数据驱动的定义：

数据驱动将**海量的数据**进行组织并**形成信息**，然后对形成的**信息进行整合和提炼**，在数据的基础上经过训练和拟合形成**自动化的决策模型**。

数据驱动式的测试：

在软件测试过程中，多元、大量的数据可以通过各种手段(收集测试日志、分析测试结果、总结以往的故障类型等)获得。然而，这些数据本身没有意义或者所含的信息量较少，需要进一步整理、分析并形成对揭示智能软件缺陷（程序中存在故障、模型训练不充分、训练数据不正确等）有帮助的信息，辅助测试人员揭示智能软件中的缺陷或者自动地预测缺陷类型和位置，具体来说：

1. **基于已有的数据产生新的数据：**智能软件的决策模型在测试的过程中易出现训练不足的情况，导致决策模型的智能程度达不到标准。在此情景下，通过在原有数据上做某种转化形成新的数据。已有的工作通过在原有的数据上做微小改动得到新的数据，本课题利用蜕变测试技术在已有的数据基础上通过智能软件的属性（蜕变关系）获取新的测试用例。
2. **判断缺陷类型：**引起智能软件的行为与预期不符的原因可能有多种：（1）模型训练不充分；（2）模型实现存在缺陷；（3）数据预处理不恰当；（4）多元异构数据协同程度低等。本课题拟采用数理统计的方式，分析大量的测试历史，对以上能引起智能软件行为异常的原因的发生概率进行估计，辅助测试工程师或者故障定位工程师揭示或者定位智能软件故障。