THANH HAI NGUYEN

Phone: +1(575) 650-8946 ⋄ Email: thanhnh2911@gmail.com ⋄ Austin, TX

LinkedIn: linkedin.com/in/thanh-hai-nguyen-320017118

Github: github.com/thanhnh-infinity ⋄ Web-site: thanhnh-infinity.github.io

PERSONAL STATEMENT AND OBJECTIVE

I graduated with my Ph.D degree in Computer Science - Artificial Intelligence at New Mexico State University. My research interest focuses on Knowledge Representation, Automated and Non-monotonic Reasoning, Automatic Planning, Scheduling, Domain Modelling, Solving Algorithms, Semantic Web and Semantic Integration, Web Services Composition Solutions, Natural Language Processing (NLP) and Natural Language Understanding (NLU). In addition, I am interested in Machine Learning and Collective Intelligence (recommendation system, discovering groups, searching, and ranking, collaborative filtering, document filtering, generative modelling, advanced classification, regression etc.) and Big-Data processing (with Map-Reduce framework). In my academic time, I had publications on top-tier conferences and journals.

I am currently employed as an **AI Researcher** and **Reseach Engineer** at **Elemental Cognition Inc.** (ec.ai). My role centers on Knowledge Representation, Non-monotonic and Automated Reasoning, Domain Modeling, Solving Algorithms, Logic Programming, Meta-programming, Compiler, and Natural Language Processing within EC's AI platforms and solutions. Additionally, I serve as a **Senior Software Engineer**, leveraging my expertise in Software Engineering and Development to contribute to EC's commercial applications.

I have 15 years of industry experience in software and AI research and development, with a broad skill set that encompasses both back-end and front-end development. I am seeking a position in Machine Learning and Artificial Intelligence within an industrial company where I can leverage my expertise to contribute to business goals, expand my knowledge, and further develop my research interests in these fields.

EDUCATION

Ph.D in Computer Science, New Mexico State University, USA — 2013 - 2021
Thesis: An ASP-Based Automatic Web Services Composition Framework and Its Applications

Master of Information Technology, James Cook University, Australia — 2010 - 2011
Major: Business Informatics. Graduated with an honours degree. GPA: 3.75

Bachelor of Computer Science, Hanoi University of Technology and Science — 2003 - 2008
Graduated with an honour degree, ranked as the second highest performance student. GPA: 3.8

WORKING EXPERIENCE

AI Research Engineer, Senior Software Engineer — July 2021 - current
Elemental Cognition Inc. — *New York, NY*

- **Knowledge Representation and Reasoning**: My primary work at EC involves developing domain-specific knowledge bases using logical reasoning and building comprehensive logical inference mechanisms to represent domain-independent knowledge bases using logical reasoning and meta- programming. I am currently working on developing a new formal representation language for domain modelling called **Cogent**, which (1) is a subset of English, (2) has formal syntactic and semantic properties of a representation language, and (3) can be executed by a Reasoning Engine. In this project, I serve as both a *Compiler Front-End Engineer* (responsible for compiling Cogent to produce valid IML) and a *Compiler Back-End Engineer* (handling the transformation of IML into low-level Cordial rules and facts). Additionally, I work as a *Knowledge Engineer* to design and implement the semantics for the language.

- **Reasoning Engine Development**: I am involved in developing the Reasoning Engine to perform reasoning tasks based on logic programming and constraint-oriented methodology.

- **Knowledge Implementation**: I work on developing Automated and Non-monotonic Reasoning algorithms to implement and model the EC's application domains based on logic programming languages such as Cordial and ASP.
- **Software Engineering/Development**: I contribute to software development for EC's commercial applications, ensuring robustness and efficiency across various projects and initiatives. For example, in Round-Travel-World project, my responsibilities include developing new theory features, implementing associated back-end features, supporting back-end team, fixing bugs related to theory, and conducting unit, integration and automated testing.

