



“PerformoBot,  
please help me!”

Chatbot-supported  
Performance Evaluation

06.11.2019

Samuel Beck  
Lasse Merz  
Christoph Zorn  
Fabian Beck  
Leonel Merino  
Dušan Okanović  
André van Hoorn

@samu\_beck

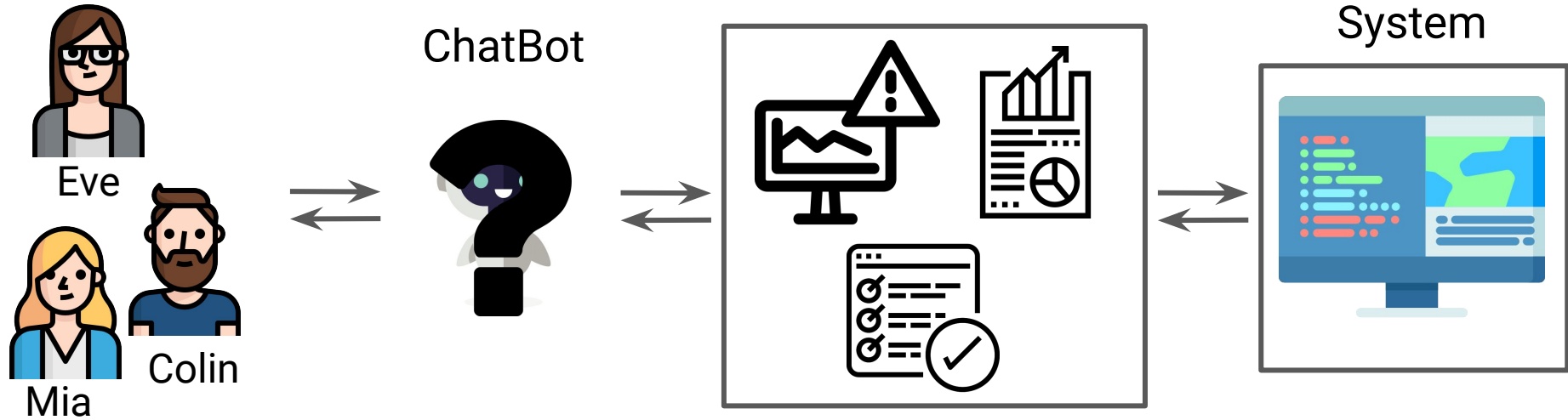
@beck\_fabian

@leonel\_merino

@okanovic\_d

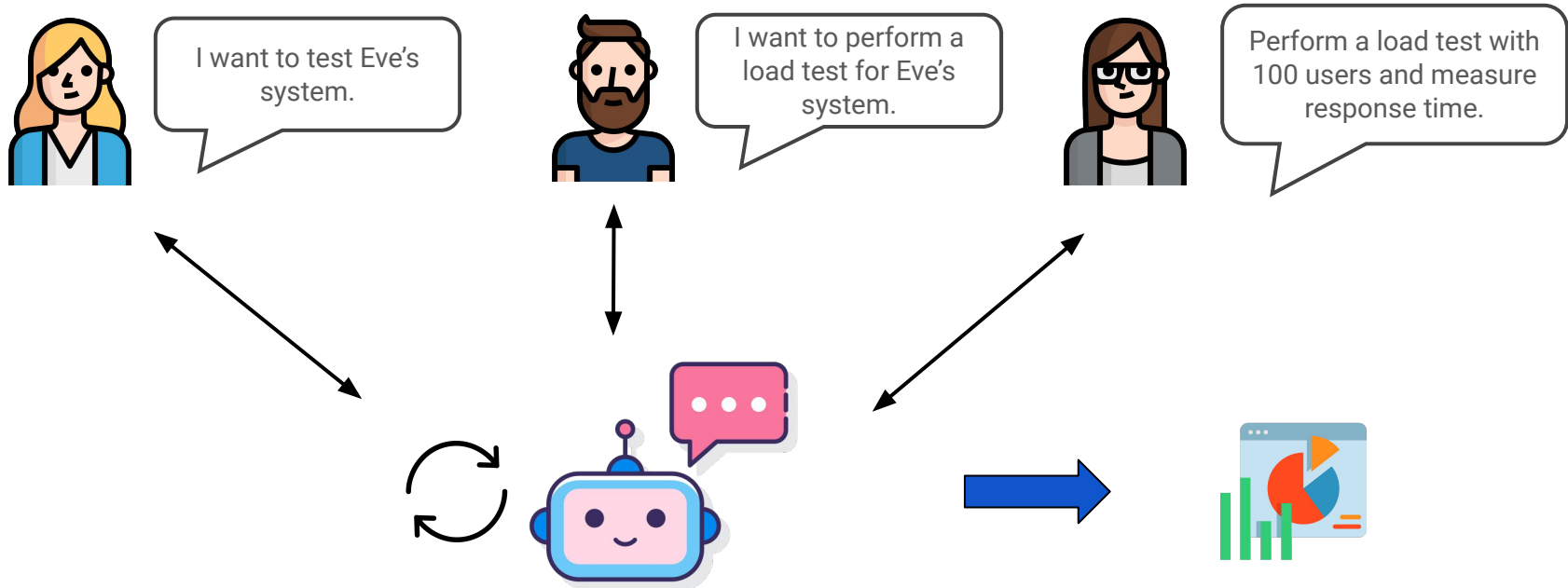
@andrevanhoorn

# Motivation



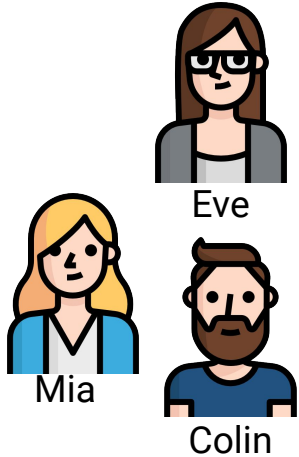
# Goals

- Evaluate the applicability of a chatbot to facilitate performance evaluations.
- Gather impressions from performance experts and novices about using a chatbot in this context.

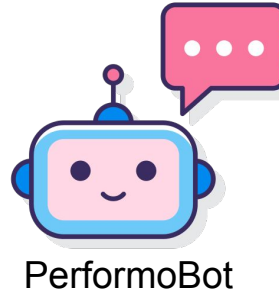


# Approach

## User Interaction



## Language Processing



## Analysis & Reporting





**Performobot** APP 21:09 Uhr

Hello! I am PerformoBot. I am here to help you with your performance evaluation. If you have any questions feel free to ask about anything you don't understand. Please state your performance evaluation concern.



**Mia** 21:09 Uhr

@Performobot What can you do?



**Performobot** APP 21:09 Uhr

I can help you to evaluate any system you want. Currently, I can create, execute, and show the results of load tests. Feel free to set up a load test regarding your concern or ask any questions.



**Mia** 21:10 Uhr

@Performobot Then let's perform a load test.



