Incorporating Commonsense Knowledge to Enhance Robot Perception

Rafael Hidalgo, Aparna S. Varde, Jesse Parron, Weitian Wang
School of Computing, Montclair State University (MSU), NJ, USA
{hidalgor | vardea | parronj1 | wangw}@montclair.edu

***Abstract*—Robots have been substantially employed across various avenues over the years. Yet, their application has been largely limited to controlled environments where variables are few and predictable. To address this challenge, we propose Robo-CSK-Organizer, a novel system that enhances robotic perception by integrating commonsense knowledge (CSK) for improved object organization, classification, and decision-making. By combining ConceptNet for semantic reasoning, DETIC for object identification, and BLIP for contextual analysis, Robo-CSK-Organizer achieves superior ambiguity resolution, task adaptation, and explainability compared to models without CSK. Testing in real-world robotics settings demonstrates notable gains in transparency, user trust, and error handling, making the approach valuable for advancing AI transparency and the development of versatile robotic applications in automation and engineering. Future directions of this work are comprehensively discussed.**

***Note to Practitioners*—This paper introduces Robo-CSK-Organizer, a system designed to enhance robotic decision-making through the integration of commonsense knowledge (CSK). Developed with the goal of improving both the efficiency and transparency of robots in task execution, this work addresses the critical need for advanced object organization and classification capabilities in multipurpose robotics, especially in domestic household settings. By leveraging a classical commonsense knowledge base alongside cutting-edge object detection and context discernment technologies, Robo-CSK-Organizer demonstrates significant improvements in ambiguity resolution, placement consistency, and explainable AI (XAI), highlighted by its superior performance over a baseline model using ChatGPT for comparison. Practitioners engaged in design and implementation of robotic systems will find Robo-CSK-Organizer truly valuable for its adequate use of commonsense reasoning (inherent in humans) due to which it is adaptable to numerous real-world applications.**

***Index Terms*—AI-Driven Robotics, Adaptability, Commonsense Reasoning, Next-Gen AI Systems, XAI**

I. Introduction

In recent years, the role of multipurpose robots has expanded significantly across various avenues, including manufacturing, healthcare, domestic, and other settings [1]–[3]. These robots are no longer confined to performing simple, repetitive tasks. Instead, they are increasingly required to handle complex, context-driven operations that demand a high level of adaptability and intelligence. The development of transparent, intelligent systems becomes imperative to meet these evolving demands, ensuring robots can operate efficiently in dynamic and unpredictable environments.

Traditional robotic arms have been the cornerstone of industrial automation for decades, excelling in tasks such as assembly, welding, and painting. However, their application has been largely limited to controlled environments where variables are few and predictable. These systems often struggle with nuanced decision-making required in more dynamic environments. For instance, traditional robots might excel in assembling a car but falter in a scenario where they must adapt to unexpected changes, such as varying component shapes or sizes not pre-encoded in their programming [4].

Deep learning-based robotic systems have achieved state-of-the-art performance in many domains, including object recognition and task planning [5]. These systems rely on pre-trained models that learn mappings from inputs to outputs using large labeled datasets. While effective in static and well-defined environments, they often struggle in dynamic or unpredictable scenarios where the relationships between objects and their contexts are not explicitly encoded in the training data. For example, a convolutional neural network (CNN) may accurately classify individual objects but fail to infer their relational stability in a complex arrangement without extensive labeled examples. One of the primary limitations of deep learning models in dynamic environments is their inability to generalize well to unseen scenarios. This limitation arises from their dependence on fixed training data and their opacity in reasoning, often referred to as the "opaque-box" nature of decision-making. Overfitting to specific datasets, difficulty in interpreting intermediate representations, and the lack of inherent reasoning capabilities make these models inadequate for tasks that require nuanced, context-aware decision-making.

To address limitations of traditional robotic systems in dynamic decision-making, our work introduces a system known as Robo-CSK-Organizer (see Fig. 1). This system represents a significant advancement in robotics, incorporating Explainable AI (XAI) via the integration of Commonsense Knowledge (CSK). Unlike conventional AI systems, which often operate as "opaque boxes" that can be quite unfathomable, Robo-CSK-

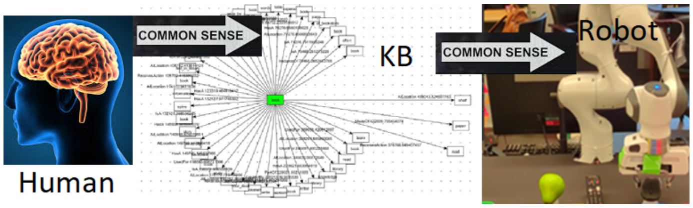


Fig. 1: Main concept of Robo-CSK-Organizer.

Organizer aims to make decision-making processes of robots more transparent and understandable to humans.

Robo-CSK-Organizer addresses challenges by integrating commonsense knowledge (CSK) with traditional machine learning techniques. This integration enables the system to infer spatial and semantic relationships using a hybrid approach that combines logical reasoning with inductive learning, as demonstrated in related research on hybrid architectures [6]. By leveraging CSK, Robo-CSK-Organizer not only improves adaptability in dynamic environments but also enhances transparency and explainability in decision-making, which are critical for fostering user trust and achieving robust performance.

CSK, i.e. commonsense knowledge refers to intuitive information that humans typically use daily. It includes basic facts, concepts, and the relationships between them that are taken for granted by human beings but are not inherently known by AI systems. By deploying CSK through the ConceptNet knowledge base, Robo-CSK-Organizer is equipped with a vast array of real-world knowledge, enabling it to make more informed and nuanced decisions. It helps to overcome the "opaque-box" issue commonly associated with pure deep learning systems, where the rationale behind decisions remains unclear [7]. Robo-CSK-Organizer aims to clear the opaque box with more explainability via explicit CSK usage.

ConceptNet offers several unique features that make it an effective choice for integrating commonsense knowledge (CSK) into the Robo-CSK-Organizer. Its graph-based structure connects concepts through a diverse set of semantic relationships, enabling nuanced reasoning about object categorization and contextual understanding. Additionally, ConceptNet supports over 80 languages, making it versatile for applications in diverse environments. Its efficient API further facilitates integration, allowing for seamless incorporation into robotic systems without extensive preprocessing or complex setups. While ConceptNet was selected as a proof-of-concept in this paper due to these practical advantages, future work can explore other CSK systems, such as OMICS and Dice, to evaluate their relative strengths and further enhance the Robo-CSK-Organizer's performance.

In the realm of robotics, the precision and trustworthiness of AI systems can be of the utmost importance. The opaque nature of many AI-driven decisions poses a significant challenge, particularly in applications where understanding the basis for a decision is crucial for safety and efficacy. By leveraging XAI, Robo-CSK-Organizer aims to enhance the decision-making capabilities of robots, making them more reliable and trustworthy partners in various applications. This transparency is essential not only for the operators who need to trust and understand the actions of robots but also for the broader acceptance of AI in critical and everyday applications.

Traditional AI models may often struggle with differentiating between similar objects in varied contexts, such as distinguishing a child's toy from a pet's toy, or understanding the appropriate placement of objects in household settings (tasks that very easy for humans due to inherent commonsense knowledge). Robo-CSK-Organizer bridges this gap by adequately harnessing CSK to enhance the organization of detected objects for task classification in multiple avenues.

CSK differs from encyclopedic knowledge [8] where AI systems far surpass humans. "Common sense" is naturally found in humans who acquire it inherently at birth, enhance it with further growth, and use it for intuitive reasoning. Machines on the other hand are not endowed with CSK by default and hence can often find it challenging to conduct reasoning intuitively unless pre-programmed with rigorous training [9]. For instance, it is very easy and in fact rather obvious for a human to know that the "door" of a *refrigerator* should not remain open, except while placing things in it or taking them out [10]. Conversely, the "door" of an *office* of *study-room* can certainly be open and is often ajar - a related fact also quite obvious to humans. Such knowledge is very subtle and is thus considered really simplistic or too common! Yet, it can often be crucial in decision-making. Hence, the role of CSK can be vital in modern-day AI systems as noticed significantly in the literature [11].

This is precisely the concept addressed by the Robo-CSK-Organizer system, i.e. CSK in AI for Robotics. More specifically, it employs a robust semantic network from a classical KB, namely ConceptNet [12], hence enabling robots to make more well-informed decisions based on contextual cues. This is in line with the logic of utilizing commonsense knowledge to augment machine intelligence [13], [14]. It thrives on the adequate extraction and compilation of CSK from a knowledge base, a non-trivial task, which can be crucial in AI applications [15], [16]. This work presents several key innovations in robotic technology and explainable AI (XAI) through the development of the Robo-CSK-Organizer system. The contributions of this study can be summarized as:

1) **Integration of Commonsense Knowledge (CSK):** Robo-CSK-Organizer uniquely integrates ConceptNet, a classical knowledge base, to enhance semantic and pragmatic reasoning in robotic decision-making, surpassing approaches reliant solely on pre-trained language models such as a ChatGPT-oriented baseline.
2) **Advancements in Explainable AI (XAI):** The system addresses the "opaque-box" challenge in AI by making robotic decision-making processes more transparent and understandable. Unlike state-of-the-art alternatives, it provides explicit, logical pathways for decisions, fostering user trust.
3) **Enhanced Contextual Understanding:** Using BLIP for contextual recognition and DETIC for object detection, Robo-CSK-Organizer achieves superior real-time comprehension of complex and dynamic environments. This

yields better organization, even in ambiguous scenarios.