Skills: *Natural Language Processing, Software Engineering, Knowledge Representation and Reasoning, Logic Programming, Natural Language Understanding, Domain Modeling, Compilers, Automatic Reasoning and Planning, Functional Programming, Scala, Scalability, Reliability, Meta-Programming.*

Senior Software Architect/Engineer   2013 - 2021
Jornada Experimental Range, USDA and New Mexico State University   *Las Cruces, NM*

- Designed and initiated the software architecture for the entire LandPKS project, ensuring seamless integration between all project components. This involved overseeing the system design from inception to final end-user experience.
- Developed data analytic systems and prediction models based on machine learning techniques to extract maximum knowledge from provided soil profiles, weather, climate conditions, water data, GIS, and other relevant datasets. These models are used to comprehensively analyze soil potential, providing valuable information and prediction data for farmers and users.
- Developed the LandPKS back-end API system, enabling clients (such as web portals and mobile apps) to connect to and access LandPKS core data, and to interact with data analytic systems and prediction models.
- Developed mobile applications for both Android and iOS platforms, enabling users to collect and interact with LandPKS data as well as view predictive data visualizations.
- Developed a big-data processing module using Hadoop to create accessible climate data and soil profiles for locations worldwide.
- Developed a data portal that analyses and displays data.

Skills: *Back-End Web Development, SQL, NoSQL databases, Scalability, Mobile Application Development, RESTful Web Services, API System.*

Senior Software Engineer   2015 - 2018
New Mexico State University, Phylotastic   *Las Cruces, NM*

- Implemented semantic integration for the Phylotastic Web Services System.
- Integrated and applied an Automatic Web Services Composition Framework to create, manipulate, and evolve phylogenetic biology workflows.
- Developed and extended Phylotastic web services.
- Developed a mobile application for generating phylogenetic trees based on Phylotastic web services, available on both Android and iOS.
- Developed Phylotastic Portal

Skills: *Formal Semantic, Mobile Application Development, Semantic Web, RESTful Web Services, API/SDK*

Deputy Director of Research & Development Center   2012 - 2013
Vivas LLC., VNPT Technology, member of Vietnam Post & Telecommunication Group (VNPT)   *Hanoi, Vietnam*

- Managed the Research and Development department's technical area.
- Researched and applied new technologies to current projects.
- Consulted and led the software architecture development for various company products.

Senior Software Architect/Engineer & Technical Team Leader   2011 - 2013
Vivas LLC., VNPT Technology, member of Vietnam Post & Telecommunication Group (VNPT)   *Hanoi, Vietnam*

- Designed and developed a multi-screen video streaming platform.
- Designed and developed CDN (Content Delivery Network) and CMS (Content Management System) systems, along with the multi-screen video streaming platform API.
- Developed a video streaming mobile application for Android, iOS, and a web portal.

Software Engineer 2008 - 2009
FPT-IS/FPT-Software, FPT Company *Hanoi, Vietnam*

- Developed a system (WTCS_TT) enabling banks to directly collect tax from taxpayers.
- Developed a system to manage and collect personal income tax for the Vietnam Tax Department.

RESEARCH EXPERIENCE

Research Assistant Aug 2013 - May 2020
Computer Science Department, New Mexico State University *Las Cruces, NM*

My primary research focused on Automatic Web Services Composition in the Semantic Web. I developed a comprehensive end-to-end AI system to collect requirements from users in natural language and automatically explore workflows to satisfy these requirements. Refinement, modification and interactions with workflows are advanced features in our framework as well as the recovery function and repairing. In this framework, Planning Engine and Re-Composition Engine are the most important components, and are developed based on Answer Set Programming ASP. This framework includes:

- Web Service Registry: A CMS (Content Management System) web-based system that collects and stores web services from providers. It details how to invoke and execute a web service.
- Ontology: An explicit specification of a conceptualization using OWL (Web Ontology Language).
- AI Core Composition Engine: This includes:
  - Planning Engine: Generates workflows based on web service composition algorithms and AI planning theory.
  - Planning Monitor: Monitors and controls generation processes, ensuring failure correctness, adaptability, scalability, and reliability.
  - Workflow Execution: Executes and invokes all web service components in the workflow to achieve the goal.
  - Re-Composition Engine: Regenerates updated workflows incorporating additional user requirements and constraints.
  - Planning Engine API: Publishes the functionalities of the AI Core Composition Engine for use in external applications via an API/SDK.
- Ontology Accessing Engine: Software based on SPARQL engine and various programming languages to access, retrieve, and interact with ontology classes, instances, and properties.
- Workflow Description Graphical Tool (WDGL): A UI software tool for end-users to build workflows based on their requirements.
- Natural Language Processing Module: Describes workflows in human-readable language.

