεμελιώδες ερέθισμα για ανθρώπινες αμυντικές και συναισθηματικές αποκρίσεις [1]. Στη ρομποτική έχουν αναπτυχθεί μέθοδοι προστατευτικού ελέγχου [2] και μεταφοράς ανθρώπινης κίνησης σε ανθρωποειδή ρομπότ [3], χωρίς όμως να λαμβάνεται υπόψη αυτή η συναισθηματική διάσταση. Στόχος της διπλωματικής είναι η μάθηση αντιδράσεων αποφυγής από ανθρώπινες επιδείξεις, ώστε ένα ανθρωποειδές ρομπότ να αποκρίνεται με ανθρώπινο τρόπο ανάλογα με το επίπεδο αιφνιδιασμού ή διέγερσης. Θα δημιουργηθεί σύνολο δεδομένων ανθρώπινων αντιδράσεων σε ελεγχόμενες, ασφαλείς προσεγγίσεις αντικειμένων με διαφορετικές κατευθύνσεις και ταχύτητες, υπό συνθήκες αναμονής και αιφνιδιασμού, καθώς και σε σενάρια αυξημένης διέγερσης (π.χ. τύπου escape room), με δυνατότητα συμπληρωματικής καταγραφής φυσιολογικών σημάτων (π.χ. ηλεκτροδερμική δραστηριότητα, καρδιακός ρυθμός). Από συγχρονισμένες καταγραφές θα εκτιμώνται η τροχιά του αντικειμένου, η τρισδιάστατη στάση του σώματος, ο λανθάνων χρόνος και το πλάτος της αντίδρασης. Θα αναπτυχθούν μοντέλα που αντιστοιχίζουν τα χαρακτηριστικά του ερεθίσματος και την κατάσταση διέγερσης σε αποκρίσεις του κορμού και των χεριών. Οι κινήσεις θα προσαρμοστούν στην κινηματική του ρομπότ (motion retargeting) και θα αξιοποιηθούν για μάθηση πολιτικών αντίδρασης εξαρτημένων από τη συναισθηματική κατάσταση, αρχικά σε προσομοίωση και στη συνέχεια σε επιλεγμένα πειράματα σε πραγματικό humanoid. Η αξιολόγηση θα εξετάζει την αποτελεσματικότητα και τον χρονισμό της αποφυγής, τη διατήρηση της ισορροπίας, την ομοιότητα με ανθρώπινες αντιδράσεις, τη γενίκευση σε νέες τροχιές, καθώς και το κατά πόσο ανθρώπινοι παρατηρητές αναγνωρίζουν το επίπεδο αιφνιδιασμού στη συμπεριφορά του ρομπότ.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΥΔ Νικόλας Χατζής)

References:
[1] Thieu et al., "Visual looming is a primitive for human emotion," iScience, 27, 109886, 2024.
[2] T. Shimizu et al., "Self-protective whole body motion for humanoid robots based on synergy of global reaction and local reflex," Neural Networks, 32, pp. 109–118, 2012.
[3] T. He et al., "OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning," CoRL, 2024.

Οραση Υπολογιστών:

V1. Three-dimensional Shape: Analysis, Modeling, Matching

References:
[1] A. Bronstein, M. Bronstein, and R. Kimmel. Numerical geometry of non-rigid shapes. Springer, 2008.
[2] M. Breuss, A. Bruckstein and P. Maragos (Eds.), Innovations for Shape Analysis: Models and Algorithms, Springer, 2013.
[3] J. Sun, M. Ovsjanikov, L. Guibas, "A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion", Eurographics Symposium on Geometry Processing 2009.
[4] P. Tzathas, P. Maragos and A. Roussos, "3D Neural Sculpting (3DNS): Editing Neural Signed Distance Functions", Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2023), Hawaii, Jan. 2023.

V2. Sign-Language Understanding using Computer Vision and Deep Learning

(*) V2.1 Sign-Language Recognition using Deep Learning (Αναγνώριση χειρονομιών και συνεχών βίντεο νοηματικής γλώσσας με την χρήση βαθιών νευρωνικών δικτύων).

Η αυτόματη αναγνώριση νοηματικής γλώσσας (Sign Language Recognition – SLR) στοχεύει στην κατανόηση νοηματικών εκφράσεων από βίντεο και στη μετατροπή τους σε γλωσσικές αναπαραστάσεις, όπως ακολουθίες glosses ή κείμενο. Η διαδικασία αυτή απαιτεί την από κοινού μοντελοποίηση των κινήσεων των χεριών, του σώματος και του προσώπου, καθώς και των χρονικών εξαρτήσεων που εμφανίζονται στη συνεχή νοηματική. Σύγχρονες προσεγγίσεις αξιοποιούν πολυτροπικές αρχιτεκτονικές, εκτίμηση 2D/3D στάσης σώματος και χεριών, graph-based μοντέλα και μηχανισμούς προσοχής, με στόχο τη βελτίωση της ακρίβειας και της γενίκευσης σε διαφορετικούς νοηματιστές και συνθήκες λήψης.

Στόχος της διπλωματικής είναι η μελέτη και ανάπτυξη σύγχρονων μεθόδων βαθιάς μάθησης για απομονωμένη και συνεχή αναγνώριση νοηματικής γλώσσας, καθώς και για συνεχή δακτυλογράφηση. Επιπλέον, θα διερευνηθεί η ενσωμάτωση των αναπτυσσόμενων μεθόδων σε εφαρμογές αλληλεπίδρασης ανθρώπου–ρομπότ, όπου το ρομπότ θα μπορεί να αναγνωρίζει νοηματικές εκφράσεις σε πραγματικό χρόνο και να τις αξιοποιεί ως φυσικό μέσο επικοινωνίας ή ως εντολές για την εκτέλεση κατάλληλων ενεργειών.

(Συνεργάτες: Δρ. Κατερίνα Παπαδημητρίου, Αν. Καθ. Γεράσιμος Ποταμιάνος, Τμήμα ΗΜΜΥ, Παν/μιο Θεσσαλίας)

References:
[1] S. Theodorakis, V. Pitsikalis, and P. Maragos, "Dynamic–static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition", Image and Vision Computing, 32 (8), pp.533–549, 2014.
[2] A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos, "Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos", J. Machine Learning Research, 2013.
[3] O. Koller, N.C. Camgoz, H. Ney, and R. Bowden, "Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos," IEEE Trans. PAMI, 2020. DOI: 10.1109/TPAMI.2019.2911077
[4] M. Parelli, K. Papadimitriou, G. Potamianos, G. Pavlakos, and P. Maragos, "Exploiting 3D Hand Pose Estimation in Deep Learning-based Sign Language Recognition from RGB Videos", Proc. ECCV Workshops (ECCVW) - Workshop on Sign Language Recognition, Translation and Production (SLRTP), 2020.
[5] M Parelli, K Papadimitriou, G Potamianos, G Pavlakos, and P. Maragos, "Spatio-Temporal Graph Convolutional Networks for Continuous Sign Language Recognition", Proc. ICASSP-2022.
[6] A. Kratimenos, G. Pavlakos and P. Maragos, "Independent Sign Language Recognition With 3D Body, Hands, And Face Reconstruction", Proc. ICASSP, 2021.
[7] C. C. de Amorim, D. Macêdo, and C. Zanchettin, "Spatial-Temporal Graph Convolutional Networks for Sign Language Recognition", https://arxiv.org/abs/1901.11164, 2019.
[8] K Papadimitriou, G Potamianos, G Sapountzaki, T Goulas, E Efthimiou, S -E Fotinea, and P. Maragos, "Greek sign language recognition for an education platform", Universal Access in the Information Society, 2023.
[9] K. Papadimitriou, P. Filntisis, G. Retsinas, G. Potamianos, and P. Maragos, "Seeing in 2D, Thinking in 3D: 3D Hand Mesh-Guided Feature Learning for Continuous Fingerspelling", *Proc. ICCV Workshops (ICCVW) - Workshop on Vision Foundation Models and Generative AI for Accessibility: Challenges and Opportunities (CV4A11Y), 2025*.