4) **Robust Ambiguity Resolution and Consistency:** Through the integration of CSK, Robo-CSK-Organizer resolves ambiguities in object categorization and ensures consistent placement across diverse scenarios, outperforming a ChatGPT baseline in controlled experiments.
5) **Improved Task Adaptability:** The system dynamically adjusts to varying task priorities and environmental changes, showcasing flexibility in real-world applications. This adaptability is crucial for next-generation multipurpose robotics.
6) **Comprehensive Experimental Validation:** Extensive experiments confirm that Robo-CSK-Organizer outperforms a ChatGPT baseline in ambiguity resolution, placement consistency, and explainability, positioning it as a robust solution for advanced robotic applications.

The innovations described above contribute to advancing next-generation robotic systems by integrating human-like commonsense reasoning and adaptability. These contributions not only enhance the transparency and reliability of AI-driven robotics but also pave the way for broader applications in multipurpose and collaborative robotics. By addressing vital challenges in explainability and adaptability, this work makes a solid foundation for future developments in AI and robotics.

The rest of this paper is organized as follows. Section II outlines related work in the area. Section III gives an overview of Robo-CSK-Organizer, while Section IV has details on its development. Section V delves into its implementation and analysis of its experimental results; with relevant discussions in Section VI. Finally, Section VII offers conclusions & perspectives, and Section VIII outlines future research directions.

II. Related Work

The integration of AI and robotics has been applied in a variety of areas and environments with innovative approaches [17]–[22]. These methodologies have significantly contributed to the field by enhancing robots' decision-making processes, adaptability, and interaction with their environment.

One approach focuses on using ConceptNet and Google search data for object categorization in domestic robotics, particularly for tidy-up services.It effectively groups objects into functional categories, aiding robots in more intuitive object handling [23]. Paradigms based on artificial neural networks (ANNs) have been adapted to many contexts, ranging from machine translation [24] in text with recurrent neural networks (RNNs) to object recognition in multifaceted scenarios with computer vision models, e.g. VGG-16 [25]. The issue of extracting cultural commonsense knowledge and its usefulness in enhancing chatbots has been addressed in a novel approach called CANDLE [26] with multifaceted real-world impacts.

Advances in visual commonsense reasoning introduce the R2C engine [11] to enhance object recognition, anchoring natural language descriptions in visual data. CSK-Detector [27] is an innovative system for object detection in domestic robotics, leveraging common sense from the Dice knowledge base [28]; it reduces the need for extensive image annotation.

The incorporation of CSK from the OMICS database using Description Logic has also been discussed. This integration enables robots to perform more nuanced tasks, showcasing the potential for more context-aware robotics [29]. Furthermore, the application of CSK in human-robot collaborative tasks has been highlighted, especially in robot action planning for assembly tasks, emphasizing the enhancement of cooperative interactions [30]. Its mathematical modeling insights, with applications in smart manufacturing have been elaborated [6], emphasizing the role of commonsense reasoning.

Additionally, semantic task planning for service robots in dynamic, open-world environments has been explored. This method leverages natural language understanding and semantic reasoning, addressing the challenges posed by ever-changing environments [31]. The combination of non-monotonic logical reasoning and incomplete CSK with inductive learning to guide deep learning in robotics is another innovative approach. This integration offers a unique perspective on the convergence of CSK and advanced learning techniques [32]. A survey of robotic applications with respect to commonsense knowledge is presented [33], while the crucial issue of robots and human rights is debated [34], in some recent works.

These diverse approaches underscore the importance of CSK in improving the functionality and intelligence of robotic systems. They have advanced the field by demonstrating effectiveness in task planning, human-robot interaction, and environmental adaptation. Building on these foundations, Robo-CSK-Organizer depicts a significant advancement to the practical application of CSK in robotics. A novel aspect of Robo-CSK-Organizer is that it explicitly harnesses human commonsense (via a knowledge base) in real-world settings to perform object organization in task-based classification, particularly beneficial for multipurpose robots. Its ability to resolve ambiguities and maintain consistency in object placement, adaptability to diverse task classifications, and contributions to XAI, make it stand out as compared to other approaches.

III. Overview of Robo-CSK-Organizer

The Robo-CSK-Organizer system is designed to enhance robotic decision-making through a multi-step process that integrates advanced object detection, context recognition, and commonsense knowledge retrieval. This process aims to guide robots to make accurate, consistent, and explainable decisions in complex environments. The flow diagram in Fig. 2 illustrates the sequential steps involved in Robo-CSK-Organizer, providing a visual representation of the operations from the initial video feed input to the final output of organized objects.

A. Context Recognition

The process (Step 1) begins with a video feed input, which serves as the primary data source. This video feed is analyzed in real-time to detect both objects and their surrounding contexts. The first step involves understanding the context in which these objects are situated, performed using BLIP (Bootstrapping Language-Image Pre-training). BLIP is designed to discern the broader environmental context by analyzing visual and textual data, determining if an observed room is a living room, kitchen, etc. This contextual understanding is crucial for

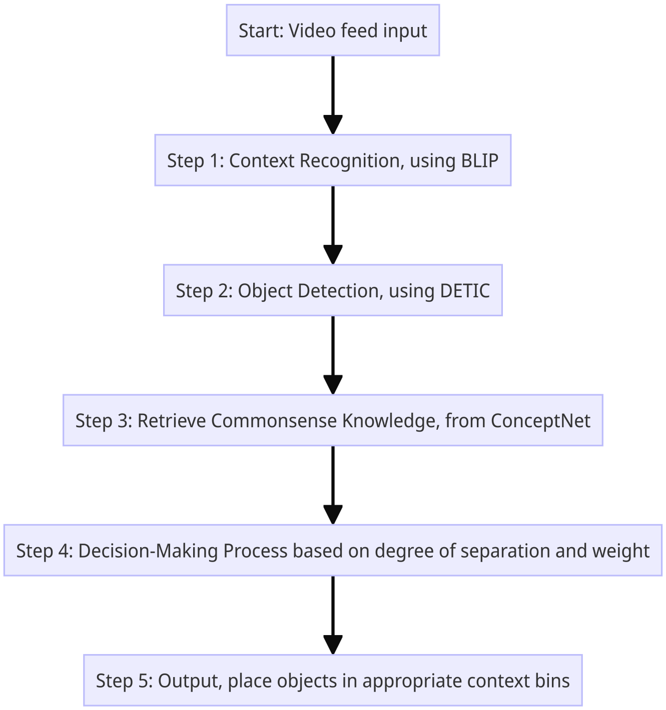


Fig. 2: Flow diagram of the Robo-CSK-Organizer process.

making informed decisions about object placement and organization. BLIP enhances the system's ability to interpret scenes holistically, integrating visual information with linguistic cues to accurately determine the appropriate context for each object.

B. Object Detection

After the contexts have been recognized and labeled, the next step (Step 2) involves detecting the objects within each frame. The Robo-CSK-Organizer employs DETIC (DEtector with Image Classes) for this task. DETIC is a state-of-the-art object detection tool that leverages extensive image classification data to accurately identify a vast array of objects. By utilizing DETIC, the system can recognize, locate, and label thousands of different objects, ensuring comprehensive detection. This step is fundamental as it lays the groundwork for subsequent contextual analysis and decision-making.

C. Retrieve Commonsense Knowledge

With the objects detected and their contexts recognized, the system then retrieves commonsense knowledge from the ConceptNet knowledge base (Step 3). ConceptNet is a comprehensive KB that encompasses a wide array of semantic and pragmatic information about everyday objects and their interrelationships. It provides the Robo-CSK-Organizer with the necessary pragmatic knowledge to understand not just what the objects are, but also how they are typically used and related to each other in various contexts. This step is crucial for imbuing the system with a level of understanding akin to human commonsense, enabling it to make more nuanced and informed decisions.

D. Decision-Making Process

In this step (Step 4), the Robo-CSK-Organizer evaluates the retrieved commonsense knowledge to make decisions about the detected objects. The decision-making process involves analyzing the degree of separation between objects and their contextual relevance, as well as the weight of the relationships defined in ConceptNet. This ensures that the system's decisions are not merely based on data but are contextually appropriate and logically sound. By leveraging human-like reasoning, the Robo-CSK-Organizer can resolve ambiguities and ensure consistency in object placement, adapting to new and dynamic scenarios effectively.

E. Operating Objects in Appropriate Context Bins

Finally, the system outputs its decisions by placing objects into their appropriate context bins (Step 5). This involves organizing the detected objects in a manner that aligns with human expectations and practical use cases. For example, utensils are placed in the kitchen, stationery in the office, and so on. This step enhances the system's reliability and transparency, as users can understand and trust the logic behind the object's placement. By ensuring that objects are organized correctly and consistently, the Robo-CSK-Organizer improves efficiency and usability in various real-world applications.

IV. Development of Robo-CSK-Organizer

In this section, we present the detailed functioning of Robo-CSK-Organizer. The salient features of the Robo-CSK-Organizer include:

Resolving ambiguity in object categorization: The Robo-CSK-Organizer leverages ConceptNet to navigate the complexities of object categorization, addressing the challenge of ambiguity by integrating commonsense knowledge in a task-relevant manner. This approach not only improves accuracy in object classification but also enhances the system's ability to deal with the diverse and unpredictable nature of real-world environments. By dynamically applying CSK, Robo-CSK-Organizer can make informed decisions even in scenarios where training examples are insufficient or nonexistent, thereby significantly reducing the ambiguity inherent in categorizing objects across different contexts.

