

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/389389984>

TORNADO: Foundation Models for Robots that Handle Small, Soft and Deformable Objects

Preprint · February 2025

DOI: 10.13140/RG.2.2.16483.67362

CITATIONS

0

READS

389

20 authors, including:



Andreas El Saer

University of West Attica

17 PUBLICATIONS 202 CITATIONS

SEE PROFILE



A. Sanfeliu

Polytechnic University of Catalonia

399 PUBLICATIONS 8,706 CITATIONS

SEE PROFILE



Anais Garrell

Polytechnic University of Catalonia

66 PUBLICATIONS 1,318 CITATIONS

SEE PROFILE



Martin Cech

University of West Bohemia

79 PUBLICATIONS 657 CITATIONS

SEE PROFILE

TORNADO: Foundation Models for Robots that Handle Small, Soft and Deformable Objects

Maria Moutousi¹, Andreas El Saer¹, Nikos Nikolaou¹, Alberto Sanfeliu², Anaís Garrell², Lukáš Bláha³, Martin Čech³, Evangelos K. Markakis⁴, Ioannis Kefaloukos⁴, Marta Lagomarsino⁵, George Margetis⁶, Emmanouil Adamakis⁶, Athanasios Poulakidas⁷, Filopoumin Lykokanellos⁷, Artemis Stefanidou⁸, Jorgen Cani⁸, Panagiotis Radoglou-Grammatikis⁹, Marios Siganos⁹, Arash Ajoudani⁵, Georgios Th. Papadopoulos⁸

¹ Research & Development Department, ITML, Athens, Greece

² Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Universitat Politècnica de Catalunya, Barcelona, Spain

³ NTIS research center, University of West Bohemia, Pilsen, Czech Republic

⁴ Department of Computer & Electrical Engineering, Hellenic Mediterranean University, Heraklion, Greece

⁵ Human-Robot Interfaces and Interaction, Istituto Italiano di Tecnologia, Genova, Italy

⁶ Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion, Greece

⁷ Netcompany-Intrasoft, Luxembourg, Luxembourg

⁸ Dept. of Informatics and Telematics, Harokopio University Athens

⁹ Department of Research and Development, K3Y Ltd, Sofia, Bulgaria

Abstract—This paper introduces TORNADO, a cloud-integrated robotics platform designed to tackle the challenges of autonomous manipulation in dynamic indoor environments, particularly those involving small, soft, or deformable objects. TORNADO integrates large-scale foundation models for perception, language comprehension, and high-level reasoning, to achieve strong zero-shot generalization across a wide range of tasks. At its core, the platform features an adaptive cognitive pipeline capable of dynamically reconfiguring its modules—including semantic 3D SLAM, people-aware navigation, dexterous manipulation, and human-in-the-loop learning—to manage uncertainty and adapt to changing conditions. Additionally, TORNADO incorporates a multi-modal Learning-from-Demonstration interface and an Explainable AI engine, enhancing transparency and easing the burden of teaching new tasks. The system is validated through three industry-relevant scenarios: (1) flexible gear and ply-sheet handling in a mechanical parts factory, (2) patient support in a hospital palliative ward, and (3) product sampling and waste management in a distribution center. TORNADO aims to significantly enhance the agility, safety, and overall task performance of mobile manipulators operating in dynamic, human-centric environments.

Keywords— *robotics, foundation models, deformable object manipulation, explainable AI, learning from demonstration, 3D SLAM, adaptive robotics, human robot interaction, healthcare, manufacturing, waste management.*

I. INTRODUCTION

Autonomous Mobile Robots (AMRs) equipped with arms and grippers are becoming increasingly common in industrial and service environments, where they perform a variety of manipulation tasks [1]. While current systems are quite effective at handling rigid objects in relatively uncluttered spaces—executing basic actions like grasping, sliding, pushing, and poking—many real-world applications demand far greater dexterity. A significant challenge arises when dealing with small, soft, or deformable objects (SSDs) in crowded, dynamic environments, particularly those shared with humans. Unlike rigid objects, SSDs are inherently unpredictable—their shapes and physical properties can change in response to interactions with the robot or surrounding environment [2], [3], [4]. Additionally, their complex internal structures and dynamic behaviors make them difficult to model and control using traditional approaches. As a result, existing AMRs struggle to meet the demands of such tasks, particularly when operating under

real-time constraints, ensuring human safety, and adapting to unforeseen variations. Overcoming these challenges is crucial for advancing robotic manipulation in more unstructured and interactive settings.

Recent advances in machine learning suggest that Foundation Models (FMs)—large-scale pretrained Deep Neural Networks (DNNs) capable of zero-shot generalization and exhibiting emergent properties—could open up new possibilities for tackling complex robotic manipulation tasks [5], [6], [7], [8], [9], [10]. These models have already transformed fields like computer vision (e.g., DINO [11], CLIP [12]) and Natural Language Processing (NLP) (e.g., Large Language Models (LLMs) such as GPT-3 [13] and GPT-4 [14]). However, when it comes to robotics, their potential remains largely untapped, with research still in its early stages.

Deploying robots in real-world environments presents additional challenges due to their dynamic and unpredictable nature. This complexity is further amplified in Human-Robot Interaction (HRI) scenarios, where robots must maintain safety, real-time responsiveness, and consistent reliability while collaborating or coexisting with humans [15], [16], [17], [18], [19]. Meeting these demands requires advanced mechanisms such as Out-of-Distribution Detection (OOD) [20], Test-Time Adaptation (TTA) [21], and Few-Shot Adaptation (FSA) [22] along with robust frameworks for assessing human trust [23] and modeling behavior [24], [25].

Without these capabilities, even the powerful reasoning and multimodal understanding of foundation models (FMs) may not be enough to ensure stable task performance or safe collaboration in environments filled with uncertainties and unexpected variations [26]. Developing these adaptive mechanisms is crucial for advancing robotic autonomy in complex, human-centric settings.

In this paper, we introduce TORNADO, a novel cloud robotics platform designed to harness the potential of FMs for real-time, adaptable, and cognitively advanced robotic manipulation, especially in the handling of SSDs within cluttered and human-populated environments. TORNADO envisions a robot agent deployed under human supervision and guided by high-level instructions (e.g., verbal commands), while most of the computationally intensive cognitive processing resides in a pool of cloud-hosted FMs

with zero-shot generalization capabilities [27]. To prevent catastrophic forgetting and maintain adaptability, TORNADO continuously monitor and adjust the active models, exploiting human feedback when available and self-supervised learning otherwise.

Central to TORNADO is an Adaptive Cognitive Pipeline Manager (ACPM) that automatically assembles an optimal pipeline based on the environment, current robot goals, and available FMs. Depending on the scenario, TORNADO may choose either a model-based approach—where FMs improve semantic task plans for low-level controllers—or an end-to-end pipeline—where Reinforcement Learning (RL) or imitation learning agents directly command actuators. In cases of task failure or entirely new skill requirements, on-the-fly learning is facilitated through novel Learning-from-Demonstration (LfD) from multi-modal inputs (e.g., vision language, and tactile) and Augmented Reality (AR) interfaces, bolstered by Explainable AI (XAI) techniques. TORNADO will be validated across three industrial use-cases—flexible small gears manipulation, palliative patient care, and product quality sampling/waste picking—each involving dynamic, unpredictable settings with complex human-robot interaction. TORNADO aims to redefine the state of the art in cognitive robotics, empowering machines to handle novel tasks in ever-changing environments with minimal human intervention or engineering overhead.

II. BACKGROUND AND RELATED WORK

A. Existing AMR Solutions and Limitations

Autonomous Mobile Robots (AMRs) are designed to operate with minimal human intervention, even in unpredictable or partially unknown environments. To achieve this, they require robust navigation systems capable of real-time obstacle avoidance and route planning. As depicted in Figure 1, there are four pillars of challenges in AMRs with respect to localization, navigation, obstacle avoidance and path planning. Depending on whether they function indoors or outdoors, AMRs utilize a range of sensors—such as sonar, inertial measurement units (IMUs), and external range finders—to perceive and map their surroundings. Over the past few decades, mobile robots have significantly increased productivity in numerous fields, including manufacturing, agriculture, military, and education. The adoption of AMRs has accelerated further due to the COVID-19 pandemic, as various sectors—particularly healthcare, security, and food services—shift toward minimizing human-to-human interaction by employing human-to-machine interfaces instead [28]. AMRs distinguish themselves through their ability to make autonomous decisions: they perceive their surroundings, interpret or recognize relevant information, and execute actions or manipulations based on the acquired knowledge. This high level of autonomy makes AMRs more promising and effective than automated guided vehicles, as they do not require physical guidance systems or centralized control for navigation [29].

AMRs have been increasingly deployed in complex missions—such as surveillance, disaster response [30], [31] and domestic applications [32], [33], [34]—where

environmental uncertainties are paramount.

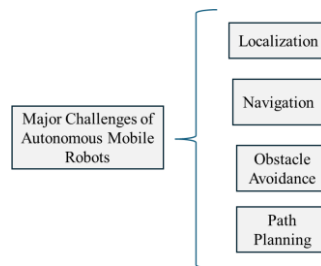


Figure 1: Challenges in AMRs include Localization, Navigation, Obstacle Avoidance and Path Planning.

AMRs can move using wheels, legs, or a hybrid combination. Wheeled robots are widely favored for their mechanical simplicity, stability, and energy efficiency [35]. However, they struggle on rough or obstructed terrain, where legged or hybrid robots offer greater adaptability. While legged robots are better suited for uneven surfaces, they come with increased mechanical complexity and control challenges. The choice of locomotion depends on factors like maneuverability, terrain conditions, and load capacity [36], [37], [38].

To navigate effectively, AMRs rely on sensors to gather data from both their internal state and external environment. Proprioceptive sensors (e.g., encoders, gyroscopes) monitor the robot’s movement, while exteroceptive sensors (e.g., cameras, sonar, radar) help detect obstacles and map the surroundings. These sensors can be active—emitting signals to measure feedback—or passive, simply capturing environmental stimuli like a camera does. Selecting the right combination of sensors is crucial for accurate object detection, data collection, and localization [37], [39].

