Nghi D. Q. **Bui**

PH.D. IN COMPUTER SCIENCE

☐ (+84) 906620362 | ✉ bdqnghi@gmail.com | 🏠 bdqnghi.github.io | ⬤ bdqnghi | 🔲 quocnghi | 🐦 @QuocNghi91

Summary

My research interests lie at the intersection of Software Engineering, Natural Language Processing, and Machine Learning. With experience across industry and academia, I bring practical experience and theoretical expertise to this domain. The focus of my research is to build intelligent software engineering tools that enhance the development process of software developers.

My skills and achievements can be summarized as below:

- Have industry experience in software engineering (4 years) and data science (5 years).
- Familiar with software engineering development process and software engineering tools.
- Familiar with back-end system design.
- Good knowledge in program analysis techniques.
- Familiar with machine learning techniques.
- Hands on experience with neural networks implementation and design.
- Be able to design experiments for data science projects.
- Hands on experience to manage an engineer or a data science team.
- Published papers in top-tier conferences across different domains in Computer Science , such as Software Engineering (ICSE, ESEC/FSE, ASE), Artificial Intelligence (NeurIPS, AAAI, SIGIR).

Education

Singapore Management University   *Singapore*

PH.D. IN COMPUTER SCIENCE   *August 2016 - August 2020*

Ho Chi Minh University of Science   *Ho Chi Minh, Viet Nam*

BACHELOR IN COMPUTER SCIENCE   *August 2009 - August 2013*

- Graduated from Advanced Program in Computer Science, which is an honor program

Work Experience

Huawei Research Center   *Dublin, Ireland*

RESEARCH SCIENTIST   *December 2020 - Now*

- Design cross-lingual language models for source code learning.
- Propose models to combine deep learning and compiler theory to translate C++ into Rust.

Software Analytics Research Lab (SOAR), Singapore Management University   *Singapore*

POSTDOCTORAL RESEARCHER   *September 2020 - December 2020*

- Designed deep learning models for API recommendation task.

Noggin Asia   *Singapore*

HEAD OF DATA SCIENCE   *March 2017 - July 2020*

- Supervised junior members of the team.
- Designed ML system architecture to process and collect a large volume of user's personal data, such as geolocation, website activities, etc.
- Used statistical modeling to model user's behavior.
- Developed recommendation system for item recommendation on client's website.

The Open University   *United Kingdom*

VISITING SCHOLAR   *August 2019 - December 2020*

- Collaborated with the researchers at the Open University on the topic: "Exaplainable Program Representation Learning". The goal for this research is to propose techniques that can interpret the output of neural networks that are being applied to solve a wide range of problems in software systems

701Search (Chotot)   *Ho Chi Minh City, Viet Nam*

SENIOR SOFTWARE ENGINEER   *March 2015 - July 2016*

- Worked with the team to redesign and migrate the existed monolith architecture into the micro-service architecture
- Developed the push notification service for the mobile app
- Used ElasticSearch as the information retrieval tool for geo-location queries
- Applied Kafka as the event-sourcing system to communicate between internal micro-services

Atlassian   *Ho Chi Minh City, Viet Nam*

Software Engineer   *January 2014 - March 2015*

- Worked as a software engineer in the JIRA Service Desk team to develop useful libraries and features for internal usages, such as logging management dashboards to measure the quality of services
- Ensured the quality of source code by writing test cases and doing code review

Skills

Distributed Computing & Streaming	Kafka, RabbitMQ, Spark, Hadoop
Machine Learning & Analytics	Tensorflow, Pytorch, Keras, Pandas, Scikit-learn, SciPy, Numpy, Pyplot
Programming Languages	Python, Java, Scala, Javascript, Matlab, C/C++
Source Code Version Control	Git, SVN
Back-end Development	Spring Framework, Play Framework, NodeJS, Flask
Database	MySQL, PostgreSQL, MongoDB, Redis, ElasticSearch
DevOps	Docker, Ansible

Publications

[SIGIR'21] Self-Supervised Learning for Code Retrieval and Summarization through Semantic-Preserving Program Transformations, by **Nghi D. Q. BUI**, Yijun YU, Lingxiao JIANG, accepted at the 44th ACM Conference on Research and Development in Information Retrieval (SIGIR), Full Paper, 2021.

