

# APPLYING MACHINE LEARNING TO SOURCE CODE

---

Tim Nieradzik

19th July 2016

Ukrainian Catholic University

**GitHub:** 2M active repositories<sup>1</sup>

---

<sup>1</sup>Q4/2014, according to [GitHub.info](http://GitHub.info)

Large proportion of commits are  
fixes!

Project	Commits (fixes)	Commits (total)	%
CAFFE	838	3737	22%
TENSORFLOW	1418	5976	24%
NODE	4329	14444	30%
RUBY	7968	43978	18%
SPARK	5779	17034	34%
ELASTICSEARCH	4059	23309	17%

```
git log --grep='fix' -i | grep '^commit' | wc -l  
git log | grep '^commit' | wc -l
```

Can we detect faulty changes  
before they get committed?

# Sentiment Analysis

# Semantic code suggestions

QUESTIONS?