For AMRs to move autonomously, they must determine their current position, target destination, and optimal path to reach it. Navigation typically involves sensor-based localization, such as computer vision for object recognition and feature matching [40], [41], [42], [43] alongside trajectory planning, which continuously updates the AMR’s route in response to obstacles. In dynamic settings—such as surveillance, disaster response, and home automation—AMRs must adapt to frequent environmental changes and incorporate real-time obstacle avoidance and path recalculation [44].

Integrating locomotion, perception, and navigation, allows AMRs to achieve fully autonomous and reliable operation. Ongoing research aims to enhance obstacle detection, sensor fusion, and adaptive learning strategies, expanding the potential applications of AMRs in fields like healthcare, mining, and education [32], [45], [46], [47], [48].

However, the degree of adaptivity to dynamically changing scenes or materials that we propose in TORNADO, by improving FMs and other recent technologies in a targeted manner, cannot be realized with existing tools. New research must be done, and novel algorithms must be developed. This requires advancements in real-time perception, sensor fusion, and adaptive control strategies, ensuring that AMRs can respond to unforeseen variations with greater precision and reliability.

B. Foundation Models (FMs) in Robotics

FMs like GPT-3, GPT-4, CLIP, DALL-E, and PaLM-E have demonstrated remarkable capabilities in vision and language processing, thanks to their training on vast, diverse datasets. In general, most FMs capitalize on transformer architecture. Transformers utilize a multi-head self-attention mechanism to capture contextual relationships between tokens efficiently, enabling significantly faster training and inference compared to Recurrent Neural Networks (RNNs) or Long Short Term Memory Networks (LSTMs). For each token, the model computes three key vectors: a query, a key, and a value. It then uses scaled dot products between queries and keys to determine the level of attention each token should give to others within the same context window. Because multiple attention heads operate in parallel, they learn different aspects of similarity, enriching the model's understanding of relationships between tokens. The outputs from these attention heads are then concatenated, passed through feedforward layers, and combined using skip connections, forming a transformer layer. Stacking multiple such layers creates the encoder-decoder architecture that powers many modern LLMs and vision-language models.

In practical applications, Transformers can efficiently manage heavy computational loads by parallelizing these attention computations across GPUs and TPUs. Several factors influence a model's capacity, including the context window size, the number of attention heads per layer, the dimensionality of each attention vector, and the depth of the model (number of layers). For example, GPT-3 features a 2048-token context window, 96 attention heads per layer, a head dimension of 128, and 96 total layers. When used autoregressively—as in text prediction—the model employs positional encodings to retain the sequence order, allowing it to generate tokens one at a time while dynamically updating its context window.

In robotics, these models present an exciting opportunity to enhance adaptability and performance across various tasks, including perception, planning, and control. Their potential spans numerous domains, from autonomous driving to household assistance, industrial automation, and assistive robotics. Foundation models offer the advantage of zero-shot learning, reducing the need for extensive task-specific training and data collection. However, integrating these models into real-world robotics comes with significant challenges. One major hurdle is the scarcity of multimodal sensor data, which is essential for effective model training. Additionally, the variability in physical environments and hardware platforms complicates deployment, while uncertainties like language ambiguity and model hallucination introduce further risks. Ensuring safety and real-time performance is another critical concern—researchers must develop rigorous evaluation frameworks and optimize model architectures to meet the stringent speed, and reliability demands of robotic systems.

Despite these obstacles, ongoing research in AI and robotics suggest a promising future. Foundation models have the potential to drive cross-domain knowledge transfer, leading to more resilient, flexible, and autonomous robotic systems.

C. Small, Soft, or Deformable Objects Manipulation (SSDOM)

Current solutions face challenges in more complex scenarios, such as dexterously handling small, soft or deformable objects (SSDs) within crowded spaces where humans are also operating and interact with the robots [49]. Many SSDs can be unpredictable to manipulate, changing shape and properties in response to contact with the robot or the environment, thus often requiring real-time adjustments.

Deformable Object Manipulation (DOM) pushes robotic grasping beyond the traditional assumption of rigid objects, recognizing that many real-world tasks—spanning from microsurgery to industrial assembly—involve materials that change shape upon contact. This expanded perspective introduces several technical challenges, including accurately sensing deformation, managing the high degrees of freedom in soft materials, and modeling their complex nonlinear behaviors. Despite these difficulties, advancing DOM is crucial for enabling autonomous robots to operate effectively in unstructured environments. Recent research has explored model-based manipulation planning [50], multi-robot collaboration for DOM [51], multi-modal sensing techniques [52], and deformable object modeling approaches [53] often categorizing solutions by material properties [54] or by advances in learning, perception, and control [55].

On the hardware side, DOM tasks frequently demand custom-designed grippers tailored to specific deformable objects. Examples range from cable-sliding end-effectors [56] and towel clips [57] to push-tap tools [58] and soft robotic hands for organ manipulation [59]. While a single, highly dexterous gripper for multiple applications is an attractive concept, practical concerns—such as hygiene, material compatibility, and task-specific constraints—often make specialized solutions more viable. Meanwhile, non-anthropomorphic soft grippers are gaining traction, offering built-in compliance for delicate items like food or biological tissue. Additionally, soft robots, which themselves deform dynamically, introduce an entirely new set of control and modeling challenges. A key research question remains: Can methodologies from soft robotics be adapted to DOM? If so, this could pave the way for a unified framework capable of manipulating both rigid and deformable objects with greater versatility and robustness.

TORNADO aims to introduce novel AI-powered algorithms contributing to AMRs with unprecedented capabilities for SSD manipulation and navigation in complex, dynamic, people-centric indoor environments adaptation to changing conditions. SSDs may have complex internal structures and their motion may be difficult to model accurately.

D. Autonomous Navigation in Dynamic, People-Centric Environments

Visual SLAM has long been a fundamental component of robotic perception, enabling simultaneous localization and mapping across various domains, including augmented reality and autonomous driving. Classical approaches, such as ORB-SLAM [60], [61], [62] and VINS-Mono [63], have demonstrated robust performance in predominantly static environments. However, real-world settings often involve moving objects and unpredictable changes, posing significant challenges for purely geometric methods. In response, recent advancements—exemplified by DS-SLAM [64], DynaSLAM

[65], and Dynam-SLAM [54]—have integrated semantic perception and multisensor fusion, leveraging deep learning to identify and segment dynamic elements. This shift toward semantic SLAM enhances localization accuracy and ensures more consistent mapping, particularly in dynamic, human-centric environments. As a result, modern SLAM systems are evolving beyond purely geometric map construction, enabling robots to interact more intelligently with the changing elements in their surroundings.

Beyond mapping and localization, navigating effectively in human-populated spaces requires robust dynamic obstacle avoidance strategies. A widely used approach in robotics is the velocity-obstacle framework (also known as the collision cone or forbidden velocity map) [66], which identifies and eliminates velocity options that could lead to collisions. Over time, this foundational method has been refined to include better trajectory prediction of surrounding agents [67], [68], [69], [70], [71], accounts for uncertainty in sensing and motion decisions [72], and distributes responsibility across multiple agents using the reciprocal velocity obstacle concept [73]. These methods are highly effective in ensuring safety in crowded, dynamic environments, with some even mirroring observed pedestrian behaviors under specific conditions [74], [75]. However, such approaches are inherently mechanistic, often prioritizing strict collision avoidance at the cost of natural, human-like motion patterns.

An alternative research direction, inspired by social and behavioral studies, aims to model human walking heuristics to enhance the legibility, comfort, and predictability of robotic motion in shared spaces. Concepts such as proxemics [76] and social-force models [77], initially developed to analyze interpersonal space and crowd behavior, have been adapted for robotic applications [78], [79]. For instance, Moussaïd et al. [2], [3] introduced a heuristic for mutual avoidance, which produces smooth and efficient paths, a desirable quality for both human pedestrians and service robots. Expanding on these sociologically inspired models, sampling-based methods [80] and global path planners [17], [81], [82], [83], [84], [85], [86], [87], [88] have incorporated dedicated cost functions and constraints to account for social comfort and clear lines of sight. Additionally, research in learning-based human intent prediction [89], [90], [91], [92], complements these approaches, enabling robots to anticipate human movements and intentions for more fluid navigation in interactive environments.

As these approaches illustrate, achieving autonomous navigation in dynamic, human-centric spaces necessitate a seamless inclusion of perception, motion planning, and social awareness. Modern visual SLAM pipelines, now augmented with semantic understanding [66], [67], [68], allow robots to recognize and localize dynamic objects [93], while sophisticated local navigation algorithms—whether based on velocity obstacles [66], [67], [68], [69], [72], [74], [75] or socially inspired heuristics—ensure motion that is safe, smooth, and human-friendly. The ongoing challenge lies in bridging these methodologies: balancing rigorous, collision-avoidance-driven models with the adaptability and nuance of human-inspired motion heuristics.

E. Human–Robot Interaction, LfD, and XAI Methods

In human–machine interaction, the ability to explain a robot’s actions or decisions plays a crucial role in building

user trust and understanding. These explanations can take various forms, including natural language descriptions [94], [95], [96], trajectory demonstrations [26], [87], [97], visualized movement paths [97], [98], [99], [100], or even decision trees [101]. While many XAI (explainable AI) techniques have been shown to enhance human–robot and human–agent interactions [102], their implementation comes with challenges.

One key trade-off is overreliance on explanations—if users begin depending too much on the robot’s justifications, they may struggle to operate independently. Additionally, certain explanation methods can negatively impact task performance, particularly when they are too complex or intrusive [103]. Another concern is cognitive overload; when explanations are too frequent or difficult to process, users may find themselves distracted rather than helped, potentially leading to a decline in task efficiency. Therefore, an effective balance must be struck—ensuring explanations enhance understanding without overwhelming the user.

Integrating XAI into Learning from Demonstration (LfD) can significantly improve the way humans teach robots new tasks, especially for users with little to no experience in programming. Research has explored various ways to make this process more intuitive.

