Vishal Sood

⌂ 22 Ch. de Grandvignes, Bussigny, CH-1030
☎ (+41) 79 942 0489
✉ vishal.chandra.sood@protonmail.com
§ Swiss immigration status : Permit C
⌨ https://github.com/visood
⌨ https://www.linkedin.com/in/physicistvishal

Profile and Professional Qualities

Research engineer with extensive experience in high-performance scientific computing and algorithm development. Specialized in parallel computing architectures, memory optimization, and I/O-intensive applications for scientific domains. Combines deep technical proficiency in modern C++ and Python with expertise in computational neuroscience, genomics, and statistical physics.

Professional Qualities

- *Adaptable and Quick Learner*: Quickly acquired new specific skills at each previous professional role.
- *Innovative Problem-Solver*: Developed comprehensive workflows, and explored cutting-edge techniques to tackle complex problems.
- *Rigorous Researcher*: Applied advanced computational and mathematical techniques to model complex systems and advance research in multiple fields.
- *Technically Proficient*: Specialized in large-scale data analysis, I/O intensive applications, and demonstrated proficiency in multiple programming languages and tools.
- *Interdisciplinary Collaborator*: Worked closely with diverse teams, effectively bridging the gap between computational capabilities and scientific research needs.
- *Proactive Contributor*: Independently added features, fixed bugs, and participated in community-based projects to improve products and advance scientific knowledge.
- *Team Player and Pleasant Collaborator*: Always willing to help and share knowledge with the team.

Professional Goal

To develop cutting-edge computational infrastructure that accelerates scientific discovery through innovative algorithms and high-performance computing solutions.

Core Skills

Programming	C/C++, Python, R, JavaScript, Scala, Haskell
HPC	SLURM, Spark, MPI, OpenMP, Dask, Joblib
Scientific	NumPy, SciPy, PyTorch, Pandas, HDF5 Matplotlib, Seaborn, Jupyter
Databases	PostgreSQL/PostGIS, MySQL, SQLAlchemy NoSQL, spatial indexing, optimization
Development	Git, CMake, Jenkins, Docker, CI/CD Test-driven development, cross-platform
Algorithms	Data structures, pattern detection, geometry Computational parsers, optimization
Statistics	Monte Carlo, Bayesian, hypothesis testing Time-series analysis, signal processing
Modeling	Stochastic processes, networks, multi-scale Virtual experiments, validation frameworks
ML / AI	Classification, ensemble methods, validation Feature engineering, deep learning
Biocomputing	Neuroscience, genomics, bioinformatics Brain circuits, variant annotation
Physics	Statistical mechanics, electrophysiology Molecular dynamics, protein interactions
Data Systems	FAIR principles, metadata, knowledge graphs Provenance tracking, controlled vocabularies
Geospatial	GIS, spatial statistics, remote sensing Cartographic modeling, urban analytics
Languages	English (native), French (intermédiaire) Hindi (native), Punjabi (mother-tongue)

Professional Timeline

Oct 2017 - Dec 2024	**Senior Scientific Software Developer**	Blue Brain Project, EPFL	**Switzerland**
Oct 2015 - Jul 2017	**Senior Product Developer**	Saphetor SA, EPFL Innovation Park	**Switzerland**
Jan 2015 - Aug 2015	**Data Scientist**	Citiviz Sárl, EPFL Innovation Park	**Switzerland**
Jan 2011 - Oct 2014	**Maître Assistant**	Lab of Statistical Biophysics, EPFL	**Switzerland**
Sep 2008 - Aug 2010	**Post-doctoral Fellow**	Niels Bohr Institute, Copenhagen	**Denmark**
Sep 2006 - Aug 2008	**Research Fellow**	Physics Dept., University of Calgary	**AB, Canada**

Education

2006	**Doctor of Philosophy** in Physics	Boston University, Boston	**MA, USA**
2000	**Bachelor of Technology** in Engr. Physics	Indian Institute of Technology Bombay	**India**

