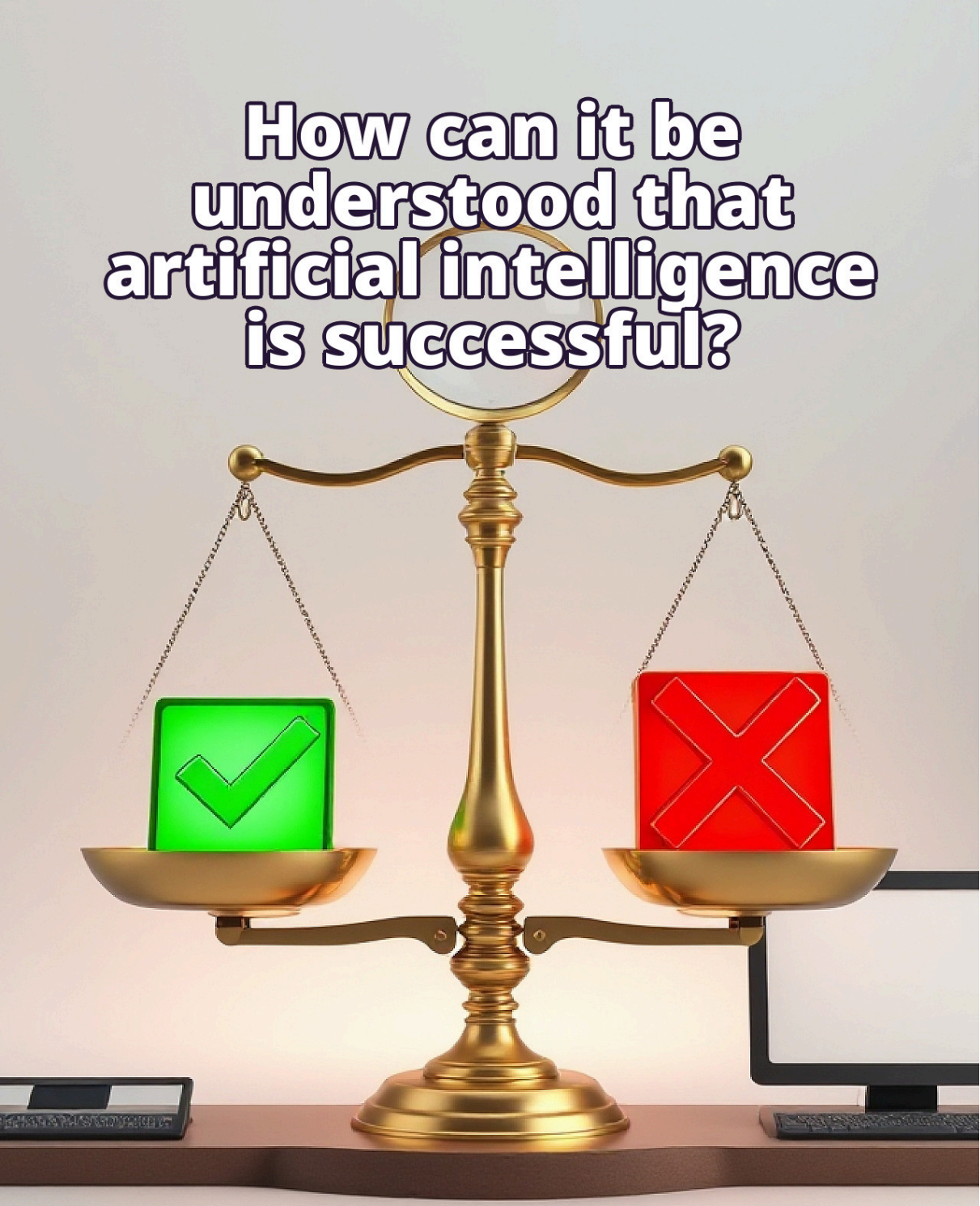
How can it be understood that artificial intelligence is successful?



www.yzokul.com.tr

How can it be understood that artificial intelligence is successful?



You want to know whether an artificial intelligence system is successful. The short answer: success is measured with tools that fit the task the system performs. A system is judged by how well its results match what it was built to do, not by how clever it seems. Different tasks need different measuring tools, just as different jobs need different rulers.



