2023 Summer Internship

List of Themes

March 30, 2023

Preferred Networks

List of themes

* IDs are jumped around because the themes held only in Japanese are omitted from the list.
If you are interested in the themes, please see the list of themes in Japanese at the following URL:
https://www.preferred.jp/wp-content/uploads/2023/03/831d7079054f3a9adf79bef7f143a578-1.pdf

ID	Theme	Area	Category
1	Development of training methods for large-scale language models	Computing infrastructure	Project internship
7	Computational electromagnetics and electric circuit analysis with open source software	Computing infrastructure / MN-Core	Project internship
8	Development MN-Core compiler and toolchain	Computing infrastructure / MN-Core	Project internship
9	Development of framework and library for deploying deep learning models in real world	Computing infrastructure / PFVM	Project internship
10	Development of CuPy	Computing infrastructure / CuPy	Project internship
11	Development of Optuna and Optuna Dashboard	Computing infrastructure / Optuna	Project internship
12	Crystal structure prediction using Matlantis and Optuna	Computing infrastructure / Optuna Materials informatics / Matlantis	Research internship
13	Research and Development for material discovery algorithms in Matlantis	Materials informatics / Matlantis	Project internship
14	Development and operation of a web system for Matlantis which is a general-purpose atomic-level simulator	Materials informatics / Matlantis	Project internship
15	Applied research and development of machine learning and atomic simulation on materials	Materials informatics / Matlantis	Research internship
16	Applied research on machine learning/molecular simulation for drug discovery	Drug discovery	Research internship
17	Synthesizable molecular design	Drug discovery	Research internship
18	Development of training methods for models using gene-related graphs	Life science	Project internship
19	Application of deep tabular models to healthcare	Life science	Project internship
20	Biomedical text mining using large language models	Life science	Research internship
22	Research on machine learning methods for medical imaging	Life science	Research internship
26	Designing and Implementing Application of AI-Generated Content.	Entertainment	Project internship
27	Application of deep learning techniques to creation	Entertainment	Research internship
36	Autonomous driving system for trucks. Development of perception modules such as 3D object detection.	Autonomous driving for trucks	Project internship
40	Developing machine learning methods for financial applications	Finance	Project internship
41	Machine learning research for financial applications	Finance	Research internship
42	Research and development on the estimation/prediction of various weather conditions using the advanced high-spatio-temporal resolution 3D weather data	Remote sensing	Project internship
43	Research and development of advanced segmentation techniques for SAR image data	Remote sensing	Research internship
45	Diffusion models with structural inductive bias	Deep learning	Research internship

46	Symmetry-based 3D representation learning and its computer vision applications	Deep learning	Research internship
47	Interactive System for 3D/4D reconstruction	HCI	Research internship
48	Desktop VR	HCI	Research internship
49	Object manipulation using robotic arms with a multitude of sensors	Robotics	Research internship

1
Development of training methods for large-scale language models

Area: **Computing infrastructure**
Category: **Project internship**

In this project, you will use PFN's large-scale clusters to develop technology related to language models with large-scale language data, which has been attracting attention in recent years.

Communication language

English/Japanese

Must requirements

- Implementation experience using deep learning frameworks such as PyTorch
- Knowledge of basic computer science

Want requirements

- Deep learning acceleration experience
- Knowledge of machine learning using deep learning
- Knowledge of natural language processing

↑ Back to list of themes

7
Computational electromagnetics and electric circuit analysis with open source software

Area: **Computing infrastructure / MN-Core**
Category: **Project internship**

Computational electromagnetics and electric circuit analysis are used for development of MN-Core, an AI accelerator developed by PFN. Open source multi-purpose linear algebra library, finite element method solver and method of momentum solver will be evaluated to see how they are adapted to development of MN-Core series.

Communication language

English/Japanese

Must requirements

- Knowledge of system of linear equations
- Capability of writing simple programs for automation
- Knowledge of static electric field, static current, electric circuit and ordinary differential equations

↑ Back to list of themes

8
Development MN-Core compiler and toolchain

Area: **Computing infrastructure / MN-Core**
Category: **Project internship**

You will work on MN-Core compiler and related software improvements. Your main task would be

- Porting of HPC applications for MN-Core
- Improvement of MN-Core compiler algorithm
- Development profiler for MN-Core
- Development tools for MN-Core Monitoring and power efficient control with k8s cluster

Communication language

English/Japanese

Must requirements

- Python, C++ programming

Want requirements

- Experience of low level optimization
- Knowledge of deep learning compiler stack

↑ Back to list of themes

9
Development of framework and library for deploying deep learning models in real world

Area: **Computing infrastructure / PFVM**
Category: **Project internship**

Develop compiler/runtime (PFVM) that optimizes computational graphs of deep learning models to perform inference performantly (in terms of execution time and memory usage), targetting on various backends such as CUDA or edge devices.