Professional Experience

Specialized Expertise

1. Scientific Computing & High-Performance Systems
    - Design and implementation of massively parallel computational workflows for multi-terabyte scientific datasets
    - Development of scalable algorithms optimized for distributed computing environments (SLURM, Apache Spark)
    - Expertise in memory-intensive applications and I/O optimization for supercomputing architectures
    - Implementation of parallel processing strategies using Python multiprocessing, Dask, and Joblib
    - Performance optimization of C++ applications for 100+ node computing clusters
    - High-throughput data processing pipelines with optimized indexing strategies using HDF5 and Parquet
    - Real-time data query systems with microsecond response times
    - Configuration-driven pipeline architectures enabling reproducible large-scale scientific computing
2. Algorithm Development & Statistical Modeling
    - Development of novel algorithms for pattern detection in high-dimensional noisy datasets
    - Implementation of statistical validation frameworks with robust hypothesis testing
    - Creation of stochastic models and Monte Carlo simulation methods for complex systems
    - Design of machine learning pipelines for scientific data classification and prediction
    - Expertise in network analysis, graph theory, and complex systems modeling
    - Development of computational geometry algorithms for 3D spatial data analysis
    - Functional programming parsers for complex query expressions and data processing
    - Advanced statistical algorithms for quality control and validation
    - Multi-scale topological analysis methods for complex network characterization
3. Data Engineering & Knowledge Management
    - Implementation of FAIR (Findable, Accessible, Interoperable, Reusable) data principles for scientific workflows
    - Design of hierarchical data structures and knowledge graphs for complex multi-modal datasets
    - Development of metadata schemas and controlled vocabularies for scientific data integration
    - Expertise in scientific data formats: HDF5, NRRD (voxelized data), Parquet, NetCDF, VCF, BAM
    - Creation of data processing pipelines integrating heterogeneous sources (experimental, literature, computational)
    - Database design and optimization for spatiotemporal and hierarchical scientific data
    - REST API design for distributed data services and high-performance applications
    - Multi-database integration frameworks for comprehensive data annotation
    - Lazy-loading data access patterns for efficient large-scale scientific dataset manipulation
4. Bioinformatics, Genomics & Computational Biology
    - Design of genomic variant annotation and interpretation systems integrating clinical guidelines
    - Implementation of high-performance genomic analysis pipelines using modern C++
    - Development of evidence-based variant classification frameworks with machine learning enhancement
    - Expertise in genomics libraries and tools: Samtools, GATK, freebayes, Bioconductor
    - Creation of REST APIs for complex genomic query processing
    - Integration of public and proprietary genomic databases for comprehensive variant analysis
    - High-performance data processing systems for large-scale biological datasets
    - Population frequency analysis systems with statistical stratification
    - Coverage analysis and alignment quality assessment algorithms
    - Coordinate mapping systems for precise genomic position translation

5. Scientific Workflow Development & Automation
    - Design of reproducible computational pipelines with parameterized scientific workflows
    - Development of command-line interfaces and automation tools for scientific applications
    - Creation of validation frameworks supporting multiple model types and experimental protocols
    - Automated workflow orchestration systems for complex data processing pipelines
    - Implementation of containerized and cloud-ready scientific applications
    - Implementation of continuous integration and testing practices for scientific software
    - Design of interactive analysis environments and Jupyter-based reporting systems
    - Quality control frameworks for data validation and error detection
    - Statistical reporting systems with detailed quality metrics and visualization
    - Multi-stage pipeline architectures with dependency management and error recovery
6. Software Architecture & API Design
    - Design of Python API interfaces for seamless integration with scientific analysis workflows
    - Development of robust data processing pipelines with comprehensive error handling and validation
    - Creation of command-line interfaces and automation tools for scientific applications
    - Implementation of continuous integration and testing practices for scientific software
    - Cross-platform compatibility layers for diverse computing environments
    - Modern C++ development (C++11/14/17) implementing functional programming paradigms
    - Plugin-based extensible framework design for collaborative scientific software development
    - Configuration management systems enabling reproducible computational research
7. Interdisciplinary Modeling & Simulation
    - Multi-scale modeling experience spanning molecular to systems-level phenomena
    - Development of biophysically detailed computational models for complex biological systems
    - Implementation of data-driven model validation and comparison frameworks
    - Creation of virtual experimental protocols for in-silico hypothesis testing
    - Integration of experimental constraints with computational model parameterization
    - Design of model adaptation interfaces enabling cross-platform model validation
    - Sophisticated algorithms for precise spatial positioning using computational geometry

Senior Scientific Developer / System Specialist at Blue Brain Project

The Blue Brain Project (BBP), launched in 2005 at the École Polytechnique Fédérale de Lausanne (EPFL), pursued the ambitious mission of creating biologically detailed digital reconstructions and simulations of the rodent, and ultimately the human brain. This interdisciplinary endeavor required breakthrough innovations in multi-scale modeling, integration of vast experimental datasets, and utilization of cutting-edge supercomputing resources for complex neural simulations. As the Blue Brain Project concluded in December 2024, it left behind a transformative legacy of advanced computational methods, detailed brain models, and openly available tools that continue to influence neuroscience research worldwide.

When I joined BBP in 2017, I rapidly acquired specialized skills in computational neuroscience and became integral to the main scientific pipeline for constructing rodent brain circuits. My role evolved from circuit validation to leading the development and maintenance of BBP's flagship brain region models. Over seven years, I solved complex technical challenges across the entire circuit construction pipeline, developed innovative analysis tools, and provided computational support to neuroscientists throughout the project.

