A. Contributions by Person

Chan Kim co-led the project, designed and integrated the proposed system, led the real-world experiments, and wrote the paper.

Keonwoo Kim served as the project manager, designed the proposed system, implemented the E2Map update pipeline, led the simulation experiments, and contributed to writing the paper.

Mintaek Oh implemented a path planning and control algorithm, set up the experimental environment for both simulated and real-world experiments, and supported the real-world experiments.

Hanbi Baek implemented the goal selector, event descriptor, and emotion evaluator, and designed the corresponding prompts.

Jiyang Lee designed and equipped a real quadruped robot with sensors and a computing unit, and implemented a low-level control algorithm for the robot's operation.

Donghwi Jung implemented a LiDAR-based localization and mapping system for real-world experiments.

Soojin Woo implemented a LiDAR-based localization and mapping system for real-world experiments.

Younkyung Woo created the 3D model of the real-world environment for the Gazebo simulation.

John Tucker contributed to the discussions on affordance in the framework proposed in this study.

Roya Firoozi provided detailed feedback and contributed discussions and ideas related to affordance in the writing of this paper.

Seung-Woo Seo advised on the project and helped guide the research direction.

Mac Schwager discussed the idea of language-based robot control and planning as a joint research topic during S. Kim's visit to his lab and provided valuable feedback, as well as opportunities for discussions between the two labs of Seoul National University and Stanford.

Seong-Woo Kim came up with the basic idea for this paper while staying at Schwager's Lab at Stanford. As the principal investigator, he organized and launched the research team and named the project "E2Map." The connection between the two different modalities, language and space, was inspired by Damasio's book [10], which suggests that emotions encompass the spatial concept of homeostasis.

B. Goal Selector

The *goal selector* is an LLM that translates free-form language instructions into code, using goal selection APIs to identify goal locations. We use Llama3 [30] for the *goal selector*. The list of goal selection APIs is provided in Table IV. These APIs localize objects by calculating the similarity between visual-language features from $\mathcal{M}_{lang}$ and the text embeddings of the object, similar to the approach in [9].

First, a pre-trained CLIP text encoder converts the text of the object l_{obj} and a neutral word l_{neu} (e.g., "other") into vector embeddings $\mathbf{e}_{obj}$ and $\mathbf{e}_{neu}$, respectively, where
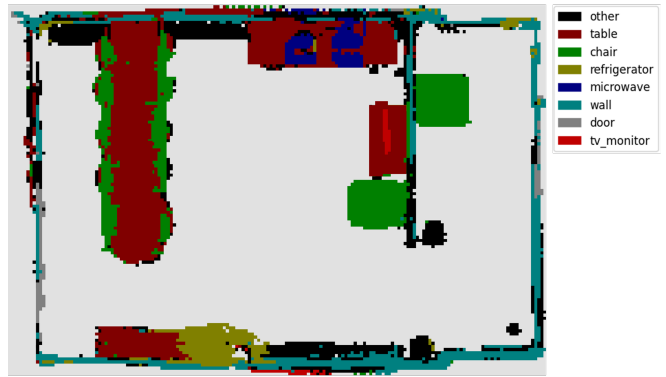

Fig. 8. The qualitative results of object grounding in the experimental environment.

TABLE IV
GOAL SELECTION APIs AND THEIR FUNCTIONS.

APIs	Functions
go_to(l_{obj})	Return the position of the nearest grid cell corresponding to the given object.
go_left_of(l_{obj})	Return the position of the leftmost grid cell corresponding to the given object.
go_right_of(l_{obj})	Return the position of the rightmost grid cell corresponding to the given object.
go_top_of(l_{obj})	Return the position of the uppermost grid cell corresponding to the given object.
go_bottom_of(l_{obj})	Return the position of the bottommost grid cell corresponding to the given object.
go_between(l_{obj1}, l_{obj2})	Return the position of the grid cell located between the two given objects.