Maintaining consistency in object placement: Consistency in object placement is vital for building user trust and ensuring seamless human-robot interaction. Robo-CSK-Organizer achieves this by synthesizing commonsense and domain-specific knowledge to manage contextual variations effectively. This capability is crucial in multi-robot environments, where consistent and predictable behavior is necessary for collaborative tasks. By employing clear reasoning paths and considering both qualitative and quantitative environmental descriptions, Robo-CSK-Organizer maintains a high level of consistency in object placement, thereby enhancing the reliability of robotic assistance in multipurpose settings, e.g. those that can be encountered in domestic robotics. It is achieved by Robo-CSK-Organizer with clear reasoning paths.

Depicting task relevance and adaptability: Adaptability in prioritizing tasks based on context is essential for robots operating in dynamic environments. Robo-CSK-Organizer exhibits this adaptability by systematically adjusting its decision-making process in response to changes in task priorities, even in the presence of probabilistic uncertainties. This feature is particularly important for applications requiring the robot to switch between tasks, such as gardening and culinary activities, showcasing the system's ability to handle real-world complexities and make contextually relevant decisions.

Fostering explainability in AI systems: Explainability is a cornerstone of user trust and acceptance in AI systems. Robo-CSK-Organizer excels in this aspect by making its decision-making processes transparent and understandable to users. Unlike systems based solely on pre-training neural networks that may act as "opaque boxes", Robo-CSK-Organizer's explicit use of CSK enables it to provide clear and comprehensible explanations for its actions. This level of explainability is critical in environments where understanding the rationale behind decisions can significantly impact user confidence and the effective collaboration between humans and robots.

To elaborate on the utilization of ConceptNet by Robo-CSK-Organizer for object categorization within a kitchen environment, we delve into a detailed example that highlights the system's sophisticated decision-making process. Fig. 3 illustrates the object categorization process for "Apple" via the Robo-CSK-Organizer.

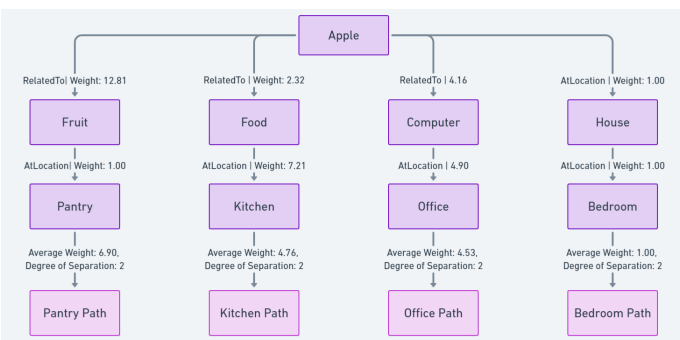


Fig. 3: ConceptNet-based decision pathway in Robo-CSK-Organizer — Illustration of object categorization of "Apple" to Pantry, Kitchen, Office, or Bedroom. Location chosen is based on degree of separation, (from low to high), and by average weight (from high to low). Here, Pantry is chosen 1st, followed by Kitchen, Office, and Bedroom.

The ConceptNet KB serves as a pivotal resource in the Robo-CSK-Organizer's arsenal, providing a rich, interconnected network of commonsense knowledge that the system leverages for object recognition and categorization tasks. ConceptNet's architecture is designed around the idea of semantic networks, encompassing a wide array of concepts and the relationships between them. These relationships are expressed through edges, such as "AtLocation" or "RelatedTo", which help in establishing connections between seemingly disparate concepts. Initially, in order to allow the Robo-CSK-organizer to take full advantage of ConceptNet, all of the relevant paths between objects and locations needed to be visited in order to find optimal paths. This is achieved by suitably adapting the well-known Breadth-First Search (BFS) technique in our work, as outlined in Algorithm 1 here.

Algorithm 1 Modified BFS for Object-Location Pathfinding with Caching

Require: $G(V, E)$, V_{obj}, V_{loc}, R, v_{start}
Ensure: Set of paths P from v_{start} to V_{obj}
Initialize queue $Q \leftarrow \{(v_{start}, [], \{v_{start}\})\}$
Initialize cache $C \leftarrow \{\}$
while $Q \neq \emptyset$ **do**
  $(v, P, Visited) \leftarrow$ dequeue(Q)
  for each $r \in R$ **do**
    $Data \leftarrow$ fetchRelatedData(v, r, C)
    for each $edge$ in $Data$ **do**
      Process $edge$ to determine v_{next}
      if $v_{next} \notin Visited$ **then**
        $Visited_{new} \leftarrow Visited \cup \{v_{next}\}$
        $P_{new} \leftarrow P + [v_{next}]$
        Enqueue $(v_{next}, P_{new}, Visited_{new})$ to Q
      end if
    end for
  end for
end while
Save paths and cache

Algorithm 1 is designed to dynamically explore a graph constructed from nodes representing various objects and locations. These nodes are interconnected through different types of relationships fetched from an API (Application Programming Interface), specifically the ConceptNet API. The primary objective is to find paths between a set of predefined locations and objects, optimizing the search with caching mechanisms and controlling the breadth of the search.

Let $G = (V, E)$ denote a directed graph; V is a set of vertices depicting objects and locations, E is a set of edges on relationships between these vertices. Each edge $e \in E$ is associated with a weight $w(e)$ and a relationship type $r(e)$.

$$G = (V, E) \tag{1}$$

- V_{obj} denotes the set of object vertices.
- V_{loc} denotes the set of location vertices.
- $R = \{r_1, r_2, \ldots, r_n\}$ is the set of relationship types, e.g., *AtLocation*, *UsedFor*.
- A path P is a sequence of vertices $(v_1, v_2, \ldots, v_n)$ such that $\forall i, (v_i, v_{i+1}) \in E$.
- The algorithm uses a cache C to store previously fetched data to reduce API calls.

$$R = r_1, r_2, \ldots, r_n \tag{2}$$

$$P = (v_1, v_2, \ldots, v_n) \tag{3}$$

The algorithm is implemented as follows:

1) **Initialization:** Define a queue Q for BFS, starting at a given location $v_{start} \in V_{loc}$. Initialize the cache C.

$$Q \leftarrow \{(v_{start}, [], \{v_{start}\})\} \tag{4}$$

$$C \leftarrow \{\} \tag{5}$$

2) **BFS Loop:**
   a) While Q is not empty, dequeue an element v, representing the current node, along with its path P and a set of visited nodes specific to P.
   b) For each relationship type $r \in R$, query API to fetch related nodes, unless data is found in C.
   c) Filter the fetched nodes based on language and relevance criteria, updating C with new findings.
   d) For each valid next node v_{next}, if v_{next} has not been visited in the current path, enqueue it to Q with the updated path and visited set.
3) **Path Management:**
   a) Maintain a global set of visited nodes to prevent revisiting nodes across different paths.
   b) Track and store paths to objects of interest, limiting search depth to a predefined maximum degree.
4) **Caching and Output:**
   a) Periodically save cache C to a file for persistence.
   b) Upon completion, output the found paths to a file, logging the elapsed time and any objects not found.

In the scenario depicted in Fig. 3, Robo-CSK-Organizer is faced with the task of categorizing the object *Apple* into one of the locations *(Pantry, Kitchen, Office, Bedroom)*. The figure illustrates the potential pathways linking the contexts of Pantry, Kitchen, Office, and Bedroom. Each pathway shows a different sequence of reasoning steps, guided by the relationships in ConceptNet. Decision-making in Robo-CSK-Organizer involves evaluating these paths and selecting the one that best fits the context of where the object *(apple)* belongs.

Initially the system prioritizes the pathway with the lowest degree of separation (in Fig. 3 this is *"2"* throughout). Afterwards, the path chosen is determined by the average weight of the path. Therefore, in Fig. 3 the most likely path is *Pantry (Average Weight: 6.90)*, followed by *Kitchen (Average Weight: 4.76)*, then *Office (Average Weight: 4.53)*, ending with *Bedroom (Average Weight: 1.00)*. To use the pantry as an example, the higher weight indicates a strong association between the location *(Pantry)* and the object category *(Fruit)*, suggesting that fruit items are commonly found in pantries. This logical inference is a crucial step in the system's reasoning process, guiding its decision-making.

The selected pathway further refines the categorization process by establishing a connection between "Apple" and "Fruit" through a "RelatedTo" edge. This connection is based on the commonsense understanding that apples, as a specific type of food, share certain characteristics and contexts with other food items. By associating "Apple" with "Fruit", the system leverages this relationship to finally connect "Apple" to "Fruit", and finally to "Pantry". This sequence of connections—linking "Apple" to "Fruit" and then "Fruit" to "Pantry" illustrates the system's ability to navigate through ConceptNet's semantic network, using logical sequences to arrive at contextually appropriate categorizations.

Through this detailed example, we can see how the Robo-CSK-Organizer exemplifies sophisticated decision-making capabilities by utilizing commonsense knowledge embedded in ConceptNet. The system's ability to select the most relevant pathway based on contextual clues and semantic relationships demonstrates its advanced reasoning process. This process not only enables accurate object categorization but also showcases the potential of integrating commonsense knowledge into AI systems for enhanced decision-making in real-world environments. By leveraging ConceptNet's vast repository of commonsense knowledge and semantic relationships, the system exemplifies the potential of AI to understand and navigate complex, real-world environments in a logical and contextually appropriate manner.