(*) V2.2. Deep Learning for Sign-Language Production and Embodied Human–Robot Interaction (Βαθιά Μάθηση για Παραγωγή Νοηματικής Γλώσσας και Ενσώματη Αλληλεπίδραση Ανθρώπου–Ρομπότ)

Η αυτόματη παραγωγή νοηματικής γλώσσας (Sign Language Production – SLP) αφορά τη μετατροπή κειμένου ή ομιλίας σε φυσικές και γλωσσικά ορθές ακολουθίες νοηματικής. Οι υπάρχουσες προσεγγίσεις βασίζονται συχνά σε τρισδιάστατα avatars, τα οποία παρουσιάζουν περιορισμούς ως προς τον ρεαλισμό της κίνησης και την ακριβή αναπαράσταση σημαντικών μη χειροκινησιακών χαρακτηριστικών, όπως οι κινήσεις της κεφαλής και οι εκφράσεις του προσώπου [1, 2, 3]. Στόχος της διπλωματικής είναι η μελέτη σύγχρονων μεθόδων βαθιάς μάθησης, όπως Transformers και diffusion models, για την παραγωγή τρισδιάστατων ακολουθιών νοηματικής από κείμενο [4, 5, 6], καθώς και η σύνθεση φωτορεαλιστικών βίντεο μέσω τεχνικών neural rendering και generative video synthesis [1, 3, 5].
Επιπλέον, θα διερευνηθεί η αξιοποίηση των παραγόμενων κινήσεων σε εφαρμογές αλληλεπίδρασης ανθρώπου–ρομπότ. Ειδικότερα, θα εξεταστεί η μεταφορά (motion retargeting) των τρισδιάστατων κινήσεων νοηματικής σε ρομποτικές πλατφόρμες, λαμβάνοντας υπόψη τους βαθμούς ελευθερίας και τους κινηματικούς περιορισμούς του ρομπότ [7, 8].

(Συνεργάτες: Δρ. Κατερίνα Παπαδημητρίου, Δρ. Παναγιώτης Φιλντίσης, Δρ. Αναστάσιος Ρούσσος, Διευθυντής Ερευνών, Ινστιτούτο Πληροφορικής, ΙΤΕ)

References:
[1] Saunders, B., Camgoz, N. C., & Bowden, R. (2022). Signing at scale: Learning to co-articulate signs for large-scale photo-realistic sign language production. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5141-5151).
[2] Forte, M. P., Kulits, P., Huang, C. H. P., Choutas, V., Tzionas, D., Kuchenbecker, K. J., & Black, M. J. (2023). Reconstructing Signing Avatars From Video Using Linguistic Priors. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR-2023).
[3] Tze, C. O., Filntisis, P. P., Dimou, A. L., Roussos, A., & Maragos, P. (2023). "Neural Sign Reenactor: Deep Photorealistic Sign Language Retargeting". CVPRW Workshop AI4CC.
[4] Saunders, B., Camgoz, N. C., & Bowden, R. (2020). Progressive Transformers for End-to-End Sign Language Production. In Proceedings of the European Conference on Computer Vision (pp. 687-705).
[5] Pratikaki, C., Fotinea, S.-E., Efthymiou, E., Filntisis, P. P., Roussos, A., & Maragos, P. (2025). Text-to-Sign Language production via intermediate skeletal representations using transformers and neural rendering. Proceedings of the 9th Workshop on Sign Language Translation and Avatar Technologies (SLTAT).
[6] Pratikaki, C., Filntisis, P. P., Efthymiou, E., Fotinea, S.-E., Roussos, A., & Maragos, P. (2025). A transformer-based framework for Greek Sign Language production and using Extended Skeletal Motion Representations. *Proc International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2025)*. ***PETRA-2025 Best Paper Award***
[7] Liang, Y., Li, W., Wang, Y., Xiong, R., Mao, Y., and Zhang, J. (2021). Dynamic Movement Primitive based Motion Retargeting for Dual-Arm Sign Language Motions. In Proceedings of the IEEE International Conference on Robotics and Automation (pp. 8195–8201).
[8] Zhang, H., Xu, H., Wang, Y., and Xiong, R. (2024). Human-Robot Sign Language Motion Retargeting from Videos. In Proceedings of the Chinese Control Conference (pp. 4736-4741).

(*) V3. 3D face (and skull) modeling/reconstruction and applications (photorealistic avatars, interactive museum installations, craniofacial reconstruction)

The human face is one of the most commonly-considered objects in Computer Vision and Graphics. Modelling and reconstructing the detailed 3D shape and dynamics of the human face has numerous applications, such as augmented reality, performance capture, computer games, visual effects, human-computer interaction, computer-aided craniofacial surgery, rehabilitation and research in psychology. During the last years we have proposed state-of-the-art methods for 3D face modelling, 3D face reconstruction from in-the-wild images and videos, facial expression recognition and photo-realistic synthesis of facial videos, see e.g. [1-9]. These methods have yielded promising results, and there are many ways they can be improved, extended, or applied to new problems.

Particularly exciting directions lie at the intersection of cultural heritage and 3D face modelling. Real-time 3D face tracking and photorealistic synthesis enable interactive museum installations: visitors can see themselves transformed in a virtual mirror [9], animate historical figures with their own expressions, or interact with expressive digital guides and avatars. These systems raise concrete research challenges in real-time robustness, rendering quality and inclusive face transformations, and they can be evaluated with real audiences. A further direction concerns the human skull and its relation to the face. Bidirectional craniofacial reconstruction means predicting the face from a skull and vice versa, and it is a highly ill-posed problem, since many faces are plausible

for a given skull. Possible projects include learned 3D models of skull shape, probabilistic/generative methods that predict a distribution of plausible faces rather than a single estimate, and robustness to damaged or incomplete skulls. Applications range from reconstructing the faces of ancient individuals for museum exhibits to forensics and craniofacial surgery.

(Συνεργάτες: Δρ. Αναστάσιος Ρούσσος, Διευθυντής Ερευνών, Ινστιτούτο Πληροφορικής, ΙΤΕ, και Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:
[1] M.C. Doukas, M.R. Koujan, V. Sharmanska, A. Roussos, and S. Zafeiriou, "Head2Head++: Deep Facial Attributes Re-Targeting", IEEE Trans. Biometrics, Behavior, and Identity Science (T-BIOM), Vol. 3 (1), pp.31-43, Jan. 2021.
[2] M.R. Koujan,…, A. Roussos, "Real-time Facial Expression Recognition "In The Wild" by Disentangling 3D Expression from Identity", Proc. Int'l Conf. Automatic Face and Gesture Recognition (FG 2020).
[3] M.R. Koujan, A. Roussos and S. Zafeiriou. "DeepFaceFlow: In-the-Wild Dense 3D Facial Motion Estimation", International Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[4] J. Booth, A. Roussos et al, "3D reconstruction of in-the-Wild Faces in Images and Videos", PAMI 2018.
[5] J. Booth, A. Roussos, …, S. Zafeiriou, "Large Scale 3D Morphable Face Models", IJCV 2017.
[6] F. Paraperas Papantoniou, P.P. Filntisis, P. Maragos, and A. Roussos. "Neural Emotion Director: Speech-preserving semantic control of facial expressions in "in-the-wild" videos", Proc. IEEE/CVF Computer Vision and Pattern Recognition (CVPR), 2022. ***Best Paper Finalist (top 33 out of 8,161 submissions).***
[7] P.P. Filntisis, G. Retsinas, F. Paraperas Papantoniou, A. Katsamanis, A. Roussos and P. Maragos. "SPECTRE: Visual Speech-Aware Perceptual 3D Facial Expression Reconstruction from Videos", Proc. IEEE/CVF CVPR 5th Workshop and Competition on Affective Behavior Analysis in-the-wild (ABAW), June 2023.
[8] G. Retsinas, P.P. Filntisis, R. Danecek, V.F. Abrevaya, A. Roussos, T. Bolkart, and P. Maragos, "SMIRK: 3D Facial Expressions through Analysis-by-Neural-Synthesis", Proc. IEEE/CVF CVPR 2024.
[9] P. Valergaki, A. Argyros, G. Giannakakis, and A. Roussos, "Reflections on Diversity: A Real-time Virtual Mirror for Inclusive 3D Face Transformations", 6th Int'l Conf. in Electronic Engineering & Information Technology (EEITE). 2025.

(*) V4. Ανάλυση Ιατρικών Εικόνων με Βαθιά Μάθηση: Κατάτμηση και Ανίχνευση Κλινικά Σημαντικών Στοιχείων

Η αυτόματη ανάλυση ιατρικών εικόνων αποτελεί σημαντικό πεδίο εφαρμογής της σύγχρονης Όρασης Υπολογιστών, με εφαρμογές στην υποβοηθούμενη διάγνωση, την ανίχνευση παθολογικών ευρημάτων και την κατανόηση χειρουργικών επεμβάσεων. Στόχος της διπλωματικής είναι η μελέτη και ανάπτυξη μεθόδων βαθιάς μάθησης για την ανάλυση διαφορετικών τύπων ιατρικών δεδομένων, όπως CT, MRI και ενδοσκοπικές εικόνες ή βίντεο. Έμφαση θα δοθεί σε δύο βασικές κατευθύνσεις:

- Semantic/instance segmentation για την κατάτμηση ανατομικών δομών ή παθολογικών περιοχών. Ενδεικτικές εφαρμογές περιλαμβάνουν την κατάτμηση όγκων ή άλλων αλλοιώσεων σε CT/MRI, καθώς και την αναγνώριση ανατομικών και χειρουργικά σημαντικών περιοχών σε ενδοσκοπικές εικόνες.

