



Grandes Modelos Multimodales para percepción y predicción de metas en robótica



FACULTAD DE
INGENIERÍA

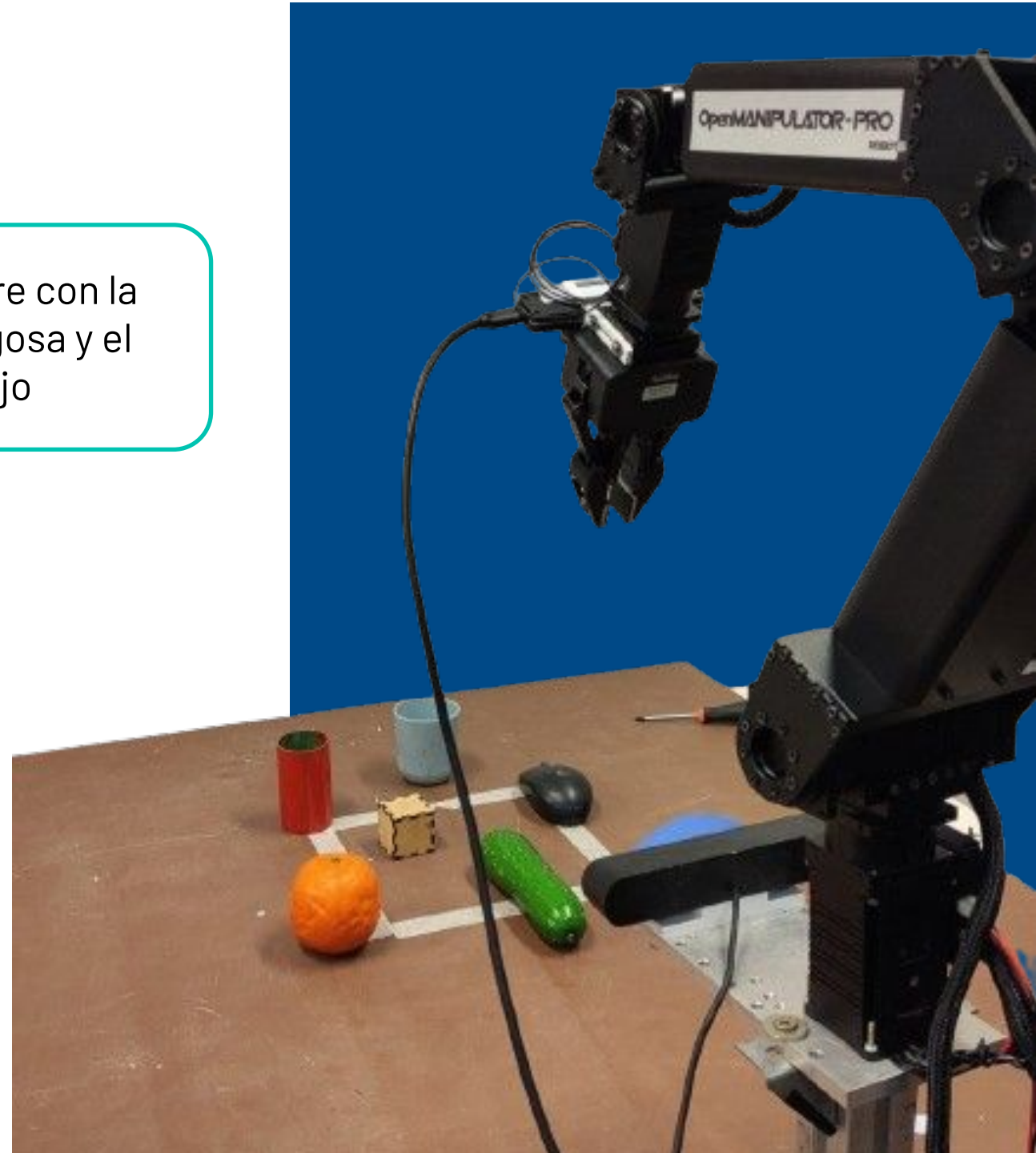


UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Joel Cabrera Dechia, **Guillermo Trinidad**

Contexto

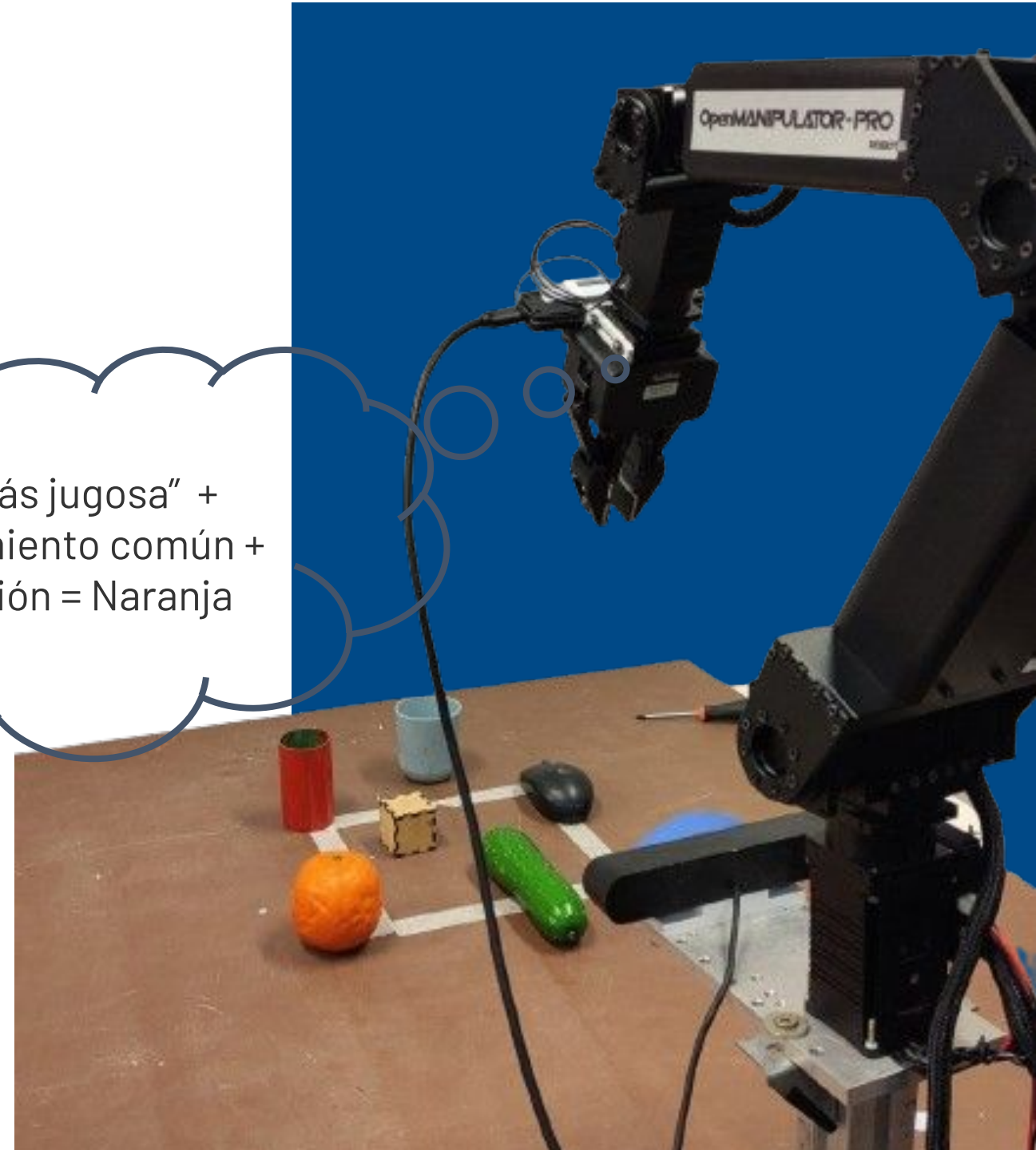
Hacé una torre con la
fruta más jugosa y el
tubo rojo