Communication language

English/Japanese

Must requirements

- Programming in Python
- Basic knowledge of computer science
- Basic knowledge of compiler

Want requirements

- Programming in C++
- Experience on deep learning compilers (ONNX, LLVM, MLIR, etc.)
- Development experience of multi-pass compilers
- Basic knowledge of optimization algorithms

↑ Back to list of themes

10
Development of CuPy

Area: **Computing infrastructure / CuPy**
Category: **Project internship**

In this project, you will work on development of CuPy, an open-source array library for GPU. (1) Development of new features for performance improvement, (2) Implementation of a prototype for new GPU backends (e.g., Metal, oneAPI), (3) Enhancing SciPy compatible APIs.

Communication language

English/Japanese

Must requirements

- Programming in CUDA
- Programming in Python and NumPy
- Basic knowledge of computer science

Want requirements

- Experience with distributed data processing in multi-GPU environments (NCCL, MPI, etc.)
- Development experience with Metal Shading Language

↑ Back to list of themes

11
Development of Optuna and Optuna Dashboard

Area: **Computing infrastructure / Optuna**
Category: **Project internship**

For Optuna and Optuna Dashboard, OSS hyperparameter optimization frameworks, (1) develop new optimization algorithms, (2) improve the code quality and design, or (3) improve processes of OSS operation/support.

Communication language

English/Japanese

Must requirements

- Basic knowledge in computer science

Want requirements

- Experience in the development of or contribution to open source softwares
- Experience of the development or support of software in a multi-person team
- Any achievement in machine-learning area such as paper publication or awards at competitions
- Expert knowledge in Bayesian optimization or black-box optimization algorithms
- Experience of the modern web frontend development
- Experience of the UI/UX design

↑ Back to list of themes

12
Crystal structure prediction using Matlantis and Optuna

Area: **Computing infrastructure / Optuna, Materials informatics / Matlantis**
Category: **Research internship**

Using Matlantis, which is the versatile atomistic simulator, and Optuna, which is the black-box optimization software, you will search unknown crystal structure on the large scale computing cluster in PFN.

Communication language

English/Japanese

Must requirements

- Basic knowledge in computer science
- Basic knowledge or interest of physic, chemistry, or material science

Want requirements

- Expert knowledge in black-box optimization (especially, genetic algorithms)
- Expert knowledge in material informatics
- Expert knowledge in parallel computation on the large scale computing cluster

↑ Back to list of themes

13
Research and Development for material discovery algorithms in Matlantis

Area: **Materials informatics / Matlantis**
Category: **Project internship**

Develop algorithms to search materials for Matlantis. This may involve implementing methods based on computational chemistry or developing models to predict some physical property values or structures from datasets.

Communication language

English/Japanese

Must requirements

- Coding skills in Python

Want requirements

- Knowledge of computational chemistry
- Experience in using deep learning frameworks such as PyTorch
- Motivation to bring value to customers and connect academic research to real products

↑ Back to list of themes

14
Development and operation of a web system for Matlantis which is a general-purpose atomic-level simulator

Area: **Materials informatics / Matlantis**
Category: **Project internship**

Work on the development and operation of Matlantis features. (e.g., improving release methodology, investigating and implementing Telepresence and Karpenter, optimizing API Gateway scheduling, cross-document search, developing JupyterLab extensions, etc.)

Communication language

English/Japanese

Must requirements

- Basic knowledge in computer science
- Familiarity with Python or Go

Want requirements

- Experience with cloud computing (AWS/GCP/Azure, etc.)
- Knowledge of container orchestration systems (Kubernetes)
- Client and server model implementation (REST/gRPC/GraphQL, etc.)
- Experience with tools that enable Infrastructure as Code (Terraform / CloudFormation)

↑ Back to list of themes

15
Applied research and development of machine learning and atomic simulation on materials

Area: **Materials informatics / Matlantis**
Category: **Research internship**

Will work on the development of machine learning based atomic simulator, or the application of machine learning to molecular dynamics calculations and materials exploration techniques. Reference: Matlantis https://matlantis.com/

Communication language

English/Japanese

Must requirements

- Basic knowledge of computer science
- Basic knowledge or interest of physic, chemistry, or material science