- Detection και localization για τον αυτόματο εντοπισμό παθολογικών ευρημάτων ή άλλων κλινικά σημαντικών στοιχείων σε ιατρικές εικόνες και βίντεο. Ενδεικτικές εφαρμογές αποτελούν η ανίχνευση πρώιμων νεοπλασιών σε ενδοσκοπικές εικόνες, καθώς και ο εντοπισμός χειρουργικών εργαλείων.

Στο πλαίσιο της εργασίας μπορούν να μελετηθούν σύγχρονες αρχιτεκτονικές CNNs, Vision Transformers και foundation models, καθώς και τεχνικές transfer learning και fine-tuning.

(Συνεργάτες: Δρ. Κατερίνα Παπαδημητρίου, Υ.Δ. Νίκος Κάρδαρης)

References:

[1] D. Gorpas, P. Maragos, and D. Yova, "A new morphological segmentation algorithm for biomedical imaging applications," In *Proceedings of SPIE*, 2009.
[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," In *Proceedings of MICCAI*, 2015.
[3] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, 2021.
[4] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, M. P. Lungren, S. Zhang, L. Xing, L. Lu, A. Yuille and Y. Zhou, "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, 2024.
[5] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, 2024.
[6] A. Zia, M. Berniker, R. G. Nespolo et al., "Surgical tool classification and localization: results and methods from the MICCAI 2022 SurgToolLoc challenge," 2023, arXiv:2305.07152.

[7] W. Yue, J. Zhang, K. Hu, Y. Xia, J. Luo, and Z. Wang, “SurgicalSAM: Efficient Class Promptable Surgical Instrument Segmentation,” In Proceedings of the AAAI Conference on Artificial Intelligence, 2024.
[8] Antonelli, M., Reinke, A., Bakas, S. et al. The Medical Segmentation Decathlon. Nat Commun 13, 4128 (2022). https://doi.org/10.1038/s41467-022-30695-9

Επεξεργασία Ακουστικών σημάτων (π.χ. μουσικής, φωνής) ή Οπτικών σημάτων σε Πολυμεσικά περιβάλλοντα:

AVM1. Υπολογιστική επεξεργασία μουσικών σημάτων και γενικότερα μουσικής πληροφορίας
(Συνεργάτες: Δρ. Νάνσυ Ζλατίντση και ΥΔ Χρήστος Γαρούφης, ΕΜΠ)

(*) AVM1.1: Αυτο-επιβλεπόμενη εκμάθηση μουσικών αναπαραστάσεων: Στόχος αυτής της διπλωματικής εργασίας είναι η ανάπτυξη πρωτοκόλλων αντιθετικής/αυτο-επιβλεπόμενης μάθησης (contrastive/self-supervised learning) για την εκμάθηση αναπαραστάσεων μουσικών σημάτων [1-8], χρήσιμων για πληθώρα προβλημάτων μουσικής κατανόησης. Ενδεικτικά, ως κατευθύνσεις και πιθανά προβλήματα-στόχους αναφέρουμε 1) τη μοντελοποίηση της φωνητικής ομοιότητας [8] μέσω κατάλληλα καθοδηγούμενης [9, 10] αυτο-επιβλεπόμενης δημιουργίας συνθετικών μουσικών κομματιών κατά την προ-εκπαίδευση, 2) την ανάπτυξη αντίστοιχων πρωτοκόλλων για μικροσκοπικά (frame-wise) προβλήματα μουσικής κατανόησης [11, 12, 13] και 3) την εφαρμογή ισαλλοίωτης (equivariant) μάθησης [6,7] για την εκμάθηση ενός ερμηνεύσιμου (explainable) ενδιάμεσου χώρου ως προς συγκεκριμένα φωνητικά ή μουσικά χαρακτηριστικά.

Datasets:

- MagnaTagATune: https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset
- MTG-Jamendo: https://github.com/MTG/mtg-jamendo-dataset
- Music4All: https://sites.google.com/view/contact4music4all (available at the lab server)

References:

[1] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G.. A simple framework for contrastive learning of visual representations, in Proc. ICML 2020
[2] Saeed, A., Grangier, D., & Zeghidour, N. Contrastive learning of general-purpose audio representations. In *Proc. ICASSP 2021*
[3] Spijkervet, Janne, and John Ashley Burgoyne. "Contrastive learning of musical representations.", in Proc. ISMIR 2021
[4] Alonso-Jiménez, Pablo, Xavier Serra, and Dmitry Bogdanov. "Music representation learning based on editorial metadata from discogs.", in Proc. ISMIR 2022
[5] Garoufis, Christos, Athanasia Zlatintsi, and Petros Maragos. "Multi-Source Contrastive Learning from Musical Audio.", in Proc. SMC 2023
[6] Quinton, Elio, “Equivariant Self-Supervision for Musical Tempo Estimation”, in Proc. ISMIR 2022
[7] Riou, Alain, Stefan Lattner, Gaëtan Hadjeres, and Geoffroy Peeters. "PESTO: Pitch Estimation with Self-Supervised Transposition-Equivariant Objective.", in Proc. ISMIR 2023
[8] Garoufis, Christos, Athanasia Zlatintsi, and Petros Maragos. "CVSM: Contrastive Vocal Similarity Modeling", under revision in IEEE Trans. on Audio, Speech and Language Proc.
[9] Ciranni, Ruben, Giorgio Mariani, Michele Mancusi, Emilian Postolache, Giorgio Fabbro, Emanuele Rodola and Luca Cosmo, “COCOLA: Coherence-Oriented Contrastive Learning of Musical Audio Representations”, in Proc. ICASSP 2025
[10] Riou, Alain, Stefan Lattner, Gaetan Hadjeres, Michael Anslow, and Geoffrey Peeters, “Stem-JEPA: A Joint-Embedding Predictive Architecture for Musical Stem Compatibility Estimation”, in Proc. ISMIR 2024
[11] Cuesta, Helena, Brian McFee, and Emilia Gómez, “Multiple F0 Estimation in Vocal Ensembles using Convolutional Neural Networks”, in Proc. ISMIR 2020
[12] Demirel, Emir, Sven Ahlback, and Simon Dixon, “MSTRE-Net: Multi-Streaming Acoustic Modeling for Automatic Lyrics Transcription”, in Proc. ISMIR 2021
[13] Y. Wu, C. Tsirigotis, K. Chen, C.Z.A Huang, A. Courville, O. Nieto, P. Seetharaman, and J. Salamon, FLAM: Frame-Wise Language-Audio Modeling, in Proc. ICML 2025

AVM1.2: Ανάπτυξη συστήματος εξατομικευμένων μουσικών προτάσεων
Αντικείμενο αυτής της διπλωματικής εργασίας είναι η δημιουργία ενός συστήματος μουσικών προτάσεων (music recommendation) ή συνέχισης playlists (playlist continuation), βάσει είτε μουσικών χαρακτηριστικών ή μουσικής ομοιότητας [1-3], είτε μεθόδους collaborative filtering [4-5] σε στατιστικά ακρόασης, με

αρχιτεκτονικές βασισμένες σε Variational Autoencoders [6-7] να πετυχαίνουν το state-of-the-art στην περίπτωση αυτή. Πιθανές επεκτάσεις των παραπάνω αφορούν την ενσωμάτωση/ανάπτυξη πολυεπίπεδων αρχιτεκτονικών VAEs ώστε να λαμβάνουν υπόψη την ιεραρχία των μεταδεδομένων (καλλιτέχνης, άλμπουμ, κομμάτι), καθώς και τη διερεύνηση της χρήσης ακουστικής πληροφορίας για την εξισορρόπηση των μουσικών προτάσεων προερχόμενων από μη-κυρίαρχες μουσικές κουλτούρες [8]. Επιπρόσθετα, μία δυνατότητα αφορά την προσθήκη εξατομίκευσης των προτάσεων, μέσω μεθοδολογιών ενεργής μάθησης (active learning) [9], ή αποσύνδεσης (decoupling) των διαφόρων μουσικών παραμέτρων, όπως η χροιά, το είδος, ή το επαγόμενο συναίσθημα [3,10].