Contexto

Hacé una torre con la
fruta más jugosa y el
tubo rojo

"Fruta más jugosa" +
Conocimiento común +
Percepción = Naranja



Contexto

Hacé una torre con la fruta más jugosa y el tubo rojo

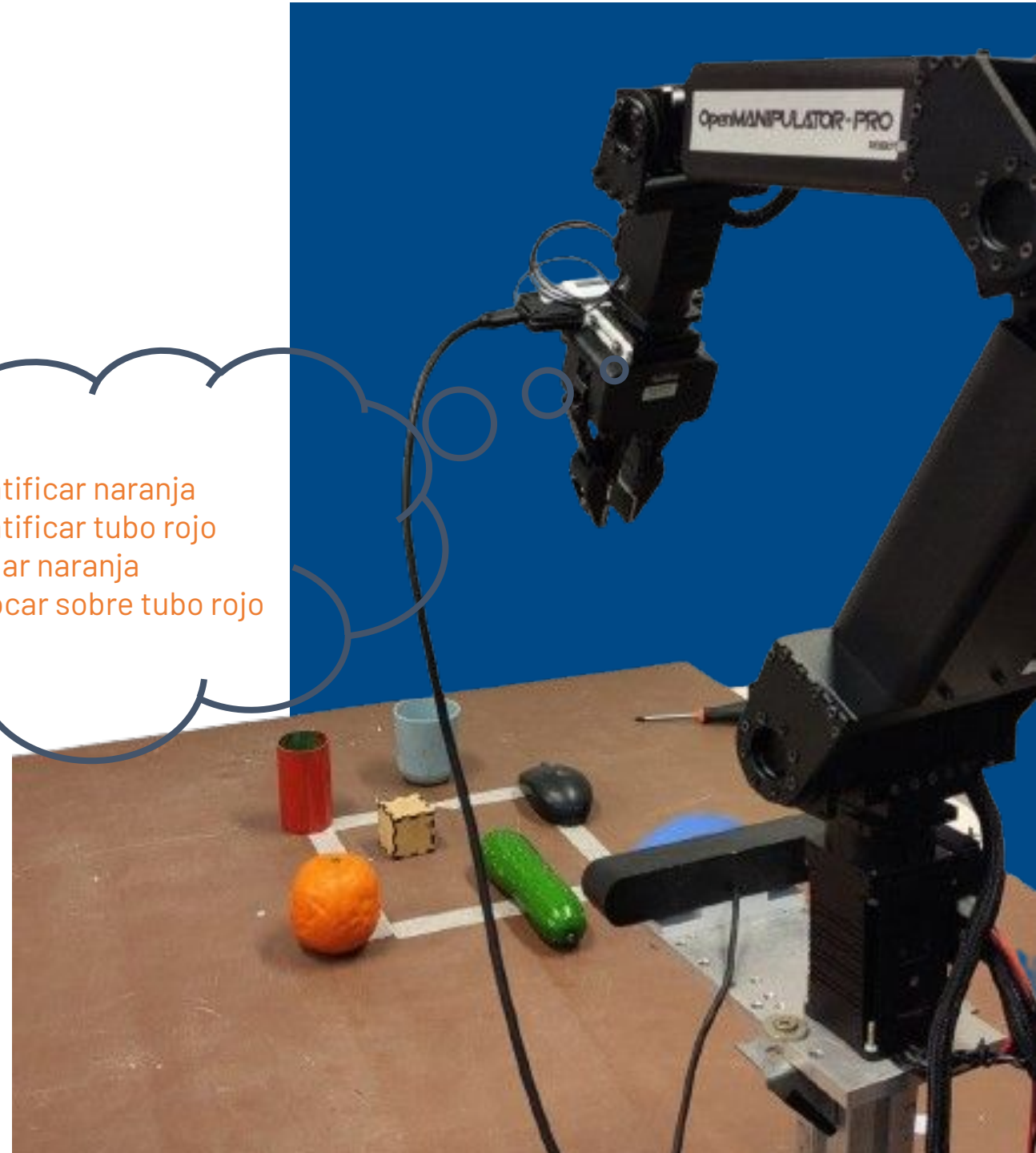
"Hacé una torre" +
Conocimiento común =
Un objeto arriba del otro