Research Assistant Aug 2018 - May 2021
Cyber-Physical System, New Mexico State University, NIST *Las Cruces, NM*

I conducted research and contributed to the development of techniques for identifying critical dependencies in complex and critical systems, as well as for evaluating, verifying, and restoring their properties. This research introduced a formal definition of Cyber-Physical Systems (CPS) aligned with the CPS Framework proposed by NIST. Using this definition, we were able to precisely formalize and address various issues within CPS through Answer Set Programming (ASP). These issues included identifying dependencies or conflicts between concerns, determining how to mitigate problems, and selecting the most suitable mitigation strategy for a given situation. The research demonstrated how ASP can be effectively used to implement solutions to these problems. My primary contributions to this project include:

- Contributing to Theoretical Development:
  - Played a significant role in defining essential concepts for the project.
  - Developed algorithm for evaluating dependencies and conflicts, identifying the most vulnerable components, computing a strategy for mitigating an issue, selecting the most suitable mitigation strategy in complex systems.
- Implementing Sophisticated Reasoning Engine Tools:
  - Designed the architecture of the tools derived from our research.
  - Implemented parts of these tools and supervised the development efforts conducted by students.

TECHNICAL EXPERTISE

Strong problem solving and critical thinking skills.

Expert knowledge, skills and experience on PL Compiler pipeline, Meta-Programming.

Expert knowledge, skills and experience of programming languages:

- *Imperative Programming Languages*: JAVA, Python, C#, VB.NET, C, C++, Objective-C, Swift, GO language, Rust, NodeJS, JavaScript and Libraries, Typescript, Perl, Ruby, PHP, Dart and Kotlin.
- *Declarative Programming Languages*: Answer Set Programming (ASP), Prolog (SWI-Prolog, GNU-Prolog, B-Prolog), Cordial, Scala, Haskell, SQL & PL-SQL, R language.

Enterprise Development Framework: Java EE(Struts, Spring MVC, Spring Boot, Hibernate, Vert.x), Python (Django, FastAPI, Flask, CherryPy, Falcon), ASP.NET, Ruby on Rails, AngularJS, ReactJS, PHP and its framework as well as Web development libraries (JQuery, Bootstrap, HTML5, CSS).

Scalability, Reliability and Performance: Content Delivery Network (CDN), Load Balancing, Caching, Database Partitioning, Parallel Programming.

Data Mining and Machine Learning Frameworks: TensorFlow, PyTorch, Scikit-Learn, NumPy, Matplotlib, Pandas, SciPy, R packages, Weka.

AI Knowledge Representation Languages: Answer Set Programming, Prolog, FOL, Ontology RDF, Actional Description Languages (PDDL, AL, ADL), Cordial, Constraints-Oriented methodology.

Natural Language Processing and Understanding: Natural Language Toolkit (NLTK), Standford CoreNLP, Grammatical Framework, Attempto Controlled NL.

Big-data processing: Map-Reduce frameworks such as Apache Hadoop, Apache Spark.

Databases: Oracle Database, SQLServer, MySQL Database, PostgreSQL; and NoSQL Database Technology especially on MongoDB, Cassandra, Redis, FireBase, CouchDB, and PouchDB.

Mobile Development: Native iOS (Swift, Objective-C), Native Android both NDK and SDK (Java, Kotlin and C), Hybrid Apps (Cordova, IONIC, PhoneGap, NativeScript, React).

Cloud Computing: Google App Engine, Google Cloud API, Google Cloud Data-store, Google Computer Engine, AWS Elastic Beanstalk, AWS EC2, FireBase.

Automatic testing/build frameworks and CI/CD: Selenium and Appium, Jenkins, Travis and Github CI/CD.

IDE: IBM Websphere Application Developer, IBM Websphere Application Server, Visual Studio .NET, Eclipse for J2SE, J2EE, Eclipse for PHP, Rational Rose, Offices, NetBean, R Studio, XCode, InteliJ.

Other skills: Git (Github, Bitbucket), Communcation, Latex, Window, Linux, Unix.