Datasets:

- Music4All: https://sites.google.com/view/contact4music4all
- MusAV: https://mtg.github.io/musav-dataset
- MPD: https://research.atspotify.com/2020/09/the-million-playlist-dataset-remastered/

References:

[1] McFee, B., L. Barrington, and G. R. G. Lanckriet, "Learning Content Similarity for Music Recommendation", *IEEE Trans. on Audio, Speech & Language Processing* 20, no.8 (2012): 2207-2218

[2] Huang, Q., A. Jansen, L. Zhang, D.P. Ellis, R.A. Saurous, and J. Anderson, "Large-scale weakly-supervised content embeddings for music recommendation and tagging". *In Proc ICASSP 2020*

[3] Lee, J., Bryan, N. J., Salamon, J., Jin, Z., & Nam, J. . "Disentangled multidimensional metric learning for music similarity." In *Proc. ICASSP 2020*

[4] W. Bendada, G. Salha-Galvan, T. Bouabça, and T. Cazenave, "A Scalable Framework for Automatic Playlist Continuation on Music Streaming Services". In *Proc. SIGIR 2023*

[5] M. Volkovs, H. Rai, Z. Cheng, G. Wu, Y. Lu, and S. Sanner, "Two-stage Model for Automatic Playlist Continuation at Scale" In Proc.of the ACM Recommender Systems Challenge 2018, 2018.

[6] V. Vančura and P. Kordík. "Deep Variational Autoencoder with Shallow Parallel Path for Top-N Recommendation (VASP)." In Proc. ICANN 2021

[7] I. Shenbin, A. Alekseev, E. Tutubalina, V. Malykh, and S. I. Nikolenko. "REC-VAE: A New Variational Autoencoder for Top-N Recommendations with Implicit Feedback." In Proc. WSDM 2020

[8] O. Lesota et al. "Traces of Globalization in Online Music Consumption Patterns and Results of Recommendation Algorithms". in Proc. ISMIR 2022.

[9] Cañón, J. S. G., Cano, E., Yang, Y. H., Herrera, P., & Gómez, E. . Let's agree to disagree: Consensus Entropy Active Learning for Personalized Music Emotion Recognition., in Proc. ISMIR 2021

[10] Wang, X., Chen, H., Zhou, Y., Ma, J., & Zhu, W. (2022). Disentangled representation learning for recommendation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 408-424.

AVM1.3: Διαχωρισμός μουσικών πηγών

Αντικείμενο αυτής της διπλωματικής εργασίας αποτελεί το πρόβλημα του διαχωρισμού μουσικών πηγών (music source separation), της απομόνωσης δηλαδή των φωνητικών, ή των επιμέρους οργάνων, που αποτελούν ένα μουσικό κομμάτι. Το τελευταίο χρονικό διάστημα παρατηρείται μία τάση αντικατάστασης, ή ενίσχυσης, των παραδοσιακών, state-of-the-art αρχιτεκτονικών [1-4] με καθαρά παραγωγικές (generative) προσεγγίσεις, όπως αυτή των diffusion models [5-6]. Μία άλλη ενεργή πιθανή κατεύθυνση του προβλήματος είναι αυτή του query-based διαχωρισμού πηγών, όπου στόχος είναι ο διαχωρισμός μίας πηγής-στόχου από ένα ηχητικό μίγμα [7-8], ενώ με την πρόσφατη άνθηση των προ-εκπαιδευμένων μοντέλων [9] για προβλήματα μουσικής ταξινόμησης, παρουσιάζει ενδιαφέρον η προσαρμογή τους στο πρόβλημα του διαχωρισμού μουσικών πηγών [10-11].

Datasets:

- musdb18: https://sigsep.github.io/datasets/musdb.html
- URMP: https://labsites.rochester.edu/air/projects/URMP.html
- MoisesDB: https://github.com/moises-ai/moises-db

References:

[1] Kong, Q., Cao, Y., Liu, H., Choi, K., & Wang, Y. Decoupling magnitude and phase estimation with deep resunet for music source separation, in Proc. ISMIR 2021
[2] Défossez, A.. Hybrid spectrogram and waveform source separation, in Proc. MDX Workshop 2021
[3] Papantonakis, P., Garoufis, C., & Maragos, P. Multi-band Masking for Waveform-based Singing Voice Separation. In Proc. EUSIPCO 2022
[4] Luo, Yi, and Jianwei Yu. "Music Source Separation With Band-Split RNN." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023).
[5] Plaja-Roglans, G., Miron, M., & Serra, X. A diffusion-inspired training strategy for singing voice extraction in the waveform domain., in Proc. ISMIR 2022
[6] Mariani, G., Tallini, I., Postolache, E., Mancusi, M., Cosmo, L., & Rodolà, E. Multi-source diffusion models for simultaneous music generation and separation. *arXiv preprint arXiv:2302.02257*.
[7] Lin, L., Kong, Q., Jiang, J., & Xia, G. A unified model for zero-shot music source separation, transcription and synthesis., in Proc. ISMIR 2021
[8] Kong, Q., Chen, K., Liu, H., Du, X., Berg-Kirkpatrick, T., Dubnov, S., & Plumbley, M. D.,. Universal Source Separation with Weakly Labelled Data. *arXiv preprint arXiv:2305.07447*.
[9] Li, Y., R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao et al. "MERT: Acoustic Music Understanding Model with Large-Scale Self-Supervised Training", in Proc. ICLR 2024
[10] Chen, Z., Kanda, N., Wu, J., Wu, Y., Wang, X., Yoshioka, T., Li, J., Sivasankaran, S. and Eskimez, S.E., "Speech Separation with Large-Scale Self-Supervised Learning", in Proc. ICASSP 2023
[11] Chen, K., Wichern, G., Germain, F. G., & Le Roux J., "PaCᗡ-HuBERT: Self-Supervised Music Source Separation Via Primitive Auditory Clustering And Hidden-Unit BERT" in Proc. ICASSP-W 2023

AVM2. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα AM-FM και αλγορίθμους πολυζωνικής ενεργειακής αποδιαμόρφωσης, και σχετικά μη-γραμμικά συστήματα.

References:
[1] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE Trans. Signal Processing*, vol.41, no.10, pp.3024-3051, Oct. 1993.
[2] A. Potamianos and P. Maragos, "Speech Processing Applications Using an AM-FM Modulation Model", *Speech Communication*, vol.28, no.3, pp.195-209, July 1999.
[3] D. Dimitriadis, P. Maragos, and A. Potamianos, "Robust AM-FM Features for Speech Recognition", *IEEE Signal Processing Letters*, vol.12, no.9, pp.621-624, Sep. 2005.
[4] J. H.L. Hansen and S. Patil, *Speech Under Stress: Analysis, Modeling and Recognition*, Springer LNAI 4343, 2007.
[5] D. Dimitriadis, P. Maragos and A. Potamianos, "On the Effects of Filterbank Design and Energy Computation on Robust Speech Recognition", *IEEE Trans. Audio, Speech and Language Processing*, vol.19, pp.1504-1516, Aug.2011.
[6] A. Zlatintsi and P. Maragos, "Comparison of Different Representations Based on Nonlinear Features for Music Genre Classification" , *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
[7] T. Chaspari, D. Dimitriadis and P. Maragos, "Emotion Classification of Speech Using Modulation Features", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
[8] I. Rodomagoulakis and P. Maragos, "On the Improvement of Modulation Features Using Multi-Microphone Energy Tracking for Robust Distant Speech Recognition", *Proc. EUSIPCO-2017*, Kos, Greece, Aug. 2017.
[9] H. B. Sailor, M. R. Kamble and H. A. Patil, "Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection", in *Proc. Interspeech 2018*.

AVM3. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα από φράκταλς και χάος, και σχετικά μη-γραμμικά συστήματα.

References:
[1] P. Maragos and A. Potamianos, "Fractal Dimensions of Speech Sounds: Computation and Application to Automatic Speech Recognition", *Journal of Acoustical Society of America*, vol.105 (3), pp.1925--1932, March 1999.
[2] V. Pitsikalis and P. Maragos, ``Filtered Dynamics and Fractal Dimensions for Noisy Speech Recognition", *IEEE Signal Processing Letters*, vol.13, no.11, pp.711-714, Nov. 2006..