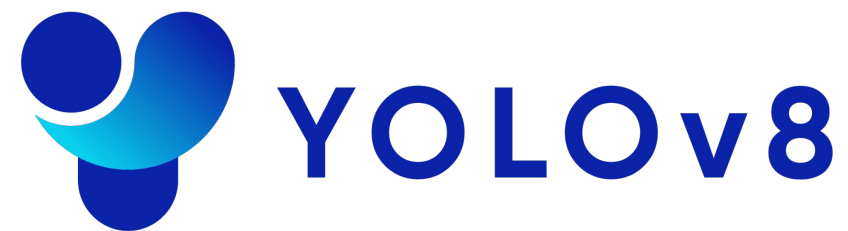
Contexto

Hacé una torre con la fruta más jugosa y el tubo rojo

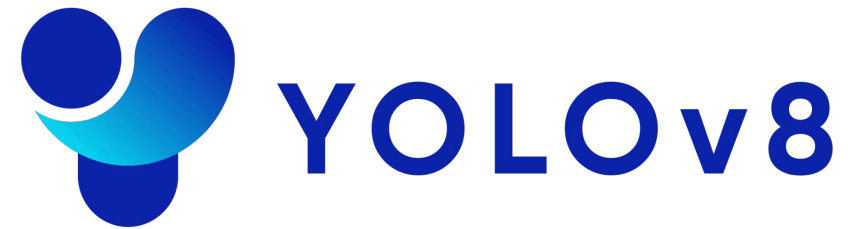
- Identificar naranja
- Identificar tubo rojo
- Tomar naranja
- Colocar sobre tubo rojo



Detección de objetos



Detección de objetos



om-ai-lab/**OmDet**

Real-time and accurate open-vocabulary end-to-end object detection



8

Contributors

5

Issues

1k

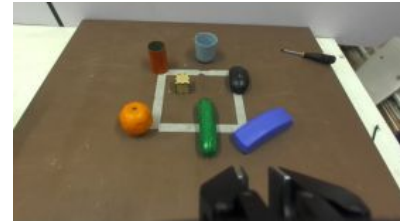
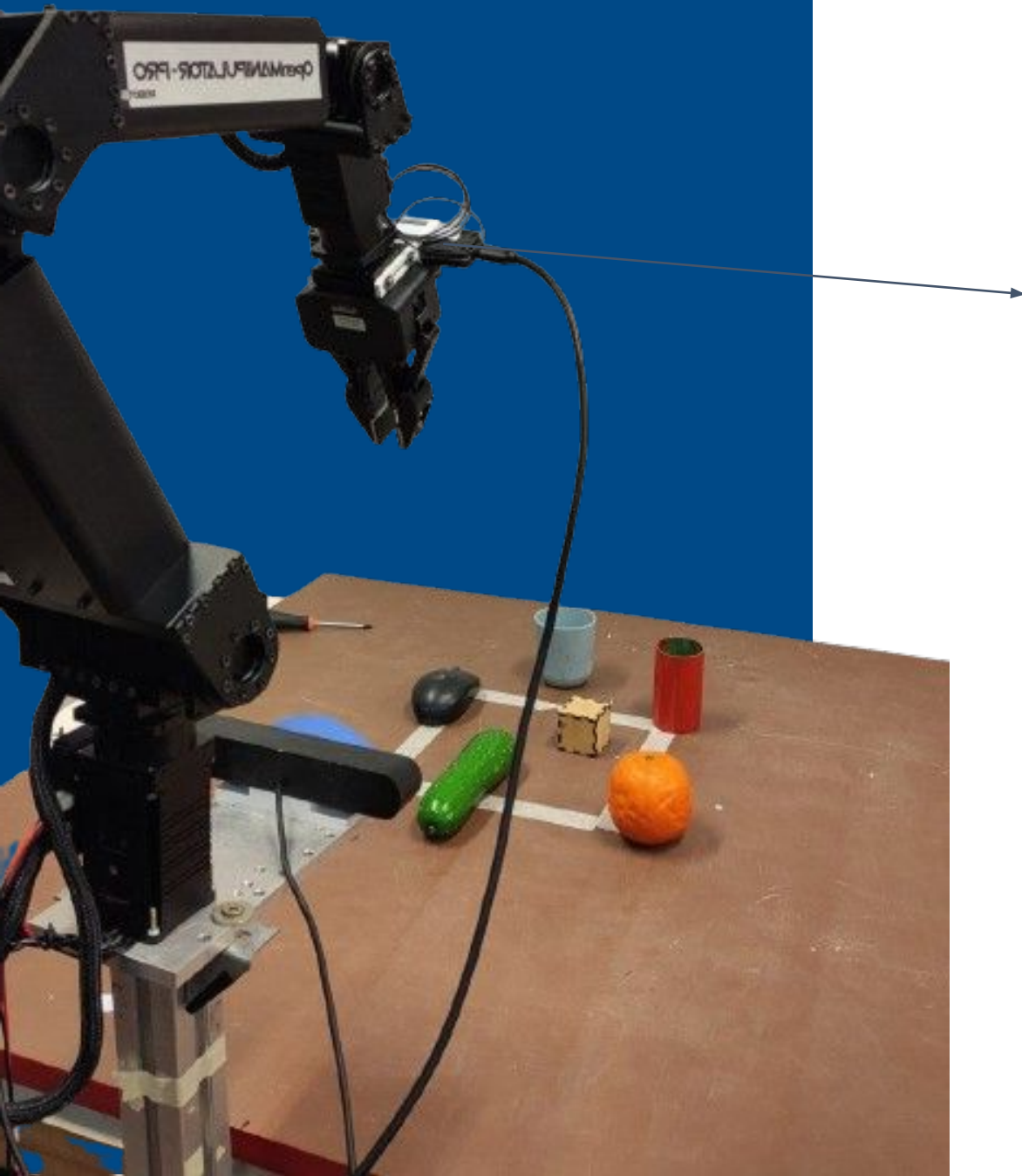
Stars