Key Technical Contributions:

1. Multi-Scale Brain Modeling & Circuit Construction

  - Led development of comprehensive computational pipelines for constructing anatomically detailed brain circuit models across multiple regions (neocortex, hippocampus, olfactory bulb, basal ganglia, cerebellum) and species
  - Implemented sophisticated algorithms for precise neuron placement within voxelized 3D spaces using computational geometry techniques guided by experimental imaging data
  - Developed custom connectivity algorithms incorporating experimental constraints for synapse placement and network topology optimization
  - Created tools for virtual experimentation enabling systematic validation of circuit models against published experimental studies
  - Contributed to community-based reconstruction of full-scale models including the rat somatosensory cortex and hippocampus CA1 region
  - Architected scalable topological analysis frameworks for multi-scale network characterization across spatial hierarchies

2. High-Performance Data Processing & Management
   - Designed and implemented parallel data processing workflows handling multi-terabyte datasets using HDF5, NRRD (voxelized data), and Parquet formats
   - Developed SLURM-based cluster computing solutions parallelizing circuit construction across 100+ computing nodes
   - Optimized memory-intensive applications for supercomputing environments, achieving significant performance improvements
   - Created hierarchical data management systems implementing FAIR principles for complex neuroscientific datasets
   - Established metadata schemas and controlled vocabularies enabling integration of experimental and computational data
   - Implemented lazy-loading data architectures enabling efficient access to petabyte-scale scientific datasets
3. Scientific Software Architecture & Development
   - Architected the DMT (Data, Models, Tests) validation framework enabling systematic model-data comparison through standardized interfaces
   - Implemented Python Abstract Base Classes (ABCs) creating reusable adapter patterns for diverse model types
   - Developed command-line interfaces and automated workflows allowing scientists to launch complex multi-node computations
   - Created comprehensive statistical validation pipelines incorporating robust hypothesis testing and uncertainty quantification
   - Built interactive analysis environments using Jupyter notebooks enabling reproducible scientific reporting
   - Designed configuration-driven pipeline architectures supporting reproducible computational research workflows
   - Developed plugin-based extensible frameworks enabling collaborative scientific software development
4. Knowledge Integration & Collaborative Infrastructure
   - Designed and implemented knowledge graphs integrating experimental data, literature findings, and computational results
   - Developed data curation workflows enabling systematic integration of heterogeneous neuroscientific datasets
   - Created visualization tools for exploring large-scale brain connectivity patterns and anatomical structures
   - Established version control and collaborative development practices ensuring reproducible scientific workflows
   - Contributed to scientific dissemination through manuscript co-authoring and web portal development
   - Built collaborative analysis platforms enabling community-driven computational neuroscience research
5. Interdisciplinary Collaboration & Problem-Solving
   - Collaborated effectively with diverse teams including neuroscientists, bridging experimental neuroscience with computational modeling requirements
   - Solved complex technical challenges spanning computational geometry, statistical analysis, database optimization, and parallel computing

- Mentored junior developers and scientists in scientific computing best practices and software development methodologies
- Led development of multi-institutional collaborative frameworks for large-scale computational projects

Technical Impact: My work directly enabled the construction and validation of some of the most detailed brain circuit models ever created, providing the computational infrastructure that supported BBP's scientific breakthroughs. I developed validation frameworks, analysis tools, and computational pipelines that enabled handling of unprecedented data volumes in neuroscience research.

Scientific Contribution: Through systematic validation of brain circuit models against experimental data, my work helped establish new standards for computational neuroscience model development. The collaborative frameworks I created facilitate ongoing scientific progress by enabling researchers to build upon previous modeling efforts and compare different approaches quantitatively.

Product Developer at Saphetor SA

Saphetor is a precision medicine startup at the cutting-edge of whole-genome-scale analysis based on Next Generation Sequencing (NGS). Their products include VarSome (varsome.com), a comprehensive genomic variant knowledge base, and enterprise-grade SaaS services providing high-performance genomic variant annotation and clinical interpretation based on ACMG guidelines.

As a Product Developer at Saphetor, I architected and implemented core components of THALIA, their flagship genomic variant annotation platform. Working within a cross-functional team of bioinformaticians, geneticists, and developers, I delivered critical features that advanced the state-of-the-art in clinical genomics.