PROJECTS

Cogent - Formal knowledge representation language. Cogent is EC's knowledge authoring tool for building modular domain models in a low- or no-code manner using intuitive natural language (instead of requiring a formal ontology language). These domain models are then used as knowledge base in EC's Reasoning Engine to reason over the problem customers are solving with EC AI Platform. Cogent combines the fluency of Large Language Models with the problem-solving precision of Logical Reasoning to automate deep expertise and help business convert customers better than ever. With Cogent, customers (companies, businesses organizations) are empowered to self-serve when navigating complex purchasing decisions or resolving tough issues where they would normally need to consult highly-trained and expensive human experts and maintain the power and flexibility of the best human agents. More information: cogent.ec.ai

Round-The-World (RTW). RTW is an EC project in partnership with OneWorld, designed to revolutionize complex travel planning and booking. RTW engages with customers to understand their preferences and constraints, then optimizes these against millions of flight combinations. It simplifies fare rules, helps customers understand their options, and proactively addresses potential issues to ensure a seamless experience.. More information: rtw-va.oneworld.com

Land-Potential Knowledge System (LandPKS). LandPKS aims to serve as the primary access and development point for free, simple to use, and locally appropriate technologies and knowledge needed for making sustainable land management decisions. The program develops innovative web-based and mobile data collection and analysis methods and tools to support local land use planning and to optimize design and implementation of food security, land restoration, climate change adaptation, and biodiversity conservation programs. Main webs site: landpotential.org; Data portal: portal.landpotential.org

Phylotastic. The Phylotastic project aims to make accessing the tree of life as easy as getting directions from a mapping app. Traditionally, obtaining a comprehensive phylogeny of species requires specialized knowledge and complex software. To simplify this, we developed a robust system that delivers phylogenetic knowledge on the fly. This system includes a web portal, a mobile app for customizable workflows, over 30 web services, and toolkits in R and Python. More information: phylotastic.org

TVOD - Multiple-screens Video Streaming Platform. This platform is built on two core systems: a Content Management System (CMS) and a Content Delivery Network (CDN). Additional components include the TVOD API, Transcoding Software, Portal, Mobile Applications, and Admin System. Together, they form a comprehensive solution for delivering video content (both Video on Demand and Live Streaming) to end-users across multiple screens and platforms. More information: vivas.vn/giai-phap-multimedia-streaming-platform

VNPT Content Delivery Network (CDN). We designed this solution to deliver content to end users reliably and efficiently. The CDN consists of a large distributed network of server nodes deployed on VNPT's backbone, with servers located in data centers across Vietnam. This network optimizes traffic and ensures the fastest download speeds and highest quality for end-users. The CDN allows content providers to deliver their content with top quality, without the need for costly hosting infrastructure. More information: vivas.vn/giai-phap-cdn

RESEARCH INTERESTS

- **Artificial Intelligence**: Knowledge Representation and Advanced Reasoning (Argumentation, Logic Programming, Answer Set Programming, Non-monotonic Reasoning), Reasoning about Actions and Changes, Automatic Planning, Scheduling, Autonomous Agents, Robotics, Control Theory, Inductive Logic Programming, Web Services Composition, Semantics Web, Natural Language Processing and Understanding.

- **Machine Learning and Collective Intelligence**: Recommendation system, discovering groups, searching and ranking, collaborative filtering, document filtering, generative modelling, advanced classification, regression.

- **Data Mining and Knowledge Discovery**: Combine Data Mining and Knowledge Discovery with available methods of classification, clustering, association rules to construct special methods to mine data. Build smart systems such as Recommendation System, Finding Similarity, Collaborating Filtering, Social networking.

PUBLICATIONS

- **The Land Potential Knowledge System: Generating site-specific estimates of land potential productivity and degradation risk using a mobile application and cloud computing**. 2015 AgMIP 5th Global Workshop. Joshua Beniston, Adam Beh, **Thanh Nguyen**, Lilian Ndungu, Jason Karl, Jeffrey Herrick.

- **Automatic Web Services Composition for Phylotastic**. PADL 2018 20th International Symposium on Practical Aspects of Declarative Languages. *Conference Full-Paper*. **Thanh Hai Nguyen**, Tran Cao Son, Enrico Pontelli. https://doi.org/10.1007/978-3-319-73305-0_13

- **Phylotastic: An Experiment in Creating, Manipulating, and Evolving Phylogenetic Biology Workflows Using Logic Programming**. ICLP 2018 34th International Conference on Logic Programming. TPLP Theory and Practice of Logic Programming Journal 2018. *Conference Full-Paper and Journal Research*. **Thanh Hai Nguyen**, Tran Cao Son, Enrico Pontelli. https://doi.org/10.1017/S1471068418000236

- **An Automatic Web Services Composition Framework over Biological Domain and Specifications**. The 15th International Conference on Logic Programming and Nonmonotonic Reasoning LPNMR 2019. *Doctoral consortium presentation*. **Thanh Hai Nguyen**. https://dbai.tuwien.ac.at/lpnmr-dc-19/