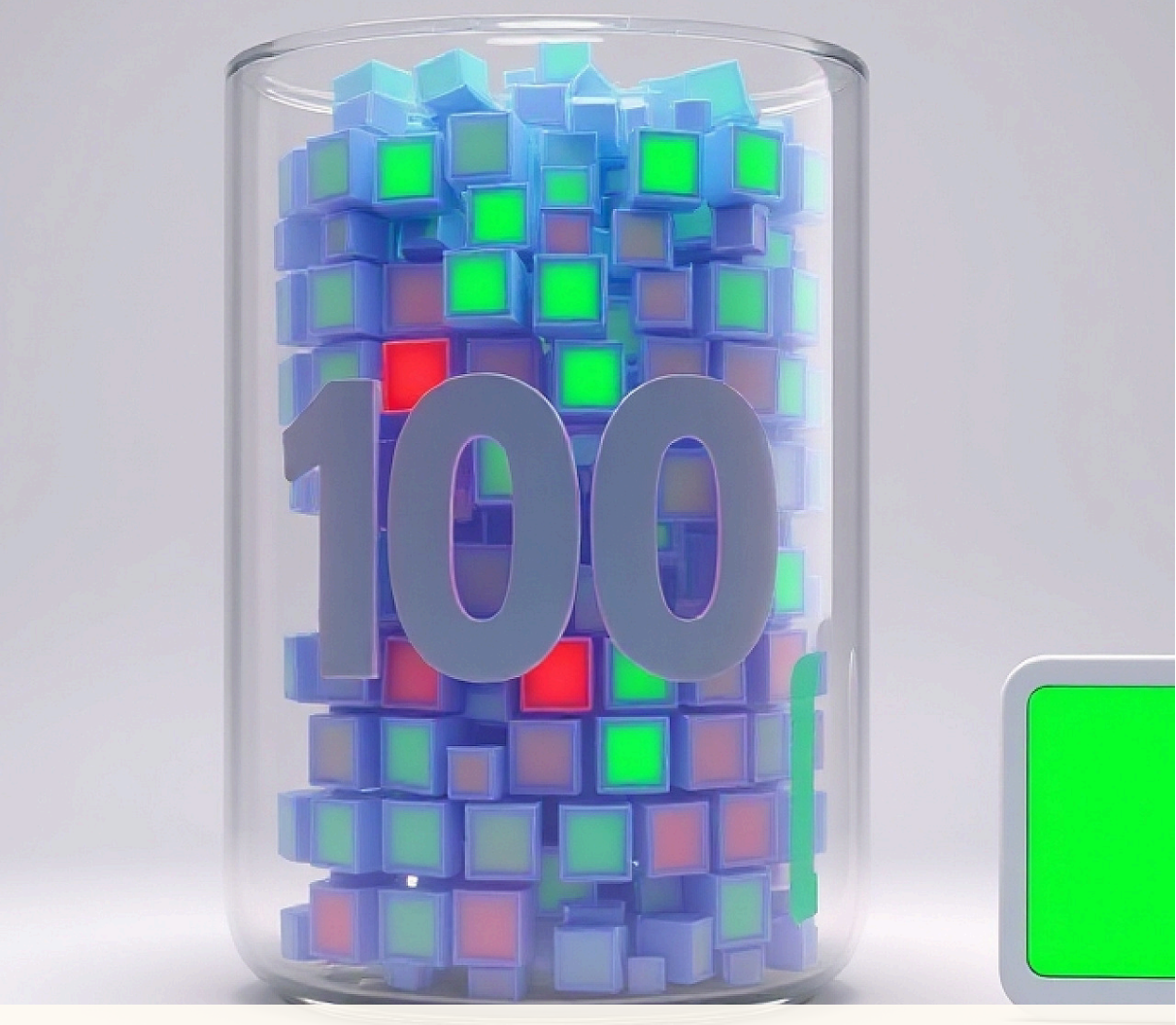
Success here means how well the system does its job. It does not mean how impressive or clever the system looks. A system is successful when its results match what you wanted it to do. If you wanted correct category labels and it gives correct category labels, it is doing well. If you wanted close number guesses and it gives close number guesses, it is doing well.



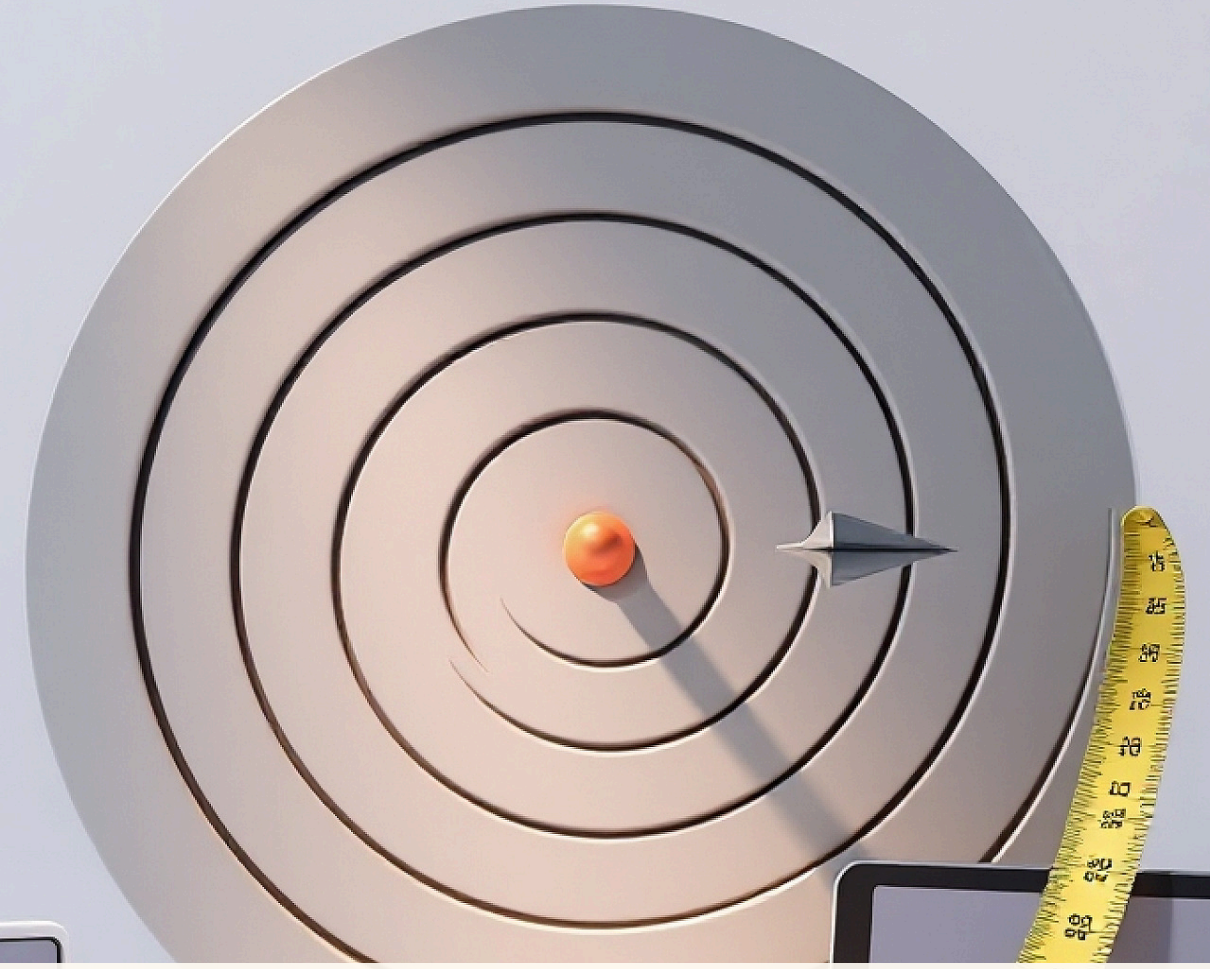
Match the measure to the task. Ask first what kind of task the system performs, because each kind of task has its own measuring tool. A classification task and a prediction task are measured in different ways. Using the wrong tool gives a misleading result. So the first step in judging success is naming the task type clearly.



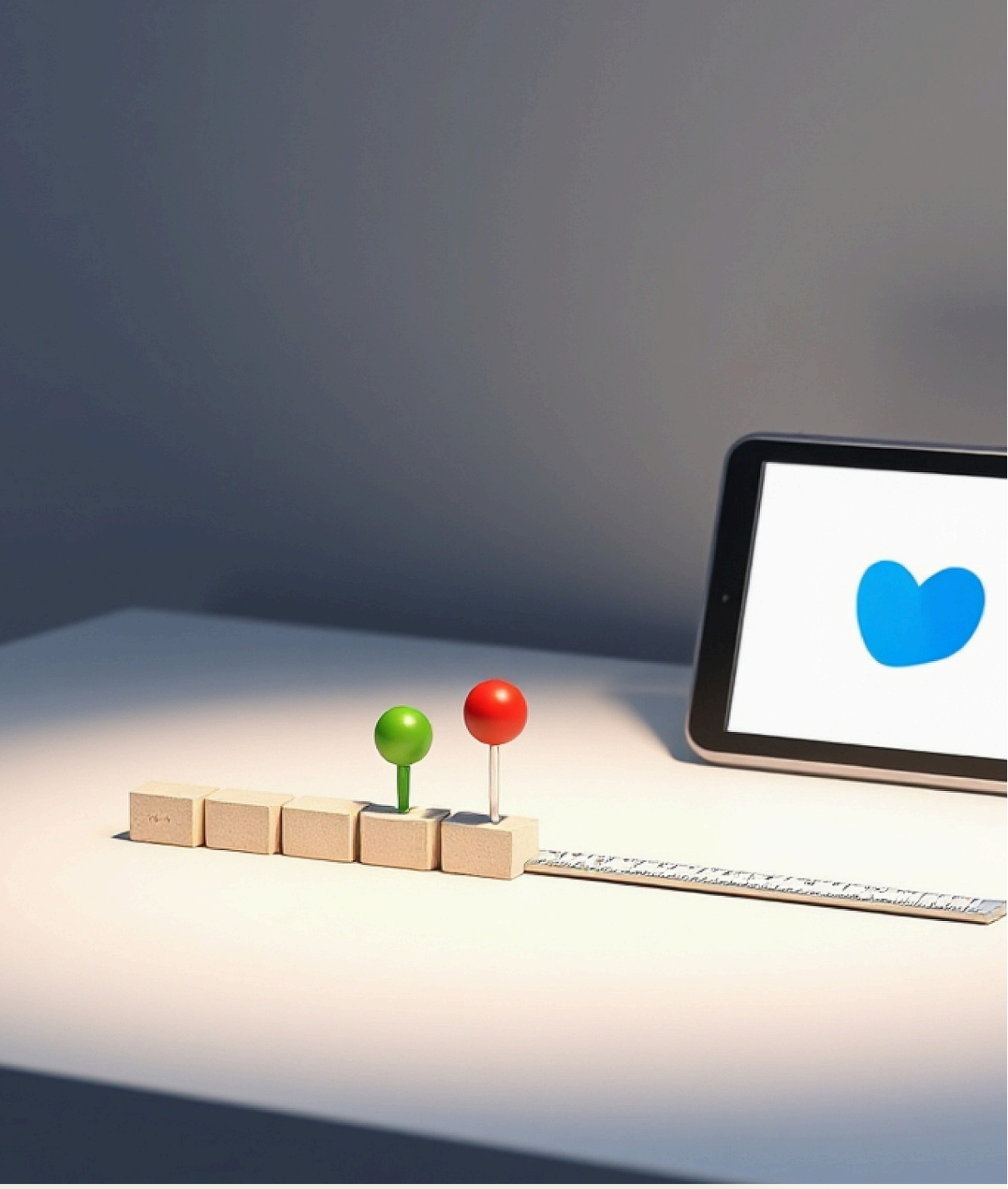
In a classification system, the system puts each example into a category. For example, it decides whether a photo shows a cat or a dog. Success is measured as the share of decisions that were correct. You count the correct decisions and divide by all the decisions made. The result is a number between zero and one, or as a percentage.



Walk through a small example. Suppose the system makes 100 decisions and 90 of them are correct. The success score is 90 correct decisions divided by 100 total decisions, which is 0.90 or 90 percent. The score is just correct decisions divided by all decisions. This simple correct-rate is the basic measure for classification tasks.



In a prediction system, the system gives a number as its answer, such as tomorrow's temperature. Success is measured by how far that answer is from the actual result. You subtract the prediction from the true value, or the true value from the prediction, and look at the size of the difference. Smaller distances mean better predictions. This distance measure is the basic tool for prediction tasks.



See a simple distance example. If the true value is 10 and the system predicts 12, the answer is off by 2. If it predicts 10, the answer is off by 0, which is perfect. If it predicts 7, it is off by 3, which is worse than being off by 2. Smaller distances mean better predictions, so compare distances to judge success.



One score is not enough. A single number can look good while hiding real problems. The overall score averages all the examples together, so strong results on many easy examples can cover weak results on a few important ones. To understand success properly, you must look deeper than the overall score. The next checks show where to look.



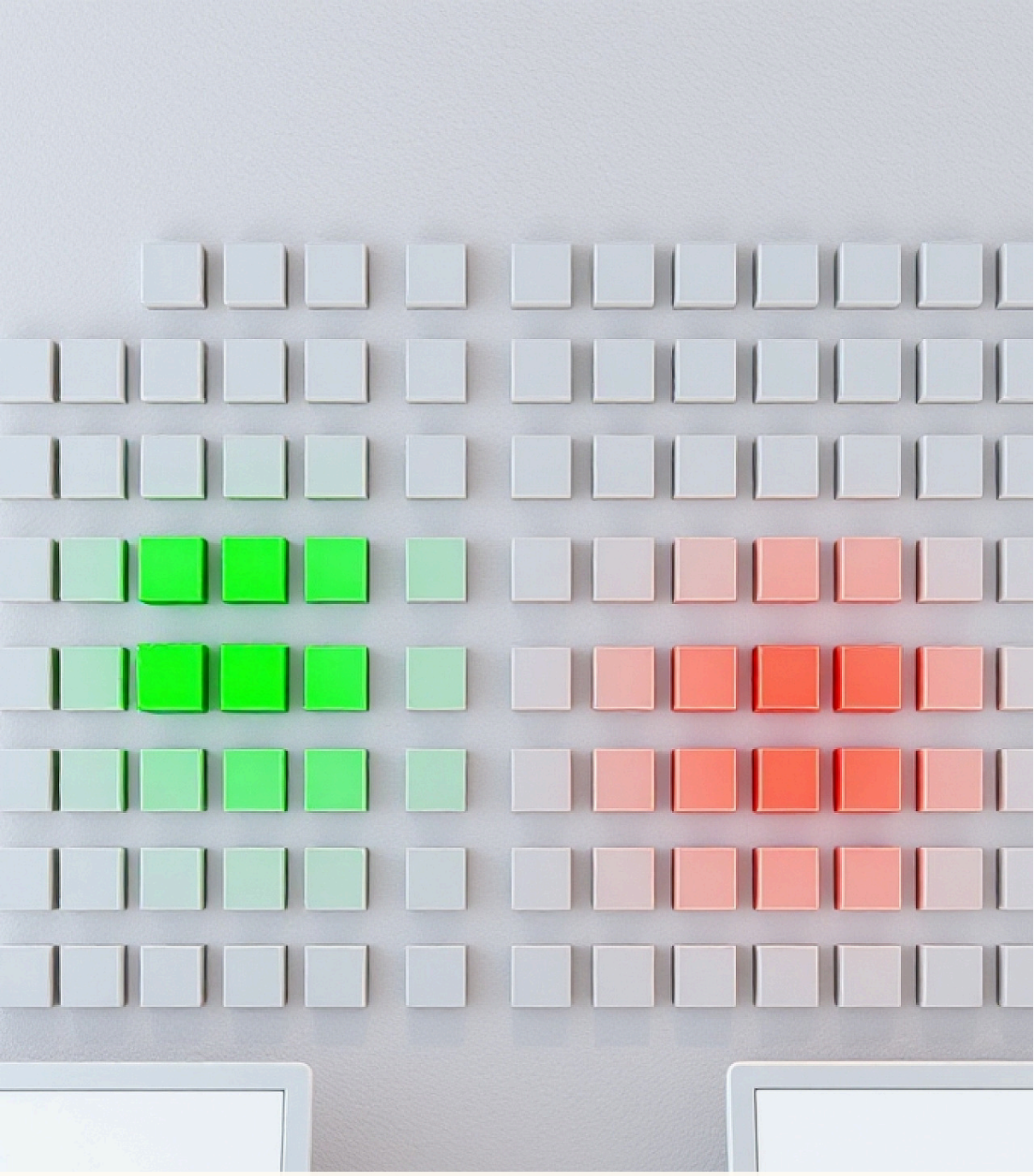
The rare-case trap is a common hidden problem. Imagine a system that always misses a rare situation but still scores high overall. It scores high because it gets the common examples right, and those are most of the examples. The high score hides the failure on the rare case. A single number does not reveal this weakness.



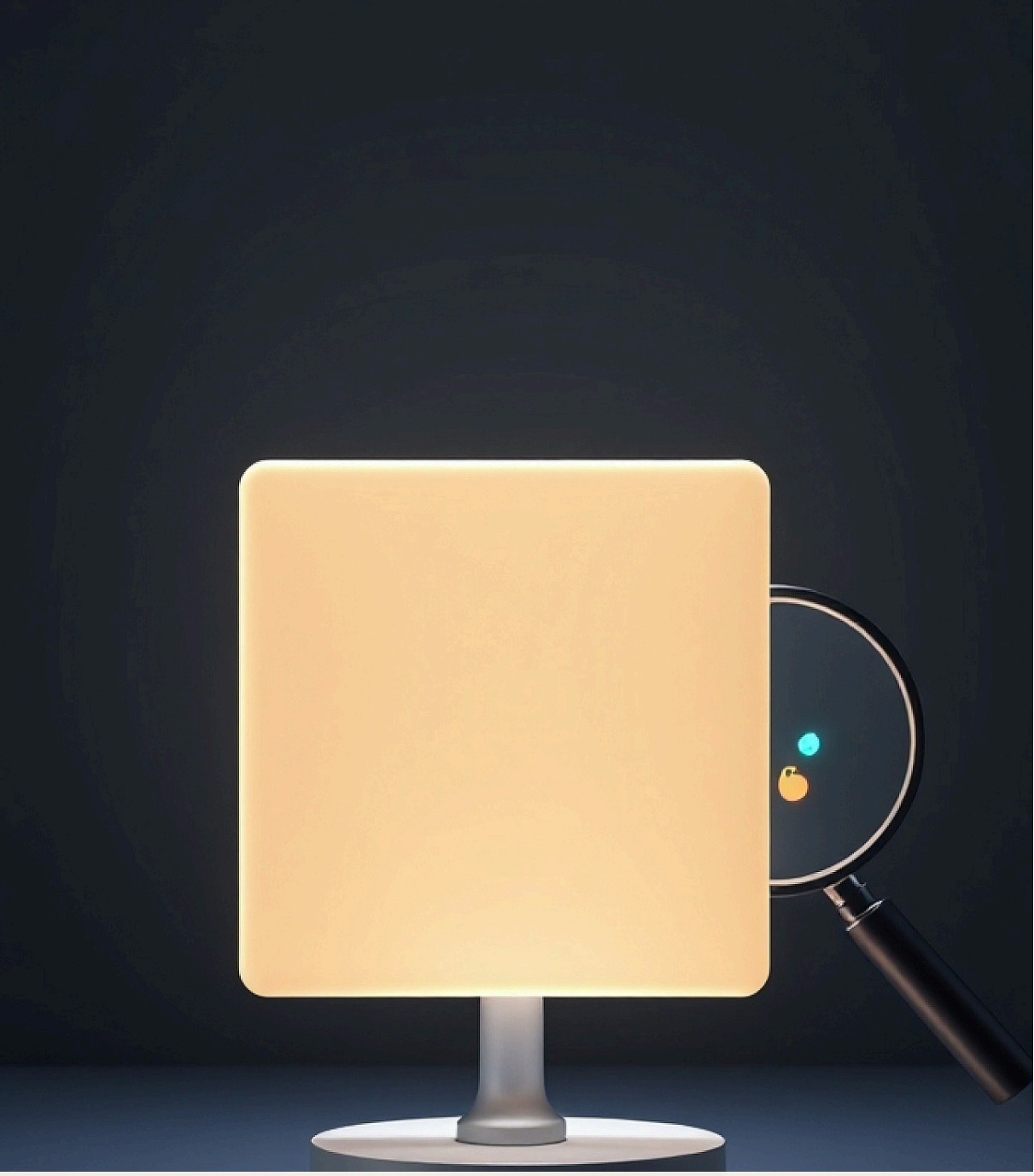
Rare cases matter because they are often the ones you care about most. A system that misses a rare but serious situation is not truly successful, even if its overall score is high. For example, missing a rare disease is worse than misreading a common cold. A good overall score does not cancel out a failure on rare cases.



Look at error types instead of only counting mistakes. Sort the mistakes into kinds, such as saying yes when the answer is no, or saying no when the answer is yes. Each kind of error has a different cost. Seeing which kinds of errors happen shows what the single overall score hid. Error types give a clearer picture than one number.



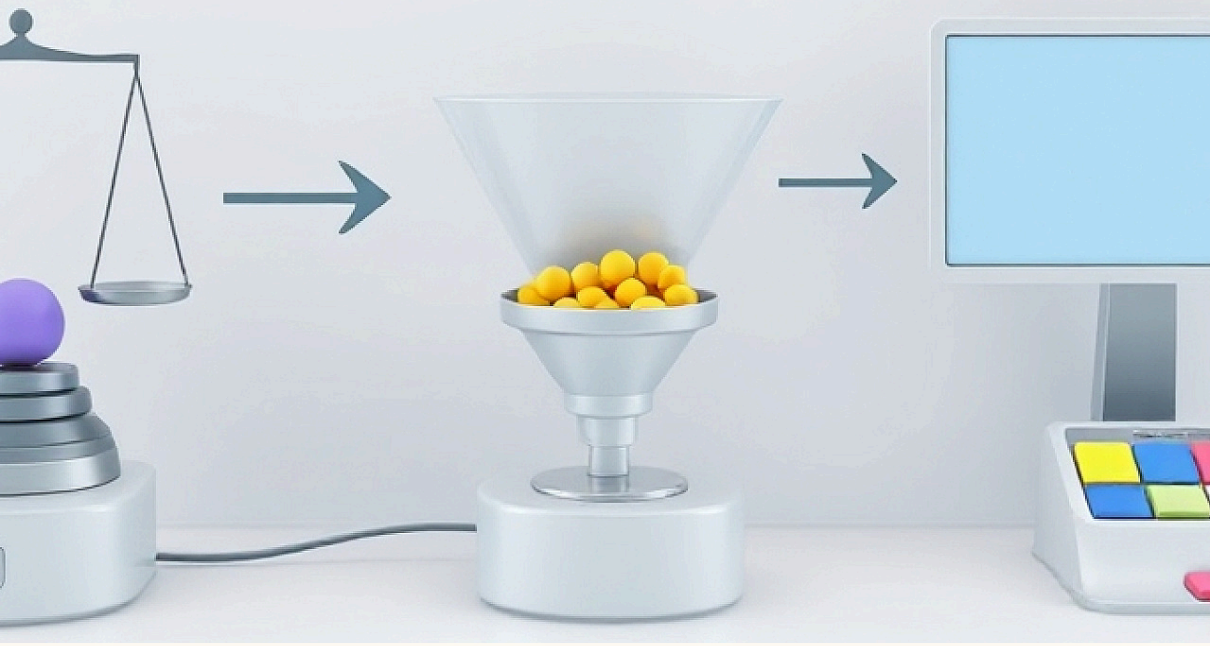
Look at example groups. Split the examples into groups, such as by category or by source, and check success in each group separately. A system can be strong in one group and weak in another. For example, it may work well on daytime photos but poorly on nighttime photos. Group-by-group scores reveal these differences that the overall score hides.



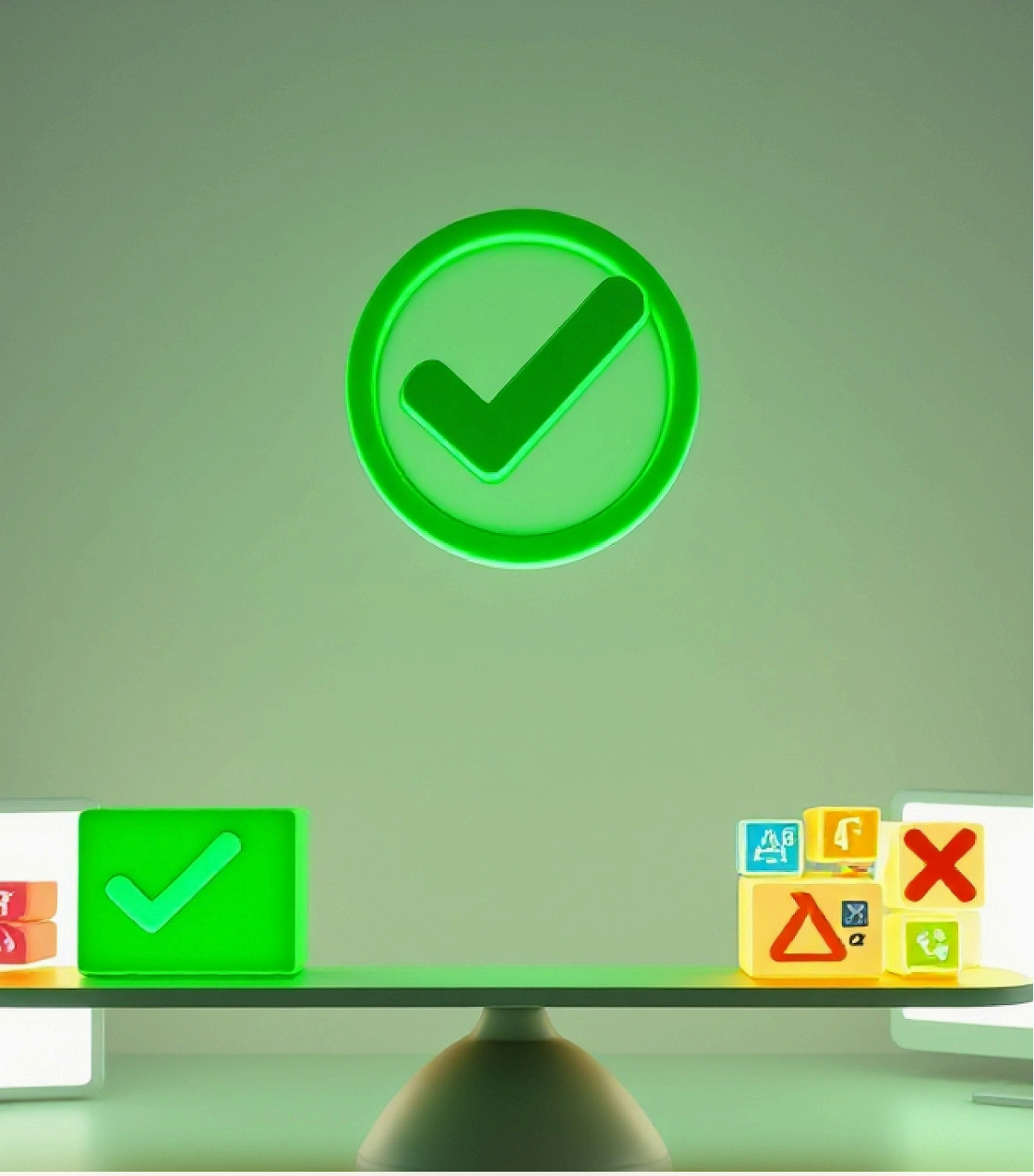
A common mistake is trusting one number alone. People often judge a system by its overall score and stop there. The fix is to always ask what that one number leaves out. Ask which error types occur and which example groups are weak. One number is a starting point, not the whole answer. Always look deeper before declaring success.



Another common mistake is ignoring the task. People sometimes use the wrong measure, such as counting correct decisions for a prediction task. Counting correct decisions does not make sense when the answer is a number. The measure must fit the task. Ask what kind of answer the system gives, then choose the matching measuring tool.



Put the checks together. First, pick the measure that fits the task: correct-rate for classification, distance for prediction. Then examine error types to see which mistakes happen. Then examine example groups to see where the system is weak. Together these three checks give a fuller picture of success than any single score alone.



What to remember: measure success with a tool that fits the task the system performs. Never rely on one score alone, because it can hide rare-case failures and weak groups. Always check error types and different example groups too. A system is truly successful when it does well on the task, on all kinds of examples, not just on average.