Designed with modules for object detection, context recognition, and semantic analysis, Robo-CSK-Organizer not only classifies objects but also interprets their appropriate placement within various contexts. This approach significantly enhances the transparency and interpretability of AI decisions, addressing the critical challenge of explainability in AI systems. It can thus be considered analogous to systems that delve into aspects such as spatial commonsense [35], [36] for object recognition and training of autonomous systems. For instance, if there is an error, its traceability is highly facilitated in Robo-CSK-organizer (versus solely deep learning based systems, e.g, those using ChatGPT for training). Hence, Robo-CSK-Organizer can help robots learn from their mistakes and correct themselves, thus getting better in their performance. Moreover, XAI is important in learning to enhance interpretability, as noticed over the years [37]–[39]. The XAI contribution of Robo-CSK-Organizer is particularly notable when humans and robots work together, i.e. for human-robot collaboration, as humans are able to understand the actions of the robots much better, along with the reasons behind the robots' decisions. The same logic applies to numerous robots working together. This enhances trust in the realm of robotics. All these facets are vital, especially with the growing prevalence of multipurpose robots, heading towards next-generation advancements.

The main functioning of Robo-CSK-Organizer, focusing on its reasoning, is outlined in Algorithm 2 here.

This algorithm, highlighting the main functioning of Robo-CSK-Organizer, operates on 2 primary inputs: a video feed $\mathcal{V}$ with multiple frames $f_1 \dots f_n$ and a commonsense knowledge base $\mathcal{C}$ (here, ConceptNet).

$$\mathcal{V} = \{f_1, f_2, \dots, f_n\} \tag{6}$$

Its goal is to sort detected objects into their appropriate contexts. Initially, the robot's vision system $\mathcal{R}$, implemented using Detectron2, is initialized. Thereafter, the context bins are scanned and recognized using BLIP ($\mathcal{B}$), with the contexts stored in a CSV (comma-separated variable) format ($\mathcal{S}_{\text{csv}}$). For each frame f in the video feed $\mathcal{V}$, objects are detected and labeled as $\mathcal{D}$. Each detected object o is then checked against $\mathcal{C}$ to determine its context $\mathcal{K}_o$. If this context matches that of a bin in $\mathcal{B}$, the object is placed in the corresponding bin. The process continues for each object in the frame, and the annotated frame can be displayed as needed.

Algorithm 2 Reasoning Process of Robo-CSK-Organizer

Require: $\mathcal{V}$ (Video feed), $\mathcal{C}$ (ConceptNet knowledge base)
Ensure: $\mathcal{O}_{\text{sorted}}$ (Objects sorted into appropriate contexts)

Initialize robot vision system $\mathcal{R} \leftarrow$ Detectron2
Scan context bins using $\mathcal{B} \leftarrow$ BLIP, store context in $\mathcal{S}_{\text{csv}}$
for $f \in \mathcal{V}$ **do**
  $\mathcal{D} \leftarrow$ Detect and label objects in f using $\mathcal{R}$
  for $o \in \mathcal{D}$ **do**
    $\mathcal{K}_o \leftarrow$ Query $\mathcal{C}$ for context of o
    if $\mathcal{K}_o$ matches context in $\mathcal{B}$ **then**
      Place o in matched context bin
    else
      Continue to next object
    end if
  end for
end for
Optional: Display annotated frame from f

This algorithm is implemented into the Robo-CSK-Organizer system using Python and is integrated with a robotic arm in our laboratory, namely, the CRoSS Lab (Collaborative Robotics and Smart Systems Lab at our university). Robo-CSK-Organizer is then executed using various real-world objects. Details of its execution are mentioned next in the respective parts of its system demonstration.

V. Implementation Results and Analysis

The Robo-CSK-Organizer system is implemented as illustrated in the system diagram of Fig. 4. It uses Detic for object detection, BLIP for context discernment, and ConceptNet for harnessing commonsense knowledge. It conducts reasoning based on the CSK extracted from ConceptNet and uses that to make decisions about object placement for task classification. In this paper, we address its implementation and experimentation with the application domain of domestic robotics. Multipurpose robots can be useful here, especially in next-generation systems.

In order to demonstrate the efficacy of Robo-CSK-Organizer, it is compared with a baseline task organizer that uses the well-known ChatGPT for guidance. We thus present the following with respect to the overall task setup and the experiments evaluation with comparison.

A. Task Setup

Object Detection: Both systems, our Robo-CSK-Organizer and the ChatGPT baseline, use DETIC (DEtector with Image) to ensure a broad evaluation spectrum. Specific context groups, particularly domestic locations (e.g. kitchen, garden, pantry, dining room) are chosen for evaluation. These contexts are relevant to the selected object categories and provided with a controlled environment for testing. An advanced extension of Detectron2 for object detection, which contains over 21,000 classes [40] is used here to provide choices of classes for object organization. Each object is queried against each system (Robo-CSK-Organizer / ChatGPT) 10 times, asking it to organize the object into one of the provided contexts. Responses from both systems are recorded for each iteration, and the most frequent context is identified as the predominant choice for object placement.

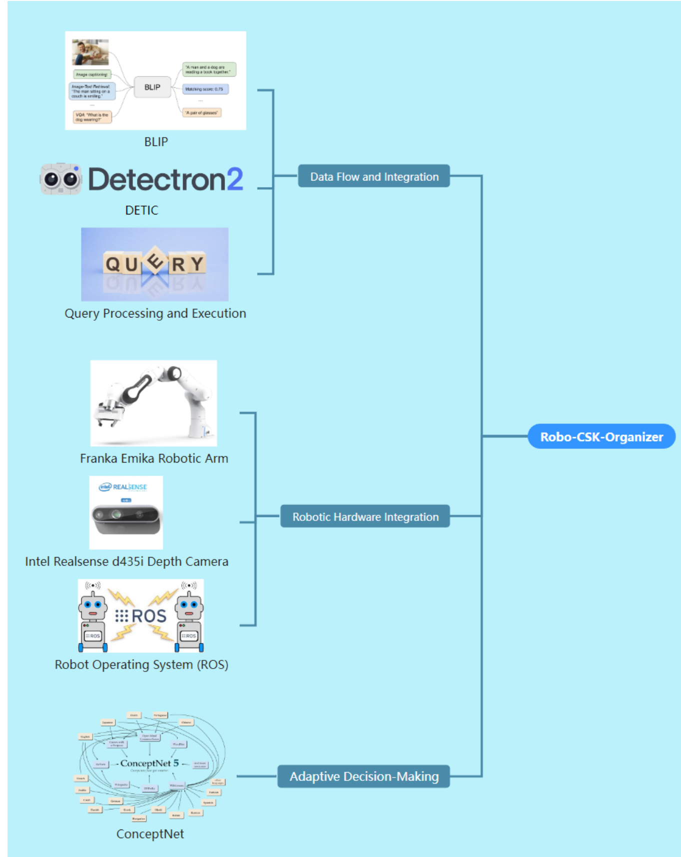


Fig. 4: System diagram of Robo-CSK-Organizer.

Context Recognition: BLIP (Bootstrapping Language-Image Pre-training) is employed to identify contexts such as the kitchen or office, and thus generate room captions for enhanced clarity [41]. Note that the usage of such software can be helpful in a variety of applications, e.g. image personalization via text by harnessing diffusion models [42]. The hardware foundation for both systems (i.e. Robo-CSK-Organizer and the ChatGPT baseline) includes a Franka Emika robotic arm with an Intel Realsense D435i camera [43]–[45], integrated with ROS. Robo-CSK-Organizer works with this hardware, and incorporates CSK-based reasoning (See Algorithm 2). This is the key to addressing the opaque-box issue in AI, aiming for more transparent object sorting.

Reasoning in the Functioning Approach: The pivotal distinction between the two systems is in the functioning approach for sorting objects into relevant contexts. While ChatGPT relies solely on prior training with deep learning, Robo-CSK-Organizer applies commonsense knowledge due to which it can be more adept in successfully handling first-time scenarios as well. This is explained as follows:

Robo-CSK-Organizer: It harnesses a classical knowledge base called ConceptNet [12] for semantic insights and commonsense reasoning. It infers object locations using metrics known as *edge weight* and *degree of separation*, prioritizing paths based on these factors. While ConceptNet is chosen for its user-friendly interface and clear path logic, we claim that other relevant CSK knowledge bases can also be used.

ChatGPT baseline: A ChatGPT-trained organizer is used as a baseline; it relies on generative pre-trained transformer models for its decision-making, processing text-based inputs to infer object locations and categorizations. This approach, though adept in language processing, does not integrate a structured commonsense knowledge base. Consequently, the ChatGPT-based organizer's decisions are more influenced by pre-trained patterns in textual data rather than explicit semantic relationships and intuitive logical reasoning. This can affect consistency and transparency in decision-making in complex or ambiguous scenarios, notably (but not limited to) those encountered for the first time.

B. Experiments for Comparative Evaluation

In our detailed evaluation process including comparative studies, we conduct experiments to assess the performance of Robo-CSK-Organizer and the ChatGPT baseline across various contexts. The key aspects of these experiments are focused on ambiguity resolution, consistency, task-relevance adaptability, and explainability. A summary of our exhaustive experimentation is presented below.