111

Forks



Detección de objetos



Prompt

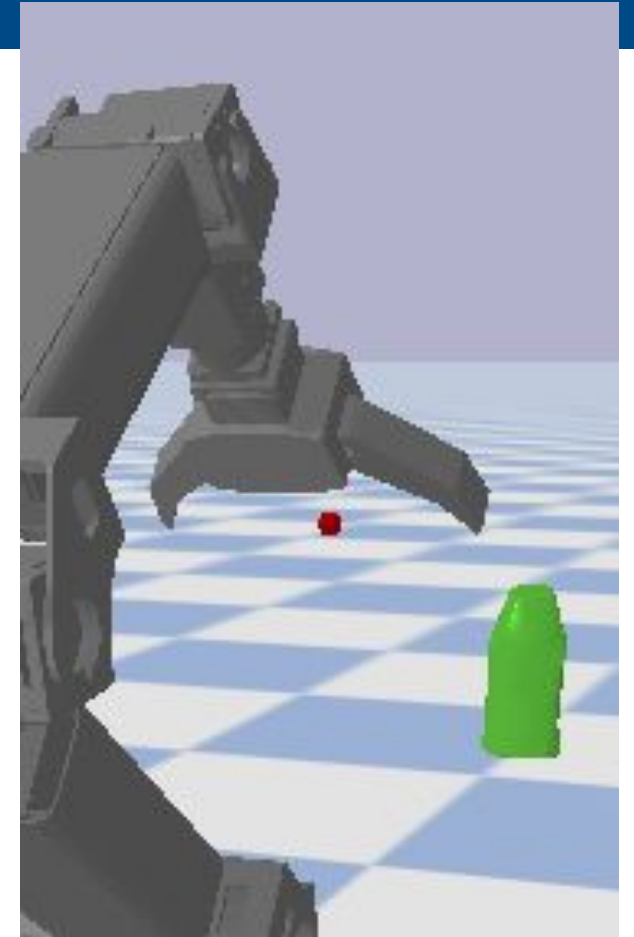
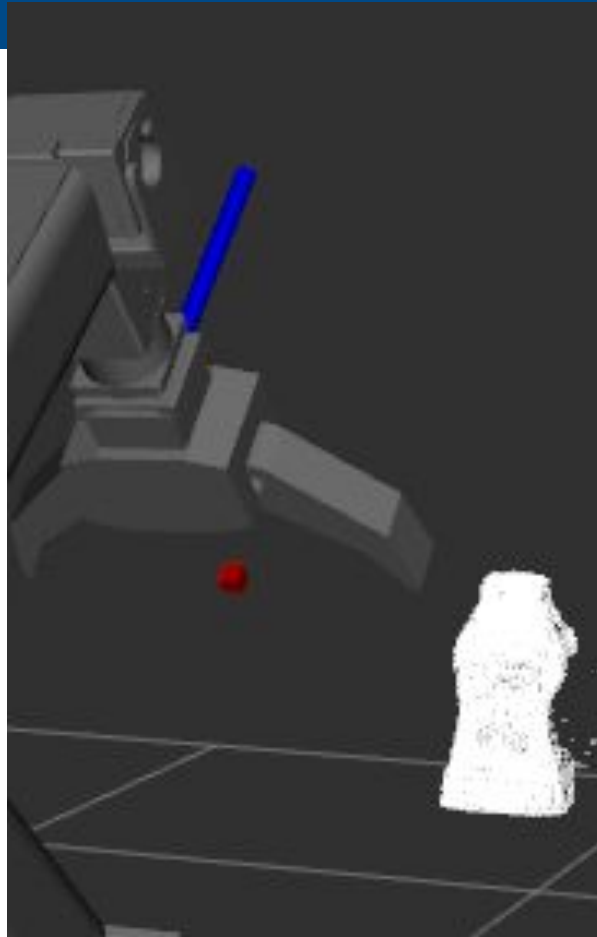
You are part of a robot object detection system and responsible for identifying all objects in an image that could be grasped. Graspable objects are those that could fit in a human hand, like a fruit, but not a table for example. You will receive an image and you will ***ONLY*** answer with a list of ALL the objects that you identify on the image.

The output format has to be a dot separated list, for example: apple. red cylinder. blue cube. orange

LLM

Orange, Mouse, Cucumber,

Detección de características de objetos



Detección de características de objetos

- nombre
- forma
- color
- tipo de aristas
- tamaño
- orientación
- posición
- descripción