- **Phylotastic: Improving Access to Tree-of-Life Knowledge with Flexible, on-the-Fly Delivery of Trees**. Evolutionary Bioinformatics Journal 2020. *Journal Research*. **Thanh H Nguyen**, Van D. Nguyen, Abu Saleh Md. Tayeen, H. Dail Laughinghouse IV, Luna L.Sanchez- Reyes, Enrico Pontelli, Dmitry Mozzherin, Brian O'Meara, Arlin Stoltzfus. https://doi.org/10.1177/1176934319899384

- **Design and Implementation of Phylotastic, a Service Architecture for Evolutionary Biology**. International Journal of Software Engineering and Knowledge Engineering IJSEKE 2020. *Journal Research*. Abu Saleh Md Tayeen, **Thanh H Nguyen**, Van Nguyen, Enrico Pontelli. https://doi.org/10.1142/S0218194020500382

- **On Repairing Web Services Workflows**. The 22nd International Conference on Practical Aspects of Declarative Languages PADL 2020. *Conference Full-Paper*. Abu Saleh Md Tayeen, **Thanh H Nguyen**, Tran Cao Son, Enrico Pontelli. https://doi.org/10.1007/978-3-030-39197-3_3

- **Specifying and Reasoning about Concerns in Cyber-Physical System Using Answer Set Programming**. KR-2020 17th International Conference on Principles of Knowledge Representation and Reasoning. *Poster presentation*. **Thanh H Nguyen**, Tran Cao Son, Matthew Bundas, Marcello Balduccini, Kathleen Campbell Garwood, Edward R. Griffo. https://kr2020.inf.unibz.it/program/

- **Reasoning about Trustworthiness in Cyber-Physical Systems Using Ontology-Based Representation and ASP**. PRIMA-2020 The 23rd International Conference on Principles and Practice of Multi-Agent Systems. *Conference Full-Paper*. **Thanh H Nguyen**, Tran Cao Son, Matthew Bundas, Marcello Balduccini, Kathleen Campbell Garwood, Edward R. Griffo. http://dx.doi.org/10.1007/978-3-030-69322-0_4

- **Specifying and Reasoning about CPS through the Lens of the NIST CPS Framework**. TPLP Theory and Practice of Logic Programming Journal 2021. *Journal Research*. **Thanh H Nguyen**, Tran Cao Son, Matthew Bundas, Marcello Balduccini, Kathleen Campbell Garwood, Edward R. Griffo. https://doi.org/10.1017/S1471068422000035

- **Towards a Framework for Characterizing the Behavior of AI-Enabled Cyber-Physical and IoT Systems**. IEEE World Forum on Internet of Things (WF-IOT 2021). *Conference Full-Paper*. Matthew Bundas, Chasity Nadeau, **Thanh H Nguyen**, Jeannine Shantz, Marcello Balduccini, Tran Cao Son. https://doi.org/10.1109/WF-IoT51360.2021.9595077

- **Facilitating stakeholder communication around AI-enabled systems and business processes**. Research Handbook on Artificial Intelligence and Communication 2023. *Book chapter*. Matthew Bundas, Chasity Nadeau, **Thanh H Nguyen**, Jeannine Shantz, Marcello Balduccini, Tran Cao Son. https://doi.org/10.4337/9781803920306

- **LandPKS (Land Potential Knowledge System): Mobile App for Extension, Land-Use and Project Planning, M&E and On-Farm Research**. USDA Agriculture Data Articles. *Book chapter*. Herrick, Jeffrey E.; Quandt, Amy; Kenny, Ciaran; Contreras, Maximilian; Neff, Jason; Jang, Won Seok; Maynard, Jonathan; Peacock, George; Salley, Shawn; Daniel, Elifadhili; Buni, Adane; Ndungu, Lilian; Herrera, Jolene M.; **Nguyen,**

Thanh; Terrones, Luis; Karl, Jason; Kimiti, David; Nauman, Travis; Uruma, Kevin; Schrader, Scott; Courtright, Ericha; Van Zee, Justin. https://hdl.handle.net/10113/AA22842

LEADERSHIP

- Deputy Director in Research and Development Center at Vivas Ltd. - A member of the Vietnam Post & Telecommunication Group (VNPT). VNPT is a leading company in telecommunications added services, multimedia streaming platforms, and content delivery networks (CDN).

- Technical Leader in Vivas Ltd., responsible to lead all software development in the company.