[3] V. Pitsikalis and P. Maragos, "Analysis and Classification of Speech Signals by Generalized Fractal Dimension Features", *Speech Communication*, vol.51, no.12, pp.1206-1223, Dec. 2009.
[4] A. Zlatintsi and P. Maragos, "Multiscale Fractal Analysis of Musical Instrument Signals with Application to Recognition", *IEEE Trans. Audio, Speech and Language Processing*, vol.21, no.4, pp.737-748, Apr. 2013.
[5] Karmele López-de-Ipina et al., "Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach", *Computer Speech and Language* 30, pp.43–60, 2015.
[6] E. Tzinis, G. Paraskevopoulos, C. Baziotis, A. Potamianos, "Integrating Recurrence Dynamics for Speech Emotion Recognition", in *Proc. Interspeech 2018*.
[7] K. Avramidis, A. Zlatintsi, C. Garoufis, and P. Maragos, "Multiscale Fractal Analysis on EEG Signals for Music-Induced Emotion Recognition", *Proc. EUSIPCO, 2021*.
[8] W. Gilpin, "Deep Reconstruction of Strange Attractors from Time Series", NIPS 2020, *arXiv:2002.05909* (2020).
[9] P. Maragos, *Topics in Computer Vision and Machine Learning*, [Postgraduate textbook], KALLIPOS, Open Academic Editions, https://doi.org/10.57713/kallipos-1097 . [Chapter 5].

AVM4. Ανίχνευση βιοακουστικών γεγονότων με few-shot learning

Η ανίχνευση και ανάλυση βιοακουστικών σημάτων από φωνήματα θηλαστικών στη φύση είναι σημαντική για την μελέτη των πληθυσμών και της συμπεριφοράς τους ενώ ταυτόχρονα μπορεί να αποτελέσει ένα πολύτιμο εργαλείο για την προστασία απειλούμενων ειδών. Ο σκοπός της διπλωματικής εργασίας είναι η ανάπτυξη αποτελεσματικών αλγορίθμων, ικανών να κάνουν ταυτοποίηση και κατηγοριοποίηση φωνημάτων θηλαστικών στα χρονικά διαστήματα που αυτά συμβαίνουν με περιορισμένο αριθμό επισημειωμένων δειγμάτων. Συγκεκριμένα, θα χρησιμοποιηθεί το σύνολο δεδομένων του διαγωνισμού DCASE (Detection and Classification of Acoustics Scenes and Events - task 5). Στο σύνολο δεδομένων εκπαίδευσης (development set), τα δεδομένα είναι πλήρως επισημειωμένα με αναφορά στα χρονικά διαστήματα που ανιχνεύονται ήχοι (POS) ή δεν ανιχνεύονται (NEG) ή υπάρχει αβεβαιότητα (UNK) για κάθε κλάση ενώ το σύνολο αξιολόγησης (evaluation set) του αλγόριθμου περιλαμβάνει επισημειώσεις μόνο για τις πέντε πρώτες ανιχνεύσεις σε κάθε αρχείο. Το σύστημα βιοακουστικής ανίχνευσης που θα αναπτυχθεί στα πλαίσια της διπλωματικής, θα γενικεύει χαρακτηριστικά των φωνημάτων που θα εξάγει από τις 5 αναφορές ανά ηχητικό αρχείο για να ανιχνεύει επακόλουθα στιγμιότυπα των φωνημάτων στο υπόλοιπο αρχείο.

(Συνεργάτες: ΥΔ Δημήτρης Μακρόπουλος)

References:
[1] O.Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching Networks for One Shot Learning, Advances in Neural Information Processing Systems, 2016.
[2] J. Snell, K. Swersky, and R. Zemel,, Prototypical Networks for Few Shot Learning, Advances in Neural Information Processing Systems, 2017.
[3] L. You, E. P. Coyotl, S. Gunturu and M. Van Segbroeck, Transformer-Based Bioacoustic Sound Event Detection on Few-Shot Learning Tasks, in Proc. IEEE Int'l Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.
[4] J Liang, I Nolasco, B Ghani, H Phan, E Benetos, D Stowell, Mind the Domain Gap: a Systematic Analysis on Bioacoustic Sound Event Detection, Eusipco, 2024.
[5] E.Cakir, S.Adavanne, G.Parascandolo, K.Drossos, T.Virtanen, Convolutional recurrent neural networks for bird audio detection, Eusipco, 2017.
[6] D.Makropoulos, A.Tsiami, A.Prospathopoulos, D. Kassis, A. Frantzis, G. Piperakis, P. Maragos, Convolutional Recurrent Neural Networks for the Classification of Cetacean Bioacoustic Patterns, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

AVM7. Παραγωγή ανθρώπινης κίνησης από χωρικό ήχο (spatial audio) με βαθιά μάθηση

Η παραγωγή ρεαλιστικής ανθρώπινης κίνησης από ήχο έχει εφαρμογές σε εικονικούς χαρακτήρες, παιχνίδια και αλληλεπίδραση ανθρώπου-υπολογιστή. Πρόσφατες μέθοδοι αξιοποιούν ομιλία, γενικότερους ήχους, αλλά και τη χωρική πληροφορία του ήχου (spatial audio). Έτσι παράγουν κινήσεις σώματος που φαίνονται φυσικές και παραμένουν συνεπείς στον χρόνο [1-3].

Η διπλωματική θα μελετήσει τη γενίκευση μοντέλων που παράγουν ανθρώπινη κίνηση από spatial audio. Αρχικά θα χρησιμοποιηθούν το SAM dataset και το μοντέλο MOSPA [1], ώστε να αναπαραχθεί και να αξιολογηθεί το

υπάρχον baseline. Η αξιολόγηση θα γίνει με πρωτόκολλα γενίκευσης σε νέα άτομα, νέους ήχους και νέες χωρικές συνθήκες.

Στη συνέχεια θα εξεταστούν καλύτερες αναπαραστάσεις του ήχου και της χωρικής πληροφορίας, όπως προεκπαιδευμένοι audio encoders και εναλλακτικές κωδικοποιήσεις spatial audio. Παράλληλα θα μελετηθούν representation spaces της ανθρώπινης κίνησης [5], μέσα στα οποία θα γίνεται η παραγωγική μοντελοποίηση με Diffusion Models ή Flow Matching [3]. Στόχος είναι η καλύτερη γενίκευση και η παραγωγή συνεπούς κίνησης με απλούστερα objectives, με μικρότερη εξάρτηση από πολλαπλά, ειδικά σχεδιασμένα losses. Αν απαιτηθεί, μπορούν να συλλεχθούν και συμπληρωματικά δεδομένα ήχου και κίνησης.

References:

[1] S. Xu et al., "MOSPA: Human Motion Generation Driven by Spatial Audio," Advances in Neural Information Processing Systems (NeurIPS), 2025.
[2] S. Xu et al., "MoSAT: Human Motion Generation from Spatial Audio and Textual Description," arXiv:2609.23797, 2026.
[3] S. Alexanderson, R. Nagy, J. Beskow, and G. E. Henter, "Listen, Denoise, Action! Audio-Driven Motion Synthesis with Diffusion Models," ACM Transactions on Graphics (SIGGRAPH), vol. 42, no. 4, 2023.
[4] H. Liu et al., "BEAT: A Large-Scale Semantic and Emotional Multi-Modal Dataset for Conversational Gestures Synthesis," Proc. European Conference on Computer Vision (ECCV), 2022.
[5] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of Motion Capture as Surface Shapes," Proc. IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΥΔ Μάριος Γλυτσος NYU, Δρ. Νάνσυ Ζλατίντση)