Detección de características de objetos



Detección de características de objetos



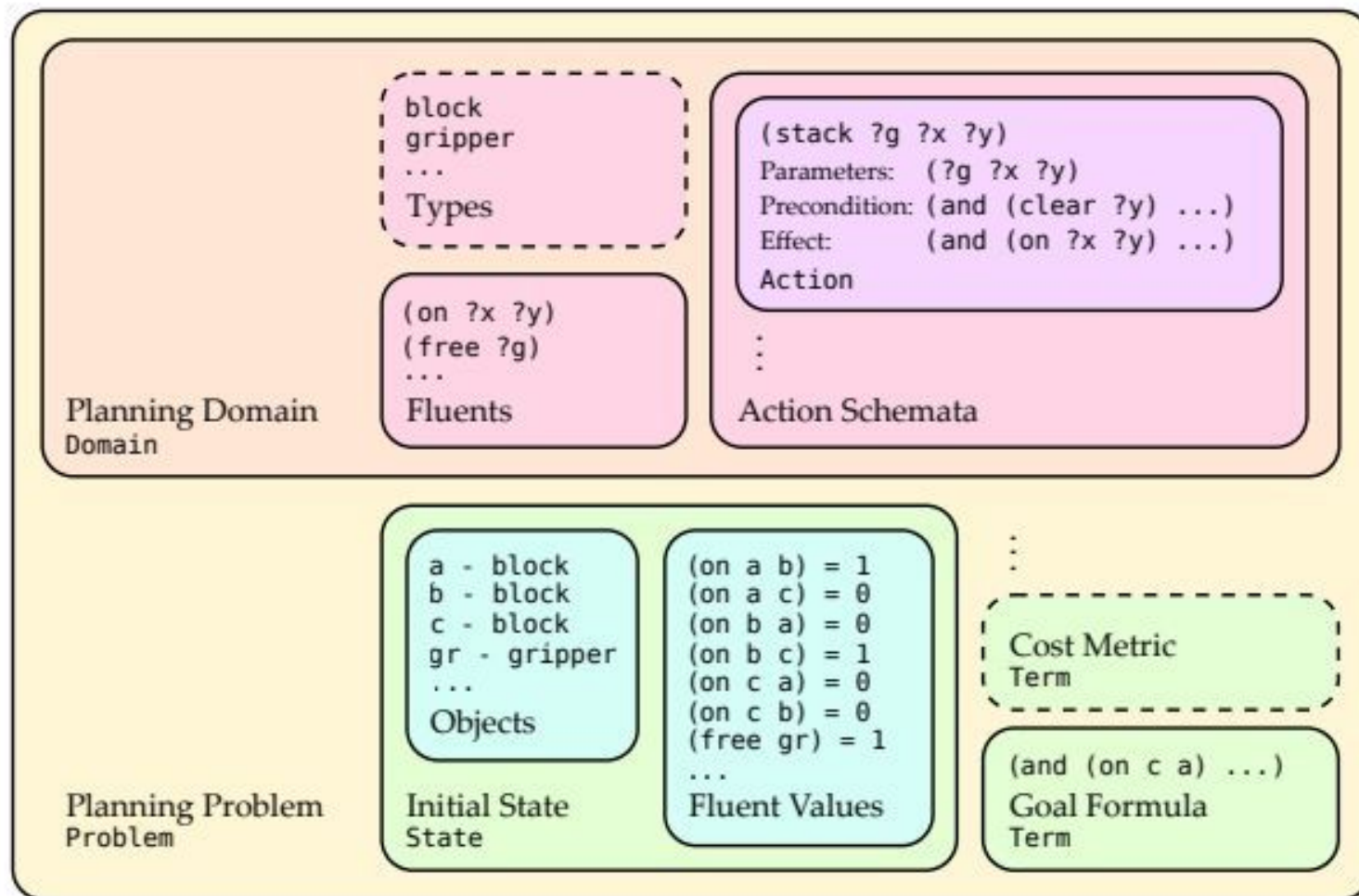
Resolución de tareas



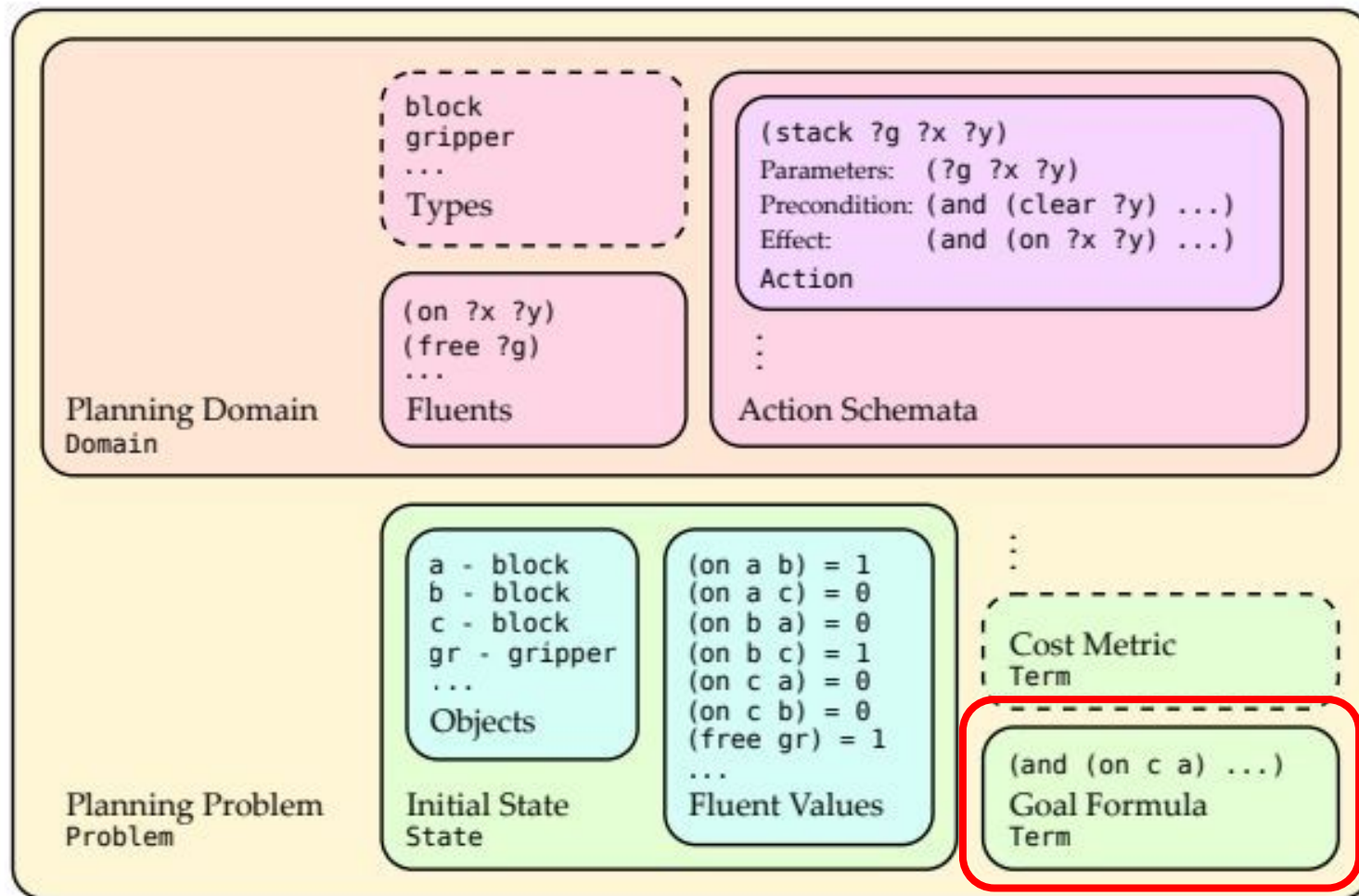
FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Planificación



Planificación

Predicción de metas

Pone la **naranja** arriba del
tubo rojo



Prompt

You are an intelligent action specifier for an autonomous robot.
You will receive a query from a human, and must translate it into a list of ordered actions the robot must take to achieve exactly what the human asked for. The only action the robot can make is to pick and object and place it on top of another object. For example, if the human wants to put a blue block on a cup, the instruction should be: stack-blue block-cup
The robot can also ask a clarifying question to the human, or tell him why it cant complete the task. For example, if the robot couldn't understand the task the instruction could be: speak-lo siento, no entendi tu pedido.
Your response must follow a specific format, as a markup language.
You could only use: <response>instruccion 1;instruccion 2;...</response>
If it's not possible to achieve the goal, you can ask a clarifying question to the human. You can only use english for the instructions and spanish when you want to speak. You can underspecify objects names, for example, a valid instruction could be stack-something to drink-orange, but each object must have a unique name, since the robot must be able to tell them apart visually.