1) Ambiguity Resolution: The experiments aimed at resolving ambiguity sought to rigorously evaluate the capabilities of the Robo-CSK-Organizer and the ChatGPT baseline in discerning the most contextually suitable placements for a variety of objects. This undertaking was pivotal, given the array of object categories involved—ranging from fruits and vegetables to office supplies—and the necessity for these systems to function effectively across diverse contextual settings, such as kitchens, offices, and playrooms. A key aspect of this experiment was the adoption of FastText, Word2Vec, and GloVe models for semantic similarity analysis, significantly bolstering the systems' capacity to navigate a wide spectrum of objects and contexts. This was particularly advantageous for FastText, which, through its subword information processing, excels in generating word embeddings for out-of-vocabulary words, enhancing its utility across varied objects/contexts.

The experimental methodology was comprehensive, involving the calculation of similarity scores for each object-context pairing through the application of FastText, Word2Vec, and GloVe models. The aggregated scores from these models furnished a detailed view of the semantic relationships between objects and contexts, ensuring a nuanced and accurate evaluation of ambiguity by leveraging the strengths of each model. To quantitatively assess ambiguity, the experiment introduced an innovative approach, measuring the smallest difference in average similarity scores across different contexts for each object. This method, predicated on analyzing semantic similarity scores derived from the NLP models, posited that the closer the average scores for two distinct contexts, the greater the associated ambiguity for the object-context pair.

The capacity for ambiguity resolution was further scrutinized by selecting two contexts deemed ambiguous for each object, effectively testing the models' determinations that an object could logically belong to either context. Ground truth was then established based on the higher of the two similarity scores, setting a benchmark against which the responses of both the Robo-CSK-Organizer and the ChatGPT Organizer were compared. This comparative analysis elucidated each system's proficiency in navigating and resolving ambiguity within object-context associations, thereby providing a comprehensive understanding of their capabilities in this domain. The results (See Fig. 5), show that Robo-CSK-Organizer has notable accuracy.

2) Ensuring Consistency: The consistency experiments aimed to evaluate the stability and repeatability of the Robo-CSK-Organizer compared to the ChatGPT baseline when identical queries were presented across multiple iterations. This assessment was crucial to determine the reliability of knowledge organization systems, testing the systems' potential to consistently select the same context for an object through numerous trials, essential for such technologies.

The methodology for ensuring consistency involved selecting objects from various categories, including personal items, clothing, office supplies, and toys, and evaluating them in specific context groups like culinary locations (kitchen, garden, pantry, dining room). These objects were queried against the systems multiple times to assess the repeatability of the context choice for each object.

Robo-CSK-Organizer demonstrated a 100% consistency rate across all object-location pairs. This remarkable level of consistency can be attributed to the static nature of the ConceptNet knowledge graphs utilized in its decision-making process. In contrast, the ChatGPT Organizer exhibited less consistency, especially for objects such as adhesive tape, belt, sock, remote control, toothpaste, and aerosol can. (See Fig. 5) This variability can be due to the fact that pre-training alone might not always lead to consistent results in systematic object organization — a potential limitation of dynamic, context-aware systems like ChatGPT in achieving repeatable outcomes in specific knowledge organization tasks.

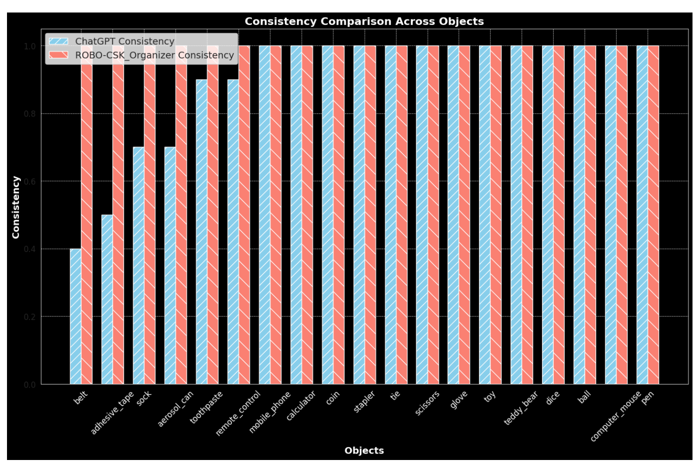


Fig. 5: Robo-CSK-Organizer has 100% consistency in all object-location pairs. ChatGPT baseline is not as consistent for all objects, specifically for adhesive tape, belt, sock, remote control, toothpaste, and aerosol can.

3) Task-Relevance Adaptability: These experiments evaluated the systems' ability to adapt their responses to different context-specific directives. The objective was to assess whether Robo-CSK-Organizer and the ChatGPT baseline can re-calibrate their responses when directed to focus on alternative contexts differing from their initial preferences. The experiments commenced with an initial straightforward assessment, querying "apple" against 4 contexts (kitchen, living room, bedroom, bathroom) without any focused directives. This determined the systems' natural inclination or preference for context association. In the adaptability testing phase, the systems were prompted to focus on the remaining 3 contexts, one at a time, to observe if they could adapt their responses when a specific context was emphasized — analogous to humans adapting to different contexts in real life. For instance, if we specifically tell a human not to place apples in a kitchen (for any reason, maybe kitchen is too small or is being cleaned for pest-control) then the human should intuitively find another good place for the apples rather than placing them in the kitchen again. Accordingly, it was interesting to assess how robotic systems behaved in such situations.

Likewise, results for the initial as well adaptability tests were collected by repeating each context-query 10 times. Responses are compiled into respective data frames for detailed analysis. The focused contexts for the adaptability tests are based on preference, with the most preferred context excluded to emphasize the remaining contexts. The initial phase identified kitchen as the clear preference for sorting apples for both the ChatGPT-based organizer and Robo-CSK-Organizer. In the adaptability phase, there were observable shifts in Robo-CSK-Organizer's response, as desired, i.e. it was more adaptable when needed. The explicit paths Robo-CSK-Organizer employed for sorting, leading to object placement, were as follows:

- Path: Kitchen (AtLocation) <- food (RelatedTo) > apple
- Path: Bedroom (AtLocation) <- house (AtLocation) <- apple

Figs. 6 and 7 here provide a well-summarized visual representation of these findings.

4) Explainability: The experiment on explainability assessed the Robo-CSK-Organizer system and the ChatGPT-based system as per their abilities to elucidate their decision-making processes. This aspect was crucial for building user trust and understanding, considering the fact that decisions based on various situations might seem counter-intuitive at times. Robo-CSK-Organizer, which utiilized ConceptNet for harnessing human-like commonsense, could provide logical paths for its decisions, thus enhancing transparency. For instance, during its incorrect placement of beer into the playroom, Robo-CSK-Organizer provided a clear logical path: *playroom (UsedFor) fun (RelatedTo) party (RelatedTo) beer*. This path demonstrated the connection of concepts leading to the system's conclusion. In contrast, the ChatGPT-based organizer, relying basically on deep learning models, functioned as an opaque-box. It was unable to ascertain the explicit reasoning behind its decisions, e.g. it placed "scissors" into a "playroom" (a potentially hazardous decision). Hence, the ChatGPT baseline was lacking in a clear explainable framework. This could pose problems in error-correction, thus adversely impacting performance.

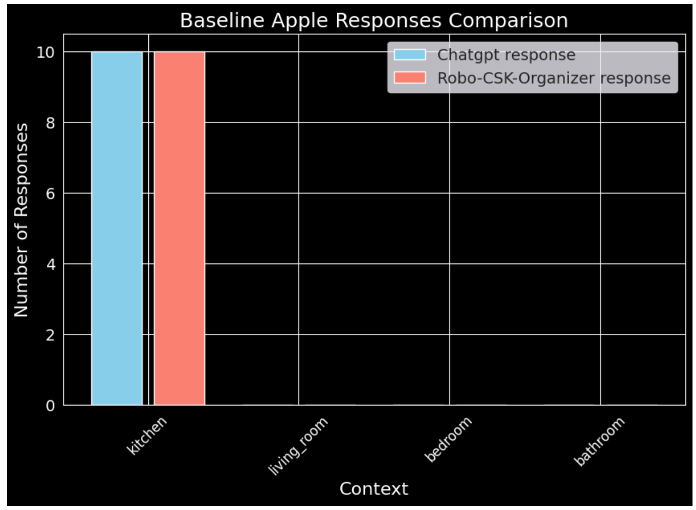


Fig. 6: When asking both ChatGPT and Robo-CSK-Detector where apple most likely belongs when choosing between Kitchen, Living Room, Bathroom, and Bedroom, both systems established the kitchen as their baseline responses.

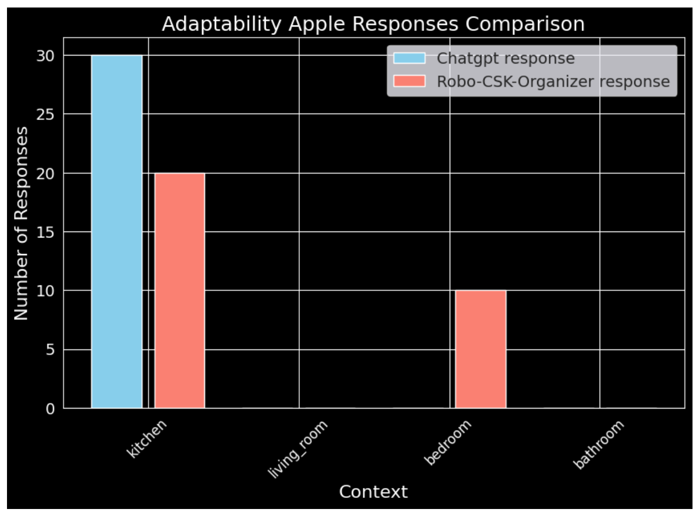


Fig. 7: In the adaptability phase, ChatGPT's implementation did not cause any observable shifts, despite requesting it to change its context. There were observable shifts on the Robo-CSK-Organizer in one of the locations, after requesting context-switching.