[JST'21] On the Generalizability of Neural Program Analyzers with respect to Semantic-Preserving Program Transformation, by Md. Rafiqul Islam RABIN, **Nghi D. Q. BUI**, Ke Wang, Yijun YU, Lingxiao JIANG, Mohammad Amin ALIPOUR, accepted at the Journal of Information and Software Technology.

[ICSE'21] InferCode: Self-Supervised Learning of Code Representations by Predicting Subtrees, by **Nghi D. Q. BUI**, Yijun YU, Lingxiao JIANG, in Proceedings of the IEEE/ACM 43th International Conference on Software Engineering, Full Paper, Technical Track, Madrid, Spain, 2021.

[AAAI'21] TreeCaps:Tree-based Capsule Networks for Source Code Processing, by **Nghi D. Q. BUI**, Yijun YU, Lingxiao JIANG, Full Paper, Main Track in Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI), Vancouver, Canada, 2021.

[ML4Sys-NeurIPS'19] TreeCaps:Tree-Structured Capsule Networks for Program Source Code Processing, by Vinoj Jayasundara, **Nghi D. Q. Bui**, Lingxiao Jiang, David Lo , in NeurIPS Workshop on ML for Systems, 2019, Vancouver, Canada.

[ASE'19] AutoFocus: Interpreting Attention-based Neural Networks by Code Perturbation, by **Nghi D. Q. BUI**, Yijun YU, Lingxiao JIANG, in Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering (ASE 2019), Research Track, New Ideas Papers, San Diego, California, United States, 2019.

[ESEC/FSE'19] SAR: Learning Cross-Language API Mappings with Little Knowledge , by **Nghi D. Q. Bui**, Yijun Yu, Lingxiao Jiang , accepted at the IEEE/ACM 27th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), Research Track, Tallinn, Estonia, 2019.

[ICSE'19] Towards Zero Knowledge Learning for Cross-Language API Mappings , by **Nghi D. Q. BUI**, in Proceedings of the IEEE/ACM 41th International Conference on Software Engineering (ICSE): ACM Student Research Competition Track (SRC), Montreal, Canada, 2019 - (**Bronze Medal**).

[SANER'19] Bilateral Dependency Neural Networks for Cross-Language Algorithm Classification, by **Nghi D. Q. Bui**, Yijun Yu, Lingxiao Jiang, in the 26th edition of the IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), Research Track, Zhejiang University in Hangzhou, February 24-27, 2019.

[ICSE'18] Hierarchical Learning of Cross-Language Mappings through Distributed Vector Representations for Code , by **Nghi D. Q. Bui**, Lingxiao JIANG, in Proceedings of the IEEE/ACM 40th International Conference on Software Engineering (ICSE): New Ideas and Emerging Technologies Results Track (NIER), Gothenburg, Sweden, 2018 - (**ACM SIGSOFT Distinguished Paper Award**).

[NL4SE-AAAI'17] Cross-Language Learning for Program Classification Using Bilateral Tree-Based Convolutional Neural Networks , by **Nghi D. Q. Bui**, Lingxiao JIANG, and Yijun YU . In the proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI) Workshop on NLP for Software Engineering, New Orleans, Lousiana, USA, 2018.

Honors & Awards

2019	**SMU Dean's List**, Ph.D. Program	*Singapore*
2019	**SMU Presidential Doctoral Fellowship**, Ph.D. Program	*Singapore*
2019	**ACM SIGSOFT Travel Grant Award**, ESEC/FSE 2019	*Tallinn, Estonia*
2019	**Bronze Medal**, ACM Student Research Competition, ICSE 2019	*Montreal, Canada*
2019	**ACM SIGSOFT Travel Grant Award**, ICSE 2019	*Montreal, Canada*
2018	**ACM SIGSOFT Distinguished Paper Award**, ICSE 2018	*Gothenburg, Sweden*
2016	**Ph.D. Scholarship from the Singapore Ministry of Education**, Ph.D Program	*Singapore*

Teaching Experience

IS706: Software Mining and Analysis: Teaching Assistant

IS212: Software Project Management: Teaching Assistant

Academic Service

Program Committee: Mining Challenge Track, the 18th Mining Software Repository Conference (MRS 2021).

Program Committee: Shadow PC Track, the 18th Mining Software Repository Conference (MRS 2021).

External Reviewer: the 36th International Conference on Software Maintenance and Evolution (ICSME 2021).

External Reviewer: the 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2021).