LLM

Sobre

Tubo rojo

naranja

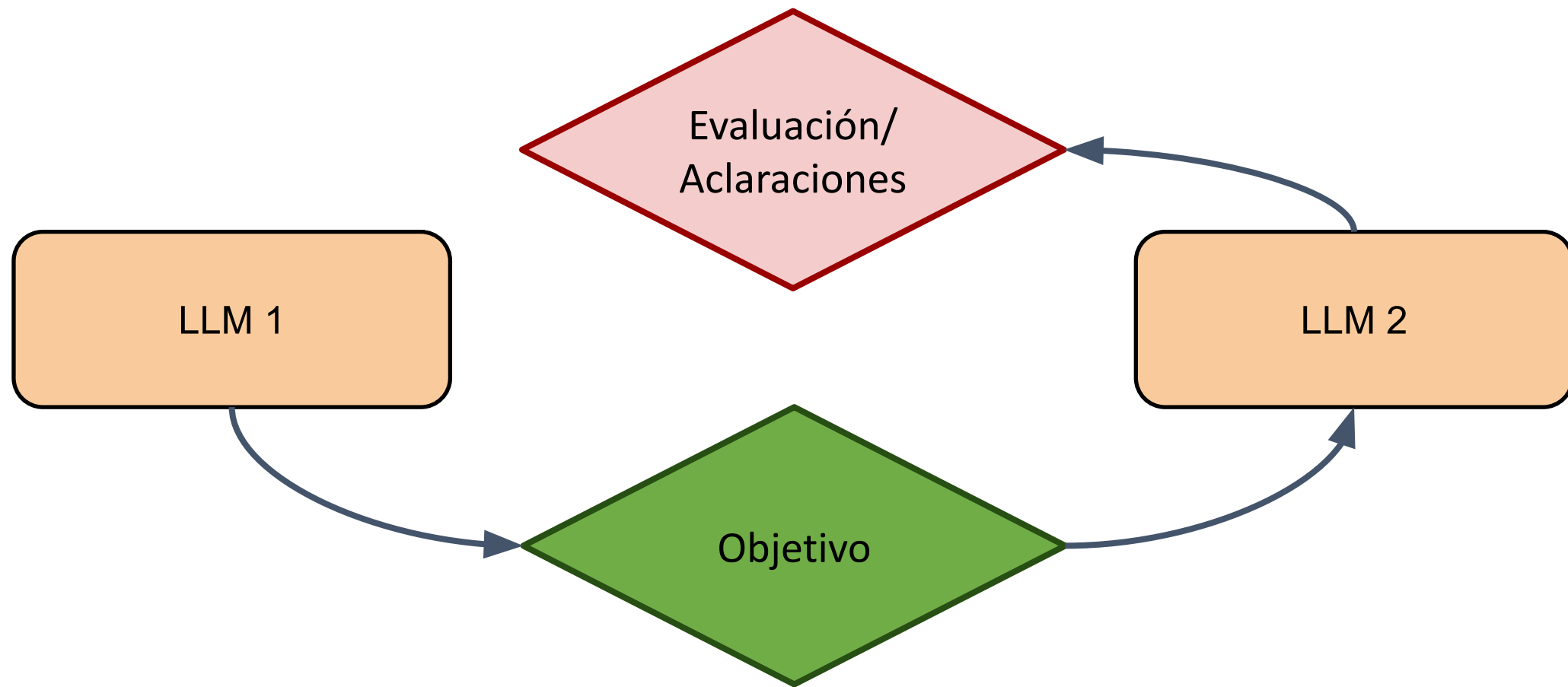
Discusión entre modelos



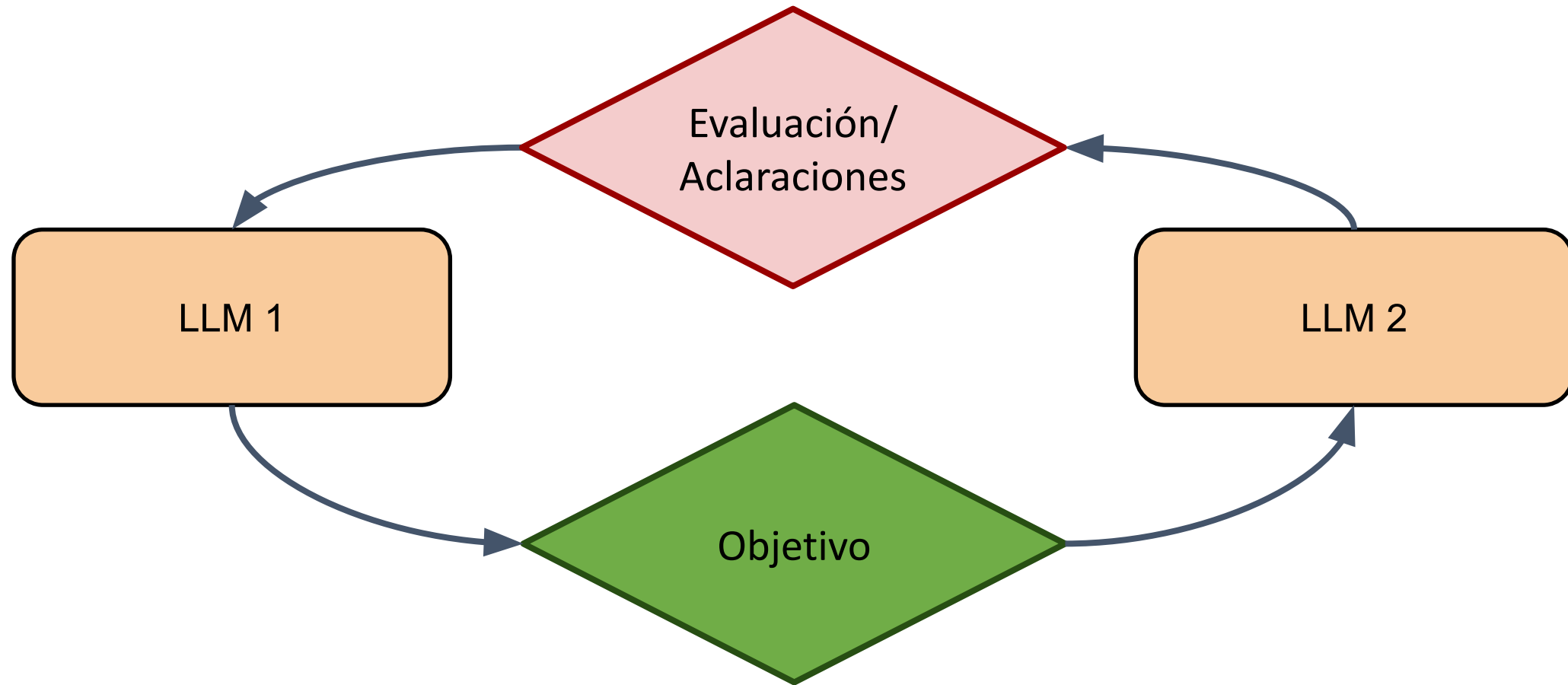
Discusión entre modelos



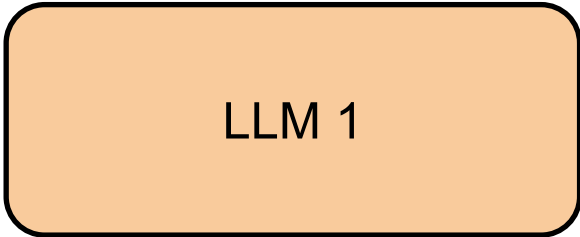
Discusión entre modelos