For example, Luebbbers et al. [104] introduced the concept of counterfactuals—which involves modifying specific conditions to show how they influence learning. By using augmented reality to display a robot’s trajectory both with and without a particular learned parameter, they helped users visually grasp the cause-and-effect relationship between their teaching and the robot’s behavior. Similarly, Sena et al. [97] examined the impact of XAI on LfD by showing users how learned policies generated movement trajectories from different points in a workspace. Their findings revealed that teaching effectiveness improved when explanations clarified how well a robot generalized its learning. However, merely replaying demonstrations from pre-selected or already-taught locations did not offer significant benefits. Through two user studies—one involving a 2D point-to-point reaching task and another focusing on pick-and-place operations—Sena et al. further found that feedback-based explanations could replace explicit rule-based guidance, making human instruction more intuitive.

Despite its benefits, a major limitation of XAI in LfD is that many current approaches rely on handcrafted explanations [97], [99]. As a robot’s operational environment expands, manually curating these explanations becomes impractical. Ultimately, integrating XAI within HRI and LfD represents a promising direction for improving human–robot collaboration. However, refining these methods to balance transparency, usability, and scalability remains a crucial challenge for future research.

III. TORNADO CONCEPT & METHODOLOGY

A. TORNADO Concept and Functional Architecture

The TORNADO project, whose overall functional architecture is illustrated in Figure 2, aims to advance autonomous mobile robot (AMR) operations by integrating state-of-the-art AI-driven mechanisms that enhance robotic perception, cognition, interaction, and action within dynamic, time-sensitive indoor environments. The project focuses on

enabling sensing, scene understanding, and dexterous manipulation (SSD) tasks that can be performed safely and efficiently with minimal human oversight, while ensuring seamless human–robot interaction.

To achieve these objectives, TORNADO encompasses a multidisciplinary approach, incorporating project management, user-centered pilot studies, core research and development, system integration, and dissemination activities. These components are closely interlinked: user requirements inform the design and implementation of new technologies, while continuous risk assessment ensures that research efforts remain both innovative and grounded in practical constraints. The various technological advancements are ultimately consolidated into a unified ecosystem and validated through pilot applications, focusing on human trust and acceptance. Additionally, continuous engagement with stakeholders fosters a collaborative research community, facilitating broader adoption and practical deployment of TORNADO’s innovations.

computations, ensuring efficient resource utilization while overcoming hardware limitations.

A core innovation within TORNADO is its adaptive cognitive pipeline, which enables the AMR to dynamically adjust to unexpected environmental changes. A key component is the dynamic semantic 3D SLAM module, which continuously updates the robot’s internal environmental representation as objects or people move. If the task context shifts—for instance, due to a new user request or an unforeseen obstacle—the mission planning system reconfigures routes or task sequences in real time. Moreover, the system incorporates online fine-tuning of foundation models, allowing it to refine its behavior based on real-world interactions. This adaptive mechanism enhances operational robustness, enabling the AMR to function reliably across a range of conditions without requiring constant human intervention.

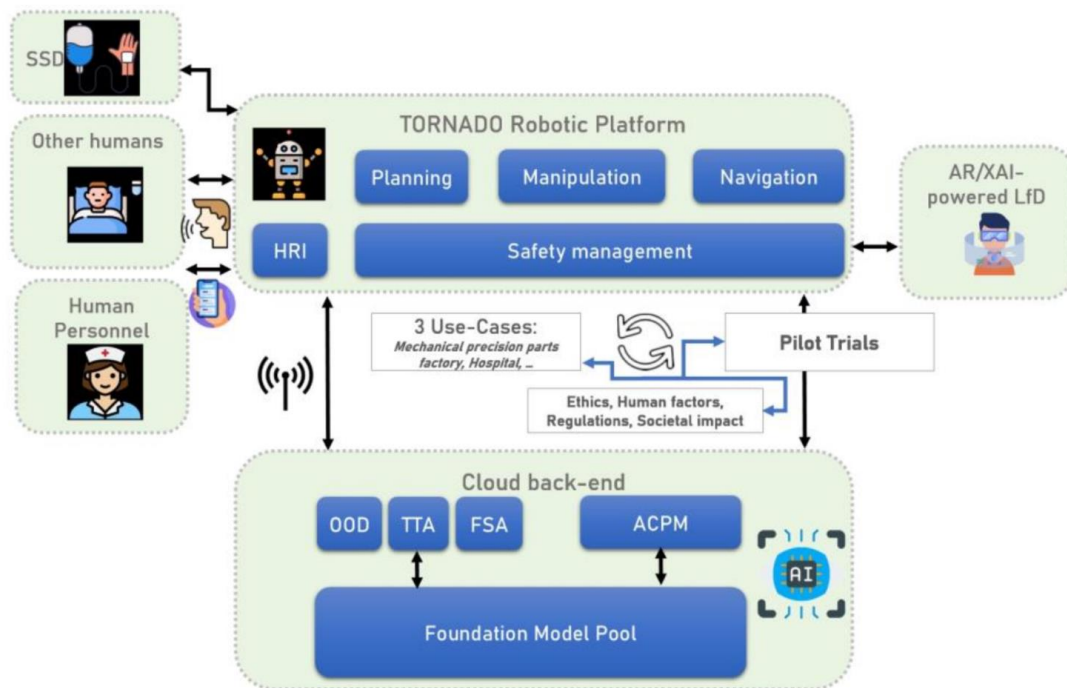


Figure 2: TORNADO system functional architecture

TORNADO envisions a flexible, modular, and cloud-connected AMR platform designed for operation in indoor environments. The AMR is equipped with specialized effectors, such as dexterous grippers, and a comprehensive suite of sensors, including standard cameras and depth sensors, to enable continuous 3D environmental perception.

Processing is distributed across on-board and off-board resources. Lightweight on-board computers manage local control and safety functions, while low-level controllers handle actuation and motion execution. For computationally intensive tasks, such as semantic 3D simultaneous localization and mapping (SLAM), the platform utilizes a secure cloud infrastructure hosting large-scale foundation models (FMs) and high-performance servers. This hybrid processing approach allows TORNADO to offload demanding

TORNADO is designed to facilitate intuitive human–robot interaction, accommodating both expert users and non-specialists. The system supports multimodal communication, enabling users to issue high-level commands through speech, gestures, or a smartphone interface. To minimize cognitive overload, the robot provides only essential status updates, such as mission completion alerts or notifications of unexpected obstacles. For deeper engagement, TORNADO incorporates augmented reality (AR) and explainable AI (XAI) subsystems, which provide transparency into the robot’s decision-making process. These systems allow users to visualize recognized objects, planned actions, and reasoning processes, thereby enhancing trust and interpretability. These mechanisms support effective human–

robot collaboration by enabling users to anticipate, understand, and influence the robot’s behavior. In scenarios where the robot encounters tasks beyond its current capabilities, an internal failsafe mechanism initiates automatic task replanning. If repeated failures occur, a remote supervisor is notified via a smartphone application, allowing for human intervention as needed.

When manual intervention is required, an operator can provide ground-truth labels for perception tasks or conduct a Learning from Demonstration (LfD) session. The LfD process is enhanced through AR-based visualization tools, which overlay real-world information onto the robot’s perceived environment, enabling the operator to demonstrate actions directly. These demonstrations are then stored as reusable action policies, allowing the robot to gradually expand its skill set over time. This iterative learning approach progressively reduces the need for human intervention, thereby increasing autonomy and efficiency in future tasks.

TORNADO relies on secure, low-latency networking, supported by next-generation wireless infrastructure, to enable seamless communication between on-board and off-board systems. This connectivity ensures that real-time perception, AI model adaptation, and human–robot interaction can occur without significant latency. TORNADO achieves a high degree of computational efficiency, allowing the system to scale effectively across diverse operational settings. Whenever additional computing power is required, intensive processing tasks—such as deploying large-scale foundation models or high-level reasoning modules—can be dynamically offloaded to cloud servers. This architecture enables robust, adaptive functionality while mitigating the computational constraints typically associated with mobile robotic platforms.

The TORNADO project represents a comprehensive integration of cutting-edge AI, advanced robotics, and human–robot collaboration frameworks. Each technological component—whether in dynamic SLAM, mission planning, or AR-enhanced interaction—is designed to function within a cohesive, scalable ecosystem. The result is an AMR platform that is capable of executing complex, real-world tasks with minimal supervision, adapting seamlessly to changing environments, and engaging intuitively with human users. Furthermore, TORNADO ensures that technological advancements align with regulatory, ethical, and societal considerations, providing a reliable and responsible foundation for the next generation of autonomous robotic systems.

B. Safe Robot Planning, Navigation, and SSD Manipulation

The TORNADO project aims to develop an advanced robotics framework that enables intelligent, autonomous operation in dynamic and unpredictable environments. The integration of intelligent motion control with industrial systems, as outlined in the IMOCO4.E reference framework [105], provides a structured approach to combining architecture, data management, AI, and digital twin technologies for resilient and adaptable automation. Building on this foundation, TORNADO enhances autonomy by integrating foundation models (FMs), state-of-the-art AI techniques, and a set of key functional modules designed to ameliorate perception, planning, execution, safety, and

communication. Tornado incorporates a) An Adaptive task planning that utilizes FMs such as large language models (LLMs) to interpret user instructions and generate real-time task hierarchies, b) Navigation and manipulation policies that dynamically adjust to environmental changes, c) A multimodal semantic 3D SLAM framework that continuously maps and updates the robot’s surroundings, d) A safety mechanism capable of evaluating and modifying robot actions on-the-fly to maintain operational integrity and e) A low-latency wireless infrastructure that optimally distributes computational workloads between on-board and cloud resources.

The Dynamic Task Planner (DTP) is responsible for processing high-level user instructions while continuously adapting the robot’s task execution based on real-time environmental feedback. It interprets natural language commands using FM-based real-time comprehension and integrates insights from dynamic 3D scene information and human-awareness cues. Building upon LLMs, the DTP can contextualize user goals and decompose them into a hierarchical task tree, which is dynamically restructured as new conditions emerge [106]. This results in a resilient, on-the-fly re-planning system capable of handling unpredictable real-world scenarios while maintaining task efficiency.