(*) AVM8. Ανίχνευση και Κατηγοριοποίηση Βιοσημάτων Θαλάσσιων Θηλαστικών

Τα τελευταία χρόνια ερευνητικές ομάδες που δραστηριοποιούνται στην βιοακουστική έχουν αναπτύξει θεμελιώδη μοντέλα (foundation models) που προεκπαιδεύουν τα βάρη τους σε μεγάλες βάσεις βιοακουστικών δεδομένων με επιβλεπόμενη μάθηση. Οι βάσεις αυτές συνίστανται σε εκατομμύρια επισημειωμένους ήχους από κελαηδίσματα ή φωνήματα από δεκάδες χιλιάδες διαφορετικά είδη ζώων, κυρίως πουλιών, θηλαστικών, εντόμων από τις βάσεις XenoCanto, i-Naturalist, FSD50K. Παραδείγματα τέτοιων μοντέλων βιοακουστικής είναι οι διάφορες εκδοχές του PERCH (1.0, 2.0, SurfPERCH), το Birdnet, το AVES, AVES-Bird. Μετά την προ-εκπαίδευσή τους σε κάποιο βασικό δίκτυο κορμού (π.χ. EfficientNetB3), τα μοντέλα χρησιμοποιούνται για γραμμικό διαχωρισμό αναπαραστάσεων (embeddings) σε ακουστικά σύνολα που προέρχονται από απομακρυσμένα βιολογικά είδη. Τα δίκτυα αυτά χρησιμοποιούν αναλλοίωτα τα προ-εκπαιδευμένα βάρη τους για να μεταφέρουν γνώση (transfer learning) σε νέα σύνολα δεδομένων (datasets) με τεχνικές μάθησης με λίγα παραδείγματα (few-shot learning). Συγκεκριμένα προσαρμόζουν στον κορμό του δικτύου μια κεφαλή ταξινόμησης ως ένα πλήρως συνδεδεμένο στρώμα (fully connected layer) υλοποιώντας έτσι ένα γραμμικό ανιχνευτή (linear probing). Ενδεικτικά οι βάσεις δεδομένων στις οποίες ελέγχεται η ικανότητα γενίκευσης των θεμελιωδών μοντέλων είναι η DCLDE 2027, η NOAA PIPAN και το Reefset. Η διπλωματική εργασία έχει δύο διακριτές θεματικές διαβαθμισμένης δυσκολίας :

1. Την χρησιμοποίηση των επισημειωμένων δεδομένων του διαγωνισμού DCLDE 2027 για την ανίχνευση φωνημάτων από όρκες (Orcinus orca) σε θορυβώδεις υποθαλάσσιες καταγραφές και επακόλουθη κατηγοριοποίηση των ανιχνεύσεων σε τύπους φωνημάτων που αντιστοιχούν σε διαφορετικούς πληθυσμούς (Southern Resident Killer Whales ή Transient Killer Whales) στα πλαίσια του διαγωνισμού (challenge) που έχει προκηρύξει το workshop DCLDE 2027.
2. Την χρησιμοποίηση υποσυνόλων από τα δεδομένα που χρησιμοποιούν τα θεμελιώδη μοντέλα στην προ-εκπαίδευσή τους για την επίτευξη συγκρίσιμων ποσοστών επιτυχίας με αυτά σε προβλήματα μεταφοράς μάθησης (transfer learning) στα βιοακουστικά σύνολα που προαναφέρθηκαν (DCLDE 2027, η NOAA PIPAN και το Reefset).

References:

[1] Palmer, K.J., Cummings, E., Dowd, M.G. et al. A Public Dataset of Annotated Orcinus orca Acoustic Signals for Detection and Ecotype Classification. Sci Data 12, 1137 (2025). https://doi.org/10.1038/s41597-025-05281-5
[2] B. Ghani, T. Denton, S. Kahl, and H. Klinck. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Scientific Reports, 13(1):22876, Dec. 2023.
[3] Burns, Andrea et al. "Perch 2.0 transfers 'whale' to underwater tasks." ArXiv abs/2512.03219 (2025): n. pag.

[4] D. N. Makropoulos, C. Garoufis, A. Tsiami, P. P. Filntisis and P. Maragos, "Representation-Diverse Self-Supervision for Cross-Domain Bioacoustic Learning in Low-Resource Settings," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2026.
[5] M. Hagiwara. "AVES: Animal vocalization encoder based on self-supervision," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.
[6] M. Hagiwara, et al. " BEANS: The benchmark of animal sounds." In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1–5. IEEE, June 2023
[7] https://www.dclde.org/challenges-and-data-sets#detection

(Συνεργάτες: ΥΔ Δημήτρης Μακρόπουλος, Δρ. Νάνσυ Ζλατίντση, Δρ. Κατερίνα Παπαδημητρίου)

Συστήματα, Δίκτυα, Γράφοι, Optimization & Learning:

S1. ***Μη-γραμμικά δυναμικά συστήματα που χρησιμοποιούν max-plus άλγεβρα και finite-state automata με εφαρμογές σε ένα πρόβλημα από τις περιοχές ανίχνευσης, βελτιστοποίησης, δικτύων, ελέγχου, θεωρία γράφων, κ.ά.***

References:
[1] R. Cuninghame-Green, Minimax Algebra, Springer-Verlag, New York, 1979.
[2] P. Butkoviˇc. Max-linear Systems: Theory and Algorithms. Springer, 2010.
[3] P. Maragos, "Dynamical Systems on Weighted Lattices: General Theory", *Mathematics of Control, Signals and Systems*, 29:21, 2017.
[4] I. Kordonis, P. Maragos, G. P. Papavassilopoulos, "Stochastic Stability in Max-Product and Max-Plus Systems with Markovian Jumps," *Automatica* 92 (2018) 123–132.
[6] A. Tsiamis and P. Maragos, "Sparsity in Max-plus Algebra", Discrete Events Dynamic Systems, 29:163–189, May 2019. https://doi.org/10.1007/s10626-019-00281-1
[7] N. Tsilivis, A. Tsiamis, P Maragos, "Toward a Sparsity Theory on Weighted Lattices", J. Mathematical Imaging and Vision, 2022.

Machine Learning:

L1. Machine Learning, Deep Neural Nets, and Geometry

(*) L1.1 Geometry, Training, Applications of DNNs with PWL Activations: Regression Problems and NN Minimization

- Τα τροπικά μαθηματικά μπορούν να δώσουν κλειστές λύσεις για piecewise-linear approximations [3]. Μετέπειτα εργασίες, επεκτείνουν το framework ώστε να γίνεται sparse approximation των πολυδιάστατων επιφανειών [4,9], ή προσέγγισή τους με εναλλασσόμενη βελτιστοποίηση [9]. Στην περίπτωση του sparse approximation, η επιλογή των στοιχείων γίνεται από μια δοσμένη "βάση" πιθανών στοιχείων. Στο προκείμενο θέμα εξετάζουμε τρόπους να *μάθουμε από τα δεδομένα* τη βάση των πιθανών στοιχείων για το sparse approximation.
- Ενώ τα τροπικά μαθηματικά μπορούν να χρησιμοποιηθούν για piecewise-linear approximations [3,4], τα approximations είναι πάντα convex. Κάτι τέτοιο περιορίζει την ευρεία εφαρμογή των ανεπτυγμένων μεθόδων, και παράλληλα δεν επιτρέπει το approximation με αυθαίρετη ευκρίνεια. Σύγχρονες εργασίες προσεγγίζουν τα approximation με tropical rational functions [9], που βοηθά να αρθούν αυτοί οι περιορισμοί. Σε αυτό το θέμα εξετάζουμε περαιτέρω τρόπους για να αρθεί αυτός ο περιορισμός (π.χ. difference of convex functions) ή τρόπους να χαρακτηριστεί το approximation error έναντι κάποιων βέλτιστων μεθόδων (π.χ. με upper και lower envelopes).
- Neural Network Compression [10,11]: Συνέχεια της γεωμετρικής μεθόδου με Ζωνότοπα [11]. Πιθανές κατευθύνσεις: 1) Αναγνώριση ομοιότητας νευρώνων με χρήση δεδομένων, ώστε η μέθοδος να γίνει data-driven. Πιο αναλυτική περιγραφή υπάρχει στο Broader Discussion section του [11]. 2) Ενδιαφέρον πρόβλημα θα ήταν επίσης η ανάλυση των δικτύων ως προς το depth και το depth compression.

References:
[1] V. Charisopoulos and P. Maragos, "Morphological Perceptrons: Geometry and Training Algorithms," *Proc. ISMM* 2017. Lecture Notes in Computer Science, vol. 10225, Springer, 2017.