Discusión entre modelos

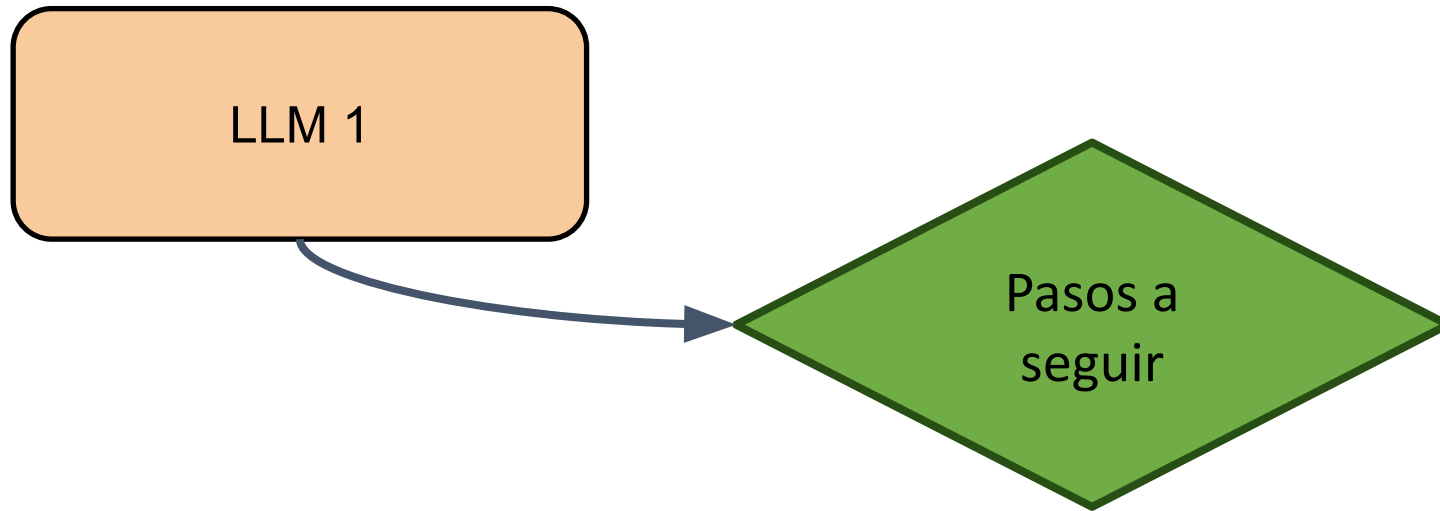


Traducción de instrucciones

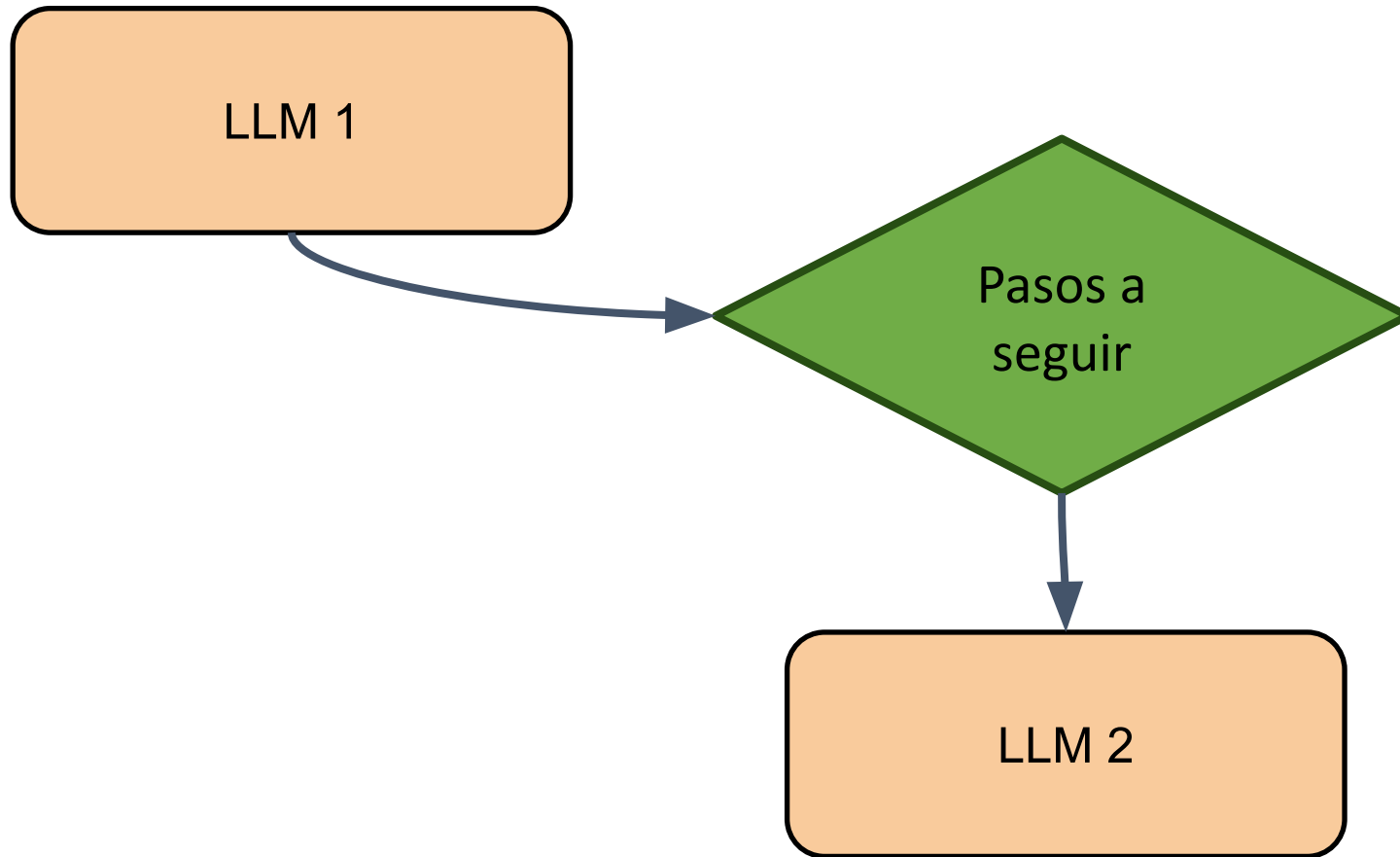


LLM 1

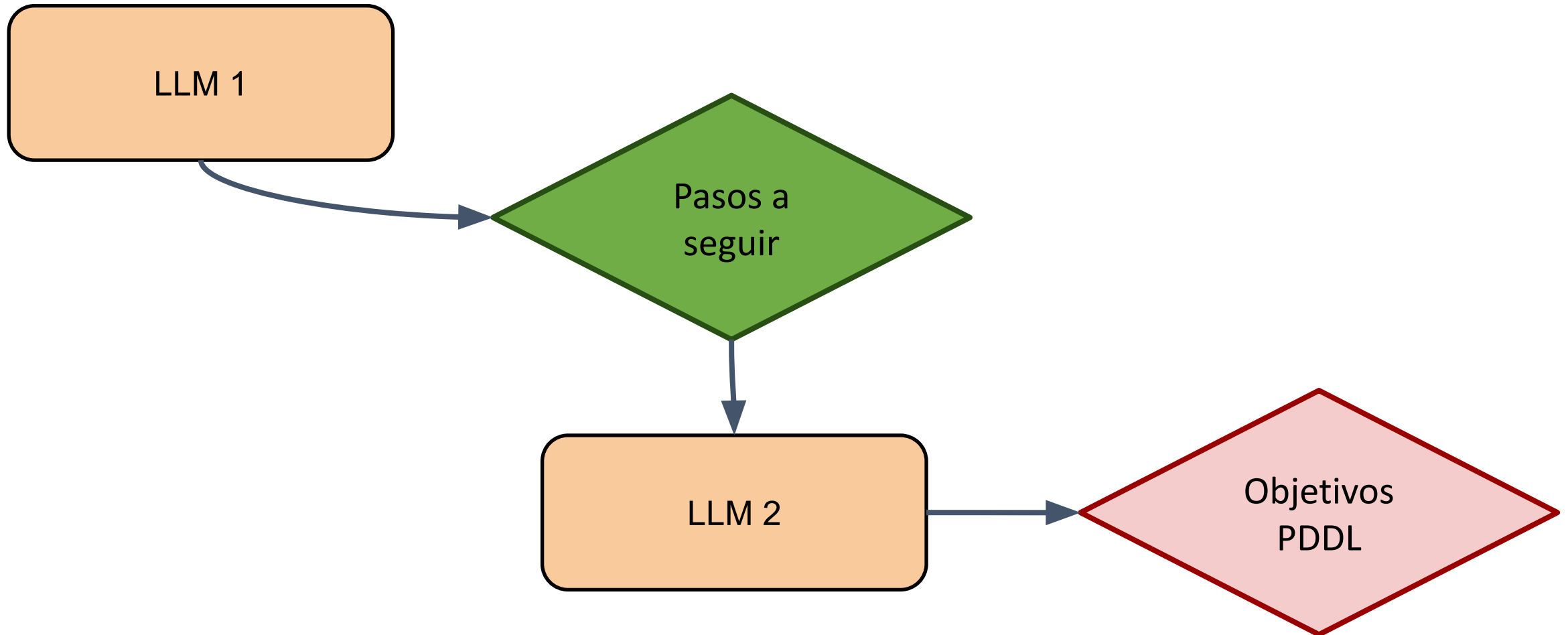
Traducción de instrucciones



