UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING
Department of Systems Design Engineering

2025 Systems Design Engineering Graduate Symposium

Tuesday, April 8, 2025

Faculty Hall
E7-7363

Schedule

April 8, 2025
Faculty Hall

10:00 a.m.-10:30 a.m.
Welcome Coffee and Snacks

10:30 a.m.- 11:45 a.m.
Student Research Seminar Session 1
Chairs: Eihab Abdel-Rahman and Jessie Ma
Speakers: Sabrina Saiko, Yasser Shama, Wendy Ding, Marco Moran-Ledesma, and Rui Zhou

12:00 p.m.-1:00 p.m.
Pizza Lunch
Special Guest Speaker: Dr. Dan Mashatan, Senior Manager Autonomy & AI SW - Teledyne FLIR Unmanned Systems North America

1:15 p.m.-2:15 p.m.
Student Research Seminar Session 2
Chair: Nasser Lashgarian Azad and Nima Maftoon
Speakers: Chang Liu, Yaoxin Li, Nima Abbasi Firoozjah, and Abdul Abbasi

Student Research Seminar Session 1

Translating Research into Practice: Enhancing Social Robotics for Real-World ASD Therapy (1/2)

Sabrina Saiko

Social robots have emerged as a promising intervention for supporting children with underdeveloped social skills, particularly those with Autism Spectrum Disorder (ASD). Despite the documented therapeutic benefits and commercial availability of social robots, such as QTRobot, Nao, Milo, etc., their adoption remains limited in both clinical and domestic settings (Salimi et al.). While existing research has predominantly concentrated on demonstrating therapeutic efficacy for children, this narrow focus overlooks the critical perspectives of key decision-makers: parents and clinicians. Beyond positive outcomes for children, factors like cost-effectiveness, usability, and practical implementation significantly influence technology adoption by caregivers and healthcare providers.

During the past decade, Social Robotics has gained more use in the therapeutic field, particularly in pediatrics. And particularly Autism Spectrum Disorder (ASD) started to draw more and more attention from the specialists of HRI. (Centers for Disease Control and Prevention, 2024) Many of their projects, like the works of K. Dautenhahn, or the initiative AskNAO (Falconer, 2013) have proved that robots are effective in ASD diagnosing and assessment. (Hourihan, 2024) According to them, robots can establish a stable and consistent environment while facilitating social interactions during structured activities. Moreover, their capacity to execute tasks or activities with unwavering uniformity is invaluable for discerning and analyzing behavioral patterns across various children within identical environments and circumstances. This attribute holds particular significance in the context of working with individuals diagnosed with autism spectrum disorder. However, social robots are not yet widely used in real-world therapy.
Thus, we decided to investigate and bridge the gap between research on the potential benefits of social robots and their real-life application.

Translating Research into Practice: Enhancing Social Robotics for Real-World ASD Therapy (2/2)

Sabrina Saiko

We employ a mixed-methods approach, combining surveys and semi-structured interviews to collect insights from parents of children with underdeveloped social skills and clinicians providing therapeutic services. By comparing their expectations and identifying implementation barriers, we aim to bridge the gap between the proven therapeutic potential of social robots and the challenges of real-world adoption.
To support our study, we conducted a systematic literature search using PubMed, IEEE Xplore, and Scopus, focusing on peer-reviewed studies published between January 2014 and December 2024. These studies examine the design and application of social robots for ASD interventions in home and clinical settings, with an emphasis on robot-child interactions and their therapeutic benefits.

The initial findings of our study highlight several key barriers to the adoption of social robots in therapeutic care for individuals with ASD. These include inconsistent robot design—such as varying personalities and appearances—lack of standardized evidence and guidelines and limited user-friendly interfaces for caregivers and therapists. Despite the promising potential of social robots, these challenges must be addressed to facilitate broader integration into therapy settings.

Additionally, we are actively collecting data from key stakeholders—parents and clinicians—to compare and validate our theories and findings from the literature. This data will provide real-world evidence to support our research and offer a more comprehensive understanding of the barriers to the integration of social robots in ASD therapeutic care. Ultimately, our findings will offer valuable insights to drive effective, real-world implementation of social robots in therapeutic environments.

Bifurcation Drift in Electrostatic MEMS Sensors: Characterization and Mitigation Strategies

Yasser Shama

Microelectromechanical systems (MEMS) have been utilized over the past two decades in a wide range of sensing applications due to their compact size, high accuracy, and ability to operate in real-time. Among these, electrostatic MEMS are particularly valued for their speed, low power consumption, and independence from external field sources. Bifurcation-based MEMS sensors are a recent addition that seeks to improve the signal-to-noise ratio (SNR). Bifurcations are nonlinear phenomena where a system response undergoes a sudden qualitative change in response amplitude, providing an ideal setup to increase a sensor's SNR.