**Performobot** APP 21:10 Uhr

Great, let us set up a load test together. You need to provide the load you want to use, the domain which should be tested and the metric to be measured.



**Mia** 21:10 Uhr

@Performobot What is a load?



**Performobot** APP 21:10 Uhr

The load defines the amount of work a load test is processing during its execution. The load is usually the number of (virtual) users that produce work on a web service (e.g 100 users).



**Mia** 21:10 Uhr

@Performobot Please use 500 users.



**Performobot** APP 21:10 Uhr

Please enter the domain (e.g. [www.google.com](http://www.google.com)).

Otherwise, you can change the parameters, add more parameters from the parameter list or reset and start over.



**Eve** 21:14 Uhr

@Performobot execute



**Performobot** APP 21:14 Uhr

I executed your load test. Well done, you successfully completed the task! I prepared a report about this problem with a very similar query. I examined 500 users at <https://registration.uni-stuttgart.edu/login> and measured the server status. After clicking the "I'm done" button you will get a link to this report. Take a look and fill out the survey afterwards. Thank you, it was a pleasure!



**Eve** 21:14 Uhr

@Performobot result

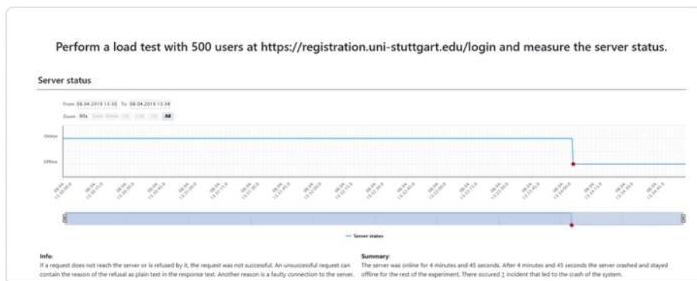


**Performobot** APP 21:14 Uhr

I created a report for your query. Click on the button below to see the details.

