目前智能系统的测试工作存在以下几个问题：

1. 得到带标签的数据困难。
2. 低的测试覆盖率：现有的智能系统测试技术没有尝试覆盖DNN的不同规则，造成了测试揭示的故障不具备多样性。这里的“覆盖率”不是传统软件测试中代码或者分支覆盖率。在智能系统中，系统的逻辑是通过大量的数据学习得到的，不是开发人员指定的。因此，将传统软件的覆盖标准运用到智能软件系统测试中可能存在问题。

产生的数据集应当具有两个方面的特征：1，尽可能多的测试系统的行为；2，尽可能多的覆盖神经元范围：High neuron coverage alone may not induce  
many erroneous behaviors while just maximizing different  
behaviors might simply identify different manifestations of  
the same underlying root cause