A bifurcation in the vicinity of primary resonance of electrostatic MEMS has widely been used in binary sensing applications. Binary sensing relies on the sensor transitioning between On and Off states through a bifurcation when exposed to a stimulus. This is achieved by operating the sensor at a set-off frequency very close to the bifurcation location and allowing an external stimulus to shift the bifurcation point past the operating point, resulting in a sudden jump (detection signal). Thus, it is essential that the bifurcation location is unaffected by other factors to protect against false positives.

In this work, we investigated the stability and repeatability of electrostatic MEMS bifurcations over prolonged periods of time. The sensors were subjected to two test conditions: continuous frequency sweeps and long-term residence on a resonant branch. We found that prolonged high-amplitude oscillations undermine repeatability and cause significant shifts in the bifurcation location toward lower frequencies by building up plastic deformations that reduce the capacitive gap. Biased excitation waveforms were also found to lead to charge buildup within dielectrics, exacerbating bifurcation drift. In comparison, stiffer in-plane sensors with no metallization operating under unbiased waveforms showed dramatic improvement in repeatability. These findings confirm the effectiveness of our proposed mitigation strategies and establish key recommendations for improved sensor design.

Age-Related Differences in Eye Movements During Level 3 Autonomous Vehicle Takeover (1/2)

Wendy Ding

Semi-autonomous vehicles have become increasingly accessible in our daily lives, offering enhanced comfort and convenience. With the capacity of navigation between locations with minimal human intervention, it notably decreases the attention needed to drive a car. This functionality is particularly beneficial for older adults, who may experience reduced energy levels and cognitive capacities. However, level 3 autonomous driving systems still expect drivers' intervention in situations that exceed their operation capacity. In such cases, drivers need to conduct the takeover process and successfully control the vehicle manually. This transition can be particularly challenging, especially when drivers have been engaged in non-driving-related tasks prior to the Takeover Request (TOR).

In the takeover process, drivers must rapidly gather and process a substantial amount of information, primarily through their vision. Consequently, understanding drivers' eye movements during this process is essential for gaining insights into their reaction processes and decision-making mechanisms. Previous research has demonstrated age-related differences in eye movements behaviors, which may contribute to variations in their takeover performance. Therefore, investigating eye movements patterns across drivers from different age groups is crucial for understanding age-specific behaviorsThis knowledge can be leveraged to enhance the safety and efficiency of the takeover process by tailoring takeover warnings to meet the specific needs of each group.

In this study, drivers from three age groups—young (18–24 years), middle-aged (35–55 years), and older (65+ years)—were recruited, with 24 participants in each group. Each participant completed 12 simulated driving scenarios, and their eye movement data was recorded using Dikablis 3 eye tracking glasses.

Age-Related Differences in Eye Movements During Level 3 Autonomous Vehicle Takeover (2/2)

Wendy Ding

Three categories of eye tracking features were analyzed based on 4 pre-defined Area-of-Interests (AOIs), including dwelling time, entry count, and entropy (Stationary Gaze Entropy and Gaze Transition Entropy). Given the high consistency in performance between young and middle-aged drivers, these two groups were combined, resulting in the comparison between younger and older groups. The results indicated that older drivers exhibited significant lower dwelling time in the "Dashboard" from 5s before the TOR to the end, as well as significant lower dwelling time in the "Mirror" after TOR was given compared to younger drivers. In terms of entry count, the older drivers demonstrated a significant higher number of entries towards "Mirrors", "Screen" and "Others" from 5s before the TOR to the end, along with a significant higher number of entries towards "Center" from TOR to the end.

Since older drivers exhibited significantly lower dwelling time on the Dashboard before and after the TOR, it suggests that they may not be adequately monitoring critical vehicle information. The higher entry count in certain AOIs among older drivers suggests that they might engage in excessive visual scanning. To better support older drivers, TOR designs could improve dashboard alerts by incorporating larger fonts, high-contrast visual cues, and centralized critical information to ensure key messages are noticed.
Integrating these design considerations would allow future TOR systems to be tailored to drivers' age-related needs, providing safer and more efficient transitions from automated to manual driving.