Summary: The server was online for 4 minutes and 45 seconds. After 4 minutes and 45 seconds, the server crashed and stayed offline for the rest of the experiment. There occurred 1 incident that led to the crash of the system

Here is the report. Click on the button for details. (57 kB) ▾



Inspect results

# Study: Research Questions

01

How can a chatbot help users do a performance evaluation through a load test?

02

How do users interact with PerformoBot?

03

What educational effect does PerformoBot have on participants?

04

What are the differences between inexperienced users and experts in the above categories?

Task

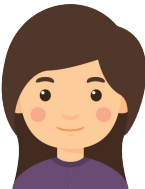


+

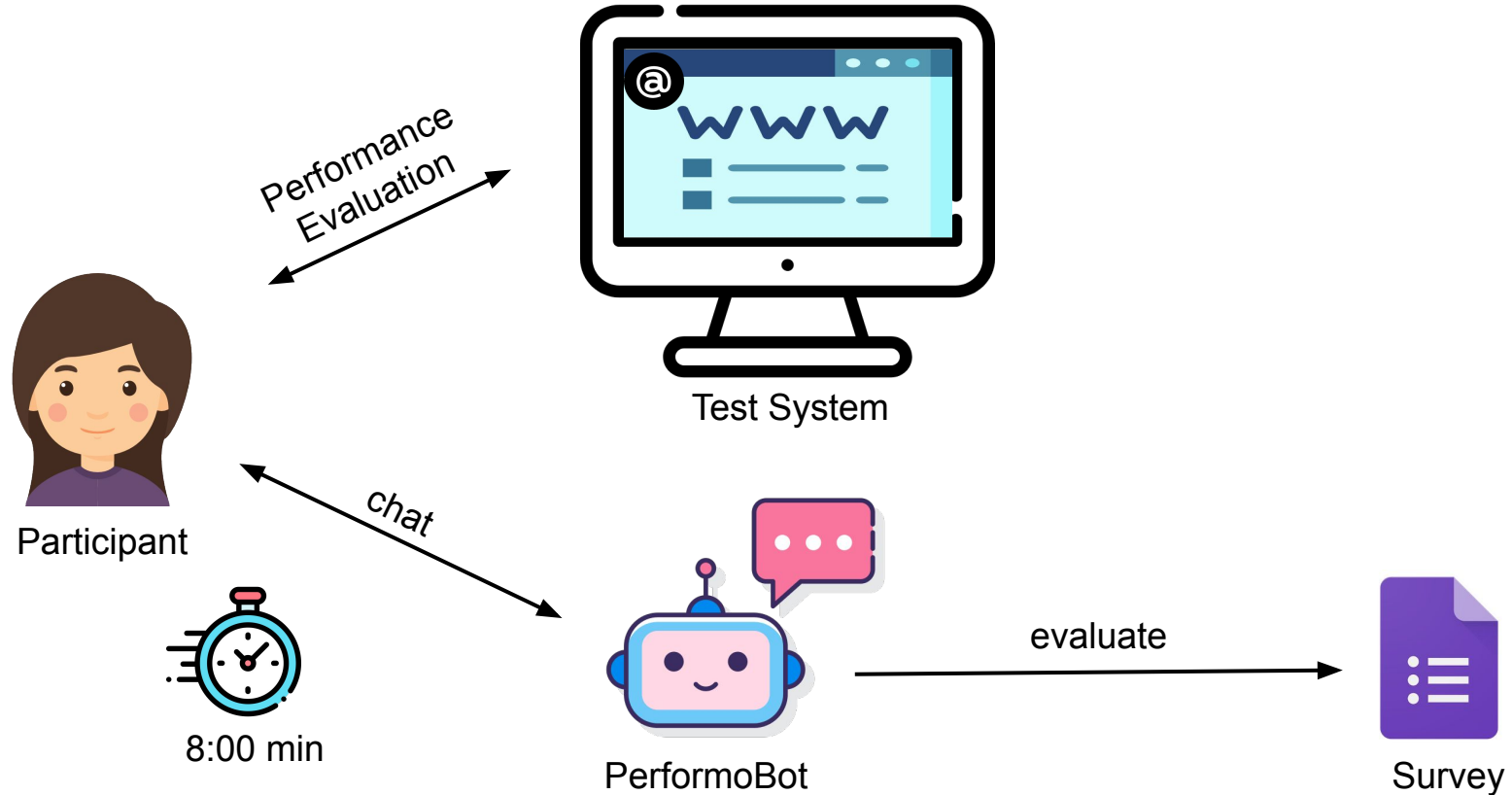
Survey



47 participants



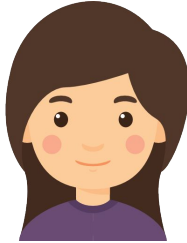
# Study: Task + Survey





# Study: Results

47 participants



Median of 10 interactions per participant



Average of 3.64 minutes session time



87.2% finished the task

No educational effect

Automatic execution & report creation



80.9% correctly answered the scenario

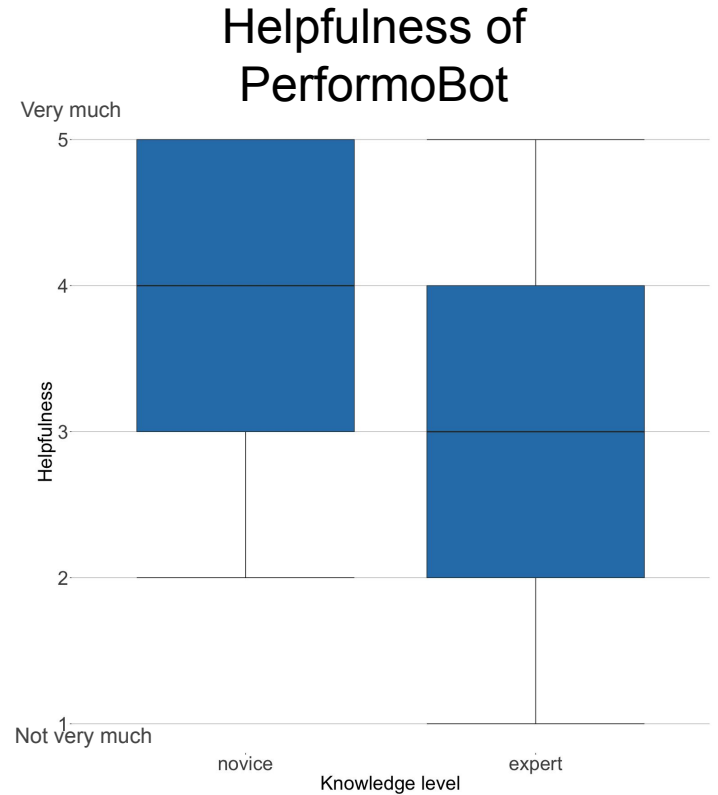
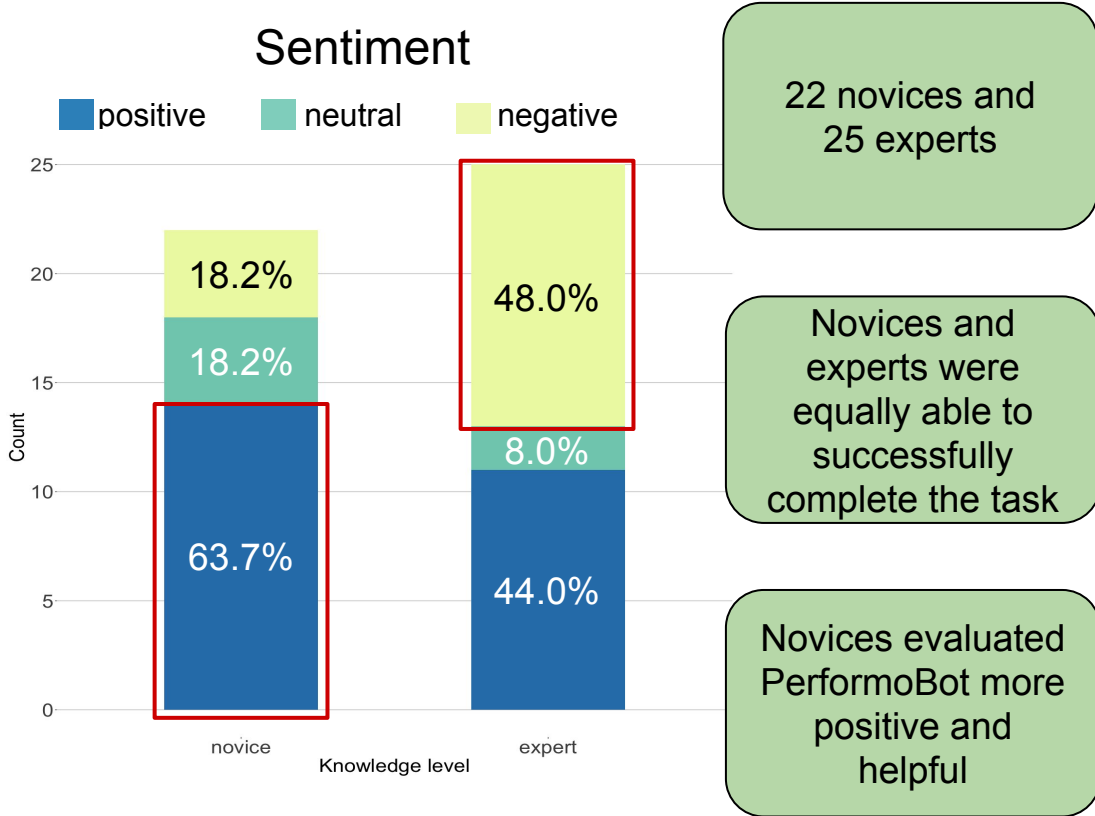