The Action Execution Manager (AEM) translates the DTP’s high-level task plans into concrete execution policies for navigation and manipulation, referencing the robot’s current 3D scene map. Unlike traditional robotic control systems that rely on predefined, rigid behaviors, the AEM incorporates continual learning mechanisms, enabling the robot to refine its actions based on user feedback and evolving environmental conditions. FMs specialized for navigation and manipulation can be dynamically deployed or fine-tuned, allowing the robot to adapt its behavior based on demonstrations or newly acquired data. In cases where AI-driven policies encounter unfamiliar conditions, traditional control systems act as a fallback mechanism, ensuring reliability and operational stability. In some scenarios, end-to-end FM-powered pipelines may be used for direct perception-to-action execution [107], enabling the robot to react dynamically to workspace changes, material variations, or human commands.

At the foundation of TORNADO’s cognitive architecture is the Scene Mapper (SM), which employs a neural implicit 3D SLAM framework [108]. This system fuses data from multiple sensory inputs, including RGB, RGB-D, stereo video streams, and neurally derived semantic or geometric cues, to construct a continuously updated 3D representation of the environment. To enhance mapping accuracy, LiDAR-based techniques are incorporated from the WOLF framework [109], which utilizes factor graphs for multi-sensor fusion. The outputs from neural and geometric SLAM pipelines are integrated with the SM produces a cohesive, real-time scene representation that supports multiple system functions; Task planning (DTP) which updates the environment model for task scheduling and replanning. Action execution (AEM) which provides localized data for fine-tuned navigation and manipulation. Knowledge representation which supplies scene semantics for improved AI reasoning and interaction. Progress monitoring which ensures that task execution aligns with dynamic workspace conditions.

Systemic safety is governed by the Safety Manager (SAM), which evaluates whether the robot can safely execute

a given task based on sensor states, actuator conditions, and high-level decision layers [110]. This module features a fault-diagnosis subsystem that detects and localizes both hardware and software faults, triggering appropriate responses based on severity levels. For minor issues, the SAM attempts automatic fault recovery, reinitializing or replacing failed modules while maintaining overall mission continuity. In the case of critical failures, it can instruct the robot to halt operations and return to a designated safe zone. Additionally, the SAM ensures that remote supervisors remain informed, transmitting decision logs and risk alerts to authorized personnel. Should communications be disrupted, locally implemented on-board failsafes ensure that the robot maintains safe behavior until connectivity is restored.

Finally, the Communications Manager (COM) provides a secure, low-latency infrastructure that facilitates seamless data exchange between the robot, edge computing resources, and cloud servers. This infrastructure is designed to a) Dynamically allocate computing resources based on task demands, b) Employ post-quantum encryption to safeguard data integrity, c) Leverage Beyond-5G wireless networks for ultra-fast, real-time connectivity. The edge networking gateway, built on FIWARE standards, enables advanced data handling, while a containerized cloud backbone (orchestrated via Kubernetes) ensures scalable and efficient processing. This allows the system to offload computationally intensive workloads to cloud servers as needed, ensuring that the AMR remains agile and responsive regardless of task complexity or mission variability.

C. Adaptive AI and Self-Adjusting Cognitive Pipelines

TORNADO introduces a three-tier framework designed to enable real-time adaptation of foundation model (FM)-driven cognition, ensuring robust and reliable performance even as conditions change, or domain shifts occur. At the core of this framework is the Adaptation Manager (AM), which integrates out-of-distribution (OOD) detection, few-shot adaptation (FSA), and test-time adaptation (TTA) to determine when and how to update active FMs autonomously [111], [112]. By incorporating human-provided ground-truth labels when available, the AM refines perception, navigation, and manipulation FMs during deployment, addressing the limitations of purely zero-shot inference. It also detects unknown or out-of-domain objects, preventing incorrect classification or inappropriate task execution.

Building on these adaptive mechanisms, the Adaptive Cognitive Pipeline Manager (ACPM) manages the selection and composition of pretrained FMs or alternative control strategies based on real-time environmental analysis and mission objectives. Using 3D scene graphs and behavior trees, the ACPM dynamically switches between perception-to-action pipelines (e.g., RT-2) and model-based approaches (e.g., CLIP, DINOv2) depending on the operational context [106]. This ensures that task execution remains flexible and context-aware, aligning task decomposition with the most effective FM or combination of models. When pre-existing models and adaptive methods fail to achieve accurate performance, the Demonstration Manager (DM) enables on-demand human teaching. Skilled operators can initiate learning-from-demonstration (LfD) sessions through an XAI- and AR-supported interface [113]. The system captures

multimodal cues, including human pose estimation and spoken instructions, to label and store new task policies, which can later be recalled for similar tasks. Meanwhile, it performs few-shot learning to encode structured motion primitives, this approach expands the robot’s skill set efficiently without requiring lengthy offline retraining.

Through this integrated adaptation framework, TORNADO ensures that AMRs can continuously refine their capabilities, respond intelligently to new challenges, and extend their learning through human interaction when necessary, enhancing both autonomy and operational reliability.

D. Advanced Robotic Perception and HRI

TORNADO develops an advanced perception and human-robot interaction (HRI) architecture that utilizes AI-driven mechanisms to enhance situational awareness, natural communication, and ergonomic collaboration in dynamic, people-centric environments. At its core, the Semantic Environment Analyzer (SEA) utilizes pretrained foundation models (FMs) to extract scene semantics in real time, performing instance segmentation, 3D object pose estimation, and person recognition [114], [115]. Beyond image-based analysis, monocular neural depth estimation provides 3D geometric cues from single RGB frames. Depending on the domain requirements, multiple FM backends—such as CLIP, DINOv2, SAM, and InternVideo—can be deployed [116]. To optimize efficiency, knowledge distillation techniques are applied to reduce model size while maintaining performance. The semantic predictions and depth estimates generated by SEA are then integrated into scene-mapping processes, ensuring a comprehensive understanding of the environment.

For verbal interaction, the Sound and Language Manager (SLM) combines pretrained audio FMs and large language models (LLMs) to facilitate speech recognition, sentiment classification, and real-time dialogue processing [117]. A dialog supervisor dynamically adjusts conversational prompts to prevent inappropriate responses and ensure context-aware interaction. When high-level user commands are detected, SLM forwards them to the robot’s planning module, while a specialized LLM version, equipped with Named Entity Recognition and advanced querying, accesses up-to-date knowledge graph (KG) data for informed decision-making. Additionally, dedicated audio FMs handle environmental sound classification, speaker identification, and voice sentiment analysis, further enriching the robot’s understanding of human communication.

The Human State Analyzer (HSA) constructs a real-time human model, integrating visual cues (such as facial expressions, body poses [118], and gestures [119], [120], [121], [122]) with audio signals, including speaker identity and sentiment analysis [123]. By incorporating optical flow, gaze estimation, and contextual knowledge, HSA can predict human movements and intentions [124], [125], [126]. These outputs, along with scene representations, populate a dynamic knowledge graph (KG) [127], supporting real-time contextual reasoning and action anticipation. To enhance interpretability, TORNADO extends existing ontologies [128] by incorporating beliefs, goals, and cultural norms, while a combination of graph neural networks (GNNs), rule-based

approaches, and temporal logic reasoning guides adaptive robot [129], [130], [131], [132].

To improve human–robot collaboration, an XAI-AR interface supports teleoperation-based learning-from-demonstration (LfD), providing real-time insights into the robot’s operational status and AI-driven decisions [133]. Advanced explainable AI (XAI) techniques, including LIME, SHAP, and ELI5, generate attribution heatmaps that highlight the factors influencing deep neural network (DNN) decisions [134]. Additionally, concept-based explanations clarify how missing or hidden information affects the robot’s behavior. Predictive and adaptive AR interfaces [135] are designed to balance situational awareness with cognitive load, ensuring smooth demonstration sessions and real-time perceptual adaptation.

Beyond technological development, TORNADO incorporates human factors research, conducting behavioral studies and ergonomic analyses [136], [137]. These investigations explore how user traits—such as attention span, working memory, and anthropomorphism tendencies— influence HRI experiences. By combining offline experimental evaluations with real-time cognitive and ergonomic monitoring, the project ensures that AR-based demonstrations and verbal interactions remain intuitive, safe, and accessible to diverse user populations.

Together, these interconnected modules create a comprehensive perception and HRI framework, enabling robust autonomous operation in dynamic indoor environments. Through the integration of AI-driven perception, multimodal communication, adaptive learning,

and human-aware interaction, TORNADO enhances the capabilities of autonomous robots, facilitating seamless collaboration and adaptability in complex real-world settings

E. Integration Strategy and Pilot Use Cases

TORNADO follows a step-by-step development and integration strategy, gradually bringing together its

framework components through a ROS-based approach. Each module undergoes unit testing and iterative validation, ensuring reliability before being deployed in three distinct industry-relevant use cases. These scenarios vary in domain, environmental complexity, and human–robot interaction requirements, demonstrating how a single TORNADO system can adapt to different conditions with minimal adjustments. In all cases, the autonomous mobile robot (AMR) operates on battery power while relying on low-latency Beyond-5G wireless networks for real-time AI processing. TORNADO will be validated through the following Use Cases, presented in Figure 3 as well:

1) Use Case 1 – Handling Small Gears and Deformable Ply-Sheets (Mechanical Parts Factory)

In a mechanical parts factory, a mobile collaborative robot with dexterous locomotion and manipulation skills autonomously handles small, delicate gears and larger, heavier ones, as well as placing deformable ply sheets between gear layers. Beyond simple grasping from blisters or crates, the robot assists human workers by assembling product sets on a worktable while ensuring safe and damage-free handling. The ability to adjust to different gear sizes and flexible materials highlights TORNADO’s strengths in precise motion execution, real-time task re-planning, and effective human collaboration.