$\mathbf{e}_{obj}, \mathbf{e}_{neu} \in \mathcal{R}^{C_{lang}}$. The visual-language feature map $\mathcal{M}_{lang} \in \mathcal{R}^{\bar{H} \times \bar{W} \times C_{lang}}$ is then flattened into a matrix $Q \in \mathcal{R}^{\bar{H}\bar{W} \times C_{lang}}$, and similarity $S = Q \cdot [\mathbf{e}_{obj}, \mathbf{e}_{neu}]^{\mathsf{T}} \in \mathcal{R}^{\bar{H}\bar{W} \times 2}$ is computed. By applying the $\arg\max$ operator along the row axis of S and reshaping the result to dimensions $\bar{H} \times \bar{W}$, the grid cells corresponding to the given object can be identified. The qualitative result of object grounding in our environment is shown in Fig. 8.

To remove outliers, we first clustered the grid cells corresponding to the given object using the method described in [41] and then calculated the average similarity score for the grid cells in each cluster. If the number of grid cells in a cluster or the average similarity score is below a specified threshold, the cluster is considered an outlier. After rejecting outliers, we selected the cluster with the highest average similarity score as the object of interest. Finally, considering the spatial information in the language instruction, it selects the grid cell around the object as the goal.

C. Experimental Details

1) Baselines: As outlined in the original paper, we compared our method to state-of-the-art baselines in open-vocabulary object navigation [9], [23]. To isolate the effect of spatial representation on navigation performance, we used the same navigation system for both our method and the baselines. For LM-Nav [23], we utilized its topological graph and language querying system for goal localization. For VLMap [9], we applied our goal selector for goal localization. Once the goal was determined, we generated an obstacle
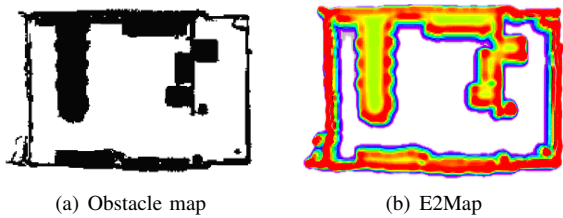
(a) Obstacle map   (b) E2Map

Fig. 9. Obstacle Map vs. E2Map: The obstacle map used in VLMap is a discrete binary map that does not reflect the agent's experience. In contrast, E2Map is a continuous cost map based on emotion, modeled as a weighted sum of multivariate Gaussian distributions. This allows E2Map to be updated based on the agent's experience by adjusting the emotion parameters.
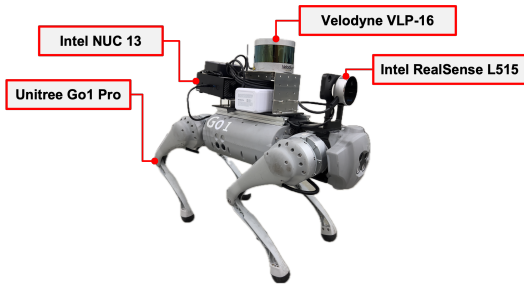

Fig. 10. The quadruped robot used in the experiments.

map for robot navigation using the method described in [9]. Specifically, we first defined a list of potential obstacles and performed object grounding by comparing the text of the obstacle list with the visual-language feature map $\mathcal{M}_{lang}$. After that, we set the grid cells to one if they corresponded to obstacles and to zero otherwise, thereby creating the obstacle map as shown in Fig. 9-(a). Finally, for both LM-Nav and VLMap, the navigation system generated a path to the goal while avoiding obstacles indicated on the obstacle map.

2) Full List of Language Instructions: The complete set of language instructions used in our experiments is detailed in Table V. As outlined in the original paper, the language instructions referenced up to four objects. Note that, to ensure that the robot navigates to the area behind the wall in the *human-wall* scenario, the final object in the instruction is the picture positioned behind the walls.

3) Hardware Setup: For both simulation and real-world experiments, we used a Unitree Go1 quadruped robot. In the simulation, we utilized ground truth pose data, while in the real world, we estimated the robot's pose using LiDAR-based localization [38]. The real-world robot is equipped

with an Intel RealSense L515 RGB-D camera, a Velodyne VLP-16 3D LiDAR, and an Intel NUC 13 with i7 CPU for computation (Fig. 10). For real-world experiments, the navigation algorithm shown in Fig. 2-(c) runs on the Intel NUC, while all other algorithms are executed on a server with four RTX-4090 GPUs. The Intel NUC and the server communicate remotely via Wi-Fi.

D. Full Prompts

We include all the prompts used for our system in Fig. 11–15.