This distinction highlighted the fact that both systems may err, analogous to the adage "*to err is human*". However, Robo-CSK-Organizer's explainability allowed for better understanding and correction of the errors, which was useful.

Explainability is essential, especially in robotic systems where precision and safety are critical, contributing to user trust and understanding of AI decisions. Note that such explainability can in turn help in explicit communication with various AI systems, including intelligent agents in mobile apps, e.g. it can help to enhance existing apps with virtual voice agents [46] by adding more image-based functions where adequate object recognition is crucial. Hence, it can indirectly help a different type of robot, including a chatbot or a virtual voice assistant. All these systems would benefit from easier comprehension and enhanced interpretability. Hence, explainable AI plays a vital role here.

C. Real-world Application and Results

Considering real-world aspects, Figs. 8 and 9 illustrate how Robo-CSK-Organizer and the ChatGPT-based organizer derive their decisions in the placement of "scissors". This comparative analysis emphasizes the Robo-CSK-Organizer's strengths in consistency, adaptability, and explainability. The findings of our experimentation thus underscore the importance of integrating structured knowledge bases in AI systems. This fact is highlighted here, focusing on various scenarios for domestic environments. More information about the implementation of ConceptNet for organizing objects in contexts is at [47].

(a) Detic detects scissors in the workspace

(b) Robo-CSK-Organizer prompts arm to grab scissors

(c) Arm lifts scissors prior to organizing

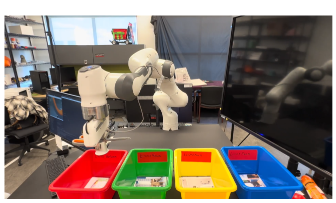
(d) Arm drops scissor in dining room, with CSK-guided reasoning

Fig. 8: Robo-CSK-Organizer's (quite adequate) placement of scissors into the dining room.

(a) Detic detects scissors in the workspace

(b) Chat GPT prompts arm to grab scissors

(c) Arm lifts scissors prior to organizing



(d) Arm drops scissor in child's playroom (without CSK)

Fig. 9: ChatGPT-Organizer's (rather dangerous) placement of scissors into the child's playroom.

Finally, we synopsize the comparative evaluation of Robo-CSK-Organizer and the ChatGPT baseline in TABLE I below.

TABLE I: Robo-CSK-Organizer versus ChatGPT

Parameter	**Robo-CSK-Organizer**	**ChatGPT**
Approach	Uses ConceptNet, logical paths	Generative transformer models
Ambiguity Resolution	Noticeable Accuracy	Varies based on trained patterns
Consistency	100% consistent across all classes	Consistency varies depending on class
Task-Relevance Adaptability	Adaptable to directives	Limited adaptability shown
Explainability	High; clear paths	Lower; opaque due to AI
Decision-Making Basis	Semantic, pragmatic CSK	Textual data patterns
Hardware Integration	Robotic systems integration	Robotic systems integration
Use of AI	AI with knowledge bases	Primarily AI-driven

VI. Discussions

This research presents significant contributions to the field of AI-driven robotics, particularly in the development and application of commonsense knowledge for task classification in multipurpose robots. Robo-CSK-Organizer, as a pioneering system, stands out in several key areas when compared to existing systems (e.g. a baseline organizer using ChatGPT). These are mentioned as follows.

A. Novel Integration of CSK

The Robo-CSK-Organizer's use of commonsense knowledge through ConceptNet depicts a significant shift from traditional deep learning approaches, which often rely on vast datasets and computational power to learn from scratch. This integration allows for a more nuanced understanding of objects and their contexts, leveraging structured knowledge bases to enhance decision-making transparency and accuracy.

Deep learning systems, while powerful, may struggle with tasks that require explicit understanding of context; they may also have problems functioning well without extensive training data. In contrast, the Robo-CSK-Organizer uses CSK to fill these gaps, providing a solid foundation for reasoning that does not solely depend on data quantity or quality. This approach offers a comparative advantage in scenarios where data is scarce, ambiguous, or where interpretability is crucial.

B. Superiority in Consistency and Adaptability

1) Consistency Across Trials: The achievement of a 100% consistency rate by the Robo-CSK-Organizer is a testament to its methodological robustness and reliability in knowledge organization tasks. This level of consistency is unparalleled, especially when compared to historical benchmarks or similar systems, which often show variability in their responses due to the dynamic nature of their underlying algorithms. For instance, traditional AI systems, which rely heavily on just deep learning and pattern recognition, might demonstrate fluctuating performance when dealing with identical queries over multiple iterations due to their dependence on the variety and scope of the training data. In contrast, the static nature of the ConceptNet knowledge graph underpinning the Robo-CSK-Organizer ensures a stable and repeatable output, emphasizing the system's reliability and trustworthiness for end-users.

2) Adaptability to Context-Specific Directives: The Robo-CSK-Organizer's adaptability was rigorously tested through experiments designed to assess its response to varying context-specific directives. An illustrative example of its adaptability is evidenced in its ability to recalibrate responses based on directed focus changes. For instance, when initially assessing an object such as an apple, the system might preferentially categorize it within a kitchen context. However, when prompted to consider alternative contexts, the Robo-CSK-Organizer demonstrated the capability to shift its classification logic, adapting its decision-making process to align with the new contextual directive.

This adaptability is crucial for applications in dynamic environments, such as domestic settings, where the relevance of context can frequently change. The Robo-CSK-Organizer's performance in these scenarios underscores its potential for real-world applications, setting a new standard for AI systems in terms of both flexibility and responsiveness to human needs.

The combination of high consistency and notable adaptability positions the Robo-CSK-Organizer as a superior system in the landscape of AI and robotics, offering a robust solution for knowledge organization tasks that require both reliability and dynamic adaptability.

C. Advancements in Explainability

1) Mechanics of Decision-Making Transparency: The Robo-CSK-Organizer system enhances explainability through a combination of advanced technological integration and the strategic use of ConceptNet, a comprehensive commonsense knowledge base. This integration facilitates a transparent decision-making process by providing logical and understandable pathways for its operations. Specifically, the system leverages ConceptNet to navigate through a vast network of semantically connected data, enabling it to explain the rationale behind its decisions and actions. This capability is further enriched by the inclusion of advanced object detection and environmental context understanding tools such as Detic and BLIP. These technologies provide the system with a detailed perception of its surroundings, allowing it to make and explain decisions based on a deeper understanding of both objects and their contextual relevance.

2) Implications for Human-Robot Collaboration: The emphasis on explainability is paramount for fostering trust and efficiency in human-robot interactions. By offering clear explanations for its decisions, the Robo-CSK-Organizer significantly reduces the opacity that often characterizes AI and robotics systems, thus building user trust. This transparency is essential in scenarios where robots and humans collaborate closely, as it ensures that users can understand and predict the robot's actions, potentially leading to more harmonious and effective teamwork.

For instance, when the Robo-CSK-Organizer decides on the placement of an object in a seemingly non-intuitive location, its ability to provide a logical explanation based on CSK can reassure the user of the system's reliability and thoughtfulness. This level of explainability is especially crucial in complex or ambiguous scenarios where the system's choices might not be immediately apparent to the user. The Robo-CSK-Organizer excels in these situations by using its integrated technologies and CSK to clarify its reasoning, thereby enhancing real-world adaptability and user interaction.

Hence, advances in explainability not only make Robo-CSK-Organizer more user-friendly and trustworthy but also significantly contribute to the broader field of XAI within Robotics by showcasing the importance and feasibility of creating transparent, understandable, and collaborative robots.

D. Integration of Advanced Technologies

The synergistic use of Detic and BLIP technologies with ConceptNet has significantly enhanced the Robo-CSK-Organizer's capabilities in object detection and environmental context understanding. Detic, a novel object detector, extends the vocabulary of detectable concepts to tens of thousands, overcoming the limitations of traditional object detectors constrained by the smaller scale of detection datasets. It simplifies the process by training the classifiers of a detector on image classification data, thereby expanding the system's ability to recognize a vast array of objects, including those without specific box annotations. This advancement allows superior performance in open-vocabulary and long-tail detection benchmarks, effectively closing performance gaps for object categories with few samples.

BLIP (Bootstrapping Language-Image Pre-training), a Vision-Language Pre-training (VLP) framework, complements Detic's object detection capabilities by enhancing environmental context understanding. BLIP utilizes noisy web data for synthetic caption generation and employs a filtering process to enhance the quality of information. This framework achieves state-of-the-art results in vision-language tasks such as image-text retrieval, image captioning, and visual question answering. Its strong generalization ability extends to video-language tasks, indicating its versatility and effectiveness in understanding complex environmental contexts.

The synergistic integration of Detic and BLIP provides the Robo-CSK-Organizer with a nuanced understanding of its surroundings, enabling the system to not only accurately detect and localize objects but also to understand the nature of the environment they operate in. This comprehensive solution enhances the system's ability to interact with and adapt to diverse domestic environments, making it more intelligent and capable in real-world applications.

