A Taxonomy of Construction Task Activities for Robot Workers

Sadman Sakib[1], Zhangyi Peng[1,3], Yujie Pang[1,3], Yu Otsuki[2], and Mohammad Abdullah Al Faruque[1]

arXiv:2608.25395v3 [cs.RO] 24 Sep 2026

***Abstract*— Recent vision–language–action models offer a path toward robots with broader capabilities than conventional task-specific systems. Deploying such systems in construction, however, requires a precise inventory of worker activities and the capabilities needed to execute them. We present TARCAT, an occupation-grounded taxonomy derived from 91 O*NET task statements across seven high-employment construction occupations and 30 instructional videos. TARCAT defines 41 primitives spanning intellectual, social, and physical categories and provides a mechanism for composing parameterized primitive sequences into reusable skills. This human-interpretable vocabulary supports the specification of robot requirements and enables coding agents to retrieve and extend skill libraries for task execution. We also demonstrate selected primitives on a DOBOT CR3 arm with a CRAFT hand. TARCAT thus provides a common vocabulary for analyzing construction work and developing general-purpose construction robots. Annotations are available at `https://github.com/AICPS/TARCAT-Taxonomy`.**

I. INTRODUCTION

Construction robotics has traditionally focused on purpose-built systems designed for specific tasks, such as ceiling installation, wall painting, and bricklaying [1]. Meanwhile, current construction workflows are largely organized around human workers, who perform diverse activities involving different tools, actions, and interactions. As a result, general-purpose robotic systems capable of performing a broader range of construction activities could better assist human workers.

Recent advances in robotics, such as vision-language-action (VLA) models, are creating new opportunities for general-purpose robots capable of performing a broader range of tasks. Emerging agentic robotic systems, such as ASPIRE [2], can also autonomously generate control programs, retain successful procedures in skill libraries, and retrieve them for new tasks. However, developing such systems still requires a systematic understanding of what tasks robots need to perform, what physical actions and skills those tasks require, and how these low-level capabilities can be composed into higher-level activities.

Structured action taxonomies provide this interface between task activities and robot learning [3]. A manipulation vocabulary developed in [4] includes 32 cooking motions characterized by contact, trajectory, duration, and manual-operation attributes. AVA [5] annotates 80 atomic actions in movies. In industrial and manufacturing robotics, manipulation work has been classified into reusable tasks, skills, and actions [6], [7]. TARMAC [8] decomposes chemistry procedures into reusable robot manipulation actions.

In the construction industry, taxonomy development for robotic tasks is still emerging. Everett and Slocum [9] organized a set of 12 basic tasks that describe construction field work. Built on these basic tasks, Li and Leicht [10] recently analyzed O*NET [11] task statements across 15 construction occupations to identify task verbs, objectives, tools, and materials, but the analysis does not consider physical robot embodiments. At the human-motion level, Rodrigues et al. [12] classified human–robot interaction according to platform, autonomy, interaction, and team structure, while the study does not connect occupation-grounded construction tasks to robotic embodiments and reusable robot actions.

We propose TARCAT (Taxonomy of Construction Task Activities for Robot Workers) to address these gaps. TARCAT provides an occupation-grounded vocabulary of construction tasks, their composition into reusable robotic actions, and video annotations (Fig. 1). TARCAT contributes (1) 41 primitives spanning intellectual, social, and physical categories; (2) annotations of 30 videos and labels covering 91 O*NET tasks across seven high-employment construction occupations; and (3) hardware demonstrations on a collaborative robotic arm with a dexterous hand.

II. OCCUPATIONAL DATA AND INSTRUCTIONAL VIDEO COLLECTION

Our taxonomy framework is grounded in two U.S. Department of Labor databases. A recent study, GDPval [13], used BLS employment data and O*NET task statements to identify economically significant knowledge-work tasks. However, GDPval did not consider construction occupations or tasks involving physical movement. TARCAT adopts a similar grounding strategy to identify economically significant construction occupations rather than relying on a manually selected set of construction tasks.

