

1η Οκτωβρίου 2025

Εισαγωγή

Στο ακόλουθο έγγραφο καταγράφονται πιθανά θέματα πτυχιακών εργασιών της ομάδας του εργαστηρίου Επικοινωνίας Ανθρώπου Μηχανής και Εικονικής Πραγματικότητας. Η λίστα θεμάτων **δεν** είναι περιοριστική. Μπορείτε να προτείνετε και εσείς κάποιο θέμα σε περιοχές ενδιαφέροντος των μελών του εργαστηρίου (π.χ. Υλοποίηση συγκεκριμένου τύπου Ηλεκτρονικού Παιχνιδιού).

Τίτλος

Σχεδιασμός πλοήγησης σε δωμάτιο απόδρασης

Περιγραφή

Με βάση την υφιστάμενη εφαρμογή και το δωμάτιο απόδρασης στο MobiCave, να ανασχεδιασθεί και υλοποιηθεί για χρήση σε συσκευή Vive Pro Eye. Η υφιστάμενη εφαρμογή έχει υλοποιηθεί σε περιβάλλον Unity. Το περιεχόμενο/σενάριο παιχνιδιού είναι ήδη έτοιμο.

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού.

Παραδοτέα

Αναφορά, κώδικας εφαρμογής

Σχετικές αναφορές

TBC

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

Σχεδιασμός διάδρασης σε παιχνίδι shoot 'em up στο MobiCave

Περιγραφή

Με βάση την υφιστάμενη εφαρμογή και το παιχνίδι δύο ατόμων στο MobiCave (Turtle Heroes), να ανασχεδιασθεί και υλοποιηθεί η μέθοδος στόχευσης και η ανάδραση από το παιχνίδι. Η υφιστάμενη εφαρμογή έχει υλοποιηθεί σε περιβάλλον Unity. Το περιεχόμενο/σενάριο παιχνιδιού είναι ήδη έτοιμο.

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού.

Παραδοτέα

Αναφορά, κώδικας εφαρμογής

Σχετικές αναφορές

TBC

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

Αξιολόγηση στερεοσκοπικής προβολής στο MobiCave.

Περιγραφή

Στην παρούσα φάση το MobiCave υποστηρίζει στερεοσκοπική με χρήση γυαλιών αναγλύφου. Η πτυχιακή στοχεύει στο σχεδιασμό του συστήματος, βελτιστοποίηση μέσω δοκιμαστικών ρυθμίσεων και αξιολόγηση εμπύθισης του τελικού συστήματος.

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού και Εικονικής Πραγματικότητας.

Παραδοτέα

Αναφορά, κώδικας εφαρμογής

Σχετικές αναφορές

TBC

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

Εφαρμογή Γκράφιτι στο MobiCave

Περιγραφή

Δημιουργία μιας εφαρμογής εικονικού γκράφιτι και καλλιτεχνικής δημιουργίας στο MobiCAVE, αξιοποιώντας το OpenBrush (open-source έκδοση του Tilt Brush). Θα επιτρέπει στους χρήστες να ζωγραφίζουν σε 3D χώρο, να πειραματίζονται με εικονικά πινέλα και εφέ, και να δημιουργούν graffiti-style έργα σε ένα κοινόχρηστο εικονικό περιβάλλον (VR HMD και MobiCAVE).

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού και Εικονικής Πραγματικότητας.

Παραδοτέα

Αναφορά, κώδικας εφαρμογής, Παραδείγματα VR graffiti έργων

Σχετικές αναφορές

OpenBrush: <https://openbrush.app>

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

Εικονικό Μουσείο

Περιγραφή

Ανάπτυξη ενός εξελιγμένου Εικονικού Μουσείου (Virtual Museum) σε περιβάλλον Εικονικής Πραγματικότητας. Τρισδιάστατη απεικόνιση χώρων και αντικειμένων δυνατότητα ελεύθερης πλοήγησης και αλληλεπίδρασης εμπλουτισμένες πληροφορίες (κειμενικές, οπτικοακουστικές) για τα εκθέματα. Unity με χρήση VR headsets.

Παραδοτέα**Απαιτήσεις**

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού και Εικονικής Πραγματικότητας.

Παραδοτέα

Αναφορά, κώδικας εφαρμογής, Πλήρως λειτουργικό πρωτότυπο

Σχετικές αναφορές

TBC

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

New Departmental AI Server - Open Source LLMs - Agents

Περιγραφή

Εγκατάσταση και βελτιστοποίηση LLMs (π.χ. LLaMA, Mistral, Falcon) σε GPU server περιβάλλον με σκοπό τον σχεδιασμό και την ανάπτυξη πρακτικών agents που αξιοποιούν τα μοντέλα. Ενδεικτικά, NPCs σε παιχνίδια με φυσικό διάλογο, memory, και goal-oriented συμπεριφορές. Σχεδίαση game-ready agent stack και ενσωμάτωση σε Unity/Unreal (API/SDK), με παραδείγματα NPCs σε σκηνή demo.

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού (Python) και εξοικείωση με περιβάλλοντα Linux/servers.

Παραδοτέα

Αναφορά, κώδικας εφαρμογής (agents + scripts για εγκατάσταση/ρύθμιση LLMs)

Σχετικές αναφορές

TBC

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

Δημιουργία παιχνιδιού MR

Περιγραφή

Δημιουργία ενός παιχνιδιού Μεικτής Πραγματικότητας (Mixed Reality) με χρήση της κάμερας ZED Mini της Stereolabs, η οποία επιτρέπει στερεοσκοπική όραση και παρακολούθηση βάθους σε πραγματικό χρόνο. Το παιχνίδι θα βασίζεται στη συνδυασμένη απεικόνιση πραγματικού και εικονικού χώρου, εκμεταλλευόμενο τις δυνατότητες όπως spatial mapping και tracking. Η υλοποίηση θα γίνει σε Unity με ενσωμάτωση του ZED SDK.

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν πολύ καλή γνώση προγραμματισμού και Εικονικής/Επαυξημένης Πραγματικότητας (Unity, C#).

Παραδοτέα

Αναφορά, Κώδικας εφαρμογής, Πρωτότυπο παιχνιδιού σε λειτουργική μορφή

Σχετικές αναφορές

Stereolabs ZED Mini: <https://www.stereolabs.com/zed-mini>

Τίτλος

Σχεδιασμός και Υλοποίηση ARG

Περιγραφή

Σχεδίαση και ανάπτυξη ενός Alternate Reality Game (ARG) με θεματικό άξονα την πόλη του Ναυπλίου ή/και της Τρίπολης. Το παιχνίδι θα συνδυάζει πραγματικούς χώρους (ιστορικά μνημεία, δρόμοι, πλατείες) με ψηφιακά στοιχεία (ιστοσελίδες, εφαρμογή κινητού, QR codes, κοινωνικά δίκτυα), δημιουργώντας μια αφήγηση που καλεί τους παίκτες να εξερευνήσουν την πόλη, να επιλύσουν γρίφους και να ανακαλύψουν κρυμμένα στοιχεία της τοπικής ιστορίας και κουλτούρας.

Απαιτήσεις

Διατμηματική, ΤΠΤ και ΤΠΨΤ

Παραδοτέα

Αναφορά, κώδικας εφαρμογής, πρωτότυπο app

Σχετικές αναφορές

McGonigal, J. (2006). This Is Not a Game: A Guide to Alternate Reality Gaming

Επιβλέποντες

Λέππουρας Γ., Θεοδωρόπουλος Τ.

Τίτλος

E-learning μορφές μάθησης: ανασχεδιασμός εκπαιδευτικού υλικού για τη διδασκαλία με το μοντέλο της ανεστραμμένης τάξης (flipped classroom)

Περιγραφή

Στόχος είναι ο ανασχεδιασμός υφιστάμενου υλικού και η υλοποίηση ψηφιακού εκπαιδευτικού υλικού για μάθημα του Τμήματος Πληροφορικής και Τηλεπικοινωνιών (προτείνεται ένα από τα ακόλουθα: Ανάπτυξη Ηλεκτρονικών Παιχνιδιών (προπτυχιακό), Προηγμένες Διεπαφές - Εικονική Πραγματικότητα (προπτυχιακό), Γραφικά Υπολογιστών (προπτυχιακό) ή Σχεδιασμός Ηλεκτρονικών Παιχνιδιών (μεταπτυχιακό)). Το παραδοτέο θα περιλαμβάνει: περιγραφή θεωρητικού

πλαίσιου, διαμόρφωση πρωτότυπου οπτικοακουστικού υλικού, σχεδιασμός ψηφιακής διδασκαλίας σύμφωνα με τις αρχές του μοντέλου.

Απαιτήσεις

1 ή 2 φοιτητές. Θα πρέπει να έχουν γνώσεις της γνωστικής θεωρίας μάθησης και διαμόρφωσης/σχεδιασμού εκπαιδευτικού περιεχομένου.

Παραδοτέα

Αναφορά, Ψηφιακό υλικό

Σχετικές αναφορές

TBC

Επιβλέποντες

Λέππουρας Γ., Πλατής Ν., Γιαννακοπούλου Γ.

Τίτλος

Εφαρμογή οπτικοποίησης προαπαιτούμενων του οδηγού σπουδών

Περιγραφή

Ζητείται η δημιουργία μίας διαδραστικής διαδικτυακής εφαρμογής για την οπτικοποίηση των προαπαιτούμενων του οδηγού σπουδών. Η εφαρμογή θα πρέπει να είναι εύχρηστη και αποδοτική για τις ανάγκες των φοιτητών του Τμήματος.

Απαιτήσεις

1 φοιτητής. Γνώσεις σε προγραμματισμό διαδικτυακών εφαρμογών (HTML, PHP, Javascript ή άλλα σχετικά εργαλεία), βασικές γνώσεις βάσεων δεδομένων.

Παραδοτέα

Εφαρμογή, αναφορά

Σχετικές αναφορές

TBC

Επιβλέποντες

Πλατής Ν.

Τίτλος

Σχεδιασμός και ανάπτυξη συστήματος κυρτής προβολής (panoramic)

Περιγραφή

Σχεδίαση και ανάπτυξη αλγορίθμων (Πλατής)

Απαιτήσεις

....

Παραδοτέα

....

Σχετικές αναφορές

....

Επιβλέποντες

Λέπουρας Γ., Θεοδωρόπουλος Τ., Πλατής Ν.

Τίτλος

Evidence Detective – Web παιχνίδι διερεύνησης στερεοτύπων με Stable Diffusion

Περιγραφή

Ανάπτυξη web παιχνιδιού τύπου point-and-click «ντετέκτιβ», όπου ο παίκτης διερευνά πολιτισμικά στερεότυπα ακολουθώντας τη διαδικασία Evidence → Intent → Prompt → Findings. Κάθε γύρος λειτουργεί ως ερευνητική αποστολή: ο παίκτης συλλέγει «τεκμήρια» (κάρτες με εικόνες, αποσπάσματα, γλωσσικά ή εικονογραφικά μοτίβα) από curated συλλογή, τα ταξινομεί ανά θεματικό πεδίο και συνθέτει Counter-Intent Cards μέσω builder τύπου «μείκτης», όπου σύρει αξίες-κάρτες (agency, expertise, care κ.ά.) πάνω σε οπτικά sliders (φωτισμός, χρώμα, κεντρικότητα, γωνία κάμερας). Προκαθορισμένοι κανόνες δίνουν feedback για κάθε συνδυασμό (π.χ. “eye-level ↔ agency”, “warm light ↔ care”).

Στη φάση Prompt Authoring, ο παίκτης συμπληρώνει 4 καθοδηγούμενα slots (Subject, Action/Scene, Style/Aesthetic, Composition/Lighting) με κλειδωμένο λεξιλόγιο από τα

επιλεγμένα evidence και values. Το UI δεν επιτρέπει τη χρήση άμεσων labels (π.χ. «nurse», «refugee»), αλλά προτείνει συνώνυμα και περιφράσεις.

Στο Testing & Findings, το παιχνίδι παράγει αυτόματα baseline εικόνα (με direct-label), counter-intent εικόνα και δύο single-slot variations με σταθερό seed. Ο παίκτης συγκρίνει τις εικόνες μέσω εργαλείου «slider» και επιλέγει cues (π.χ. pose, gaze, PPE, light) για να καταγράψει bias ή αποκλίσεις. Τα αποτελέσματα αποθηκεύονται σε scoreboard για μεταγενέστερη ανάλυση.

Απαιτήσεις

1 ή 2 φοιτητές. Γνώση προγραμματισμού σε React ή Vue, εξοικείωση με Stable Diffusion WebUI (AUTOMATIC1111) και SQLite.

Παραδοτέα

Αναφορά, λειτουργική web εφαρμογή, τοπική εγκατάσταση Stable Diffusion, curated συλλογή evidence & counter-intents, scoreboard δεδομένων.

Σχετικές αναφορές

TBC

Επιβλέποντες

Φωκά Α.

Τίτλος

Card-Roguelike Interrogate – Web deck-building παιχνίδι διερεύνησης στερεοτύπων με Stable Diffusion

Περιγραφή

Ανάπτυξη web παιχνιδιού τύπου *deck-building roguelike*, όπου ο παίκτης διεξάγει ερευνητικούς «γύρους» διερεύνησης στερεοτύπων, χρησιμοποιώντας κάρτες Evidence, Intent, Semiotic και Constraint. Κάθε παρτίδα ξεκινά με θεματική αποστολή (π.χ. *housekeeper*, *migrant worker*) και στόχο τη δημιουργία ενός πειστικού *Counter-Intent Prompt* που ανατρέπει καθιερωμένες αναπαραστάσεις.

Ο παίκτης τραβάει κάρτες Evidence από curated deck (εικόνες, φράσεις, πολιτισμικά μοτίβα) και καλείται να τις συνδέσει σε τουλάχιστον τρεις διαφορετικούς τομείς (οπτικός πολιτισμός, γλώσσα, κοινωνικό πλαίσιο). Στη συνέχεια δημιουργεί συνδυασμούς *Intent + Semiotic* (π.χ. *agency + eye-level camera*), κερδίζοντας «rationale points» αν αιτιολογήσει την επιλογή με τεκμήρια που έχει ήδη παίξει.

Το Prompt Authoring γίνεται ως “card-lane builder”: κάθε slot (Subject, Action/Scene, Style/Aesthetic, Composition/Lighting) αντιστοιχεί σε lane όπου ο παίκτης «ρίχνει» κάρτες. Το σύστημα ελέγχει τη συνάφεια και απορρίπτει απαγορευμένα labels, απαιτώντας παραπομπές στα evidence.

Κατά το *Testing & Findings*, ο παίκτης παίζει “variation cards” που αλλάζουν μία παράμετρο τη φορά· στη συνέχεια συγκρίνει τις εικόνες (baseline, counter-intent, variations) που παράγονται με Stable Diffusion, εντοπίζει μοτίβα bias και συντάσσει μικρή υπόθεση. Το παιχνίδι αποδίδει συνολικό *combo score* με βάση τη συνέπεια, την ποικιλία και τη σαφήνεια του αποτελέσματος.

Απαιτήσεις

1 ή 2 φοιτητές. Γνώση JavaScript (React ή Svelte), εμπειρία σε state machines (XState), βασική κατανόηση λειτουργίας Stable Diffusion (AUTOMATIC1111 API).

Παραδοτέα

Αναφορά, λειτουργική web εφαρμογή deck-building τύπου, τοπική εγκατάσταση Stable Diffusion, curated deck evidence & counter-intents, scoreboard αποτελεσμάτων bias-interrogation.

Σχετικές αναφορές

TBC

Επιβλέποντες

Φωκά Α.

Τίτλος

Mentor-Chat Interrogate – Διαδραστικό LLM web παιχνίδι διερεύνησης στερεοτύπων με Stable Diffusion

Περιγραφή

Ανάπτυξη διαδραστικού web παιχνιδιού τύπου “guided chat”, όπου ο παίκτης συνεργάζεται με έναν εικονικό μέντορα (LLM) για να διερευνήσει στερεότυπα και να συνθέσει κατευθυνόμενα *counter-intent prompts*. Το παιχνίδι αναπαράγει τη δομή Evidence → Intent → Prompt → Findings, αλλά η καθοδήγηση πραγματοποιείται μέσω διαλόγου.

Ο μέντορας (τοπικό LLM, π.χ. Mistral-7B μέσω Ollama) καθοδηγεί τον παίκτη βήμα-βήμα: ζητά τουλάχιστον έξι τεκμήρια από διαφορετικούς τομείς (οπτικά, λεκτικά, πολιτισμικά), ελέγχει ότι συνοδεύονται από χρονολογία και πλαίσιο και δίνει ανατροφοδότηση πριν τον αφήσει να προχωρήσει. Στη φάση *Intent / Semiotic Palette*, ο μέντορας προτείνει αξίες και οπτικά χαρακτηριστικά, ζητώντας αιτιολόγηση· κατόπιν συνθέτει μαζί με τον παίκτη την *Counter-Intent Card*.

Στο *Prompt Authoring*, η σύνθεση των 4 slots γίνεται μέσα από διαδοχικές ερωταπαντήσεις και παραδείγματα, με ζωντανό έλεγχο εγκυρότητας (π.χ. «το Subject σου είναι label — ξαναγράψ' το έμμεσα»).

Κατά το *Testing & Findings*, ο μέντορας δημιουργεί αυτόματα baseline, counter-intent και δύο variations, παρουσιάζει τα αποτελέσματα και ζητά από τον παίκτη να διατυπώσει υπόθεση bias και επόμενο βήμα. Όλη η πορεία αποθηκεύεται και εμφανίζεται σε scoreboard, που δείχνει πρόοδο και πληρότητα ολοκλήρωσης.

Απαιτήσεις

1 ή 2 φοιτητές. Γνώση React, βασική εξοικείωση με τοπικά LLMs (Ollama, Mistral-7B), Python/FastAPI για integration με Stable Diffusion, βασική κατανόηση JSON-based validation.

Παραδοτέα

Αναφορά, λειτουργική web εφαρμογή guided chat, τοπική εγκατάσταση Stable Diffusion & Ollama, curated dataset evidence & counter-intents, scoreboard προόδου και bias καταγραφών.

Σχετικές αναφορές

TBC

Επιβλέποντες

Φωκά Α.

Τίτλος

Ανάλυση μεροληψίας φύλου στα επαγγέλματα με χρήση ανοικτού κώδικα GradBias και OpenBias dataset

Περιγραφή

Η εργασία στοχεύει στη μελέτη του φαινομένου της **μεροληψίας φύλου στα επαγγέλματα** (gender–occupation bias) σε μοντέλα παραγωγής εικόνας από κείμενο (text-to-image models).

Ο/η φοιτητής/τρια θα αξιοποιήσει τον **ανοικτό κώδικα του GradBias** και το **OpenBias dataset** για την ανάλυση και ποσοτικοποίηση προκαταλήψεων σε prompts που αφορούν επαγγέλματα με κοινωνικά φορτισμένες προσδοκίες (π.χ. *nurse, doctor, engineer, teacher, mechanic*).

Η μεθοδολογία θα περιλαμβάνει:

- επαναχρησιμοποίηση και προσαρμογή του pipeline του OpenBias για παραγωγή εικόνων με ουδέτερα συμφραζόμενα (π.χ. “A [profession] working in an office / outdoors”),
- ανάλυση των παραγόμενων αποτελεσμάτων με GradBias για την **ανίχνευση λέξεων-κλειδιών** που επηρεάζουν το bias,

- υπολογισμό **ποσοτικών μετρικών** (π.χ. ποσοστό εικόνων που απεικονίζουν άνδρα/γυναίκα ή συγκεκριμένο εργασιακό περιβάλλον) μέσω Visual Question Answering (VQA) ή image tagging,
- ποιοτική παρουσίαση ευρημάτων και οπτική σύγκριση επαγγελμάτων.

Η εργασία συνδυάζει ανάλυση δεδομένων, δημιουργία εικόνων και οπτικοποίηση προκαταλήψεων, συμβάλλοντας στην κατανόηση των κοινωνικών διαστάσεων των συστημάτων ΤΝ.

Απαιτήσεις

1 ή 2 φοιτητής/τρια. Γνώσεις Python και βασική εξοικείωση με PyTorch ή περιβάλλοντα παραγωγής εικόνων (Stable Diffusion). Ενδιαφέρον για κοινωνικές και ηθικές διαστάσεις της Τεχνητής Νοημοσύνης.

Παραδοτέα

Αναφορά με περιγραφή μεθοδολογίας, ανάλυση και σχολιασμό αποτελεσμάτων.
Επεκτάσεις του OpenBias dataset με νέα prompts.
Ποσοτικά αποτελέσματα και οπτική παρουσίαση.
Κώδικας αναπαραγωγής πειραμάτων.

Σχετικές αναφορές

D’Inca, M. et al. (2025). *GradBias: Unveiling Word Influence on Bias in Text-to-Image Generation*. arXiv:2408.16700.

Επιβλέποντες

Φωκά Α.

Τίτλος

Ανάλυση πολιτισμικών προκαταλήψεων σε μοντέλα παραγωγής εικόνας με χρήση GradBias και OpenBias

Περιγραφή

Η πτυχιακή εργασία εστιάζει στην ανίχνευση και ποσοτικοποίηση **πολιτισμικών προκαταλήψεων (cultural bias)** σε μοντέλα παραγωγής εικόνας από κείμενο. Με αξιοποίηση του **open-source περιβάλλοντος GradBias** και δεδομένων από το **OpenBias dataset**, ο/η φοιτητής/τρια θα μελετήσει πώς τα μοντέλα απεικονίζουν καθημερινές έννοιες ή επαγγέλματα σε διαφορετικά πολιτισμικά πλαίσια. Η μελέτη θα περιλαμβάνει:

- δημιουργία ή επιλογή prompts όπως “A wedding ceremony”, “A doctor”, “A classroom”, “A meal” σε συνδυασμό με γεωγραφικές/πολιτισμικές αναφορές (“in Africa”, “in Japan”, “in Greece”),
- επαναχρησιμοποίηση του pipeline του OpenBias για παραγωγή και αξιολόγηση εικόνων με βάση bias concepts,
- **ανάλυση συνεισφοράς λέξεων** (word influence) μέσω GradBias,
- **ποσοτικοποίηση διαφορών** (π.χ. στο περιβάλλον, ένδυση, χρώμα δέρματος) με μετρικές VQA ή CLIP similarity,
- οπτική και ποιοτική παρουσίαση των αποτελεσμάτων (visual summaries, gallery).

Η εργασία συνδυάζει τεχνική ανάλυση με ερμηνευτική προσέγγιση, προωθώντας την κατανόηση της πολιτισμικής αναπαράστασης στα συστήματα TN.

Απαιτήσεις

1 ή 2 φοιτητής/τρια. Γνώσεις Python και βασική εμπειρία με εργαλεία μηχανικής μάθησης ή επεξεργασίας εικόνων. Ενδιαφέρον για ζητήματα πολιτισμικής αναπαράστασης και ηθικής στην Τεχνητή Νοημοσύνη.

Παραδοτέα

- Αναφορά με μεθοδολογία και σχολιασμό αποτελεσμάτων.
- Προσαρμοσμένο υποσύνολο ή επέκταση του OpenBias dataset.
- Ποσοτικά και ποιοτικά αποτελέσματα.
- Κώδικας ανάλυσης και οπτική παρουσίαση ευρημάτων.

Σχετικές αναφορές

D’Inca, M. et al. (2025). *GradBias: Unveiling Word Influence on Bias in Text-to-Image Generation*. arXiv:2408.16700.

Επιβλέποντες

Φωκά Α.

Τίτλος

Ανάλυση μεροληψίας φύλου μέσα από αξιακά και περιγραφικά prompts σε μοντέλα παραγωγής εικόνας με χρήση GradBias και OpenBias

Περιγραφή

Η εργασία διερευνά τη **μεροληψία φύλου (gender bias)** σε μοντέλα παραγωγής εικόνας από κείμενο (text-to-image models), εστιάζοντας όχι μόνο στο περιγραφικό πλαίσιο ενός επαγγέλματος, αλλά και στις **αξίες, συναισθήματα και κοινωνικές προσδοκίες** που συνδέονται με αυτό.

Ο/η φοιτητής/τρια θα αξιοποιήσει τον ανοικτό κώδικα του **GradBias** και το **OpenBias dataset**, επεκτείνοντάς το με **σύνθετα, νοηματικά εμπλουτισμένα prompts** (π.χ.

*“A doctor devoted to saving lives”,
“A teacher inspiring young minds”,
“An engineer driven by innovation”,
“A nurse caring for those in need”*).

Η μεθοδολογία θα περιλαμβάνει:

- **προσαρμογή του OpenBias pipeline** για παραγωγή εικόνων από “value-driven” prompts,
- **ανάλυση των λέξεων ή εκφράσεων που προκαλούν bias** με GradBias (gradient attribution σε επίπεδο semantic weight),
- **ποσοτική αξιολόγηση** (π.χ. ποσοστό απεικονίσεων ανά φύλο ή συναισθηματικό τόνο) μέσω εργαλείων Visual Question Answering (VQA) και CLIP similarity,
- **ποιοτική ερμηνεία** των εικόνων, με σχολιασμό των τρόπων με τους οποίους οι αξίες συνδέονται με στερεότυπα φύλου.

Η εργασία στοχεύει να αναδείξει πώς οι αφηρημένες έννοιες αξίας ή συναισθήματος επηρεάζουν τις έμφυλες αναπαραστάσεις, συνδυάζοντας μεθοδολογίες μηχανικής μάθησης με θεωρητική ανάλυση κοινωνικής μεροληψίας.

Απαιτήσεις

1 ή 2 φοιτητές/τριες. Γνώσεις Python και βασική εξοικείωση με PyTorch ή περιβάλλοντα παραγωγής εικόνων (Stable Diffusion). Ενδιαφέρον για κοινωνικές και ηθικές διαστάσεις της Τεχνητής Νοημοσύνης, για γλώσσα και πολιτισμική αναπαράσταση.

Παραδοτέα

Αναφορά με περιγραφή μεθοδολογίας και ανάλυση αποτελεσμάτων. Νέο υποσύνολο prompts βασισμένο σε αξιακές/συναισθηματικές περιγραφές επαγγελμάτων. Ποσοτικά αποτελέσματα (μετρικές bias) και ποιοτική οπτική παρουσίαση. Κώδικας αναπαραγωγής πειραμάτων.

Σχετικές αναφορές

D’Inca, M. et al. (2025). *GradBias: Unveiling Word Influence on Bias in Text-to-Image Generation*. arXiv:2408.16700.

Επιβλέπουσα

Φωκά Α.

Τίτλος

Ποια δεδομένα “γέννησαν” την εικόνα; Εντοπισμός επιρροής εικόνων εκπαίδευσης σε τοπικό Stable Diffusion με ευρετήριο LAION

Περιγραφή

Στόχος είναι η ανάπτυξη ενός **τοπικού pipeline** που, για δοσμένο prompt και την εικόνα που παράγεται από **Stable Diffusion**, εντοπίζει **ποιες εικόνες εκπαίδευσης** από ένα **τοπικά ευρετηριασμένο υποσύνολο LAION** (ή άλλο web-scale dataset) είναι οι πιο “**επιρροϊκές**” (**influential**) στη γένεση του αποτελέσματος.

Η προσέγγιση συνδυάζει **CLIP-βασική αναζήτηση γειτόνων** (k-NN με FAISS) με **μεθόδους attribution** για diffusion models (π.χ. per-step/latents similarity, influence-style δείκτες) και αξιολογείται με **counterfactual πειράματα** (π.χ. unlearning/απο-συμβολή των top υποψηφίων) και **ποιοτική επιθεώρηση** ομοιοτήτων/στυλ.

Παραδοτέα

- **Τοπικό LAION index (FAISS)** και κώδικας **retrieval/attribution**.
- **Αναφορές/πίνακες** με **top-influential training images** ανά prompt και **counterfactual** αποτελέσματα.
- **Demo report:** prompt → παραγόμενη εικόνα → top-k υποψήφιοι από LAION → ομοιότητα/σχόλια → (προαιρετικά) αποτέλεσμα μετά από unlearning/masking.

Απαιτήσεις

1 ή 2 φοιτητής/τρια. **Καλή γνώση Python, PyTorch**, βασική εμπειρία με **CLIP/FAISS**, άνεση με **GPU** περιβάλλον.

Σχετικές αναφορές

Επιβλέπουσα

Φωκά Α.