2) Use Case 2 – Patient Care in a Palliative Ward (Hospital)

In a hospital palliative care ward, an existing two-armed mobile robotic platform is deployed to perform various support tasks, such as attendance tracking of patients and staff, patient assistance (e.g., handing over tissues or water bottle, picking up dropped items, or calling the nurse), medical support (e.g. monitoring and replacing about catheter bags or closing IV bag valves). It also assists with clinical administration and provides psychological support for caregivers. A key challenge in this setting is the ability to

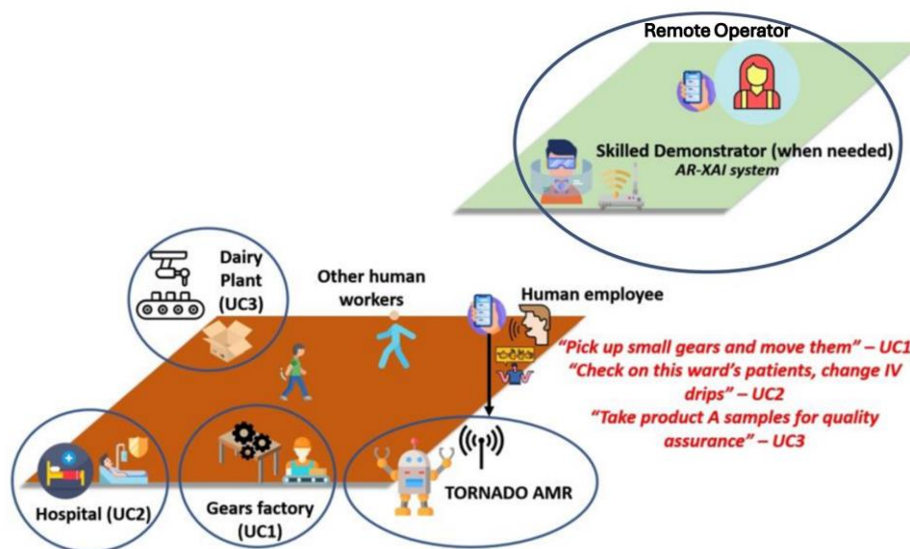


Figure 3: Use Cases in TORNADO

recognize and adapt to unexpected situations, such as respiratory crises or psychological distress. Through short verbal exchanges, the robot can offer timely assistance and reassurance, demonstrating TORNADO's advanced human-robot interaction capabilities and real-time perceptual adaptation.

3) Use Case 3 – Product Sampling and Waste Collection (Distribution Center)

In a distribution center, a mobile robotic platform takes on quality assurance and waste disposal responsibilities. It periodically checks for hazards like floor spills, clears movable obstacles, and samples raw or partially damaged materials from shelves, delivering them to a testing area. Additionally, it identifies and collects deformable packaging waste for proper disposal. These tasks highlight TORNADO's ability to navigate complex, high-traffic environments, dynamically respond to user commands, and adapt workflows based on changing conditions. Across all three use cases, TORNADO's performance is assessed using both quantitative metrics (e.g., task accuracy) and qualitative feedback from professionals. The final hardware setup for each scenario—including robotic arms, grippers, sensors, and embedded computing units—is progressively integrated over the course of the project. This ensures that TORNADO's AI-driven architecture remains flexible, scalable, and adaptable to a wide range of industrial applications and real-world challenges.

IV. EXPECTED OUTCOMES

TORNADO will introduce novel AI-powered algorithms contributing to AMRs with unprecedented capabilities for SSD manipulation and navigation in complex, dynamic, people-centric indoor environments, in the following ways: (i) equipping robots with ground-breaking AI technologies for autonomous and efficient robotic perception, cognition, interaction and action; (ii) introducing a new generation of AI-powered robots able to perform non-repetitive functionalities with unprecedented success and limited-to-no human supervision requirements, in complex dynamic environments; (iii) targeting to launch a new line of interactive, human-centric autonomous robots with improved capabilities to assist humans; and via (iv) unveiling robots with safe, fast and dexterous autonomous SSD manipulation capabilities under changing conditions.

Furthermore, TORNADO will deliver innovations that will result in the development of smarter, safer AMRs with unprecedented cognitive autonomy and robustness, focusing on the following key aspects: (i) augmenting robots with sophisticated adaptive cognition for safe and natural physical and verbal interaction with humans and/or the environment in social/collaborative settings; (ii) significantly increased levels of safety in AMRs in uncontrolled, time-varying, dynamic environments; and via (iii) development of AMRs with advanced SSD manipulation capabilities for various high-impact industries, by concretely implementing the single TORNADO system on three different industrial use-cases.

V. DISCUSSION

TORNADO's impact is expected to cover and interact with European society, economy and the scientific landscape. This impact is expected to be characterized as wider while having long-term effect. TORNADO will contribute to revolutionizing AMR technology, by inducing the following: (i) accelerating European robotics innovation by timely incorporation of cutting-edge AI research; (ii) facilitating the spread of Europe-made AMRs to new sectors with significant societal impact, thus increasing their productivity and reducing relevant costs and via (iii) Enabling the gradual reduction of AI's carbon footprint.

TORNADO elegantly combines novel/emerging ideas, approaches and technologies to enhance Europe's open strategic autonomy goals. TORNADO will contribute to this impact in the following manner: (i) radically new adaptive AI-enabled AMRs with unprecedented cognitive capabilities enabling new functionalities; (ii) integration of cloud and Beyond-5G technologies for remarkably enhancing current robot capabilities; and via the (iii) incorporation of a "human-centric" design for multifunctional, interactive AMRs, as foreseen by the so-called Industry 5.0 concept.

TORNADO directly contributes to the needs of the European industry regarding innovative and efficient approaches across the digital supply chain, by concentrating on providing robust AMR solutions in industrial automation and healthcare robotics for dynamic and unpredictable environments, to improve performance, quality and human satisfaction. TORNADO will introduce novel contributions regarding the following disruptive technologies: (i) integration of advanced, multifunctional and highly adaptive AMR solutions; (ii) development of a toolkit for on-the-fly/on-line robotic AI adaptation during robot deployment; (iii) incorporation of FM technologies for more reliable AMR perception; and via the (iv) adoption of user-friendly and intuitive AR technologies.

Finally, TORNADO contributes to greener digital supply chains by developing advanced, lower-complexity AI algorithms for a new generation of multifunctional robots with higher cognitive autonomy and adaptivity. In a way, TORNADO is expected to have an impact on the fulfillment of Green AI goals.

ACKNOWLEDGMENT

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The work was supported by the European Union through the research and innovation programme under Grant No. 101189557 (TORNADO).

REFERENCES

- [1] M. U. Farooq, A. Eizad, and H.-K. Bae, "Power solutions for autonomous mobile robots: A survey," *Robotics and Autonomous Systems*, vol. 159, p. 104285, 2023, doi: <https://doi.org/10.1016/j.robot.2022.104285>.
- [2] B. F. G. Silva, "Objects and furniture manipulation in domestic environment using a service robot," masterThesis, 2024. Accessed: Feb. 20, 2025. [Online]. Available: <https://repositorium.sdum.uminho.pt/handle/1822/93743>
- [3] D. Mukherjee, K. Gupta, L. H. Chang, and H. Najjaran, "A Survey of Robot Learning Strategies for Human-Robot Collaboration in

