

# CLEON LIEW

[linkedin.com/in/cleon-liew](https://www.linkedin.com/in/cleon-liew) | [srcliew@uwaterloo.ca](mailto:srcliew@uwaterloo.ca) (226-201-3572) | [cleon01.github.io](https://cleon01.github.io) (My Portfolio)

## SKILLS

---

**Programming Languages:** Python, R, SQL, Bash Shell Scripting, VBA, MATLAB, LaTeX

**Libraries:** scikit-learn, Keras, PyTorch, TensorFlow, NumPy, SciPy, Pandas, Matplotlib, ggplot2, Cuda, Hugging Face, BeautifulSoup, Scrapy, Splash, Selenium, Appium

**Technologies:** Git, Bitbucket, Jira, Confluence, Power BI, Tableau, Microsoft Office Suite, Excel (Pivot Tables, VBA, Macros), AWS Batch, AWS CloudWatch, Jupyter Notebook, Linux, PuTTY, MobaXterm, Wireshark

## EDUCATION

---

University of Waterloo

Waterloo, ON

Bachelor of Mathematics, Major in Computational Mathematics (Co-op), Minor in Computing

Sep 2020 – Jun 2024

- Cumulative GPA: 85/100; Graduated with Distinction
- Relevant Coursework: Designing Functional Programs; Elementary Algorithm Design and Data Abstraction; Data Types & Structures; Computational Statistics and Data Analysis; Neural Networks; Data Visualization; Applied Linear Models

## WORK EXPERIENCE

---

Data Science Research Assistant

Waterloo, ON

University of Waterloo

May 2023 – Feb 2024

**FHIR-Formatted Healthcare Data Encoding Project** under Prof. Bryan Tripp

- Improved the **Natural Language Model Transformer's** accuracy by 10.7% by applying **data encoding methods** such as Spatial Semantic Pointers and Sinusoidal Encoding in **Python** to enhance the quality of the data inputs to the model.
- Used **SQL** to **cleanse, normalize, and structure FHIR-formatted datasets**, optimizing them for model input.
- Utilized **NumPy, SciPy, and Pandas** for data transformation and analysis to support research insight for team members.

**Metastatic Cancer Medical Image Analysis Project** under Prof. Subha Kalyanamoorthy

- Achieved 78.28% accuracy for nerve cell image classification using **PyTorch** by implementing **feature engineering and hyperparameter tuning** to fine-tune and optimize **Computer Vision Models** such as **ViT, ResNet, and CNNs**.
- Visualized and presented findings on model performance to stakeholders using **R (ggplot2), Matplotlib, and Pandas**.
- Trained models on Compute Canada's High-Performance Computing (HPC) clusters, managing jobs via **Bash scripting**.

DPI Automation Data Analyst Intern

Ottawa, ON

Nokia

Sep 2022 – Dec 2022

- **Automated** network traffic detection QA for 20+ applications (Web, Android, iOS) using **Python, Selenium, and Appium**.
- Analyzed traffic data captured by 60+ application filters by performing packet-level analysis with **Wireshark** to **troubleshoot issues** and **submit bug reports** for 4 filters to improve classification accuracy and enforce proper QoS policies.
- Scheduled recurring automated testing tasks using **Linux Crontab** to collect testing data.

Data Quality Assurance Intern

Calgary, AB (Remote in Toronto, ON)

Ontopical

Jan 2022 – Apr 2022

- Developed over 50 web scrapers to aid the ETL process for multimedia **data collection** using **Scrapy, Splash, and Lua**.
- Executed web scrapers on **AWS Batch**, monitored their performance using **AWS CloudWatch**, collected and aggregated document type counts from each scraper post-execution, and leveraged **Power BI** to visualize document distribution trends.
- Did **version control** and task management within an **Agile Scrum** framework using **Git, Bitbucket, Jira, and Confluence**.

## PROJECTS/ PUBLICATION

---

**Beam - An Algorithm For Detecting Phishing Link** 2022 APSIPA Annual Summit and Conference ([Github code](#))

- Applied six different NLP AI models, including fine-tuned BERT, CNN, and LSTM networks, for phishing URL detection.
- Utilized various **NLP-based tokenizations** (character & subword) and evaluated their performance on processed datasets.

**Use Of Subword Tokenization For Domain Generation Algorithm (DGA) Classification** Cybersecurity 6, 49 (2023).

- Developed an integrated scheme consisting of various NLP models for better classification of domains generated by DGAs.