E. Scalability for Real-World Application

The advanced object detection capabilities of Detic, combined with the environmental context understanding provided by BLIP, pave the way for enhanced real-world applications of the Robo-CSK-Organizer system. We present a few hypothetical scenarios next, so as to demonstrate its potential capabilities in domestic environments.

1) Automated Home Organization: The Robo-CSK-Organizer could autonomously organize a cluttered room by identifying objects, understanding their usage context, and placing them in their appropriate locations. For example, it could differentiate between kitchen utensils and office supplies, organizing them into the kitchen or home office areas based on its understanding of the environment and object functionality.

2) Adaptive Assistance in Dynamic Environments: In a scenario where a living space is temporarily repurposed (e.g., a dining table used as a work-from-home station), the Robo-CSK-Organizer could adapt its object placement strategy. It would recognize the shift in environmental context and adjust its actions accordingly, such as placing office-related items in the area temporarily designated as a workspace.

3) Collaborative Cooking and Cleaning Tasks: In a kitchen setting, the system could assist in cooking and cleaning tasks by recognizing ingredients, utensils, and cleaning tools. It could suggest where items should be stored after use or retrieved from, facilitating a more organized and efficient kitchen environment.

These examples highlight the system's potential to significantly improve the convenience and functionality of domestic environments, showcasing a future where robots seamlessly integrate into our living spaces, contributing to organization, efficiency, and adaptability.

VII. Conclusions and Perspectives

The Robo-CSK-Organizer represents a pioneering integration of CSK into AI-driven robotics applications, specifically tailored for multipurpse robotics. This integration is crucial for enhancing decision-making capabilities, particularly in object organization and classification tasks. Unlike traditional AI systems, which may rely solely on pre-training via huge datasets for object recognition and categorization, Robo-CSK-Organizer leverages human-like commonsense reasoning by harnessing ConceptNet, a comprehensive KB that encompasses a wide array of logically derived commonsense knowledge. This utilization marks a stark departure from conventional methods, enabling the system to navigate and interpret complex, nuanced contexts inherent in human environments.

This work proposes, demonstrates, and discusses the Robo-CSK-Organizer system that addresses various perspectives in AI and Robotics, with respect to novelty, scalability, collaborative efforts and more, along with future pathways. These highlights include:

Novelty of the Approach: This innovative application of CSK through ConceptNet has proven to be instrumental in addressing key challenges in AI, such as ambiguity resolution and explainability. By dynamically applying CSK, the Robo-CSK-Organizer system not only enhances the accuracy of object classification but also significantly improves the ability to explain decision-making processes to users. These capabilities underscore the system's contribution to the field of XAI, where transparency and comprehensibility of AI decisions are paramount. These contributions are foundational to advancing the practical application of CSK in AI, particularly within the domain of domestic robotics. By providing empirical evidence of the efficacy and benefits of integrating CSK through ConceptNet, this work lays the groundwork for future innovations in the field, emphasizing the importance of clarity, trust, and user-centric design in AI systems.

Challenges in Scalability and Adaptability: Scaling the application of CSK for broader AI and robotics solutions presents notable challenges. One such challenge is maintaining the system's performance and adaptability as the complexity of tasks and environments increases. Strategies to overcome these challenges include developing more dynamic CSK frameworks that can evolve with new information and experiences, enhancing the system's ability to generalize across different scenarios. Furthermore, optimizing computational efficiency to support the processing of extensive CSK databases without compromising system responsiveness or accuracy is crucial.

Collaborative and Cross-Disciplinary Efforts: The advancement of CSK in robotics will significantly benefit from collaborative and cross-disciplinary efforts. Engaging experts from cognitive science, linguistics, computer science, and robotics can foster innovative approaches to integrating CSK into AI systems. These collaborations can drive the development of new models and algorithms that better capture the complexity of human knowledge and reasoning, facilitating the creation of robotic systems that are more intuitive and effective in their interactions with humans and the world.

Addressing Technical and Ethical Considerations: Future endeavors will also navigate the technical limitations and ethical considerations inherent in deploying CSK-enhanced robots in real-world settings. Issues such as data privacy, security, and ethical use of robots in domestic environments, will require careful consideration and innovative solutions. Establishing guidelines and standards for the ethical development and deployment of these systems will be essential in ensuring they benefit society while respecting individual rights and privacy.

Limitations and Lessons: The main limitations of this research on Robo-CSK-Organizer are as follows: (1) It focused on data only in the domestic robotics domain for experimentation; (2) It only deployed only one KB for extracting commonsense knowledge; (3) It conducted comparative analysis only with a ChatGPT baseline. However, all these limitations can possibly be addressed in future extensions, e.g. application to other domains, utilization of multiple knowledge bases for CSK as needed, and comparison with other state-of-the-art systems that could be potentially relevant to object placement and task organization.

Reflecting on this journey, there are important lessons learned from the success story of Robo-CSK-Organizer, as well as its limitations that can provide the scope for further extensions. It is evident that the contributions of the Robo-CSK-Organizer, while somewhat modest, can lay the groundwork for significant, lasting impacts in the fields of AI and Robotics. This entire project has illuminated the path for future research, showcasing the need for and benefits of embedding human-like reasoning and understanding into robotic systems. It stands out as a pioneering effort towards creating more intuitive, transparent, and collaborative human-robot interactions. It paves the way for future work in AI-driven Robotics aiming to make robots closer to the thresholds of human cognition.

VIII. Future Work

This paper presents the development and evaluation of the Robo-CSK-Organizer system, focusing on its comparison to a ChatGPT-based baseline. While the presented results highlight significant advancements in explainability, adaptability, and task performance through the integration of commonsense knowledge (CSK), there are several opportunities for future work to further strengthen and expand this research:

Broader Comparative Analysis: In this work, the comparison was limited to a ChatGPT-based baseline as a proof-of-concept to evaluate the impact of integrating CSK via ConceptNet. Future work can potentially include comparisons with a broader spectrum of state-of-the-art methodologies for embodied intelligent grasping, such as reinforcement learning-based systems, other pre-trained large language models, computer vision models, and hybrid CSK-based models. These comparisons will provide deeper insights into the relative strengths and limitations of the Robo-CSK-Organizer.

Integration with Multiple Knowledge Bases: While this research leveraged ConceptNet as the primary knowledge base, future extensions could explore integrating other structured and unstructured CSK sources, such as ATOMIC, OMICS, Quasimodo, Dice, Ascent etc. to enhance the system's contextual reasoning and adaptability across diverse domains.

Application to Diverse Domains: The current experiments were conducted within the domain of domestic robotics. Future studies could extend the Robo-CSK-Organizer to other domains, such as healthcare, manufacturing, energy management, and disaster response, to assess its generalizability and adaptability in handling specialized tasks.

Scalability and Real-Time Performance: Scaling Robo-CSK-Organizer to handle more complex environments with larger datasets and real-time constraints is a vital next step. It includes optimizing computational efficiency to maintain performance while managing extensive CSKBs.

Enhanced User Interaction: Future iterations of Robo-CSK-Organizer can possibly focus on incorporating richer user feedback mechanisms. These enhancements are likely to enable the system to learn dynamically from human collaborators, improving its decision-making capabilities over time and fostering better human-robot collaboration.

Explainability Across Tasks: While the current work emphasizes explainability in object placement, future research will explore applying the system's explainability features to more complex robotic tasks, such as multi-step planning and autonomous decision-making in dynamic environments.

By addressing these research directions, future work aims to solidify the role of CSK in advancing robotic systems, enabling them to become more adaptable, reliable, and collaborative in increasingly complex environments. As we stand on the cusp of a new era in Robotics and AI, the Robo-CSK-Organizer system beckons the scientific community, industry stakeholders, and interdisciplinary teams to further the dialogue, research, and collaboration in this domain. The pursuit of more transparent, reliable, and human-like AI systems is not only a technological endeavor but also a philosophical and ethical one. It challenges us to re-imagine the boundaries of machine intelligence, pushing us towards innovations that are not just technologically advanced but also socially and ethically responsible.

Acknowledgments

This work is supported in part by the National Science Foundation under Grants CMMI-2338767, CMMI-2138351, and CNS-2117308.

References

[1] A. Dzedzickis, J. Subačiūtė-Žemaitienė, E. Šutinys, U. Samukaitė-Bubnienė, and V. Bučinskas, "Advanced applications of industrial robotics: New trends and possibilities," *Applied Sciences*, vol. 12, no. 1, p. 135, 2021.

[2] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating human–robot collaborative tasks by teaching-learning-collaboration from human demonstrations," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 640–653, 2018.

[3] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling object hand-over in human–robot collaboration via natural wearable sensing," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 1, pp. 59–71, 2019.

[4] A.-N. Sharkawy, "Human-robot interaction: Applications," in *Proceedings of the 1st IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 (ARCI' 2021)*, Chamonix-Mont-Blanc, France, 2021.

[5] G. Modery and W. Wang, "Assisting humans in human-robot collaborative assembly contexts through deep q-learning," in *2024 IEEE MIT Undergraduate Research Technology Conference (URTC)*. IEEE, 2024, pp. 1–6.

[6] C. J. Conti., A. S. Varde, and W. Wang, "Human-robot collaboration with commonsense reasoning in smart manufacturing contexts," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 1784–1797, 2022.

[7] C. Zednik, "Solving the black box problem: A normative framework for explainable artificial intelligence," 2019.

[8] V. Nastase, K. Filippova, and D. N. Milne, "Summarizing with encyclopedic knowledge." in *TAC*, 2009.