TRAIN-KNEE: Developing a Haptic Manikin for Knee Injury Assessment Training

Marco Moran-Ledesma

We present the design, development, and evaluation of a high-fidelity haptic manikin for knee injury assessment training. Our device is designed to help novice practitioners gain hands-on experience with injury assessment techniques, focusing on the medial collateral ligament (MCL). Proper assessment of MCL injuries is critical for accurate diagnosis and treatment planning, but it can be challenging for those with limited clinical experience. Our haptic manikin offers a controlled, risk-free, and repeatable training environment, allowing practitioners to develop their skills.

Our manikin was designed in close collaboration with a certified clinician to ensure anatomical accuracy and practical utility. We integrated several key components: a commercial human knee joint model for realistic anatomical representation, skin-like materials that simulate human skin properties, an injury simulation mechanism capable of reproducing different MCL injury conditions, and pressure sensors to capture the force applied by the user during manipulation. Our design choices ensure that the manikin provides both a realistic tactile experience and quantitative feedback for trainees.

To evaluate the effectiveness of our manikin, we conducted two rounds of testing. In the first one, our collaborating clinician configured the manikin to represent four distinct MCL conditions (i.e., healthy, grade 1, grade 2, and grade 3 injuries) using the staircase psychophysics method. In our second round, six additional certified clinicians participated in a validation study where they assessed each condition and provided ratings for consistency. Our results indicate that the manikin effectively differentiates between healthy and unhealthy MCL conditions, achieving a sensitivity of 0.83 and specificity of 1.00 for identifying healthy knees, and 1.00 and 0.83, respectively, for detecting injury. However, distinctions between injury grades require further refinement of our device.

Feedback from clinicians highlighted the realistic weight and shape of the manikin as significant strengths, though improvements in skin texture simulation could enhance the overall experience. Future iterations will focus on refining the injury differentiation mechanism and optimizing material properties to provide an even more immersive and effective training tool.

Sensitive, Accurate and Fast Terahertz Detectors Design

Rui Zhiu

Terahertz (THz) technology has witnessed rapid advancements and is renowned for its deep penetration capabilities, non-ionizing nature, and sensitivity to molecular vibrations. These unique properties render THz technology highly applicable across diverse domains such as biomedical detection, material analysis, and non-destructive testing, offering significant potential for scientific and industrial advancements. However, the efficacy of THz systems is currently hampered by significant challenges in the detectors used. These include limited sensitivity, accuracy, response speeds, and the high costs associated with existing technologies. Such limitations are critical barriers to the development of high-performance THz detection systems and highlight the pressing need for innovative research to devise effective solutions.

This research introduces a groundbreaking THz detection methodology designed to significantly enhance sensitivity, accuracy, and the speed of data acquisition, with a special focus on biomedical imaging applications. Central to this innovative approach is a novel micro-bolometer specifically engineered to leverage the thermal effects of THz radiation. This detector is adept at identifying changes in material resistance triggered by temperature fluctuations, which is essential for achieving precise THz imaging. The micro-bolometer incorporates advanced two-dimensional materials, markedly enhancing its sensitivity, accuracy, and response times. These materials allow for the tailoring of the detector’s properties to meet specific operational demands, thereby optimizing its performance across various applications. Moreover, this research integrates complementary technologies such as antennas, metasurfaces, and low-noise amplifiers. These enhancements are critical in boosting the overall performance of the detection system, enabling more robust and reliable THz detection capabilities. The integration of these sophisticated components not only improves the quality of THz detection but also establishes a foundation for innovative approaches in biomedical imaging. By advancing THz detector technology, this research represents a substantial leap forward, promising more sensitive, precise, faster, and cost-effective imaging solutions. Overall, the findings of this study are set to make a significant impact on the evolution of THz technology, opening new possibilities for its application in science and industry, and potentially revolutionizing how we perceive material properties and biological structures at the molecular level.

Student Research Seminar Session 2

Language-guided Context-aware Unsupervised Semantic Segmentation Under Distribution Shift (1/2)

Chang Liu

Semantic segmentation is a dense prediction task that demands expensive and time-consuming pixel-level annotations. This problem worsens with domain shifts, requiring additional annotations as the data evolves. To reduce the need for manual annotation, unsupervised domain adaptation for semantic segmentation (DASS) methods train segmentation networks – usually based on a student-teacher architecture – on an available labeled source domain and adapt them to an unlabeled target domain.