The BLS Occupational Outlook Handbook [14] reports 2024 employment data for Construction and Extraction occupations. Based on the BLS data, we selected the seven construction occupations with the largest number of employed workers. O*NET [11] is another database that provides SOC-aligned task statements for each occupation. Using the BLS-to-SOC occupation mapping provided by the BLS [14], we selected the most relevant O*NET SOC occupation for each of the seven BLS construction occupations. We identified 169 O*NET task statements corresponding to these seven O*NET SOC occupations: (i) Construction Laborers, (ii)

[1,2]Center for Resilient Autonomous Systems, University of California, Irvine (UCI), Irvine, CA, USA.

[1]Nhu Department of Electrical Engineering and Computer Science. {ssakib,znpeng,yujiep2,alfaruqu}@uci.edu

[2]Department of Civil and Environmental Engineering. yu.otsuki@uci.edu

[3] The authors contributed during the UCInspire program.

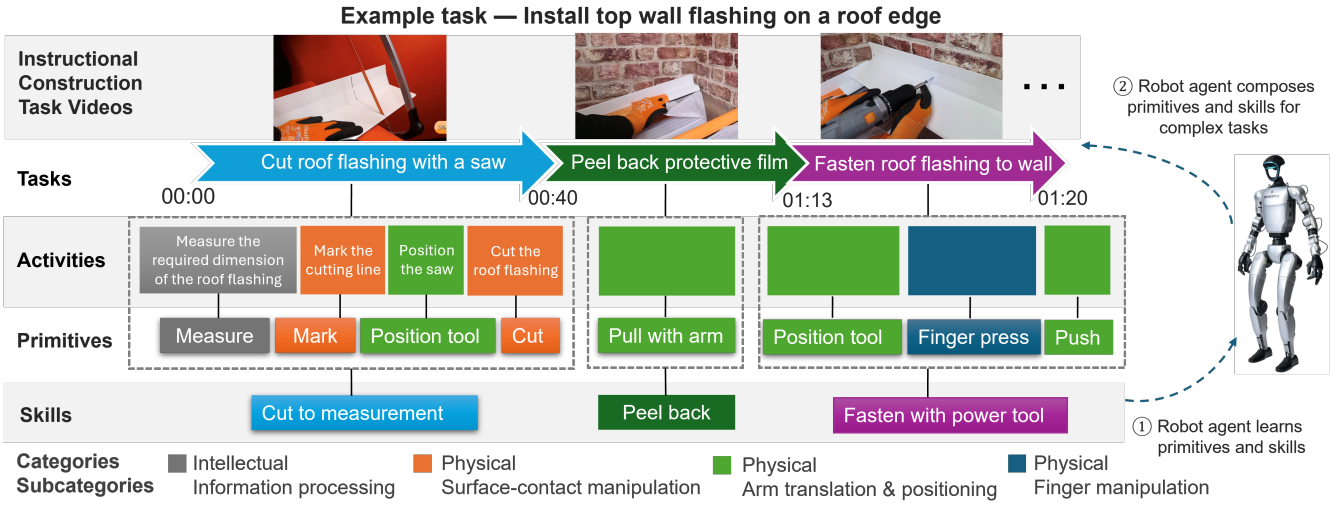


Fig. 1. **Overview of TARCAT.** Instructional construction videos are segmented into observable activities, labeled with reusable primitives, and organized into skills; activity and primitive colors indicate the category and subcategory in Table I, and each skill's color matches the task segment it accomplishes. Using the developed taxonomy and annotations, robot agents can (i) learn primitives and skills and (ii) compose them to execute complex tasks.

Carpenters, (iii) Electricians, (iv) Plumbers, Pipefitters, and Steamfitters, (v) Painters, Construction and Maintenance, (vi) Brickmasons and Blockmasons, and (vii) Roofers. Note that the heavy machinery operator occupation was excluded because it falls outside the scope of this study.

We classified 137 of the 169 O*NET task statements as involving physical movement and the remaining 32 as non-movement tasks. Using ChatGPT with GPT-5.6 [15], we searched YouTube for instructional demonstrations of the movement tasks. We manually retained 30 videos with observable, task-relevant actions and mapped each video to every task statement it demonstrated. These videos cover 59 movement tasks. We annotated all 32 non-movement tasks directly from their task statements, yielding coverage of 91 O*NET tasks across the seven occupations.

III. TAXONOMY OF CONSTRUCTION TASKS

A. Task and Activity

This section clarifies the terminology used in our taxonomy. First, O*NET defines task statements as "*task statements are exemplars of the types of activities workers in an occupation may perform and are typically conceptualized as the smallest unit of activity with a meaningful outcome*"[11]. Hence, O*NET uses *activity* to broadly describe work performed by a human worker, not as a level below a task.

While O*NET task statements are adopted in this study, we use the term *activity* in a more specific sense to denote a finer-grained unit of task execution. Accordingly, we define the relationship between a *task* and an *activity* as follows: A task is an externally specified unit comprising an activity or a sequence of activities to achieve a goal. This distinction in the definition of activity is necessary because O*NET task statements, while sufficiently detailed for human occupational analysis, may remain too coarse for robot execution and therefore require further decomposition. Note that a random sequence of activities without an intended goal does not constitute a task.

We further introduce a nested task structure, in which a task can be decomposed into smaller tasks, each with its own goal. The level of decomposition depends on the agent's capabilities. For example, O*NET task statements suffice for human workers, whereas robot agents may require smaller, short-horizon tasks compatible with their capabilities.

As an example, consider the task shown in Fig. 1, "Install top wall flashing on a roof edge", which is decomposed into three smaller tasks. The first task, "Cut roof flashing with a saw" shown in Fig. 1, can be decomposed into four activities: (i) measure the required dimension of the roof flashing, (ii) mark the cutting line, (iii) position the saw, and (iv) cut the roof flashing. Together, these activities achieve the goal of the first task: the roof flashing is cut to the required dimension.

B. Category and Subcategory

We classify the activities associated with the 91 O*NET tasks into three categories: (1) Intellectual, (2) Social, and (3) Physical, following the worker-task taxonomy proposed by [16]. The taxonomy [16] further divides each category into subcategories. We retain all subcategories defined for (1) Intellectual and adopt the Coordinating subcategory from (2) Social as relevant to construction tasks. The definitions of these categories and subcategories follow [16].

For the (3) Physical category, the study [16] defines three subcategories: strength, dexterity, and navigation based on the physical capability required of workers. To expose finer robot-action requirements, TARCAT instead distinguishes the body segment primarily involved and the form of motion or contact, thereby defining a new set of fine-grained subcategories under the Physical category as follows:

- **Finger manipulation:** Manipulate primarily through finger motion or force, with little gross arm movement.
- **Arm translation and positioning:** Grab or move an object through coordinated hand and arm motion.
- **Arm and wrist rotation:** Rotate or reorient the arm, wrist, tool, or object to apply force.

TABLE I

TARCAT PRIMITIVES ORGANIZED BY CATEGORY AND SUBCATEGORY, WITH EXAMPLE ACTIVITIES.

Category	Subcategory	Primitives	Example Activities
Intellectual	Information processing	Take measurements	Read pressure gauge; measure roof width
		Study plan, instruction or specifications	Read compressor label; check breaker rating
		Listen to or observe instructions or signals	Follow spoken directions; observe crane signals
		Prepare records, documents, and reports	Write project report; record inspection result
	Problem solving	Determine tool-application location	Locate drill point; locate nail-gun cap
		Inspect to check state or find issues	Inspect roof damage; check ladder locks
		Estimate time, materials, labor and cost	Estimate roofing labor; calculate mortar quantity
		Select and procure tools and materials	Order lumber; choose paint finish
		Prepare design, layout or installation plans	Plan wiring layout; sketch equipment locations
Social	Coordinating	Talk to convey information	Discuss work plan; explain safety issue
		Use gestures to signal other workers	Signal crane operator; direct team lift
		Schedule and coordinate work with others	Schedule work crew; coordinate subcontractors
Physical	Finger manipulation	Pull or push an object with fingers	Open nail-gun cap; connect air hose
		Press an object with fingers	Press laser button; squeeze nail-gun trigger
		Open fingers	Release plumb bob; release stud finder
		Rotate an object with fingertips	Adjust pressure valve; tighten spray-can straw
	Arm translation and positioning	Grab an object	Catch plumb-bob string; grab panel cover
		Pull or push an object with arm	Open cabinet door; push breaker into slot
		Pick or place an object	Pick up trowel; place electrical box
		Position an object	Align roof flashing; position screwdriver
		Move tool along a linear path	Apply joint sealant; dip brush into paint
	Arm and wrist rotation	Twist an object using wrist rotation	Open cement bottle; twist wire with pliers
		Pivot a tool about a fixed point	Tighten bolt with wrench; turn valve key
		Swing arm while holding an object	Hammer a nail; chip mortar with chisel
		Reorient an object	Flip shingle; turn brick over
		Pour liquid from a container	Pour paint into tray; pour water into bucket
		Scoop material with a tool	Scoop mortar with trowel; scoop soil with shovel
	Mixing motion	Shake an object	Shake spray can; shake mixing container
		Stir liquid and solid mixture	Mix mortar with trowel; stir paint with stick
		Wrap around, tie or untie	Tie rope around ladder; untie plumb-bob knot
	Surface-contact manipulation	Mark on a surface	Mark stud line; mark screw point
		Cut an object	Saw wooden stud; slice roofing shingle
		Spread object on surface	Spread mortar; smooth joint compound
		Stroke tool on surface	Brush paint on wall; roll paint on ceiling
	Whole-body locomotion	Walk to a location	Walk to inspection area; approach electric outlet
		Carry an object to a location	Carry measuring rod; carry roll of felt
		Climb up or down	Climb ladder; descend scaffold
	Whole-body load handling	Pull or push a heavy object	Align wall cabinet; slide loaded pallet
		Pick or place a heavy object	Pick up ladder; place wall cabinet
		Carry a heavy object	Carry large pipe; transport ladder
		Bend down or straighten up	Pick up leaf blower from ground; lift drill and stand

- **Mixing motion:** Use repeated, circular, or wrapping motions to mix or secure materials.
- **Surface-contact manipulation:** Move a tool while maintaining controlled contact with a work surface.
- **Whole-body locomotion:** Use the legs and torso to move between locations or transport an object.
- **Whole-body load handling:** Use whole-body force and postural control to manipulate heavy objects or bend and straighten.

Table I summarizes the classification of example activities derived from the O*NET task statements into three categories and their corresponding subcategories.

C. Activity, Primitive, and Parameter

This section introduces the terms *primitive* and *parameter* and relates them to *activity*. A *primitive* is the smallest atomic unit, such as a physical action, perception, or cognitive process, that performs a distinct operation and can be reused across contexts. *Parameters* configure a primitive for a specific activity or task and may specify objects, tools, targets, or other context-dependent information. We then define an *activity* as an instantiation of a primitive with context-specific parameters. For example, the activity "Read pressure gauge" can be represented by the primitive "Take measurements" with "pressure gauge" bound to the `to_measure` parameter.

Table I lists the 41 unique primitives identified from activities in the covered O*NET task statements and instructional videos. A human annotator labeled time-stamped video segments with the corresponding primitives. Figure 2 summarizes the resulting distribution of 643 primitive labels from the annotated videos across the subcategories. "Arm translation and positioning" is the most frequent subcategory (278 occurrences), followed by "Problem solving" (114).

D. Skill

For the application of this taxonomy, we then define a *skill* as follows: A skill is a unit of agent capability, acquired

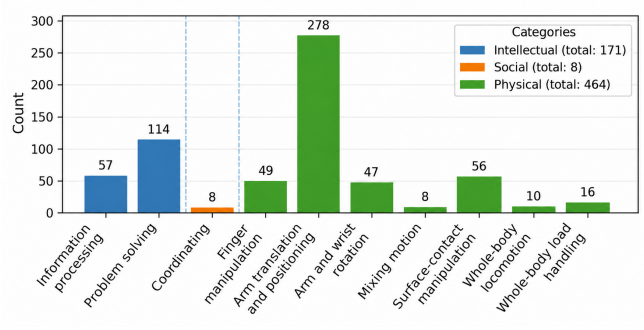


Fig. 2. **Distribution of primitive occurrences across labeled O*NET tasks and instructional videos.** Bars denote sub-categories, and colors denote their parent categories.

through learning or programming, to execute an activity or a sequence of activities to achieve a goal. While the definitions of task and skill are similar in that both involve achieving a goal through an activity or sequence of activities, the key distinction is that a task is externally specified to an agent, whereas a skill represents the agent's own capability. Therefore, a skill solves a task. In a scheduling framework, an agent can be selected for a given task based on whether it possesses the skill required to perform that task.

As an example of a skill, consider a robotic agent that has mastered the skill "Fasten an object with a power tool", which comprises four activities with associated primitives and parameters: (1) Position an object (`to_position` = fixing hardware), (2) Position an object (`to_position` = power tool), (3) Press an object with fingers (`to_press` = power-tool trigger), and (4) Pull or push an object with arm (`to_pull_push` = power tool).

Similar to the nested task structure, skills can also be recursively composed, allowing lower-level skills to be combined into higher-level skills.

IV. PHYSICAL DEMONSTRATION

To assess executability, we selected four activities from the Physical category: positioning a power screwdriver, placing a hammer in a box, hammering a nail, and brushing objects aside. We mapped each activity to its TARCAT primitive and implemented scripted demonstrations on a six-axis DOBOT CR3 arm [17] with a 15-actuator, tendon-driven CRAFT hand [18]. Figure 3 shows two representative executions. Collectively, the demonstrations cover arm translation and positioning, arm and wrist rotation, and surface-contact manipulation, providing initial evidence that these taxonomy elements can be instantiated as executable physical behaviors. The demonstration video is available at `https://youtu.be/Dc-bcWGwTLc`.

Future work will expand the corpus to evaluate long-horizon skills and develop data-centric methods for classifying primitives [19].

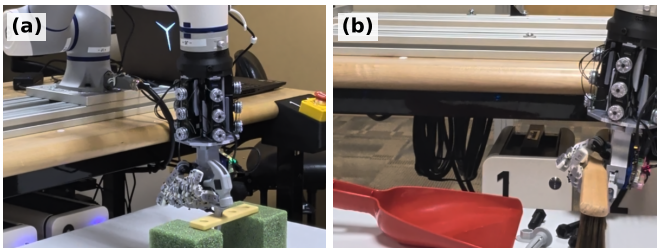


Fig. 3. **Physical demonstrations of TARCAT primitives.** (a) Hammering a nail with a primitive "*Swing arm while holding an object*". (b) Brushing objects toward a dustpan with a primitive "*Stroke tool on surface*".

REFERENCES

[1] K. S. Saidi, T. Bock, and C. Georgoulas, "Robotics in construction," in *Springer handbook of robotics*. Springer, 2016, pp. 1493–1520.

[2] R. Lu, Y. Wu, E. Kou, L. Fu, W. Xiao, A. Mandlekar, Y. Xu, G. Shi, K. Goldberg, A. Chen *et al.*, "Aspire: Agentic/skills discovery for robotics," *arXiv preprint arXiv:2607.00272*, 2026.

[3] M. Pantano, T. Eiband, and D. Lee, "Capability-based frameworks for industrial robot skills: a survey," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2022, pp. 2355–2362.

[4] D. Paulius, Y. Huang, J. Meloncon, and Y. Sun, "Manipulation motion taxonomy and coding for robots," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 5596–5601.

[5] C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. Sukthankar, C. Schmid, and J. Malik, "AVA: A video dataset of spatio-temporally localized atomic visual actions," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6047–6056.

[6] R. K.-J. Lee, H. Zheng, and Y. Lu, "Human-robot shared assembly taxonomy: A step toward seamless human-robot knowledge transfer," *Robotics and Computer-Integrated Manufacturing*, vol. 86, p. 102686, 2024.

[7] L. Johannsmeier, S. Schneider, Y. Li, E. Burdet, and S. Haddadin, "A process-centric manipulation taxonomy for the organization, classification and synthesis of tactile robot skills," *Nature Machine Intelligence*, vol. 7, no. 6, pp. 916–927, 2025.

[8] K. Huang, J. Pipe, A. E. Martin, T. Wang, B. A. Franklin, A. M. Tyrrell, I. J. S. Fairlamb, and J. Zhu, "TARMAC: A taxonomy for robot manipulation in chemistry," *arXiv preprint arXiv:2510.19289*, 2025.

[9] J. G. Everett and A. H. Slocum, "Automation and robotics opportunities: Construction versus manufacturing," *Journal of Construction Engineering and Management*, vol. 120, no. 2, pp. 443–452, 1994.

[10] F. Li and R. M. Leicht, "Exploring construction tasks that can be performed by robots," in *Construction Research Congress 2026*. American Society of Civil Engineers, Aug. 2026, pp. 120–129.

[11] National Center for O*NET Development, "O*NET database," 2026, accessed: 2026-08-20. [Online]. Available: https://www.onetcenter.org/database.html

[12] P. Rodrigues, R. Singh, M. Oytun, P. Adami, P. Woods, B. Becerik-Gerber, L. Soibelman, Y. Copur-Gencturk, and G. Lucas, "A multidimensional taxonomy for human–robot interaction in construction," *Automation in Construction*, vol. 150, p. 104845, 2023.

[13] OpenAI, "GDPval: Evaluating AI model performance on real-world tasks," OpenAI, Tech. Rep., 2025. [Online]. Available: https://openai.com/index/gdpval/

[14] U.S. Bureau of Labor Statistics, "Construction and extraction occupations," Occupational Outlook Handbook, 2025, accessed: 2026-08-20. [Online]. Available: https://www.bls.gov/ooh/construction-and-extraction/home.htm

[15] OpenAI, "GPT-5.6: Frontier intelligence that scales with your ambition," Jul. 2026, accessed: 2026-08-23. [Online]. Available: https://openai.com/index/gpt-5-6/

[16] E. Fernández-Macías and M. Bisello, "A comprehensive taxonomy of tasks for assessing the impact of new technologies on work," *Social Indicators Research*, vol. 159, no. 2, pp. 821–841, 2022.

[17] Dobot, "DOBOT CR3 specifications," 2026, accessed: 2026-08-20. [Online]. Available: https://www.dobot.us/cr3-specs/

[18] L. Lin, S. Patel, J. Moon, S. Lazebnik, and U. Jain, "CRAFT: A tendon-driven hand with hybrid hard–soft compliance," *arXiv preprint arXiv:2603.12120*, 2026.

[19] D. Liconti, Y. Zhou, Y. Toshimitsu, R. Hinchet, and R. K. Katzschmann, "A benchmark of dexterity for anthropomorphic robotic hands," *arXiv preprint arXiv:2604.09294*, 2026.