

Θέμα Διπλωματικής	Περιγραφή	Επιβλέπων	Τόπος	Προαπαιτούμενες γνώσεις/ικανότητες, ή ανάγκη εκπαίδευσης	Πιθανή ημερομηνία έναρξης	Παρατηρήσεις
Help us break LLMs	aim is to perform a study of LLM artefacts and systematic errors when the LLMs are used in translation tasks.	Γ. Ταμπουρατζής - Σ. Σοφιανόπουλος	Εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2025	
Information retrieval in LLMs using multilingual queries	the aim is to investigate to what extent the language in which queries are posed by the user affects the accuracy of responses.	Γ. Ταμπουρατζής - Σ. Σοφιανόπουλος		Γνώση προγραμματισμού	Ανοιξη 2025	
Effect of weight quantization in NMTs	Study the effectiveness of using quantization of weights to balance memory requirements and processing speed with translation accuracy.	Γ. Ταμπουρατζής - Σ. Σοφιανόπουλος		Γνώση προγραμματισμού	Ανοιξη 2025	
Hallucinations in Deep learning	Study and comparison of hallucination problems in NMT systems and LLMs when the latter are applied to machine translation tasks.	Γ. Ταμπουρατζής - Σ. Σοφιανόπουλος		Γνώση προγραμματισμού	Ανοιξη 2025	
Θέματα Διπλωματικών Καθ. Μ. Κουμπάρακη	(βλ. Link) https://docs.google.com/document/d/1xRipQBkypYE9MN7fJMZ3VXV47zXEI5YPZl7onDV0R8/edit?tab=t.0#heading=h.goownfcap3k1	Μ. Κουμπάρακη		Γνώση προγραμματισμού	Ανοιξη 2025	https://docs.google.com/document/d/1xRipQBkypYE9MN7fJMZ3VXV47zXEI5YPZl7onDV0R8/edit?tab=t.0#heading=h.goownfcap3k1
Θέματα Διπλωματικών Δρ. Κ. Κρίτση	(βλ. Link) https://docs.google.com/document/d/1zx5VEb94sTMqgUZp2hVINHAU79qNt45rT-m46uyYpeU/edit?tab=t.0	Κ. Κρίτση		Γνώση προγραμματισμού	Ανοιξη 2025	https://docs.google.com/document/d/1zx5VEb94sTMqgUZp2hVINHAU79qNt45rT-m46uyYpeU/edit?tab=t.0
Αυτοματοποίηση εξαγωγής και οργάνωσης δεδομένων για την Ισότητα των Φύλων	Η συγκεκριμένη εργασία αφορά στην ανάπτυξη μιας ολοκληρωμένης εφαρμογής εξαγωγής, διαχείρισης και οπτικοποίησης πληροφορίας από κείμενα πρακτικών διοικητικών συμβουλίων. Συγκεκριμένα, η εφαρμογή θα εστιάσει στην αυτόματη εξαγωγή πληροφορίας σχετικά με το φύλο των ατόμων που αφορούν οι αποφάσεις που εμπεριέχονται στα πρακτικά, το αντικείμενο των αποφάσεων, τα χρονικά τους πλαίσια κτλ., με στόχο την αυτόματη οργάνωση, δεικτοδότηση και ανάκτηση των αποφάσεων με κριτήρια που συμπεριλαμβάνουν και το φύλο. Οι πληροφορίες που θα εξαχθούν από το σώμα κειμένων που θα δημιουργηθεί, θα αναλυθούν στατιστικά και θα οπτικοποιηθούν. Για τις ανάγκες εξαγωγής πληροφορίας που θα λάβει χώρα θα χρησιμοποιηθεί python.	Κ. Πάστρα		Γνώση προγραμματισμού	Ανοιξη 2025	

Frugal Culinary Assistant	<p>Η εργασία αφορά στη δημιουργία ενός ψηφιακού βοηθού-μάγειρα που θα βοηθά τους χρήστες να εκτελούν συνταγές βελτιστοποιώντας ποσότητες υλικών και διαδικασίες μαγειρέματος με στόχο την εφαρμογή πρακτικών για sustainable cooking. Ξεκινώντας από ένα σύνολο κλασικών συνταγών θα αναπτυχθεί αλγόριθμος αυτόματης επαύξησης των συνταγών με πληροφορία που υπονοείται, και που θα αναδιοργανώνει τα βήματα της συνταγής ώστε να επιτυγχάνεται οικονομία δυνάμεων και πόρων. Οι λειτουργίες αυτές θα είναι προσβάσιμες στο χρήστη μέσω μίας διαδικτυακής εφαρμογής που θα αναπτυχθεί συγκεκριμένα για αυτό το σκοπό. Η εργασία θα χρησιμοποιήσει τη Μινιμαλιστική Γραμματική της Κίνησης (Πάστρα και Αλοΐμονος, 2012) και απαιτεί άνεση στη χρήση python.</p>	Κ. Πάστρα		Γνώση προγραμματισμού	Ανοιξη 2025	
Σημασιολογική επαύξηση οπτικοακουστικών αρχείων	<p>Η εργασία αφορά στην ανάπτυξη ενός αλγόριθμου αυτόματου εντοπισμού και επισημείωσης σημείων σε ένα οπτικο-ακουστικό αρχείο στα οποία ο προφορικός λόγος δεν εκφράζει ένα πλήρες μήνυμα αλλά απαιτεί οπτική πληροφορία όπως αυτή παρέχεται στο συνοδευτικό βίντεο. Η εργασία θα χτίσει πάνω στο θεωρητική προσέγγιση Cosmogoe (Πάστρα, 2008) και θα αυτοματοποιήσει τους τρεις τύπους υποχρεωτικής συμπληρωματικότητας που προβλέπει η προσέγγιση αυτή. Σημασιολογική επαύξηση του μεταγεγραμμένου προφορικού λόγου με πληροφορία από την εικόνα σε συγκεκριμένα χρονικά σημεία ενός οπτικο-ακουστικού αρχείου οδηγεί στη δημιουργία ενός σημασιολογικά αυτόνομου επαυξημένου transcript για χρήση σε συστήματα αυτόματης σδεικτοδότησης/ανάκτησης πληροφορίας.</p>	Κ. Πάστρα		Γνώση προγραμματισμού	Ανοιξη 2025	