Too little guidance fallback mechanisms & missing keywords

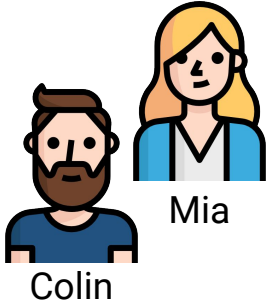
Explanations

Complex load testing capabilities

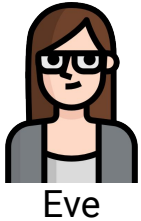
# Study: Inexperienced Users vs Experts



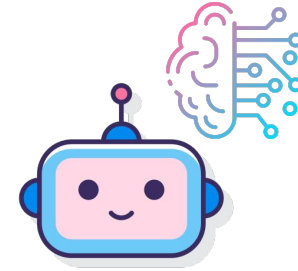
# Conclusion



Helpful for novice/inexperienced users



Needs advanced functionality for experts



- More performance concerns
- Improved usability features

<https://declare-project.github.io/PerformoBotSurveyWebsite/index.html>



# Load test intents and entities

**Intent:** Load test Newbie

Output context: loadtest

Entities:

- Analysis type

Training Phrases:

- I want to create a load test.
- Help me to define a load test.
- I want to perform a load test.

Please perform a load test for me.

**Intent:** Define a load test

Output context: loadtest

Entities:

- Load
- Domain
- Metric
- Analysis type (optional)
- Duration (optional)

Training Phrases:

- Loadtest with 55 users
- I want one with 3 hours, 22 users and response time.
- Create a load test with 30 users at www.myservice.de and measure server status.

Record the response time for 100 users on www.example.com.

Great, let us set up a load test together. I need to know what load you want to use, the domain to be tested and the metric which should be measured.

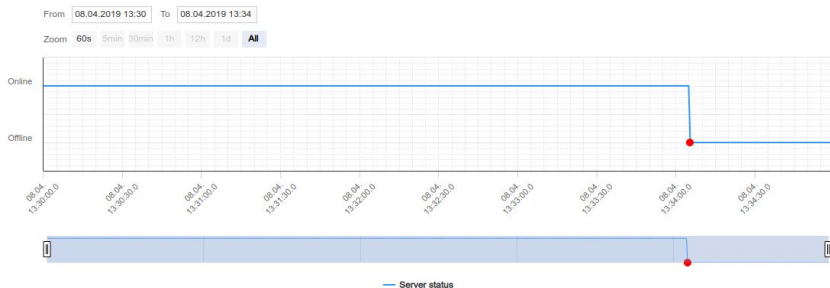
Alright I created a load test for you with the following parameters: load: 100 users, domain: www.example.com, metric: response time. If you are satisfied I can execute the load test. Otherwise, you can change the parameters, add more parameters from the parameter list or reset and start over.

# Performance Report

1

Perform a load test with 500 users at <https://registration.uni-stuttgart.edu/login> and measure the server status.

## Server status



**Info:** If a request does not reach the server or is refused by it, the request was not successful. An unsuccessful request can contain the reason of the refusal as plain text in the response text. Another reason is a faulty connection to the server.

