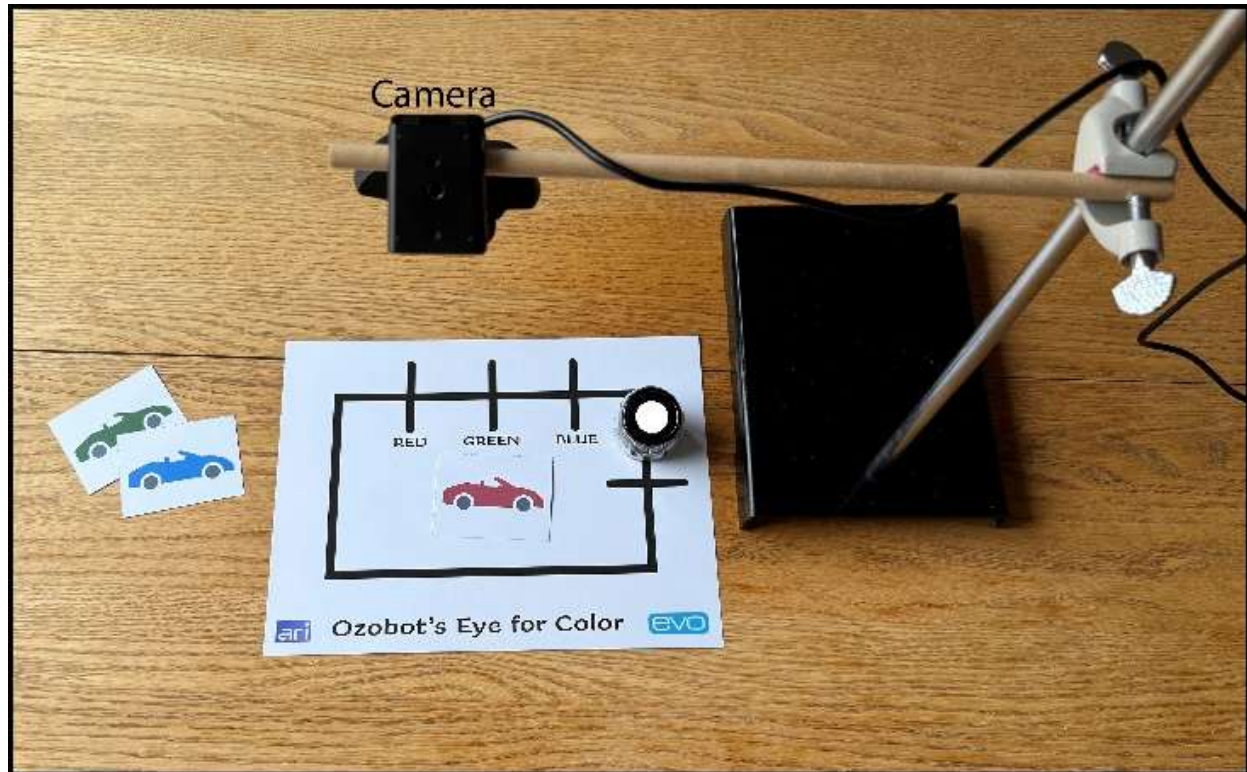


OzoBlockly Editor—**Challenge: Ozobot's Eye for Color**—Student Guide

by Richard Born, Associate Professor Emeritus, Northern Illinois University

Introduction

In this lesson, you will make use of Ari (or Evo) and an LLM Bot with camera. Your challenge is to program Ari to determine the color (either red, green, or blue) of the image of a car. Ari will start at the location shown in the figure below and then move to the intersection on the far-right side of the Ozomap. While Ari is at that intersection, the camera will snap a picture of the car and the LLM with the camera will respond with the color of the car. Ari will then navigate to and stop at the intersection at the top of the Ozomap corresponding to car's color and beep twice. After that, Ari will move to the intersection on the far-right to identify the color of the next car. When Ari reaches this intersection, the user has 5 seconds to place the image of the next car on the Ozomap.



The camera is suspended from a ring stand as shown in the figure, so that it can see the car clearly. Of course, you can just hold the camera and view the image in the Ozobot Editor to make sure that the car is clearly seen.

A video that accompanies this lesson shows Ari identifying the color of three different cars in sequence. You should view this video now to get a good feel for how your program should work.

How is Ari Able to Identify Color?

Before you begin working on this challenge, let's address the question of how Ari is able to identify the color of a car by looking at its image.

Traditionally, an LLM is only capable of processing text. But when it also makes use of a camera it becomes a **MLLM—Multimodal Large Language Model**. Sometimes, such a model is referred to as a **VLM—Vision-Language Model**. The process involved consists of four steps:

1. Image Capture and Pre-processing. The image is broken down into a large grid of numbers, with each pixel identified by its RGB color. For example, when viewing a lemon, a large number of pixels will be similar to $(R,G,B) = (255,255,0)$, which is decimal for yellow.
2. A vision encoder then identifies the object as a fruit whose dominant color is yellow.
3. Since the LLM understands that the user is interested in knowing only the color of the object, the LLM ignores characteristics of the object such as shape and size.
4. The LLM connects this concept of yellow to the word "Yellow" and responds with something like "The object I see is yellow."

Keep in mind that the model has been trained on extremely large datasets that contain both images and text descriptions. Here are some examples:

- A picture of a lime might have the label "a green lime".
- An image of a sporty red car might be labeled "a red car".
- An image of a Queen of Hearts card could have a label "The suit of this card is Hearts".
- A picture of the planet Mars is labelled "The red planet".

Working with the Program

Ready to bring Ari to life? Dive into the OzoBlockly Editor and program your Ozobot to brilliantly identify car colors. This challenge combines logic, robotics, and AI in a fun way. Get started, explore, and enjoy the satisfaction of seeing your code guide Ari through this exciting visual quest!