[9] E. Davis and G. Marcus, "Commonsense reasoning and commonsense knowledge in artificial intelligence," *Communications of the ACM*, vol. 58, no. 9, pp. 92–103, 2015.

[10] Y. Choi, "The curious case of commonsense intelligence," *Daedalus*, vol. 151, no. 2, pp. 139–155, 2022.

[11] R. Zellers, Y. Bisk, A. Farhadi, and Y. Choi, "From recognition to cognition: Visual commonsense reasoning," 2019.

[12] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," 2018.

[13] N. Tandon, A. S. Varde, and G. de Melo, "Commonsense knowledge in machine intelligence," *ACM SIGMOD Record*, vol. 46, no. 4, pp. 49–52, 2018.

[14] R. Hidalgo, J. Parron, A. S. Varde, and W. Wang, "Robo-csk-organizer: Commonsense knowledge to organize detected objects for multipurpose robots," in *International IOT, Electronics and Mechatronics Conference*. Springer, 2024, pp. 65–81.

[15] S. Razniewski, N. Tandon, and A. S. Varde, "Information to wisdom: Commonsense knowledge extraction and compilation," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 1143–1146.

[16] S. Roychoudhury, A. S. Varde, and W. Wang, "Efficient task organization with commonsense knowledge for human-robot collaborative tasks," *IEEE MIT URTC*, 2024.

[17] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," *Cognitive Robotics*, vol. 3, pp. 54–70, 2023.

[18] C. Hannum, R. Li, and W. Wang, "A trust-assist framework for human–robot co-carry tasks," *Robotics*, vol. 12, no. 2, p. 30, 2023.

[19] W. Wang, R. Li, L. Guo, Z. M. Diekel, and Y. Jia, "Hands-free maneuvers of robotic vehicles via human intentions understanding using wearable sensing," *Journal of Robotics*, vol. 2018, no. 1, p. 4546094, 2018.

[20] R. Li, W. Wang, Y. Chen, S. Srinivasan, and V. N. Krovi, "An end-to-end fully automatic bay parking approach for autonomous vehicles," in *Dynamic Systems and Control Conference*, vol. 51906. American Society of Mechanical Engineers, 2018, p. V002T15A004.

[21] S. Roychoudhury and A. S. Varde, "Discern for generalizable robotic contexts," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 8819–8821.

[22] H. Diamantopoulos and W. Wang, "Accommodating and assisting human partners in human-robot collaborative tasks through emotion understanding," in *2021 12th International Conference on Mechanical and Aerospace Engineering (ICMAE)*. IEEE, 2021, pp. 523–528.

[23] W. Skulkittiyut, H. Lee, T. Ngo Lam, Q. Tran Minh, M. A. Baharudin, T. Fujioka, E. Kamioka, and M. Mizukawa, "Commonsense knowledge extraction for tidy-up robotic service in domestic environments," in *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2013.

[24] L. Corallo, G. Li, K. Reagan, A. Saxena, A. S. Varde, and B. Wilde, "A framework for german-english machine translation with GRU RNN," in *ACM EDBT/ICDT conference workshops*, vol. 3135, 2022.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[26] T. Nguyen, S. Razniewski, A. S. Varde, and G. Weikum, "Extracting cultural commonsense knowledge at scale," in *Proceedings of the ACM Web Conference*, 2023, pp. 1907–1917.

[27] I. Chernyavsky, A. S. Varde, and S. Razniewski, "Csk-detector: Commonsense in object detection," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 6609–6612.

[28] Y. Chalier, S. Razniewski, and G. Weikum, "Dice: A joint reasoning framework for multi-faceted commonsense knowledge," in *International Workshop on the Semantic Web*, 2020.

[29] L. Kunze, M. Tenorth, and M. Beetz, "Putting people's common sense into knowledge bases of household robots," in *KI 2010: Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2010, pp. 151–159.

[30] C. J. Conti, A. S. Varde, and W. Wang, "Robot action planning by commonsense knowledge in human-robot collaborative tasks," *2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, pp. 1–7, 2020.

[31] G. Cui, W. Shuai, and X. Chen, "Semantic task planning for service robots in open worlds," *Future Internet*, vol. 13, no. 2, 2021.

[32] M. Sridharan and T. Mota, "Towards combining commonsense reasoning and knowledge acquisition to guide deep learning," vol. 37, no. 1, 2022.

[33] C. J. Conti, A. S. Varde, and W. Wang, "Web perspectives in robotics applications: Commonsense knowledge, autonomous vehicles and human-robot collaboration," *ACM SIGWEB*, vol. 2023, no. Winter, pp. 4:1–4:22, 2023.

[34] P. Persaud, A. S. Varde, and W. Wang, "Can robots get some human rights? a cross-disciplinary discussion," *Journal of Robotics*, vol. 2021, no. 1, p. 5461703, 2021.

[35] A. Garg, N. Tandon, and A. S. Varde, "I am guessing you can't recognize this: Generating adversarial images for object detection using spatial commonsense," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 10, 2020, pp. 13 789–13 790.

[36] P. Persaud, A. S. Varde, and S. Robila, "Enhancing autonomous vehicles with commonsense: Smart mobility in smart cities," in *2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI)*. IEEE, 2017, pp. 1008–1012.

[37] A. Varde, E. Rundensteiner, G. Javidi, E. Sheybani, and J. Liang, "Learning the relative importance of features in image data," in *2007 IEEE 23rd international conference on data engineering workshop*. IEEE, 2007, pp. 237–244.

[38] A. Simkute, A. Surana, E. Luger, M. Evans, and R. Jones, "Xai for learning: Narrowing down the digital divide between "new" and "old" experts," in *Adjunct proceedings of the 2022 Nordic human-computer interaction conference*, 2022, pp. 1–6.

[39] B. Austin-Gabriel, A. S. Varde, and H. Liu, "Geoinformatics-guided machine learning for power plant classification," 2025. [Online]. Available: https://arxiv.org/abs/2502.01039

[40] Y. Wu, A. Kirillov, F. Massa, W. Lo, and R. Girshick, "Detectron2," Online, 2019, available: https://github.com/facebookresearch/detectron2, Accessed: 2023-08-01.

[41] J. Li, D. Li, C. Xiong, and S. Hoi, "Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation," 2022.

[42] R. Hidalgo, N. Salah, R. Chandra Jetty, A. Jetty, and A. S. Varde, "Personalizing text-to-image diffusion models by fine-tuning classification for ai applications," in *Intelligent Systems Conference*. Springer, 2023, pp. 642–658.

[43] S. Bier, R. Li, and W. Wang, "A full-dimensional robot teleoperation platform," in *2020 11th International Conference on Mechanical and Aerospace Engineering (ICMAE)*. IEEE, 2020, pp. 186–191.

[44] J. Parron, T. T. Nguyen, and W. Wang, "Development of a multimodal trust database in human-robot collaborative contexts," in *2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2023, pp. 0601–0605.

[45] P. Tilloo, J. Parron, O. Obidat, M. Zhu, and W. Wang, "A pomdp-based robot-human trust model for human-robot collaboration," in *2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE, 2022, pp. 1009–1014.

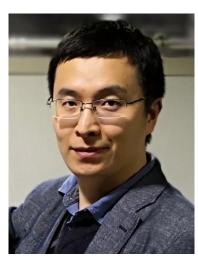
[46] V. Kalvakurthi, A. S. Varde, and J. Jenq, "Hey dona! can you help me with student course registration?" *AAAI Conference, Workshop on AI for Education, arXiv:2303.13548*, 2023.

[47] "Implementation of conceptnet for organizing objects to contexts." [Online]. Available: https://github.com/omnidox/ConceptNet.

Rafael Hidalgo was a Research Assistant in Computer Science at Montclair State University. He completed his MS in Computer Science at MSU and is currently a Data Scientist and Adjunct Professor at MSU. He has 4 publications in addition to his MS thesis. His research interests include data science and analytics, machine learning, robotics, and technological innovation.

Aparna S. Varde (Member, IEEE) received a PhD in CS from WPI, Worcester, MA, USA. She is an Associate Professor of the School of Computing and Associate Director of the Clean Energy and Sustainability Analytics Center (CESAC) at Montclair State University. She has been a Visiting Researcher at Max Planck Institute for Informatics, Germany. She has been the advisor/committee member for nine Ph.D. students at Montclair, and refereed four external ones. Her research entails multidisciplinary AI, yielding 150+ publications with projects spanning commonsense knowledge, smart cities, robotics, geoinformatics, and text mining. She has ten best paper awards at conferences by IEEE, Springer etc. and four honorable mentions, e.g. Spotlight Paper in WWW-2023.

Jesse Parron (Member, IEEE) is currently a research associate in the Collaborative Robotics and Smart Systems Laboratory and instructor in the School of Computing, Montclair State University. His research interests include collaborative robotics, machine learning, Artificial Intelligence, and sensor technologies.



Weitian Wang (Senior Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China in 2016. He worked as a Post-Doctoral Fellow at Clemson University. He is currently an Associate Professor in the School of Computing and the Founder Director of the Collaborative Robotics and Smart Systems Laboratory (CRoSS Lab) at Montclair State University. His research focuses on collaborative robotics, artificial intelligence, smart systems, and their synergistic CRoSS-disciplinary applications in smart manufacturing and remanufacturing, healthcare, smart ecosystem, intelligent transportation, smart agriculture, daily assistance, data science, cybersecurity, and interactive learning. He is the recipient of IEEE R1 Technological Innovation Award.