

# Benchmarking Deep Learning Testing Techniques: A Methodology and Its Application

Half Report Presentation, March 2020

**Group : HimanshuChuphal\_KristiyanDimitrov**

**DATX05, DIT910 SE**

**CHALMERS | GU**




# Outline

- Overview
- Key Idea
- Half Report - Overview
  - Results and Progress
  - Challenges
  - Time plan
- RQs
- Benchmarking Design
- Related Work
  - Timeline



## Key Idea

- Deep Learning (DL) Systems popularity
  - Security- and safety-critical systems
- Crucial role of testing
- Lack of Benchmarking in testing tools
- Benchmarking Design and tasks



## Overview - Half Report

Task	Status
Q1 <a href="http://www.robertfeldt.net/advice/master_thesis/feldt_thesis_halfway_questions_and_format.html">http://www.robertfeldt.net/advice/master_thesis/feldt_thesis_halfway_questions_and_format.html</a>	Results section in Report
Q2	Covered
Q3	Covered
Q4	Covered

# Q1, Q3: Results and Progress

Table 5.1: Results and Progress

Section/Task	Status	ETA	Comments
Introdcution	Complete	NA	NA
Related work and Background	Near to Complete	3 Days	Minor details in some sub-sections, cleanup and 'final review
Research Methods	50 %	2 weeks	Major work
Evaluation	3 DL Tools Simulation - Done Benchmarking Scripts - 20-30 %	2-3 weeks	Major work
Results	RQ 1 : Done, results published in RQ1 section	2 weeks	Hypothesis 1: Based on tasks - yes, might have to rephrase it.
	RQ 2 : 30 - 40 %		Hypothesis 2: Can be explained this based on our study of different tools.
	RQ 3 : Rephrasing needed to test properties, not based on feedback		Possibility to get feedback from industry experts : To be removed
Conclusion and Future work	Based on Results	3 days	NA

## Q2 : Main Challenges



- Making tools work and tools simulation
- Understanding the tools - very different from software testing
- Tools are in prototype and early framework state
- Identifying tools properties and performance measures
- Identifying test scenarios of DL testing tools
- Getting feedback from Industry

# Q4, Q5 : Time Plan Feasibility



- Results and Progress
- Very likely to be completed by end of May 2020
- Writing Seminar 2 on 4th May 2020



## RQs

- RQ 1 : Done, results published in the report
- RQ 2 : 30 - 40 %
- RQ 3 : Rephrasing needed to test properties, not based on feedback

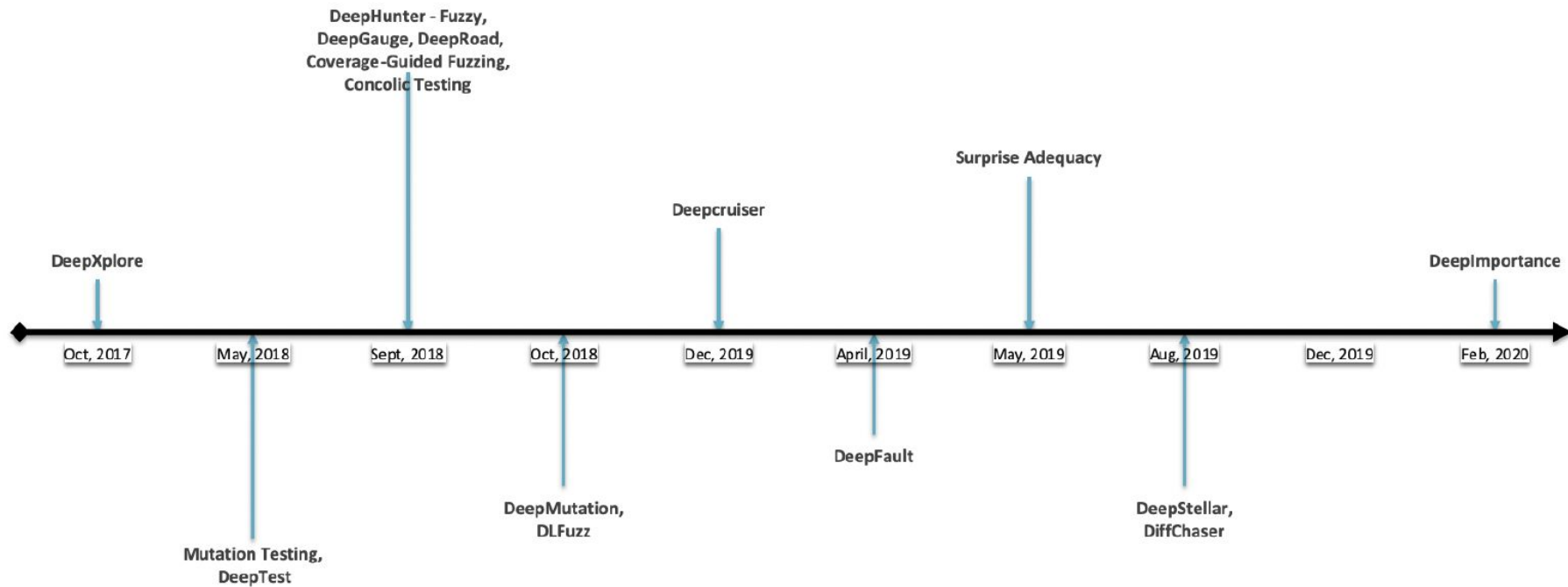




# Benchmarking Design

- DL testing tools requirements
- DL tests Scenarios
- Benchmarking tasks and sub-tasks
- Benchmarking method

# Related Work - Timeline





**Thank You!**

**Questions?**