- Industrial Settings,” *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102231, Feb. 2022, doi: 10.1016/j.rcim.2021.102231.
- [4] T. Jin and X. Han, “Robotic arms in precision agriculture: A comprehensive review of the technologies, applications, challenges, and future prospects,” *Computers and Electronics in Agriculture*, vol. 221, p. 108938, Jun. 2024, doi: 10.1016/j.compag.2024.108938.
- [5] D. Li et al., “What Foundation Models can Bring for Robot Learning in Manipulation : A Survey,” Dec. 02, 2024, arXiv: arXiv:2404.18201. doi: 10.48550/arXiv.2404.18201.
- [6] K. Kawaharazuka, T. Matsushima, A. Gambardella, J. Guo, C. Paxton, and A. Zeng, “Real-world robot applications of foundation models: a review,” *Advanced Robotics*, vol. 38, no. 18, pp. 1232–1254, Sep. 2024, doi: 10.1080/01691864.2024.2408593.
- [7] Y. Jia et al., “Lift3D Foundation Policy: Lifting 2D Large-Scale Pretrained Models for Robust 3D Robotic Manipulation,” Dec. 14, 2024, arXiv: arXiv:2411.18623. doi: 10.48550/arXiv.2411.18623.
- [8] H. Naderi, A. Shojaei, and L. Huang, “Foundation Models for Autonomous Robots in Unstructured Environments,” Jul. 22, 2024, arXiv: arXiv:2407.14296. doi: 10.48550/arXiv.2407.14296.
- [9] Y. Hu et al., “Toward General-Purpose Robots via Foundation Models: A Survey and Meta-Analysis,” Oct. 01, 2024, arXiv: arXiv:2312.08782. doi: 10.48550/arXiv.2312.08782.
- [10] M. Ahn et al., “AutoRT: Embodied Foundation Models for Large Scale Orchestration of Robotic Agents,” Jul. 02, 2024, arXiv: arXiv:2401.12963. doi: 10.48550/arXiv.2401.12963.
- [11] H. Zhang et al., “DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection,” Jul. 11, 2022, arXiv: arXiv:2203.03605. doi: 10.48550/arXiv.2203.03605.
- [12] A. Radford et al., “Learning Transferable Visual Models From Natural Language Supervision,” in *Proceedings of the 38th International Conference on Machine Learning, PMLR*, Jul. 2021, pp. 8748–8763. Accessed: Feb. 20, 2025. [Online]. Available: <https://proceedings.mlr.press/v139/radford21a.html>
- [13] T. Brown et al., “Language Models are Few-Shot Learners,” in *Advances in Neural Information Processing Systems, Curran Associates, Inc.*, 2020, pp. 1877–1901. Accessed: Feb. 20, 2025. [Online]. Available: <https://papers.nips.cc/paper/2020/hash/1457c0d6bfc64967418bfb8ac142f64a-Abstract.html>
- [14] OpenAI et al., “GPT-4 Technical Report,” Mar. 04, 2024, arXiv: arXiv:2303.08774. doi: 10.48550/arXiv.2303.08774.
- [15] N. Abdulazeem and Y. Hu, “Human Factors Considerations for Quantifiable Human States in Physical Human–Robot Interaction: A Literature Review,” *Sensors*, vol. 23, no. 17, Art. no. 17, Jan. 2023, doi: 10.3390/s23177381.
- [16] M. Lagomarsino, M. Lorenzini, E. De Momi, and A. Ajoudani, “PRO-MIND: Proximity and Reactivity Optimisation of robot Motion to tune safety limits, human stress, and productivity in INdustrial settings,” 2024, doi: 10.48550/ARXIV.2409.06864.
- [17] C.-P. Lam, C.-T. Chou, C.-F. Chang, and L.-C. Fu, “Human-centered robot navigation — Toward a harmoniously coexisting multi-human and multi-robot environment,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2010, pp. 1813–1818. doi: 10.1109/IROS.2010.5652214.
- [18] G. Th. Papadopoulos, M. Antona, and C. Stephanidis, “Towards Open and Expandable Cognitive AI Architectures for Large-Scale Multi-Agent Human-Robot Collaborative Learning,” *IEEE Access*, vol. 9, pp. 73890–73909, 2021, doi: 10.1109/ACCESS.2021.3080517.
- [19] G. Th. Papadopoulos, A. Leonidis, M. Antona, and C. Stephanidis, “User Profile-Driven Large-Scale Multi-agent Learning from Demonstration in Federated Human-Robot Collaborative Environments,” vol. 13303, pp. 548–563, 2022, doi: 10.1007/978-3-031-05409-9_40.
- [20] J. Yang, K. Zhou, Y. Li, and Z. Liu, “Generalized Out-of-Distribution Detection: A Survey,” *Int J Comput Vis*, vol. 132, no. 12, pp. 5635–5662, Dec. 2024, doi: 10.1007/s11263-024-02117-4.
- [21] J. Liang, R. He, and T. Tan, “A Comprehensive Survey on Test-Time Adaptation Under Distribution Shifts,” *Int J Comput Vis*, vol. 133, no. 1, pp. 31–64, Jan. 2025, doi: 10.1007/s11263-024-02181-w.
- [22] F. Liu et al., “Few-shot adaptation of multi-modal foundation models: a survey,” *Artif Intell Rev*, vol. 57, no. 10, p. 268, Aug. 2024, doi: 10.1007/s10462-024-10915-y.
- [23] G. Campagna, M. Lagomarsino, M. Lorenzini, D. Chrysostomou, M. Rehm, and A. Ajoudani, “Promoting Trust in Industrial Human-Robot Collaboration Through Preference-Based Optimization,” *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 9255–9262, Nov. 2024, doi: 10.1109/LRA.2024.3455792.
- [24] M. Lagomarsino, M. Lorenzini, E. De Momi, and A. Ajoudani, “An Online Framework for Cognitive Load Assessment in Industrial Tasks,” *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102380, Dec. 2022, doi: 10.1016/j.rcim.2022.102380.
- [25] G. Solak, G. J. G. Lahr, I. Ozdamar, and A. Ajoudani, “Context-aware collaborative pushing of heavy objects using skeleton-based intention prediction,” presented at the *IEEE International Conference on Robotics and Automation*,
- [26] “(PDF) A Hybrid Learning and Optimization Framework to Achieve Physically Interactive Tasks With Mobile Manipulators,” *ResearchGate*, Dec. 2024, doi: 10.1109/LRA.2022.3187258.
- [27] F. Pourpanah et al., “A Review of Generalized Zero-Shot Learning Methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4051–4070, Apr. 2023, doi: 10.1109/TPAMI.2022.3191696.
- [28] X. V. Wang and L. Wang, “A literature survey of the robotic technologies during the COVID-19 pandemic,” *Journal of Manufacturing Systems*, vol. 60, pp. 823–836, 2021, doi: <https://doi.org/10.1016/j.jmsy.2021.02.005>.
- [29] G. Fragapane, R. de Koster, F. Sgarbossa, and J. O. Strandhagen, “Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda,” *European Journal of Operational Research*, vol. 294, no. 2, pp. 405–426, Oct. 2021, doi: 10.1016/j.ejor.2021.01.019.
- [30] D. Di Paola, A. Milella, G. Cicirelli, and A. Distante, “An Autonomous Mobile Robotic System for Surveillance of Indoor Environments,” *International Journal of Advanced Robotic Systems*, vol. 7, no. 1, p. 8, Mar. 2010, doi: 10.5772/7254.
- [31] A. J. Sathyamoorthy, U. Patel, M. Paul, Y. Savle, and D. Manocha, “COVID surveillance robot: Monitoring social distancing constraints in indoor scenarios,” *PLoS One*, vol. 16, no. 12, p. e0259713, Dec. 2021, doi: 10.1371/journal.pone.0259713.
- [32] K. Berns and S. A. Mehdi, “Use of an Autonomous Mobile Robot for Elderly Care,” in *Proceedings - 2nd Advanced Technologies for Enhanced Quality of Life, ATEQUAL 2010*, Aug. 2010, pp. 121–126. doi: 10.1109/ATEQUAL.2010.30.
- [33] N. S. Ahmad, N. L. Boon, and P. Goh, “Multi-Sensor Obstacle Detection System Via Model-Based State-Feedback Control in Smart Cane Design for the Visually Challenged,” *IEEE Access*, vol. 6, pp. 64182–64192, 2018, doi: 10.1109/ACCESS.2018.2878423.
- [34] J. Zhong, C. Ling, A. Cangelosi, A. Lotfi, and X. Liu, “On the Gap between Domestic Robotic Applications and Computational Intelligence,” *Electronics*, vol. 10, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/electronics10070793.
- [35] F. Rubio, F. Valero, and C. Llopis-Albert, “A review of mobile robots: Concepts, methods, theoretical framework, and applications,” *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, p. 1729881419839596, Mar. 2019, doi: 10.1177/1729881419839596.
- [36] A. Saudabayev, F. Kungozhin, D. Nurseitov, and A. Varol, “Locomotion Strategy Selection for a Hybrid Mobile Robot Using Time of Flight Depth Sensor,” *Journal of Sensors*, vol. 2015, pp. 1–14, Apr. 2015, doi: 10.1155/2015/425732.
- [37] R. Y. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, “Introduction to Autonomous Mobile Robots,” 2004. [Online]. Available: <https://api.semanticscholar.org/CorpusID:107033282>
- [38] G. Endo and S. Hirose, “Study on Roller-Walker (multi-mode steering control and self-contained locomotion),” in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, 2000, pp. 2808–2814 vol.3. doi: 10.1109/ROBOT.2000.846453.
- [39] B. Silva, R. M. Fisher, A. Kumar, and G. P. Hancke, “Experimental Link Quality Characterization of Wireless Sensor Networks for Underground Monitoring,” *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1099–1110, 2015, doi: 10.1109/TII.2015.2471263.
- [40] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