[2] V. Charisopoulos and P. Maragos, "A tropical approach to neural networks with piecewise linear activations", arXiv:1805.08749, 2018.
[3] P. Maragos and E. Theodosis, "Multivariate Tropical Regression and Piecewise-Linear Surface Fitting", *Proc. 45th IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020)*, May 2020.
[4] P. Maragos, V. Charisopoulos, E. Theodosis, "Tropical Geometry and Machine Learning", Proceedings of the IEEE, 109 (5), pp. 728-755, 2021. https://ieeexplore.ieee.org/document/9394420
[5] G. Smyrnis and P. Maragos, "Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation", *Proc. 37th Int'l Conf. on Machine Learning (ICML-2020)*, July 2020.
[6] I. Kordonis and P. Maragos, "Revisiting Tropical Polynomial Division: Theory, Algorithms, and Application to Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 9, pp. 15978-15992, Sep. 2025.
[7] N. Dimitriadis and P. Maragos, "Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints", *Proc. ICASSP-2021*, June 2021.
[8] N. Tsilivis, A. Tsiamis and P. Maragos, "Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression", *Proc. ICASSP-2021*, June 2021.
[9] A. Dunbar and L. Ruthotto, "Alternating minimization for regression with tropical rational functions", Algebraic Statistics, 15 (1), pp. 85-111, 2024. *arXiv:2305.20072: https://arxiv.org/pdf/2305.20072*
[10] P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, "Neural Network Approximation based on Hausdorff distance of Tropical Zonotopes", *Proc. ICLR 2022*, April 2022 https://openreview.net/pdf?id=oiZJwC_fyS .
[11] K. Fotopoulos, P. Maragos and P. Misiakos, "TropNNC: Structured Neural Network Compression Using Tropical Geometry", *Transactions on Machine Learning Research, June 2026. https://openreview.net/forum?id=u7DRq1icmY*

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, ΕΜΠ, Δρ. Μάνος Θεοδόσης)

Additional References:
[12] A. Magnani and S. P. Boyd, "Convex piecewise-linear fitting," Optimization and Engineering, vol. 10, no. 1, pp. 1–17, 2009, https://www-leland.stanford.edu/~boyd/papers/pdf/cvx_pwl_fit.pdf

[13] L. Zhang, G. Naitzat, and L.-H. Lim. "Tropical geometry of deep neural networks", *Proc. 35th Int'l Conf. on Machine Learning (ICML-2018)*, July 2018, https://arxiv.org/pdf/1805.07091.
[14] N. Tsilivis, A. Tsiamis and P. Maragos, "Sparse Approximate Solutions to Max-Plus Equations", In: J. Lindblad et al. (Eds), *Discrete Geometry and Mathematical Morphology:* DGMM-2021, LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39
[15] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T.Y. Hou, and M. Tegmark,2024 "KAN: Kolmogorov-Arnold Networks," *arXiv:2404.19756 https://arxiv.org/pdf/2404.19756*.
[16] M.-C. Brandeburg, G. Loho and G. Montufar, "The Real Tropical Geometry of Neural Networks": https://arxiv.org/abs/2403.11871 .
[17] G. Montufar et al., "Sharp Bounds for the Number of Regions of Maxout Networks and Vertices of Minkowski Sums", SIAM J. Applied Algebra and Geometry, vol. 6(4), pp.618-649, 2022.
[18] S. Gaubert, "Tropical convexity and games: regression, separation, and beyond", CaLISTA Workshop: Geometry Informed Machine Learning , Paris, 2024.
[19] P. Maragos, "Tropical Algebra and Geometry for Machine Learning and Optimization", EUSIPCO-2025 Tutorial.

(*) L2. Προβλήματα Προσέγγισης στην Τροπική Άλγεβρα

Η τροπική άλγεβρα και γεωμετρία μπορούν να χρησιμοποιηθεί για την αναπαράσταση σε συμπαγή μορφή και την ανάλυση πολλών νευρωνικών δικτύων (με κατά τμήματα γραμμική ενεργοποίηση π.χ. ReLU, Maxout, κλπ), καθώς και την περιγραφή συνήθων αλγορίθμων όπως ο Viterbi [1]. Επιπλέον, στην τροπική άλγεβρα, ορισμένα προβλήματα μπορούν να λυθούν πολύ αποδοτικά (π.χ. η λύση γραμμικού συστήματος είναι πολύ πιο σύντομη στην τροπική άλγεβρα από ότι στη συνήθη [2]). Η διπλωματική αυτή ασχολείται με προβλήματα προσέγγισης στην Τροπική Άλγεβρα, δηλαδή περιπτώσεις στις οποίες θέλουμε να προσεγγίσουμε μια τροπική απεικόνιση από δεδομένα. Σε αυτό περιλαμβάνονται ενδεικτικά ένα ή περισσότερα από τα παρακάτω:

- **Προβλήματα κυρτής/τροπικής παλινδρόμησης με outliers**. Υπάρχει αρκετή προηγούμενη δουλειά σε κυρτή/τροπική παλινδρόμηση ([3],[4],[5]). Στα γραμμικά μοντέλα υπάρχουν πολύ επιτυχημένοι αλγόριθμοι για γραμμική παλινδρόμηση με outliers (π.χ. iteratively reweighted least squares, RANSAC). Στόχος εδώ είναι να επεκταθούν/τροποποιηθούν οι αλγόριθμοι της γραμμικής παλινδρόμησης στην περίπτωση της κυρτής/τροπικής.

- **Παραγοντοποίηση/Προσέγγιση Πινάκων στην Τροπική Άλγεβρα**. Η προσέγγιση χαμηλής τάξης των πινάκων καθώς και η παραγοντοποίηση με πίνακες με μη αρνητικούς όρους (Nonnegative Matrix Factorization) έχουν βρει πολλές εφαρμογές, όπως για παράδειγμα η ανάλυση κειμένων, η συμπίεση και τα recommender systems. Έχει ενδιαφέρον να επεκταθεί η προσεγγιστική παραγοντοποίηση στην τροπική περίπτωση. Κάποιες

προκαταρκτικές εργασίες είναι οι [6],[7],[8]. Στόχος της διπλωματικής θα είναι να προταθούν πιο αποτελεσματικοί αλγόριθμοι, και να επεκταθεί το πεδίο εφαρμογής σε περιπτώσεις που οι πίνακες που αναζητούμε έχουν επιπλέον ιδιότητες (π.χ. χαμηλή τάξη).

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, ΕΜΠ, Μάνος Θεοδόσης, Harvard University)

References:
[1] P. Maragos, V. Charisopoulos, E. Theodosis, "Tropical Geometry and Machine Learning", Proceedings of the IEEE, 109 (5), pp. 728-755, 2021. https://ieeexplore.ieee.org/document/9394420
[2] P. Butkovič. Max-linear systems: theory and algorithms. Springer Science & Business Media, 2010.
[3] A. Magnani and S. P. Boyd, "Convex piecewise-linear fitting," Optimization and Engineering, vol. 10, no. 1, pp. 1–17, 2009, https://www-leland.stanford.edu/~boyd/papers/pdf/cvx_pwl_fit.pdf
[4] P. Maragos and E. Theodosis, "Tropical geometry and piecewise-linear approximation of curves and surfaces on weighted lattices," ArXiv preprint arXiv:1912.03891, 2019, https://arxiv.org/pdf/1912.03891
[5] J. Hook, "Linear regression over the max-plus semiring: algorithms and applications," arXiv preprint arXiv:1712.03499, 2017, https://arxiv.org/pdf/1712.03499.
[6] S. Karaev and P. Miettinen, "Capricorn: An algorithm for subtropical matrix factorization," in Proceedings of the 2016 SIAM international conference on data mining. SIAM, 2016, pp. 702–710, https://cs.uef.fi/~pauli/papers/karaev16capricorn.pdf
[7] S. Karaev and P. Miettinen "Cancer: Another algorithm for subtropical matrix factorization," in Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II 16. Springer, 2016, pp. 576–592
[8] A. Omanovic, P. Oblak, and T. Curk, "Faststmf: Efficient tropical matrix factorization algorithm for sparse data," arXiv preprint arXiv:2205.06619, 2022. https://arxiv.org/pdf/2205.06619
[9] I. Kordonis, E. Theodosis, G. Retsinas, P. Maragos, "Matrix Factorization in Tropical and Mixed Tropical-Linear Algebras Inproceedings", Proc. ICASSP 2024, Seoul, Korea, 2024,

L3. Μοντέλα Διάχυσης για Χρονοσειρές και Εφαρμογές

Τα τελευταία χρόνια υπάρχει μεγάλο ενδιαφέρον για generative μοντέλα. Μία από τις πιο επιτυχημένες κατηγορίες είναι τα μοντέλα διάχυσης [1-3]. Τα μοντέλα διάχυσης έχουν πρόσφατα επεκταθεί σε δεδομένα χρονοσειρών [4,5]. Αυτό επιτρέπει να δειγματοληπτούμε πιθανές μελλοντικές τροχιές από χρονοσειρές. Μερικά παραδείγματα χρονοσειρών είναι η ζήτηση ενός προϊόντος (π.χ. ενέργειας), η χρήση δεδομένων, η κίνηση στους δρόμους, οι μετρήσεις διαφόρων αισθητήρων κ.λπ. Στη διπλωματική αυτή προτείνουμε τη μελέτη generative μοντέλων για χρονοσειρές με έμφαση στη χρήση τους για βελτιστοποίηση. Ένα συγκεκριμένο παράδειγμα είναι η αποθήκευση ηλεκτρικής ενέργειας. Π.χ. μοντέλο που δημιουργεί δείγματα από πιθανές μελλοντικές τροχιές τιμών και κατασκευή scenario trees όπως στις εργασίες [10,11].