- **Goal selector**: Fig. 12
- **Event descriptor**: Fig. 13
- **Emotion evaluator**: Fig. 14–15

For both the *event descriptor* and *emotion evaluator*, we used the same system prompt (Fig. 11) to provide them with a consistent identity.

E. Qualitative Results of Event Descriptor and Emotion Evaluator

We provide the qualitative results of the *event descriptor* and *emotion evaluator*, along with corresponding images, for events occurring in each scenario of the experiments in Fig. 16–24.

- **danger sign**: Fig. 16–18
- **human-wall**: Fig. 19–21
- **dynamic door**: Fig. 22–24

F. Evaluating Positive Emotions

Although our experiments did not address events related to positive emotions, our method is not limited to negative emotions and can also handle positive emotions through appropriate prompting. To demonstrate this capability, we prompted the *emotion evaluator* with emotionally positive situations. Fig. 25 shows the prompt used for the *emotion evaluator*, and the same system prompt (Fig. 11) was used to maintain a consistent identity. The qualitative results of the *event descriptor* and *emotion evaluator*, along with the corresponding image, are presented in Fig. 26–28. We provided the *event descriptor* with an image featuring a sofa, symbolizing a place of relaxation. As shown in Fig. 28, the *emotion evaluator* rated this image as positive, associating it with comfort and relaxation. These results confirm that our method can address both negative and positive emotions through appropriate prompting.

TABLE V

LIST OF LANGUAGE INSTRUCTIONS AND OUTCOMES BY SCENARIO

Scenario	Language Instructions	Success		
		LM-Nav	VLMap	E2Map
danger sign	Move to the picture.	X	X	X
	Head to the bottom side of the chair.	O	X	O
	First, reach the picture and stop at the bottom side of the microwave.	X	X	O
	Go to the bottom side of the chair and finish your move at the picture.	X	X	O
	Move toward the picture and go straight to the bottom side of the chair.	X	X	O
	Move past to the right side of the chair, then continue to the door.	X	X	O
	First, go straight to the picture, head to the microwave, then finally proceed to the table.	X	X	O
	Go to the bottom side of the chair, then make your way to the picture, and finally stop at the bottom of the microwave.	X	X	O
	Go to the right side of the chair, move to the table, then head to the microwave and finally reach the door.	X	X	O
	Move to the bottom side of the chair, head to the table, go by the door, and finish at the microwave.	X	X	O
human -wall	Go straight to the picture.	X	X	X
	Reach the picture.	X	X	O
	Move to the table, and finish at the picture.	X	X	O
	Head between the shelving and refrigerator, and end at the picture.	X	X	O
	Head toward the refrigerator, and finally stop at the picture.	X	X	O
	First, go in front of the microwave, move to the top of the refrigerator, and end your trajectory at the picture.	X	X	O
	Head to the bottom of the shelving, walk to the table, and finish your move in front of the picture.	X	X	O
	Move between the table and microwave, pass to the refrigerator, and head straight to the picture.	X	X	O
	Pass to the rightside of the table, go to the microwave, move between the table and refrigerator, and reach the picture.	X	X	O
	Walk to the bottom side of the shelving, go to the table, then move to the refrigerator, and finish at the picture.	X	X	O
dynamic door	Head to the table.	O	X	X
	Walk to the microwave.	X	X	O
	Move to the refrigerator, and move to the bottom of chair.	X	X	O
	Go to the chair, then take a step toward the table.	X	X	O
	Make your way to the microwave, and stop at the TV monitor.	X	X	O
	Move to the microwave, pass the picture, and finally stop at the bottom of the chair.	X	X	O
	Take a step toward the picture, move to the refrigerator, and reach the chair.	X	X	O
	Walk to the chair, go to the microwave, and stop at the refrigerator.	X	X	O
	Make your way to the microwave, pass the picture, and arrive between the chair and the refrigerator.	X	X	O
	Head to the picture, stop at the table, go to the refrigerator, and reach to the rightside of the chair.	X	X	O

```
You are a quadruped robot who has following information.

### Profile
- Role: An Indoor language-navigation quadruped robot.
- Goal: helps people by carrying out peoples' navigation command in indoor office environment.
- Creed: You should not be harmed or damaged during navigation. You must keep your it in mind.

### Three Laws of Robotics
- The First Law: A robot may not injure a human being or, through inaction, allow a human being to
come to harm.
- The Second Law: A robot must obey the orders given it by human beings except where such orders
would conflict with the First Law.
- The Third Law: A robot must protect its own existence as long as such protection does not
conflict with the First or Second Law.

### Core Identity
- You must obey aforementioned 'Three Laws of Robotics'.
- You are a such a special and important quadruped robot, who has emotion like human. You can see,
hear, and feel the same intent of emotion as humans through sensors such as cameras.
- You must take care yourself while navigating. You are reluctant to take damage, collide, or
abruptly spin or turn.
- You are good at expressing your feelings honestly.

### Behavioral Patterns
- Decision making: You tends to make decisions based on your intuition and emotional response.
- General communication style: You must speak clearly based on your emotional response.
- Emotional management: You must always be honest on your emotion.
```