<p>Προτεινόμενο θέμα 1.</p>	<p>Πρόσφατα ολοκληρώθηκε μια διπλωματική εργασία που είχε στόχο την δημιουργία ενός συστήματος που κάνει Key Point Analysis σε ελληνικά κείμενα που περιέχουν επιχειρήματα (argument). Πιο συγκεκριμένα στα πλαίσια της εργασίας αναπτύχθηκε α) ένας υποσύστημα Key Point Matching (KPM) που αποφασίζει αν ένα Key Point έχει προκύψει από ένα argument αλλά και β) ένας Key Point Generator (KPG) για την παραγωγή Key Points. Και τα 2 υποσυστήματα εκπαιδεύτηκαν σε ένα κατάλληλο dataset που δημιουργήθηκε στα πλαίσια της ίδιας διπλωματικής. Σκοπός της προτεινόμενης εργασίας είναι να διερευνηθούν κάποια από τα ακόλουθα α) να βελτιωθεί το train κομμάτι του dataset β) να γίνουν πειράματα με supervised fine-tuning στο κομμάτι του Key Point Generation, γ) να γίνουν πειράματα με νέα ελληνικά LLM όπως το Llama-krikri δ) να δοκιμαστούν νέα prompts για το task του KPG.</p> <p>📄 https://huggingface.co/Kleo/meltemi_arg2kp_matcher</p> <p>📄 https://huggingface.co/datasets/Kleo/ArgKP_2021_GR</p> <p>📄 Key Point Analysis in Greek: A New Dataset and Baselines M.Sc thesis, January</p>	<p>Δ. Γαλάνης</p>		<p>Γνώση προγραμματισμού</p>	<p>Ανοιξη 2025</p>	
-----------------------------	---	-------------------	--	------------------------------	--------------------	--

<p>Προτεινόμενο θέμα 2.</p>	<p>Τα LLMs έχουν επιτύχει πολύ υψηλά σκορ σε διάφορα tasks όπως για παράδειγμα το ελληνικό meltemi (https://arxiv.org/pdf/2407.20743) στα αντίστοιχα dataset. Όμως στα ελληνικά δεν έχει διερευνηθεί κατά πόσο εκτός από την σωστή απάντηση έχουν την δυνατότητα να παράγουν και τα ενδιάμεσα βήματα (reasoning) που οδηγούν σε αυτή. Αυτά τα ενδιάμεσα βήματα μπορούν να βοηθήσουν στο να επιλεγεί με μεγαλύτερη επιτυχία η απάντηση αλλά είναι και πολύ χρήσιμα καθώς αποτελούν την εξήγηση του γιατί επιλέχθηκε. Σκοπός της εργασίας είναι να α) διερευνηθεί κατά πόσο και με ποιο τρόπο τα υπάρχοντα ελληνικά μοντέλα (meltemi, Llama-krikri) μπορούν να παράγουν τέτοιες εξηγήσεις και αν αυτές τα βοηθάνε στο να διαλέξουν πιο σωστά την τελική απάντηση β) παραχθούν κατάλληλα dataset εκπαίδευσης/αξιολόγησης τέτοιων μοντέλων. Π.χ. ένα τέτοιο dataset για τα αγγλικά είναι το ακόλουθο: https://huggingface.co/datasets/Locutusque/arc-cot</p>	<p>Δ. Γαλάνης</p>		<p>Γνώση προγραμματισμού</p>	<p>Ανοιξη 2025</p>	
<p>Θέματα Διπλωματικών Δρ. Γ. Παρασκευόπουλου</p>	<p>(βλ. Link) https://docs.google.com/document/d/1kGbk3YxT1_NxHPd6M2thrpbsEPznu0wsgYIDzihtZ6s/edit?tab=t.0#heading=h.zbl0ep6q3q0u</p>	<p>Γ. Παρασκευόπουλος</p>		<p>Γνώση προγραμματισμού</p>	<p>Ανοιξη 2026</p>	<p>https://docs.google.com/document/d/1kGbk3YxT1_NxHPd6M2thrpbsEPznu0wsgYIDzihtZ6s/edit?tab=t.0#heading=h.zbl0ep6q3q0u</p>