Key Technical Contributions:

1. **Genomic Variant Annotation Infrastructure**
   - Architected comprehensive HGVS notation parsing system supporting c.dot and p.dot nomenclature
   - Implemented multi-database integration framework combining ClinVar, gnomAD, ExAC, HGMD, and dbNSFP
   - Developed ACMG evidence classification system for automated clinical variant interpretation
   - Created population frequency annotation pipelines with ethnic stratification
2. **High-Performance Data Processing**
   - Built a local datbase integration framework for high-throughput genomic data queries
   - Implemented REST client infrastructure enabling distributed genomic data services
   - Developed DataFrame utilities for large-scale genomic data manipulation
   - Architected BAM file processing systems for coverage analysis and alignment quality assessment
3. **Advanced Algorithm Development**
   - Implemented functional programming parsers for complex genomic query expressions
   - Developed strand bias detection algorithms for variant calling quality control
   - Created transcript coordinate mapping systems for accurate genomic position translation
   - Built comprehensive variant classification algorithms combining multiple ML prediction tools
4. **Software Architecture and API Design**
   - Designed Python API interfaces for seamless integration with genomic analysis workflows
   - Implemented user permission systems for secure database access control
   - Developed robust VCF processing pipelines with comprehensive error handling
   - Created statistical coverage reporting systems with detailed quality metrics
5. **Clinical Genomics Innovation**

- Explored machine learning models for improved genomic variant pathogenicity prediction
- Developed zygosity analysis frameworks for inheritance pattern determination
- Built annotation pipelines supporting both germline and somatic variant analysis
- Implemented clinical decision support features for variant interpretation workflows

Technical Excellence: Mastered modern C++ (11/14/17) implementing functional programming paradigms, integrated cutting-edge genomics libraries (Samtools, GATK, freebayes), and performed statistical analysis using R and Bioconductor.

Data Scientist / Software Developer at Citiviz Sárl

Citiviz was an early-stage startup that developed a computing and geo-visualization ecosystem for urban intelligence, monitoring, and decision-making. Leaning on my excellent learning skills, I quickly develop an expertise in geospatial data engineering, analysis and visualization. Using `Python` and the `SQAlchemy` library, I integrated geographically indexed public bike system and population data into the application's database. I then leveraged my data analysis skills to understand the raw data, clean it up, and create visualizations and machine learning models to analyze usage patterns, developing a data driven approach. My work was instrumental in the development of an analysis-framework layer that sat atop the core database and visualization functionality of the application. This layer enabled seamless reporting with visualizations and the integration of predictive model plugins, thus bridging the gap between data, analytics and end-user applications.

Key Contributions:

1. Developed data analytics products for bike-share monitoring
2. Built and maintained geospatial databases using SQLAlchemy and PostGIS
3. Performed data cleaning, analysis, and visualization of urban data
4. Created machine learning models for bike-usage analysis using R and Python
5. Contributed to the development of an analysis framework for traffic monitoring applications
6. Integrated predictive model plugins and seamless reporting with visualizations

Technical Skills Demonstrated:

- Languages: Python, R, JavaScript, SQL, YAML
- Databases: PostgreSQL, PostGIS
- Libraries: SQLAlchemy, GeoAlchemy2, Shapely, GDAL, ijson, pandas, scikit-learn
- Tools: Git, GitHub, Jupyter Notebooks
- Visualization: Matplotlib, Seaborn, Mayavi

Academic work

Throughout my research career at prestigious institutions like Niels Bohr Institute, University of Calgary, and Boston University, I consistently applied advanced computational and mathematical techniques to model complex systems in various scientific domains. My work primarily focused on stochastic processes, random walks, complex networks, and dynamical systems, utilizing both analytical and computational approaches to solve challenging problems in fields ranging from DNA-protein interactions to opinion dynamics and evolutionary processes.

At *EPFL* I developed a novel statistically rigorous methodology for measuring the amount of misfolded proteins in a cell. At the *Niels Bohr Institute* I developed a computational model for DNA repair in a cell, and a stochastic model of innovation. At *Calgary* I developed several stochastic and computational models for complex networks, and graph random walks. At *Boston University* I developed theoretical and computational models in statistical mechanics of complex networks for my PhD theses titled *Interacting Particle Systems*.

Key Contributions:

- Developed stochastic models for DNA-protein interactions in DNA repair
- Extended mathematical theorems for random walks on general graphs
- Applied complex network analysis to protein interaction networks
- Modeled innovation processes using branching-process theory
- Investigated first-passage properties of random walks on graphs
- Studied consensus formation in opinion dynamics using network analysis
- Analyzed evolutionary dynamics of mutations in structured populations
- Technical Skills:
- Programming Languages: C, C++, Python, Java, Matlab, Maple
- Data Analysis and Visualization: Python matplotlib, Gnuplot, Java applets
- Scientific Computing: Stochastic modeling, Monte Carlo simulations, Network analysis
- Mathematical Modeling: Probabilistic analysis, Differential equations, Graph theory
- Document Processing: LaTeX
- Bioinformatics: Data mining, Protein interaction network analysis
- High-Performance Computing: Large-scale simulations on complex systems