- [41] R. Lagisetty, P. N. K., R. Padhi, and M. S. Bhat, "Object detection and obstacle avoidance for mobile robot using stereo camera," in *Proceedings of the IEEE International Conference on Control Applications*, Aug. 2013, pp. 605–610. doi: 10.1109/CCA.2013.6662816.
- [42] S. Konstantakos et al., "Self-supervised visual learning in the low-data regime: A comparative evaluation," *Neurocomputing*, vol. 620, p. 129199, Mar. 2025. doi: 10.1016/j.neucom.2024.129199.
- [43] P. Alimisis, I. Mademlis, P. Radoglou-Grammatikis, P. Sarigiannidis, and G. Th. Papadopoulos, "Advances in diffusion models for image data augmentation: a review of methods, models, evaluation metrics and future research directions," *Artif Intell Rev*, vol. 58, no. 4, p. 112, Jan. 2025. doi: 10.1007/s10462-025-11116-x.
- [44] G. Mester, "Motion Control of Wheeled Mobile Robots," 2006.
- [45] B. Crnokić, M. Grubišić, and T. Volaric, "Different Applications of Mobile Robots in Education," *International Journal on Integrating Technology in Education*, vol. 6, pp. 15–28, Sep. 2017. doi: 10.5121/ijite.2017.6302.
- [46] P. Corke, J. Roberts, J. Cunningham, and D. Hainsworth, "Mining Robotics," 2008, pp. 1127–1150. doi: 10.1007/978-3-540-30301-5_50.
- [47] J. Laplaza, F. Moreno, and A. Sanfeliu, "Enhancing Robotic Collaborative Tasks Through Contextual Human Motion Prediction and Intention Inference," *Int J of Soc Robotics*, Jul. 2024. doi: 10.1007/s12369-024-01140-2.
- [48] J. E. Domínguez-Vidal, N. Rodríguez, and A. Sanfeliu, "Perception–Intention–Action Cycle in Human–Robot Collaborative Tasks: The Collaborative Lightweight Object Transportation Use-Case," *Int J of Soc Robotics*, Mar. 2024. doi: 10.1007/s12369-024-01103-7.
- [49] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019. doi: 10.1126/science.aat8414.
- [50] P. Jiménez, "Survey on model-based manipulation planning of deformable objects," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 2, pp. 154–163, Apr. 2012. doi: 10.1016/j.rcim.2011.08.002.
- [51] R. Herguedas, G. Lopez-Nicolas, R. Aragues, and C. Sagues, "Survey on multi-robot manipulation of deformable objects," 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 977–984, Sep. 2019. doi: 10.1109/ETFA.2019.8868987.
- [52] F. Nadon, A. Valencia, and P. Payeur, "Multi-Modal Sensing and Robotic Manipulation of Non-Rigid Objects: A Survey," *Robotics*, Nov. 2018. doi: 10.3390/robotics7040074.
- [53] V. Arriola-Rios, P. Guler, F. Ficuciello, D. Kragic, B. Siciliano, and J. Wyatt, "Modeling of Deformable Objects for Robotic Manipulation: A Tutorial and Review," *Frontiers in Robotics and AI*, vol. 7, Sep. 2020. doi: 10.3389/frobt.2020.00082.
- [54] J. Sanchez, J. A. Corrales Ramon, B. C. BOUZGARROU, and Y. Mezouar, "Robotic Manipulation and Sensing of Deformable Objects in Domestic and Industrial Applications: A Survey," *The International Journal of Robotics Research*, vol. 37, pp. 688–716, Jun. 2018. doi: 10.1177/0278364918779698.
- [55] H. Yin, A. Varava, and D. Kragic, "Modeling, learning, perception, and control methods for deformable object manipulation," *Science Robotics*, vol. 6, 2021, [Online]. Available: <https://api.semanticscholar.org/CorpusID:235204115>
- [56] J. Zhu, B. Navarro, R. Passama, P. Fraisse, A. Crosnier, and A. Cherubini, "Robotic manipulation planning for shaping deformable linear objects with environmental contacts," *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 16–23, Jan. 2020. doi: 10.1109/LRA.2019.2944304.
- [57] Z. Hu, T. Han, P. Sun, J. Pan, and D. Manocha, "3-D Deformable Object Manipulation Using Deep Neural Networks," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4255–4261, 2019. doi: 10.1109/LRA.2019.2930476.
- [58] A. Cherubini, V. Ortenzi, A. Cosgun, R. Lee, and P. Corke, "Model-free vision-based shaping of deformable plastic materials," *The International Journal of Robotics Research*, vol. 39, no. 14, pp. 1739–1759, 2020. doi: 10.1177/0278364920907684.
- [59] H. Liu et al., "The MUSH Hand II: A Multifunctional Hand for Robot-Assisted Laparoscopic Surgery," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 1, pp. 393–404, 2021. doi: 10.1109/TMECH.2020.3022782.
- [60] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015. doi: 10.1109/TRO.2015.2463671.
- [61] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017. doi: 10.1109/TRO.2017.2705103.
- [62] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021. doi: 10.1109/TRO.2021.3075644.
- [63] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018. doi: 10.1109/TRO.2018.2853729.
- [64] C. Yu et al., "DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 1168–1174. doi: 10.1109/IROS.2018.8593691.
- [65] B. Bescos, J. M. Fàcil, J. Civera, and J. Neira, "DynaSLAM: Tracking, Mapping and inpainting in Dynamic Scenes," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4076–4083, Oct. 2018. doi: 10.1109/LRA.2018.2860039.
- [66] P. Fiorini and Z. Shiller, "Motion Planning in Dynamic Environments Using Velocity Obstacles," *The International Journal of Robotics Research*, vol. 17, no. 7, pp. 760–772, Jul. 1998. doi: 10.1177/027836499801700706.
- [67] Y. Abe and M. Yoshiki, "Collision avoidance method for multiple autonomous mobile agents by implicit cooperation," in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180)*, Oct. 2001, pp. 1207–1212 vol.3. doi: 10.1109/IROS.2001.977147.
- [68] D. Wilkie, J. Van Den Berg, and D. Manocha, "Generalized velocity obstacles," 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5573–5578, Oct. 2009. doi: 10.1109/IROS.2009.5354175.
- [69] J. Van Den Berg, J. Snape, S. J. Guy, and D. Manocha, "Reciprocal collision avoidance with acceleration-velocity obstacles: 2011 IEEE International Conference on Robotics and Automation, ICRA 2011," 2011 IEEE International Conference on Robotics and Automation, ICRA 2011, pp. 3475–3482, Dec. 2011. doi: 10.1109/ICRA.2011.5980408.
- [70] "Adaptive social planner to accompany people in real-life dynamic environments." Accessed: Feb. 27, 2025. [Online]. Available: <https://upcommons.upc.edu/handle/2117/400230?show=full>
- [71] M. Linardakis, I. Varlamis, and G. T. Papadopoulos, "Distributed maze exploration using multiple agents and optimal goal assignment," May 30, 2024, arXiv: arXiv:2405.20232. doi: 10.48550/arXiv.2405.20232.
- [72] B. Kluge, "Recursive agent modeling with probabilistic velocity obstacles for mobile robot navigation among humans," *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, vol. 1, pp. 376–381, 2003. doi: 10.1109/IROS.2003.1250657.
- [73] J. van den Berg, M. Lin, and D. Manocha, "Reciprocal Velocity Obstacles for real-time multi-agent navigation," in *2008 IEEE International Conference on Robotics and Automation*, May 2008, pp. 1928–1935. doi: 10.1109/ROBOT.2008.4543489.
- [74] J. Snape, J. van den Berg, S. J. Guy, and D. Manocha, "The Hybrid Reciprocal Velocity Obstacle," *IEEE Transactions on Robotics*, vol. 27, no. 4, pp. 696–706, Aug. 2011. doi: 10.1109/TRO.2011.2120810.
- [75] S. J. Guy, M. C. Lin, and D. Manocha, "Modeling collision avoidance behavior for virtual humans," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 2 - Volume 2*, in AAMAS '10. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, May 2010, pp. 575–582.
- [76] E. T. Hall, "A System for the Notation of Proxemic Behavior1," *American Anthropologist*, vol. 65, no. 5, pp. 1003–1026, Oct. 1963. doi: 10.1525/aa.1963.65.5.02a00020.
- [77] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, p. 487, 2000.

- [78] Y. Tamura, T. Fukuzawa, and H. Asama, "Smooth collision avoidance in human-robot coexisting environment," 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3887–3892, Oct. 2010, doi: 10.1109/IROS.2010.5649673.
- [79] M. Luber, J. A. Stork, G. D. Tipaldi, and K. O. Arras, "People tracking with human motion predictions from social forces," 2010 IEEE International Conference on Robotics and Automation, pp. 464–469, May 2010, doi: 10.1109/ROBOT.2010.5509779.
- [80] R. A. Knepper and D. Rus, "Pedestrian-inspired sampling-based multi-robot collision avoidance," in 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France: IEEE, Sep. 2012, pp. 94–100. doi: 10.1109/ROMAN.2012.6343737.
- [81] Dongqing Shi, E. G. Collins Jr, B. Goldiez, A. Donate, Xiuwen Liu, and D. Dunlap, "Human-aware robot motion planning with velocity constraints," in 2008 International Symposium on Collaborative Technologies and Systems, Irvine, CA, USA: IEEE, May 2008, pp. 490–497. doi: 10.1109/CTS.2008.4543969.
- [82] "[PDF] Navigation in the presence of humans | Semantic Scholar." Accessed: Feb. 20, 2025. [Online]. Available: <https://www.semanticscholar.org/paper/Navigation-in-the-presence-of-humans-Sisbot-Alami/409a2a1f3a4484a5bc3bac677cf6563d9ebdf4cf>
- [83] E. Sisbot, L. Marin, R. Alami, and T. Simeon, "A mobile robot that performs human acceptable motions," 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1811–1816, Oct. 2006, doi: 10.1109/IROS.2006.282223.
- [84] T. Kruse, P. Basili, S. Glasauer, and A. Kirsch, "Legible robot navigation in the proximity of moving humans," in Advanced Robotics and its Social Impacts (ARSO), 2012 IEEE Workshop on, in Advanced Robotics and its Social Impacts (ARSO), 2012 IEEE Workshop on. Munich, Germany, May 2012, pp. 83–88. doi: 10.1109/ARSO.2012.6213404.
- [85] R. Kirby, R. Simmons, and J. Forlizzi, "COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation," in RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, Sep. 2009, pp. 607–612. doi: 10.1109/ROMAN.2009.5326271.
- [86] "(PDF) Navigating between People: A Stochastic Optimization Approach." Accessed: Feb. 20, 2025. [Online]. Available: https://www.researchgate.net/publication/244483508_Navigating_between_People_A_Stochastic_Optimization_Approach
- [87] M. Svenstrup, T. Bak, and H. J. Andersen, "Trajectory planning for robots in dynamic human environments," 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4293–4298, Oct. 2010, doi: 10.1109/IROS.2010.5651531.
- [88] L. Scandolo and T. Fraichard, "An anthropomorphic navigation scheme for dynamic scenarios," in 2011 IEEE International Conference on Robotics and Automation, Shanghai, China: IEEE, May 2011, pp. 809–814. doi: 10.1109/ICRA.2011.5979772.
- [89] "Unfreezing the robot: Navigation in dense, interacting crowds | IEEE Conference Publication | IEEE Xplore." Accessed: Feb. 20, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/5654369>
- [90] P. Agarwal et al., "Feature-Based Prediction of Trajectories for Socially Compliant Navigation," in Robotics: Science and Systems VIII, MIT Press, 2013, pp. 193–200. Accessed: Feb. 20, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/6577979>
- [91] B. D. Ziebart et al., "Planning-based prediction for pedestrians," 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3931–3936, Oct. 2009, doi: 10.1109/IROS.2009.5354147.
- [92] C. Fulgenzi, A. Spalanzani, and C. Laugier, "Probabilistic motion planning among moving obstacles following typical motion patterns," in Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems, in IROS'09. St. Louis, MO, USA: IEEE Press, Oct. 2009, pp. 4027–4033.
- [93] A. Pumarola, A. Vakhitov, A. Agudo, A. Sanfeliu, and F. Moreno-Noguer, "PL-SLAM: Real-time monocular visual SLAM with points and lines," 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4503–4508, May 2017, doi: 10.1109/ICRA.2017.7989522.
- [94] A. Silva, M. Schrum, E. Hedlund-Botti, N. Gopalan, and M. Gombolay, "Explainable Artificial Intelligence: Evaluating the Objective and Subjective Impacts of xAI on Human-Agent Interaction," International Journal of Human-Computer Interaction, vol. 39, no. 7, pp. 1390–1404, Apr. 2023, doi: 10.1080/10447318.2022.2101698.
- [95] A. Silva, P. Tambwekar, M. Schrum, and M. Gombolay, "Towards Balancing Preference and Performance through Adaptive Personalized Explainability," in Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, in HRI '24. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 658–668. doi: 10.1145/3610977.3635000.
- [96] D. Das, S. Banerjee, and S. Chernova, "Explainable AI for Robot Failures: Generating Explanations that Improve User Assistance in Fault Recovery," Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, pp. 351–360, Mar. 2021, doi: 10.1145/3434073.3444657.
- [97] "Quantifying teaching behavior in robot learning from demonstration | Semantic Scholar." Accessed: Feb. 20, 2025. [Online]. Available: <https://www.semanticscholar.org/paper/Quantifying-teaching-behavior-in-robot-learning-Sena-Howard/c0618da60be294c8b944e5973c05c5f924fd5e72>
- [98] E. Merlo, M. Lagomarsino, E. Lamon, and A. Ajoudani, "Automatic Interaction and Activity Recognition from Videos of Human Manual Demonstrations with Application to Anomaly Detection," 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 1188–1195, Aug. 2023, doi: 10.1109/RO-MAN57019.2023.10309311.
- [99] M. B. Luebbbers, C. Brooks, C. L. Mueller, D. Szafir, and B. Hayes, "ARC-LfD: Using Augmented Reality for Interactive Long-Term Robot Skill Maintenance via Constrained Learning from Demonstration," 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 3794–3800, May 2021, doi: 10.1109/ICRA48506.2021.9561844.
- [100] C. Mueller, A. Tabrez, and B. Hayes, "Interactive constrained learning from demonstration using visual robot behavior counterfactuals," in Proceedings of the Accessibility of Robot Programming and Work of the Future Workshop at RSS, 2021. Accessed: Feb. 20, 2025. [Online]. Available: <https://aaquibtabrez.github.io/assets/pdf/publications/rss21w.pdf>
- [101] R. Paleja, M. Ghuy, N. R. Arachchige, R. Jensen, and M. Gombolay, "The Utility of Explainable AI in Ad Hoc Human-Machine Teaming," Sep. 08, 2022, arXiv: arXiv:2209.03943. doi: 10.48550/arXiv.2209.03943.
- [102] N. Rodis, C. Sardanios, P. Radoglou-Grammatikis, P. Sarigiannidis, I. Varlamis, and G. Th. Papadopoulos, "Multimodal Explainable Artificial Intelligence: A Comprehensive Review of Methodological Advances and Future Research Directions," IEEE Access, vol. 12, pp. 159794–159820, 2024, doi: 10.1109/ACCESS.2024.3467062.
- [103] "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) | IEEE Journals & Magazine | IEEE Xplore." Accessed: Feb. 20, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/8466590>
- [104] M. Luebbbers, A. Tabrez, K. Ruvane, and B. Hayes, "Autonomous Justification for Enabling Explainable Decision Support in Human-Robot Teaming," Robotics: Science and Systems XIX, Jul. 2023, doi: 10.15607/RSS.2023.XIX.002.
- [105] S. Mohamed et al., "The IMOCO4.E reference framework for intelligent motion control systems," in 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2023, pp. 1–8. doi: 10.1109/ETFA54631.2023.10275410.
- [106] Y. Zhen et al., "Robot Task Planning Based on Large Language Model Representing Knowledge with Directed Graph Structures," Jun. 08, 2023, arXiv: arXiv:2306.05171. doi: 10.48550/arXiv.2306.05171.
- [107] A. Brohan et al., "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," Jul. 28, 2023, arXiv: arXiv:2307.15818. doi: 10.48550/arXiv.2307.15818.
- [108] S. Zhu et al., "SNI-SLAM: Semantic Neural Implicit SLAM," Mar. 28, 2024, arXiv: arXiv:2311.11016. doi: 10.48550/arXiv.2311.11016.
- [109] J. Solà et al., "WOLF: A Modular Estimation Framework for Robotics Based on Factor Graphs," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 4710–4717, Apr. 2022, doi: 10.1109/LRA.2022.3151404.
- [110] M. L. Visinsky, J. R. Cavallaro, and I. D. Walker, "Robotic fault detection and fault tolerance: A survey," Reliability Engineering & System Safety, vol. 46, no. 2, pp. 139–158, Jan. 1994, doi: 10.1016/0951-8320(94)90132-5.

