



Q: Please guide a player from the start to the goal.

GPT 5: \boxed{D,D,D,D,D,L,L,L,U,U,U,U}



Ours:

<think>The first step is to locate the start point I will use the 'Point' tool.</think>

<tool\_call> Point </tool\_call>

<think>After located the start point, I need to locate the target... </think>

<tool\_call> Point </tool\_call>

<think>After located the start point and target, I need to locate the ice holes ... </think>

<tool\_call> Point </tool\_call>

<think>I will call Astar to find the correct path... </think>

<tool\_call> Astar </tool\_call>

<think>I will call Draw2DPath to verify the path </think>

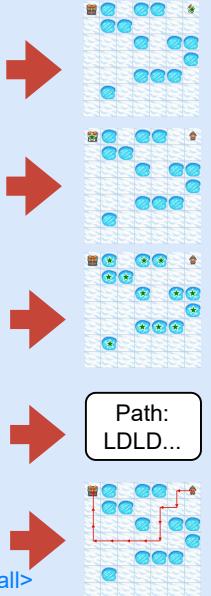
<tool\_call> Draw2DPath </tool\_call>

<think>Based on observations through tools, I can confirm that the path is correct </think>

<response>

\boxed{L,D,L,D,D,L,L,L,U,U,U}

</response>



(a) VSP-Navigation



Q: which one is the missing part?

A.



B.



Ours:

<think>I will use the 'DetectBlackArea' tool to find the bounding box of this missing region </think>

<tool\_call> DetectBlackArea </tool\_call>

<think>I will now attempt to insert the first candidate image... </think>

<tool\_call> InsertImage </tool\_call>

<think>does not align seamlessly ... I will now attempt to insert 'img\_3' (Option B) into ..</think>

<tool\_call> InsertImage </tool\_call>

<think>I have now inserted both images into the missing region. Upon visual inspection of the resulting images, **neither insertion perfectly resolves the jigsaw puzzle**, as the tool's output ... **However, I must still determine the most suitable option based** on visual evidence ... 'img\_2' (Option A) provides a more plausible and natural completion of the original image.</think>

<response>

Based on my visual inspection, the best fit is option A. \boxed{A} </response>

BBOX:  
[210, 136,  
426, 284]



(b) Jigsaw

Q: How can the user access the buying option from the webpage?

GPT5: \boxed{Click the green "Buy Now >" button. You can use either of the two on the page:- Top button: [560, 214, 654, 248]}



Ours:

<think> I can see a distinct section on the page that features a book cover and a prominent green button... To accurately describe how a user would access this, I should focus on this specific area of the page.. </think>

<tool\_call> Crop </tool\_call>

<think>To understand how a user would access this buying option, I need to analyze the text ... Now, I will use the OCR tool...</think>

<tool\_call> OCR </tool\_call>

<think>I have successfully used the OCR tool on the cropped image. The button is situated directly beneath the book's title and description, making it a direct and intuitive way for a user to access the purchasing interface.</think>

<response>

To access this buying option, a user would likely need to be logged in ... Once logged in, the user would click on the "Buy Now" button ... </response>

**Buy Now >**

{'label': 'Buy',  
'label': 'Now'}



(b) GUIQA