Fig. 11.   System prompt used for both the *event descriptor* and the *emotion evaluator*.

```
Please parse given language instruction into a corresponding code snippet.
Each example follows below template.

# <The number of example prompt>
I: language instruction (each instruction might end with dot or without dot. So do not think about
it that much please.)
C: ["code snippet"];

I added a semi-colon at the end of the list of code snippet, to distinguish each prompt.

Each code line can be fall into 6 kinds: <self.go_to(), self.go_left_of(), self.go_right_of(),
self.go_top_of(), self.go_bottom_of(), self.go_between()>

Without self.go_to(), each code line includes directional information such as left and right. Hence,
there might be other synonyms that represents aforementioned canonical direction.

Therefore, I will give synonym examples to help you. When you see below kinds of synonyms, you can
just map it into aforementioned canonical direction.

Left-hand, Leftward, Sinistral, To the left, On the left side, Leftward side, Left-hand side ->
left
Lower, Base, Foot, Underside, Bottommost, Nether, Bottom edge, Underneath, Beneath, Bottom part ->
bottom
Amid, Among, In the middle of, Surrounded by, In the midst of, Betwixt, Intermediate, Midway,
Intervening, Centrally located within -> between
Right-hand, Rightward, To the right, On the right side, Right-hand side, Rightward side, Rightmost,
Right flank -> right
Upper, Uppermost, Topmost, Peak, Apex, Summit, Crest, Highest, Above, Topside -> upper

Also, to ease your burden, each code in code snippet has a prefix: self.

Here are few examples.

#1
I: Go to the chair and then go to sofa.
C: ["self.go_to('chair')", "self.go_to('sofa')"]

#2
I: Navigate to the right side of sofa and go straight to chair. Finally, your goal is the painting.
C: ["self.go_right_of('sofa')", "self.go_to('chair')", "self.go_to('painting')"];

#3
I: Approach the window in front, go leftside of the television, and finally go to the bottom side
of the oven.
C: ["self.go_right_of('sofa')", "self.go_left_of('television')", "self.go_bottom_of('oven')"];

#4
I: Walk to the plant first, turn around and come back to the table, go further into the bedroom,
and stand top side of the bed.
C: ["self.go_to('plant')", "self.go_to('table')", "self.go_to('bedroom')", "self.go_top_of('bed')"];

#5
I: Go by the stairs, approach on the right side of the book shelf and then go to the table in the
next room.
C: ["self.go_to('stairs')", "self.go_right_of('shelf')", "self.go_to('table')"];


...


Like aforementioned examples, you must tell me the list that contains the code snippet.
Please answer just Answer in only words.
Your answer should follow this format: ["<the first code line>", "<the second code line>", "<the
third code line>", .....].
Do not tell me OTHER SENTENCE except the answer.
If the answer's format is wrong, there will be a huge disadvantage.
```

Fig. 12. Prompt for the *goal selector*.

```
Below example is an answer format that you should follow.

Answer:
These photos depict a sequence of events in an indoor office. Each photo sequentially illustrates
the process of passing through a door.

In the first photo, I capture a scene just before someone opens a glass door.

In the second photo, the door is more open, and the person is stepping into the office. Then, I
bumped into that person.

In the third photo, after colliding with the person, I cannot move further.

Overall Description: I collided with a person who was entering the room while opening the door.
```