- [111] K. Zhou, J. Yang, C. C. Loy, and Z. Liu, "Conditional Prompt Learning for Vision-Language Models," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022, pp. 16795–16804. doi: 10.1109/CVPR52688.2022.01631.
- [112] J. Silva-Rodríguez, S. Hajimiri, I. B. Ayed, and J. Dolz, "A Closer Look at the Few-Shot Adaptation of Large Vision-Language Models," Mar. 25, 2024, arXiv: arXiv:2312.12730. doi: 10.48550/arXiv.2312.12730.
- [113] V. Prasad, D. Koert, R. Stock-Homburg, J. Peters, and G. Chalvatzaki, "MILD: Multimodal Interactive Latent Dynamics for Learning Human-Robot Interaction," in 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids), Nov. 2022, pp. 472–479. doi: 10.1109/Humanoids53995.2022.10000239.
- [114] C. Papaioannidis, I. Mademlis, and I. Pitas, "Fast Semantic Image Segmentation for Autonomous Systems," in 2022 IEEE International Conference on Image Processing (ICIP), Oct. 2022, pp. 2646–2650. doi: 10.1109/ICIP46576.2022.9897582.
- [115] S. Thermos, G. Th. Papadopoulos, P. Daras, and G. Potamianos, "Deep sensorimotor learning for RGB-D object recognition," *Computer Vision and Image Understanding*, vol. 190, p. 102844, Jan. 2020, doi: 10.1016/j.cviu.2019.102844.
- [116] L. Yuan et al., "VideoGLUE: Video General Understanding Evaluation of Foundation Models," Oct. 24, 2024, arXiv: arXiv:2307.03166. doi: 10.48550/arXiv.2307.03166.
- [117] T. Wang et al., "What Language Model Architecture and Pretraining Objective Works Best for Zero-Shot Generalization?," in Proceedings of the 39th International Conference on Machine Learning, PMLR, Jun. 2022, pp. 22964–22984. Accessed: Feb. 20, 2025. [Online]. Available: <https://proceedings.mlr.press/v162/wang22u.html>
- [118] C. Papaioannidis, I. Mademlis, and I. Pitas, "Fast CNN-Based Single-Person 2D Human Pose Estimation for Autonomous Systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 3, pp. 1262–1275, Mar. 2023, doi: 10.1109/TCSVT.2022.3209160.
- [119] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," *IEEE Transactions on Multimedia*, vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.
- [120] D. Makrygiannis, C. Papaioannidis, I. Mademlis, and I. Pitas, "Optimal video handling in on-line hand gesture recognition using Deep Neural Networks," in 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Dec. 2021, pp. 1–7. doi: 10.1109/SSCI50451.2021.9660038.
- [121] "Survey on Hand Gesture Recognition from Visual Input The research leading to these results received funding from the European Commission under Grant Agreement No. 101168042 (TRIFFID)." Accessed: Feb. 25, 2025. [Online]. Available: <https://arxiv.org/html/2501.11992>
- [122] M. Peral, A. Sanfeliu, and A. Garrell, "Efficient Hand Gesture Recognition for Human-Robot Interaction," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10272–10279, Oct. 2022, doi: 10.1109/LRA.2022.3193251.
- [123] "Deep learning based multimodal emotion recognition using model-level fusion of audio-visual modalities - ScienceDirect." Accessed: Feb. 20, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0950705122002593>
- [124] J. Tanke, C. Zaveri, and J. Gall, "Intention-based Long-Term Human Motion Anticipation," in 2021 International Conference on 3D Vision (3DV), Dec. 2021, pp. 596–605. doi: 10.1109/3DV53792.2021.00069.
- [125] S. Li, L. Zhang, and X. Diao, "Deep-Learning-Based Human Intention Prediction Using RGB Images and Optical Flow," *J Intell Robot Syst*, vol. 97, no. 1, pp. 95–107, Jan. 2020, doi: 10.1007/s10846-019-01049-3.
- [126] Y. Liang, P. Zhou, R. Zimmermann, and S. Yan, "DualFormer: Local-Global Stratified Transformer for Efficient Video Recognition," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., Cham: Springer Nature Switzerland, 2022, pp. 577–595. doi: 10.1007/978-3-031-19830-4_33.
- [127] W. Liu, A. Daruna, M. Patel, K. Ramachandruni, and S. Chernova, "A survey of Semantic Reasoning frameworks for robotic systems," *Robotics and Autonomous Systems*, vol. 159, p. 104294, Jan. 2023, doi: 10.1016/j.robot.2022.104294.
- [128] J. I. Olszewska et al., "Ontology for autonomous robotics," in 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Aug. 2017, pp. 189–194. doi: 10.1109/ROMAN.2017.8172300.
- [129] Z. Ye, Y. J. Kumar, G. O. Sing, F. Song, and J. Wang, "A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs," *IEEE Access*, vol. 10, pp. 75729–75741, 2022, doi: 10.1109/ACCESS.2022.3191784.
- [130] Z. Zeng, Q. Cheng, and Y. Si, "Logical Rule-Based Knowledge Graph Reasoning: A Comprehensive Survey," *Mathematics*, vol. 11, no. 21, Art. no. 21, Jan. 2023, doi: 10.3390/math11214486.
- [131] S. Ji, S. Pan, E. Cambria, P. Martinen, and P. S. Yu, "A Survey on Knowledge Graphs: Representation, Acquisition, and Applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494–514, Feb. 2022, doi: 10.1109/TNNLS.2021.3070843.
- [132] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human-robot interaction: An implementation," *Artificial Intelligence*, vol. 247, pp. 45–69, Jun. 2017, doi: 10.1016/j.artint.2016.07.002.
- [133] K. Lotsaris, C. Gkourmelos, N. Fousekis, N. Kousi, and S. Makris, "AR based robot programming using teaching by demonstration techniques," *Procedia CIRP*, vol. 97, pp. 459–463, Jan. 2021, doi: 10.1016/j.procir.2020.09.186.
- [134] W. Jin, X. Li, and G. Hamameh, "Evaluating Explainable AI on a Multi-Modal Medical Imaging Task: Can Existing Algorithms Fulfill Clinical Requirements?," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, Art. no. 11, Jun. 2022, doi: 10.1609/aaai.v36i11.21452.
- [135] V. Li et al., "Super Resolution for Augmented Reality Applications," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2022, pp. 1–6. doi: 10.1109/INFOCOMWKSHPS54753.2022.9798101.
- [136] Z. Arkouli, G. Michalos, and S. Makris, "On the Selection of Ergonomics Evaluation Methods for Human Centric Manufacturing Tasks," *Procedia CIRP*, vol. 107, pp. 89–94, Jan. 2022, doi: 10.1016/j.procir.2022.04.015.
- [137] Z. Xu, G. Wang, S. Zhai, and P. Liu, "When Automation Fails: Examining the Effect of a Verbal Recovery Strategy on User Experience in Automated Driving," *International Journal of Human-Computer Interaction*, vol. 0, no. 0, pp. 1–11, doi: 10.1080/10447318.2023.2176986.