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, ΕΜΠ)

References:
[1] Sohl-Dickstein, Jascha, et al. "Deep unsupervised learning using nonequilibrium thermodynamics." Proc. International Conference on Machine Learning, 2015 https://proceedings.mlr.press/v37/sohl-dickstein15.html.
[2] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models" Advances in Neural Information Processing Systems 33 (2020).
[3] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." arXiv preprint arXiv:2011.13456 (2020) https://openreview.net/pdf/ef0eadbe07115b0853e964f17aa09d811cd490f1.pdf
[4] Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In Proc. International Conference on Machine Learning (ICML 2021), https://proceedings.mlr.press/v139/rasul21a/rasul21a.pdf.
[5] Lin, Lequan, et al. "Diffusion models for time series applications: A survey." arXiv preprint arXiv:2305.00624 (2023).
[6] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM 2021.
[7] Zhang, Xinjing, et al. "Arbitrage analysis for different energy storage technologies and strategies." Energy Reports 7 (2021): 8198-8206.
[8] Kordonis, Ioannis, Alexandros C. Charalampidis, and Pierre Haessig. "Optimal operation of a grid-connected battery energy storage system over its lifetime." Optimal Control Applications and Methods 44.2 (2023): 739-757.
[9] S. Zarifis, I. Kordonis, P. Maragos, "Diffusion-Based Forecasting for Uncertainty-Aware Model Predictive Control", *Proc. 33rd European Signal Processing Conf. (EUSIPCO 2025)*, Palermo, Italy, 2025.
[10] S. Zarifis, I. Kordonis, P. Maragos, "Diffusion-Based Scenario Tree Generation for Multivariate Time Series Prediction and Multistage Stochastic Optimization", *Proc. EUSIPCO 2026*, Bruges, Belgium, 2026.

(*) L4. Morphological (Max/Min-plus) Networks for Machine Learning Problems

Since 2013, there has been an explosion of machine learning research in recent years, which has now provided models in the form of deep neural networks (NNs) with near-human or even human performance. Recent advances indicate a shift towards nonlinearity, but the ``multiply-accumulate" (= linear) operations of the perceptron are still ubiquitous.
Our Questions:

- *Biology*: Are linear inner products and convolutions the only biologically plausible models of neuronal computation?
- *Applications*: How can we train NNs with nonlinear nodes (e.g. max-plus) that may also have some biological plausibility to learn from data?
- *Theory*: Can we use results and tools from nonlinear mathematics (in particular max-plus algebra and tropical geometry) to analyze the behavior and complexity of NNs whose nodes include max-plus or similar nonlinearities?

Several research groups world-wide, inspired by the capabilities of nonlinear image analysis systems using morphological systems, have developed Morphological Neural Nets (MNNs), i.e. networks whose nodes perform max-plus operations. These NNs with morphological nodes can train faster, are amenable to higher percentages of compression, and have similar or better performance than NNs with linear nodes. Recently there is a growing interest for research on these nonlinear networks.

Στην διπλωματική αυτή προτείνουμε την μελέτη σύγχρονων MNNs με έναν από τους ακόλουθους στόχους: 1) καλύτερους αλγορίθμους για την εκπαίδευση τους, 2) βαθειές αρχιτεκτονικές, 3) εφαρμογές σε μεγαλύτερα datasets.

L4.1: MNNs με θεωρητικά guarantees εκπαιδευσιμότητας: Στο [10] αναγνωρίζεται το πρόβλημα των sparse gradients και της έλλειψης του 1st order universal approximation ως λόγος των δυσκολιών στην εκπαίδευση των MNNs. Για να επιλύσει αυτό το πρόβλημα, η προτεινόμενη MPM αρχιτεκτονική χρειάζεται depth ώστε να λειτουργεί ως 1st order universal approximator. Είναι ανοιχτό πρόβλημα αν υπάρχει shallow MNN architecture που να είναι 1st order universal approximator.

L4.2: Επίλυση του προβλήματος των sparse & uninformative gradients των min-max MNNs βασισμένων στην διακριτοποίηση του representation theorem με εναλλακτικές μεθόδους εκπαίδευσης: Για shallow min-max δίκτυα, ο αλγόριθμος του [11], βασισμένος στο Multiplicative Weights Update scheme, μπορεί να αντιμετωπίσει το πρόβλημα των sparse & uninformative gradients. Έχει όμως περιορισμούς στα deep networks. Είναι ανοιχτό πρόβλημα η εύρεση αλγορίθμου εκπαίδευσης ειδικά για deep networks.

L4.3: Υλοποίηση CUDA kernels (πειραματικό): Η έρευνα στα MNNs περιορίζεται σημαντικά από την μη-υποστήριξη του μορφολογικών πολλαπλασιασμών πινάκων και συνελίξεων από το hardware. Τα γραμμικά matmuls & convolutions έχουν λάβει σημαντικά incremental optimizations για να φτάσουν στην αποδοτική μορφή που έχουν σήμερα. Δίνεται σαν πρόβλημα η βελτίωση των naive μορφολογικών kernels, ώστε να γίνουν ανταγωνιστικά με τα γραμμικά ή/και να υποστηρίζουν sparsity.

(Συνεργάτες: Κων/νος Φωτόπουλος, ETH)

References:

[1] V. Charisopoulos and P. Maragos, "Morphological Perceptrons: Geometry and Training Algorithms," *Proc. ISMM* 2017. Lecture Notes in Computer Science, vol. 10225, Springer, 2017.
[2] N. Dimitriadis and P. Maragos, "Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints", Proc. IEEE ICASSP 2021.
[3] P.-F. Yang and P. Maragos, "Min-Max Classifiers: Theory, Design, and Application", Pattern Recognition, 28 (6), pp.879-899, 1995.
[4] L. F.C. Pessoa and P. Maragos, "Hybrid Neural Networks with Morphological/Rank/Linear Nodes: A Unifying Framework with Application to Handwritten Character Recognition", Pattern Recognition, vol.33, pp.945-960, 2000.
[5] R. de A. Araújo, A. L.I. Oliveira and S. Meira, "A morphological neural network for binary classification problems", Engineering Applications of Artificial Intelligence, vol.65, pp.12-28, 2017.
[6] G. Franchi and A. Fehri and A. Yao, "Deep Morphological Networks", Pattern Recognition, 102, 107246, 2020.
[7] S. Velasco-Forero and J. Angulo, "MorphoActivation: Generalizing ReLU Activation Function by Mathematical Morphology", Proc. DGMM 2022, LNCS, vol. 13493, pp.449-461, Springer, 2022. Also https://arxiv.org/abs/2207.06413.

[8] M. E. Valle, "Bridging Vector-Valued Mathematical Morphology and Deep Learning", CaLISTA Workshop: Geometry Informed Machine Learning , Paris, 2024.
[9] J. G. Angulo, "Lattice theory and algebraic models for deep learning based on mathematical morphology", CaLISTA Workshop: Geometry Informed Machine Learning , Paris, 2024.
[10] K. Fotopoulos and P. Maragos, "Training Deep Morphological Neural Networks as Universal Approximators", https://arxiv.org/abs/2505.09710, 2025.
[11] K. Fotopoulos and P. Maragos, "Correlational Training of Morphological Neural Networks". https://arxiv.org/abs/2610.11740, 2026.
[12] P. Maragos, *Topics in Computer Vision and Machine Learning*, [Postgraduate textbook], KALLIPOS, Open Academic Editions, https://doi.org/10.57713/kallipos-1097 . [Chapter 4].

Επεξεργασία κειμένων και φυσικής γλώσσας, LLMs:

T1. Επεξεργασία κειμένων και φυσικής γλώσσας με αλγεβρικές και γεωμετρικές μεθόδους (vector spaces, graphs, tropical geometry and max-plus algebra).

(Συνεργάτες: Δρ. Ιωάννης Βλασσόπουλος, Ε.Κ. Αθηνά)

References:
[1] D. Widdows, "Geometry and Meaning", Vol. 773. Stanford: CSLI publications, 2004.
[2] P. Gärdenfors, "Conceptual spaces: The Geometry of Thought", MIT Press, 2004.
[3] T. Mikolov et al. *"Efficient estimation of word representations in vector space"*, ICLR 2013.
[4] Stéphane Gaubert and Yiannis Vlassopoulos, "Directed Metric Structures arising in Large Language Models", https://arxiv.org/abs/2405.12264 , 2024.