**Summary:** The server was online for 4 minutes and 45 seconds. After 4 minutes and 45 seconds the server crashed and stayed offline for the rest of the experiment. There occurred 1 incident that led to the crash of the system.

## Analysis Summary

Analysis [JMeter](#)  
Tool:  
Duration: 4 m 59 s 322 ms  
Collected samples: 500  
Metrics: [elapsed time](#), [request status](#), [server status](#), [received traffic](#), [latency](#), [connection time](#), [response time](#)

**Experiment summary:**  
From 08.04.2019 13:30:00 to 08.04.2019 13:34:59 [500 values](#) were recorded. JMeter was used as the load testing tool. During the execution of this experiment (4 m 59 s 322 ms) [500 virtual users](#) produced load to the tested domain (<https://registration.uni-stuttgart.edu/login>).

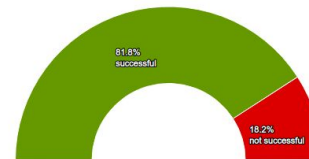
**Tool description:**  
Apache JMeter is an Apache project that can be used as a load testing tool for analyzing and measuring the performance of a variety of services, with a focus on web applications. JMeter can be used as a unit-test tool for JDBC database connections, FTP, LDAP, Webservices, JMS, HTTP, generic TCP connections and OS native processes. One can also configure JMeter as a monitor, although this is typically... [continue reading](#)<sup>[1]</sup>.

2

## Other metrics

Elapsed time

Request status



### Info:

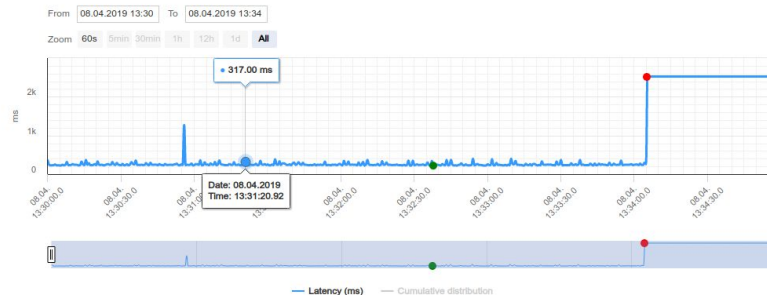
If a request does not reach the server or is refused by it, the request was not successful. An unsuccessful request can contain the reason of the refusal as plain text in the response text. Another reason is a faulty connection to the server.

### Summary:

From a total of [500](#) requests 81.8% (409) were successful. 18.2% (91) of all requests were not successful.

3

## Latency



	Min.	Max.	Mean.	Med.
Time:	after 2 m 37 s 723 ms	after 4 m 5 s 322 ms	-	-
Value:	220 ms	2523 ms	667.45 ms	244 ms

### Info:

Latency is the amount of time a message takes to traverse a system. In a computer network, it is an expression of how much time it takes for a packet of data to get from one designated point to another. It is measured as the time required for a request to be sent to the server and returned to its sender. Latency depends on the speed of the transmission medium and the delays in the transmission by devices along the way. A low latency indicates a high network efficiency.

### Summary:

The overall minimum was [220 ms](#) (after 2 m 37 s 723 ms) and the overall maximum was [2523 ms](#) (after 4 m 5 s 322 ms). The maximum of 2523 ms was recorded [91 times](#) while the minimum of 220 ms was recorded [once](#).

4

# References

Icons from flaticon (<https://www.flaticon.com/>) made by dave-gandy (question mark), monkik (people), freepik, eucalyp, smashicons (analysis, computers, clocks), flat-icons (performobot), roundicons (check mark, x, smileys)

chatbot gif: [https://cdn.dribbble.com/users/37530/screenshots/2937858/drib\\_blink\\_bot.gif](https://cdn.dribbble.com/users/37530/screenshots/2937858/drib_blink_bot.gif)

Messages gif: [https://i.gifer.com/origin/98/98447b873b927d46f752e9e0fc9c2910\\_w200.gif](https://i.gifer.com/origin/98/98447b873b927d46f752e9e0fc9c2910_w200.gif)

Okanović et al., <under review>:

Dušan Okanović, Samuel Beck, Lasse Merz, Christoph Zorn, Leonel Merino, André van Hoorn, Fabian Beck: Can a Chatbot Support Software Engineers with Load Testing? Approach and Experiences. Under review