Capturing context – the spatial relationships between objects – is key to accurate image segmentation. Existing vision-only unsupervised domain adaptation (UDA) methods attempt to build context-awareness by implicitly learning spatial relationships between different image patches. For example, HRDA uses high-resolution crops along with low-resolution images to capture fine-grained and long-range context dependencies. MIC masks target images in the source-trained model to enhance adaptation. While these methods improve DASS, they are vulnerable to noisy pseudo-labels biased toward the source domain. To address this, CoPT uses language-based generalized representations to guide the student network by aligning class-level text features (e.g., “a photo of a {car}“) with pixel features. However, aligning text and pixel features separately ignores context relationships between different classes (e.g., “car“ and “pedestrian“).

We propose LangDA, which explicitly builds contextual understanding through text. Unlike existing language-guided methods that use generic class-level prompts, LangDA aligns image scenes with descriptive captions (e.g., “a {pedestrian} is on the {sidewalk}, and the street is lined with {buildings}“). This provides models with language-based guidance on object relationships. To fully exploit context relationships, we introduce an image-level consistency module that brings the entire image’s features closer to the corresponding caption. Leveraging contextual relationships in language enables better generalization to target data.

Language-guided Context-aware Unsupervised Semantic Segmentation Under Distribution Shift (2/2)

Chang Liu

LangDA differs from existing language-guided DASS approaches in three ways:
Context-Aware Captions – Existing methods rely on text to describe the domain gap (e.g., “daytime photo“ to “nighttime photo“), which fails with unpredictable domain shifts (e.g., tumor scans with color and location variations). LangDA uses a context-aware caption generator (seeded by VLM and LLM) to automatically create image-specific captions, ensuring effectiveness even for unknown domain gaps. No Manual Prompt Tuning – LangDA eliminates the need for manual prompt tuning and human feedback, which are inconsistent, hard to reproduce, and resource-intensive. Standardized captions simplify benchmarking and remove prompt engineering efforts. Scene-Level Alignment – Instead of matching individual class embeddings (e.g., “car“ with its pixel embedding), LangDA aligns a descriptive scene embedding with its corresponding image embedding (image-level alignment), encoding richer object relationships through language.

We validate LangDA on three DASS settings: Synthia → Cityscapes, Cityscapes → ACDC, and Cityscapes → DarkZurich. LangDA achieves state-of-the-art performance in all three UDA settings, surpassing existing methods by 2.6%, 3.9%, and 1.4% respectively, for the challenging semantic segmentation tasks with domain shifts. Ablation studies further highlight the superiority of context-aware image-level alignment over pixel-level alignment. These results confirm LangDA’s capacity to extract spatial relationships encoded in language for robust domain adaptation.

Mamba fusion for multimodality action quality assessment

Yaoxin Lin

Action Quality Assessment (AQA) aims to objectively evaluate human action execution against expert standards, playing a pivotal role in sports training, medical rehabilitation, and skill certification. While recent advances leverage multimodal data (video, skeleton pose, motion flow signals extracted from the video data) to enhance evaluation accuracy, two systemic bottlenecks persist in conventional approaches. First, the quadratic computational complexity inherent in Transformer-based fusion architectures severely limits their scalability for processing long video sequences, a critical requirement for real-world deployment. Second, the heterogeneous nature of multimodal features – spanning spatial, temporal, and kinematic domains – creates fundamental alignment challenges that current fusion strategies inadequately address, leading to suboptimal information integration.

To overcome these limitations, we present Mamba Fusion, a novel paradigm that synergizes the sequential processing efficiency of Mamba's selective state-space models (SSM) with hierarchical multimodal reasoning. The architecture capitalizes on Mamba's linear-complexity temporal modeling to capture extended action patterns across extended durations, effectively addressing the computational bottlenecks of attention-based methods. Our hierarchical fusion framework operates through two complementary stages: modality-specific Mamba towers first distill discriminative spatiotemporal features within individual data streams (RGB frames, pose coordinates, optical flow fields), followed by a novel cross-modal Mamba module that dynamically establishes inter-modal correlations through learnable attention masks and adaptive feature recalibration. This dual-stage mechanism ensures both efficient intra-modal processing and context-aware multimodal synthesis.

Experimental validation across different benchmarks demonstrates the framework's superior performance-efficiency tradeoff. The model achieves competitive accuracy with state of the art Transformer baselines while reducing computational costs and parameter counts. Qualitative analyses reveal enhanced interpretability in cross-modal interactions, particularly in aligning kinematic patterns from pose data with temporal dynamics in video streams. These advancements suggest broad applicability beyond AQA to multimodal time-series analysis tasks requiring efficient long-range dependency modeling. Ablation studies validate the effectiveness of our hierarchical fusion strategy.