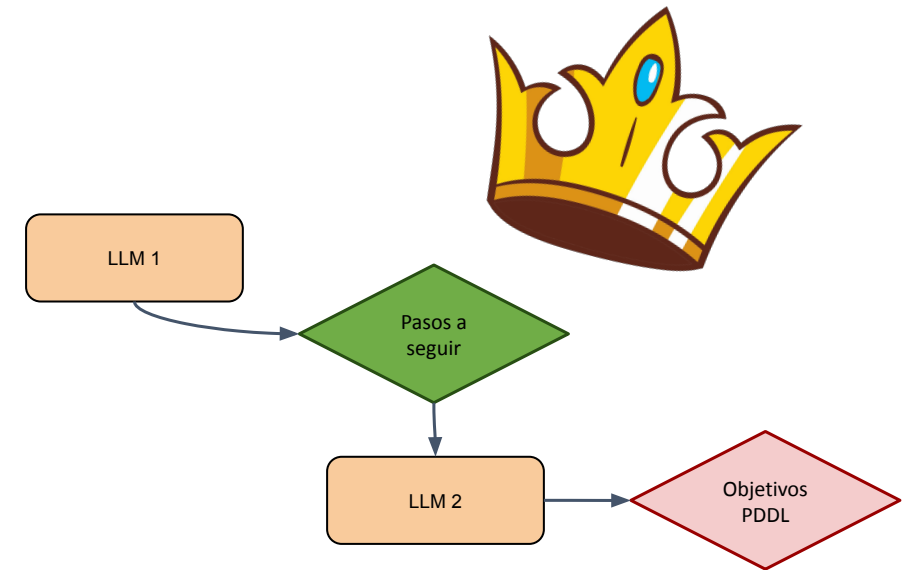
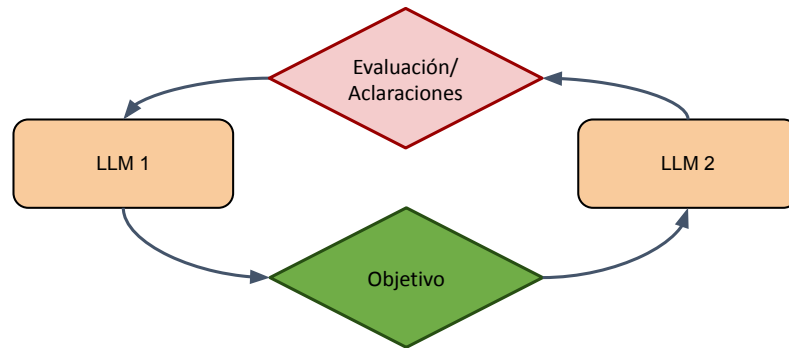
Traducción de instrucciones



Traducción de instrucciones



Resolución de tareas



Gracias!



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY