As everything, an AI must be tested. There might happen that a model makes an error, like human does. Sometimes we are unable to recognize a person or we see a dog instead of a cat. If a PC isn’t always right about the classification of a dog, that might be fatal for an autopilot by not seeing a pedestrian. If human can classify an emotion in the face correctly at the 70%, an AI cannot go a lot further.  
If we have a lot of data, we should always split it. The 4/5 in training data and 1/5 in test set. This is useful in order that the AI can work well on new data.