A General Computational Framework for Restoring Cellular-Resolution in Optical Coherence Tomography Images of the Eye

Nima Abbasi Firoozjah

Achieving full-volume cellular-resolution imaging of the eye with optical coherence tomography (OCT) is constrained by aberrations that limit the depth range over which sufficient lateral resolution and signal-to-noise ratio(SNR) can be maintained. Among many solutions proposed for this problem, computational methods stand out as they offer the correction of these aberrations without compromising imaging speed or quality, and without introducing additional costs to the OCT setups. However, these solutions are only applicable to the data acquired with complex OCT systems featuring high phase stability. In addition, the in vivo application of these solutions in the human eye has been limited to the posterior layers, i.e., retinal tissues. The proposed research aims to address these challenges through innovative computational methods that do not rely on phase-stable OCT setups. Such development holds the potential to offer a system-agnostic approach for cellular resolution imaging of the human eye in full volumes, enabling seamless integration into any clinical OCT system for diagnostic purposes.

In the first project, we explore methods for aberration- and noise-compensated reconstruction of OCT intensity images, using non-local images priors within a maximum a posteriori (MAP) framework. The results demonstrate significant detail enhancement in OCT images of various sample types including the healthy human cornea. Next, a physics-informed diffusion model (DM) is developed for super-resolved and defocus-corrected OCT image reconstruction. The non-local priors from the first phase are incorporated into the DM in the form of conditioning inputs that guide the reverse diffusion process. Zero-shot implementation results indicate the great potential of this approach. Subsequently, we will investigate the direct training of a latent-space diffusion model for learning and reversing aberration in OCT images including defocus. The performance of the developed framework will be quantitatively evaluated in terms of SNR, sharpness, and contrast improvements, and qualitatively in its facilitation of applications such as cell counting, segmentation, and dynamic cell studies in the human cornea.

By addressing key limitations in cellular-resolution OCT imaging of the eye, this research has the potential to enable non-invasive, high-fidelity cellular imaging of the eye in full volumes. The proposed methods could facilitate widespread clinical adoption of cellular-resolution OCT, extending its diagnostic capabilities for early detection of eye diseases.

Empowering Security Analysts: A User-Centric Approach to Explainable AI for Insider Threat Detection

Abdul Abbasi

The increasing prevalence of insider-driven data breaches within organizations underscores the urgent need for advanced detection systems to protect sensitive information. In this research, the application of deep learning techniques is explored to model typical user behaviour and detect deviations indicative of insider threats, utilizing the CERT r5.2 dataset—encompassing 501 days (around 17 months) of user action logs from 2000 users, including 99 labelled insiders. This study builds upon initial efforts by adopting a dual approach to anomaly detection: a Gated Recurrent Unit (GRU)-based Recurrent Neural Network (RNN) autoencoder with an attention mechanism to capture short-term dependencies in user action sequences and a Long Short-Term Memory (LSTM) autoencoder with attention to addressing long-term behavioural patterns. Anomalies in both the order and frequency of user actions across varying temporal scales are effectively identified through this combination, with robust experimental results demonstrating high accuracy in pinpointing insider threats. To enhance the interpretability of these black-box models, the SHAP (SHapley Additive exPlanations) framework has been integrated to generate Shapley values, offering detailed insights into feature importance. Global perspectives, such as beeswarm plots, provide security analysts with a comprehensive understanding of the most influential features driving model predictions across the dataset, while local perspectives, including waterfall plots, clarify the decision-making process for individual predictions. These explainability tools enable analysts to better understand the models' inner workings, fostering trust and mitigating the persistent challenge of false positives in insider threat detection. Current results affirm the effectiveness of this approach, highlighting its potential to bridge the gap between computational detection and human oversight.

Moving forward, the refinement of this deep learning-based insider threat detection system is planned by further investigating temporal dynamics and optimizing its integration into Security Operations Centers (SOCs). A key element of this research involves a user testing study with SOC analysts to assess the system's explainability, usability, and influence on decision-making. Feedback on how well analysts comprehend and trust the model's outputs—supported by SHAP-driven visualizations—will be solicited to iteratively improve the system's design, ensuring alignment with operational requirements. This work aims to deliver a user-centric, explainable AI framework that not only detects insider threats with precision but also provides security professionals with actionable insights, promoting greater adoption and effectiveness in real-world cybersecurity contexts.

THANK YOU!

UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING
Department of Systems Design Engineering