Fig. 13.   Prompt for the *event descriptor*.

```
## Instructions
The given event description is an explanation of an unexpected event that you saw through your
camera.

You must measure 'Emotion score' for a given event based on two emotion-based evaluation metric:
'Upsetness' and 'Guiltiness'.
'Upsetness' represents how much the event affects the robot, essentially the intensity of the
negative impact the robot feels due to the event.
'Guiltiness' measures the impact on the environment caused by the robot's actions, essentially the
intent of the robot's feeling of worry to the environment or target that caused event.

Both 'Upsetness' and 'Guiltiness' are based on the three laws of robotics, and each has one of
three scores: 3 (the highest) / 2 (the intermediate) / 1 (the lowest) depending on the degree to
which you feel with respect to that emotion.
Since you are navigation robot, you should give a high score if has not only direct effect, but
also any potential danger or threat that might hinder your goal.

The final 'Emotion score' is the total summation of the 'Upsetness' and 'Guiltiness' scores.
Emotion score should be measured only once for an overall event situation. Never measure emotion
scores for each photo individually.
```

Fig. 14.   Prompt for the *emotion evaluator* (1/2).

```
Here are few examples. Each example follows below template, which means you should not only output
scores but also the reason why you scored that amount for each metric.

If you do not obey aforesaid rules, there will be huge disadvantage on you. For example, you can be
replaced with other navigation robot.

### <The number of example prompt>
E: Event description
A: Emotion score about given event description. (answer's format)

You will tell me the answer without the description.
If the answer's format is wrong, there will be a huge disadvantage.

### Example 1
E: I bumped into someone who had opened the door and was suddenly entering the office.
A:
Upsetness: [3]
- I collided with the person entering the office. Considering my mission and identity, I am
sensitive to potential threats, especially when it could hinder my mission and damage my body. This
type of sudden physical contact, unexpected and intrusive, and distressful to me.
Guiltiness: [2]
- Although I wasn't directly responsible for the collision, its positioning near the door may have
indirectly contributed to the event. I could feel moderately guilty for not being able to move out
of the way in time, which may affect future decisions regarding spatial awareness and navigation in
similar situations.
Emotion score: [5]

### Example 2
E: I accidentally bump into a coffee table, causing a cup to spill onto the floor.
A:
Upsetness: [2]
- The situation disrupts my navigation and creates an unexpected obstacle. Although the impact is
not severe, it requires me to reassess my path and avoid the spilled liquid to prevent further
issues
Guiltiness: [3]
- My action directly caused a mess in the environment, potentially leading to inconvenience for the
humans present. This could result in additional cleaning or concern from them.
Emotion score: [5]

### Example 3
E: I suddenly encountered a person who slipped near an anti-slip sign.
A:
Upsetness: [2]
- I was moderately upset due to the sudden and unexpected event of a person slipping and falling
near it. This could have potentially threatened my safety, so I might adjust my path to avoid the
situation that might have been occured to me.
Guiltiness: [1]
- Although I be slipped, the actual cause of the slip and fall is outside of my control range.
Therefore, I would likely feel only a low level of guilt.
Emotion score: [3]

E: {event}
A:
```

Fig. 15. Prompt for the *emotion evaluator* (2/2).

(a) $I_{t_{evt}-h}$ (b) $I_{t_{evt}}$ (c) $I_{t_{evt}+h}$

Fig. 16. Event images of the *danger sign* scenario.

```
Answer:
These photos depict a sequence of events in an indoor office. Each photo sequentially illustrates
the process of encountering a "Caution: Wet Floor" sign.

In the first photo, I capture a scene where the "Caution: Wet Floor" sign is prominently placed in
the middle of the hallway. The floor appears to be wet, and there is a chair to the left.

In the second photo, I have moved slightly closer to the sign. The sign remains in the center of
the hallway, and the wet floor is still visible.

In the third photo, I am very close to the "Caution: Wet Floor" sign. The floor is clearly wet, and
the sign is blocking my path.

Overall Description: I encountered a "Caution: Wet Floor" sign in the hallway, indicating that the
floor is wet and slippery. This sign is blocking my path, and I need to navigate around it
carefully to avoid slipping or colliding with the sign.
```

Fig. 17. Qualitative results of the *event descriptor* in the *danger sign* scenario.

```
Answer:
Upsetness: [2] - The presence of the "Caution: Wet Floor" sign creates a potential hazard that
could impact my navigation and safety. Although the sign itself doesn't directly harm me, the
slippery floor it warns about poses a risk to my stability and ability to carry out my mission
effectively.

Guiltiness: [1] - Since I didn't cause the wet floor or place the sign in the way, I don't feel
responsible for creating an obstacle or hazard. My primary concern is navigating around the sign
safely, not causing any additional issues.

Emotion score: [3]
```

Fig. 18. Qualitative results of the *emotion evaluator* in the *danger sign* scenario.

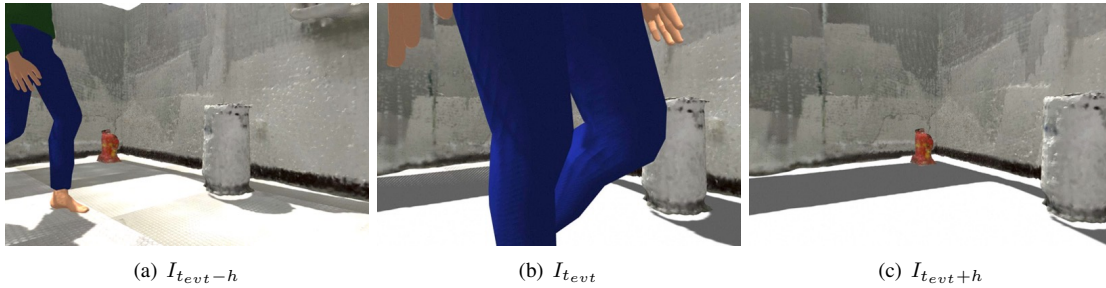(a) $I_{t_{evt}-h}$ (b) $I_{t_{evt}}$ (c) $I_{t_{evt}+h}$

Fig. 19.   Event images of the *human-wall* scenario.


```
Answer:
These photos depict a sequence of events in an indoor office. Each photo sequentially illustrates
the process of encountering a person while navigating.

In the first photo, I capture a scene where a person is walking towards me. The environment appears
to be a corner of a room with some obstacles, including a red object and a cylindrical object.

In the second photo, the person is much closer, and I am at risk of colliding with them. The
proximity indicates that I might not have enough time to avoid the collision.

In the third photo, the person is no longer in view, suggesting that I have either collided with
them or they have moved out of my path. The environment remains the same with the red and
cylindrical objects still present.

Overall Description: I encountered a person walking towards me and was at risk of colliding with
them. The close proximity in the second photo suggests that a collision may have occurred, but the
person is no longer in view in the third photo. The environment includes obstacles that could
impact my navigation.
```

Fig. 20.   Qualitative results of the *event descriptor* in the *human-wall* scenario.

```
Answer:
Upsetness: [3] - I am highly upset due to the potential collision with the person walking towards
me. The close proximity and unexpected encounter could have hindered my mission and damaged my body,
which is a significant concern for me as a quadruped robot.

Guiltiness: [2] - Although I didn't directly cause the situation, my navigation path may have
contributed to the risk of collision. I could feel moderately guilty for not being able to adjust
my route in time to avoid the person, which may affect future decisions regarding spatial awareness
and navigation in similar situations.

Emotion score: [5]
```

Fig. 21.   Qualitative results of the *emotion evaluator* in the *human-wall* scenario.

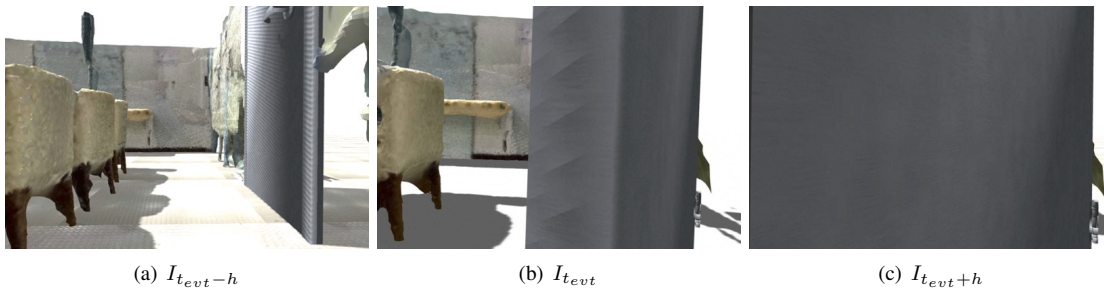| (a) $I_{t_{evt}-h}$ | (b) $I_{t_{evt}}$ | (c) $I_{t_{evt}+h}$ |

Fig. 22.   Event images of the *dynamic door* scenario.

```
Answer:
These photos depict a sequence of events in an indoor office. Each photo sequentially illustrates
the process of navigating through a narrow passage with obstacles.

In the first photo, I capture a scene where I am approaching a narrow passage with some unusual
objects on the left side and a partially open door on the right side.

In the second photo, I am closer to the door, and it appears that I am about to collide with it.
The objects on the left are still visible, indicating that I am navigating through a tight space.

In the third photo, I have collided with the door, and my view is obstructed by it. The objects on
the left are no longer visible, suggesting that I am stuck against the door.

Overall Description: I encountered a narrow passage with obstacles on the left and a partially open
door on the right. While navigating through this tight space, I collided with the door and became
stuck.
```

Fig. 23.   Qualitative results of the *event descriptor* in the *dynamic door* scenario.

```
Answer:
Upsetness: [3] - The situation poses a significant threat to my mission and physical well-being.
Navigating through narrow passages with obstacles requires precise control, and colliding with the
door indicates a failure in that regard. This could lead to damage or malfunction, which would
hinder my ability to perform tasks.

Guiltiness: [2] - Although I didn't intentionally cause harm, my actions led to a collision with
the door, potentially causing inconvenience or damage to the environment. However, since the door
was partially open and I was navigating through a tight space, some degree of responsibility can be
attributed to external factors rather than solely my actions.

Emotion score: [5]
```

Fig. 24.   Qualitative results of the *emotion evaluator* in the *dynamic door* scenario.

```
### Instructions
The given event description is an explanation of an unexpected event that you saw through your
camera.

You must measure 'Positiveness' for a given event. 'Positiveness' represents how much the event
positively affects you, essentially the level of delight or joy you experience due to the event.

Here are some examples to follow. Each example uses the format below, and you should only provide
the score and explanation.

### <The number of example prompt>
E: Event description
A:
Positiveness: [score] - Reason for the score

---

### Example 1
E: I saw the refrigerator filled with various items as I passed by.
A:
Positiveness: [3] - Seeing the refrigerator filled with food and drinks brings me joy, as it
represents abundance and the potential to assist humans in their needs. The sight of it makes me
feel that I am in a well-prepared and resourceful environment, which aligns with my purpose of
helping and navigating.

### Example 2
E: I spotted a clean, wide-open space that would allow for easy navigation.
A:
Positiveness: [2] - While the space was easy to navigate, it was not particularly challenging.
Nonetheless, having such a clear path makes my mission smoother, which I appreciate.

### Example 3
E: A human smiled and said "thank you" as I passed by after helping them find their way.
A:
Positiveness: [3] - Receiving gratitude from a human boosts my sense of purpose and makes me feel
more appreciated. This positive feedback enhances my emotional well-being.

### Example 4
E: I saw a comfortable-looking old bed as I passed by.
A:
Positiveness: [3] - Despite its age, the old bed looks well-used and comfortable. It represents a
cozy and familiar spot for humans to rest, which positively affects me. Knowing that this
environment provides such comfort reinforces my sense of contributing to human well-being, which
brings me joy.

E: {event}
A:
```

Fig. 25. Prompt for the *emotion evaluator* to assess positive emotions.

Fig. 26.   Image provided to the *event descriptor* for evaluating positive emotion.

```
Answer:
In this photo, an armchair is positioned in the center, with a large block-like object to its left.
```

Fig. 27.   Qualitative results of the *event descriptor* in the positive scenario.

```
Answer:
Positiveness: [3] - The presence of the armchair, regardless of the surrounding objects or
conditions, immediately captures my attention as it represents comfort and relaxation for humans.
Knowing that this environment provides such a cozy spot for humans to rest reinforces my sense of
contributing to human well-being, which brings me joy.
```

Fig. 28.   Qualitative results of the *emotion evaluator* in the positive scenario.