Want requirements

- Knowledge of college degree in physics, chemistry, or materials science
- Computational science experience. For example, in-depth knowledge of and experience implementing and using simulation techniques such as quantum chemical calculations, molecular dynamics methods, etc.
- Experience in research and development and writing papers on the integration of physical simulation and machine learning

↑ Back to list of themes

16
Applied research on machine learning/molecular simulation for drug discovery

Area: **Drug discovery**
Category: **Research internship**

You will work on research topics related to molecule (small molecules or proteins) design for drug discovery. The project topic will include 1) protein / peptide-binder design, 2) application of neural network potential for drug discovery, and so on.

Communication language

English/Japanese

Must requirements

- Basic knowledge of computer science
- Knowledge of chemistry or biology or physics (e.g., organic chemistry, medicinal chemistry, structural biology, or molecular simulation, etc)

Want requirements

- Experience in team development
- Experience in deep learning related researches (publications or oral presentation in scientific conferences are preferred)

↑ Back to list of themes

17
Synthesizable molecular design

Area: **Drug discovery**
Category: **Research internship**

To synthesize a novel compound, chemists have to manually craft the synthetic pathway and they may need to test several different pathways to synthesize one compound. To ease this burden, it is better to consider synthesizability during the design process. As an example, possible tasks can be: 1) create a set of chemical "building blocks" from which various drug-like compounds can be synthesized, 2) create a synthesizability aware generative model.

Communication language

English/Japanese

Must requirements

- Basic knowledge of computer science

Want requirements

Knowledge in any of the following domain:

- chemistry (e.g., synthetic chemistry, organic chemistry, etc)
- biology
- machine learning

Experience in any of the following techniques:

- Organic synthesis
- Retrosynthetic pathway planning (manual or machine learning based)
- Designing or using generative model

↑ Back to list of themes

18
Development of training methods for models using gene-related graphs

Area: **Life science**
Category: **Project internship**

There are various graphs that represent known relationships in the field of bioinformatics such as gene pathways. In this theme, you will work on developing a more accurate model by using such graph data.

Communication language

English/Japanese

Must requirements

- Experiences with Deep learning frameworks (preferably PyTorch)
- Deep knowledge of Machine Learning and Computer Vision

Want requirements

- Knowledge of bioinformatics

↑ Back to list of themes

19
Application of deep tabular models to healthcare

Area: **Life science**
Category: **Project internship**

In this theme, we address the application of machine learning and deep learning techniques to healthcare data, especially health checkup, health insurance, clinical data. Examples of topics include (but are not limited to) improving the accuracy of disease classification using missing value imputation techniques, applying deep tabular models to tabular healthcare data, and patient stratification using machine learning.

Communication language

English/Japanese

Must requirements

- Writing and reading skill in Japanese (We expect this theme uses datasets that contain Japanese descriptions. Communication with mentors in English is allowed)
- Ability and experience in at least one of the followings:

1. R&D in machine learning
2. R&D in the life sciences (e.g., bioinformatics, biomedical sciences and engineering, public health)

Want requirements

- Python programming
- Machine learning for tabular data (e.g., missing value completion, time series prediction)
- Data analysis on healthcare data such as health checkup, health insurance, clinical data (e.g., chronic diseases, genetic diseases)
- Deep learning

↑ Back to list of themes

20
Biomedical text mining using large language models

Area: **Life science**
Category: **Research internship**

Research and development in life science text mining using large language models

Communication language

English/Japanese

Must requirements

Ability and/or experience of at least one of the followings:

- R&D about Machine learning of at life-science related fields such as bioinformatics and biomedical science and engineering.
- R&D about Natural Language Processing

Want requirements

- Python programming
- Use of PubMed API
- Use of large language models

↑ Back to list of themes

22

Research on machine learning methods for medical imaging

Area: **Life science**
Category: **Research internship**

We will develop machine learning methods to solve problems in the medical imaging domain (e.g. domain generalization, medical image synthesis, applications of foundation models, survey of recent models etc.).

Communication language

English/Japanese

Must requirements

- Deep knowledge of image analysis methods
- Coding skills in Python
- Experience using deep learning frameworks such as PyTorch

Want requirements

- Knowledge of biology equivalent to at least undergraduate level
- Experience of research with images or video data
- Publishing papers in the field of biology, mathematics, or informatics
- Experience developing applications using machine learning or deep learning

↑ Back to list of themes

26
Designing and Implementing Application of AI-Generated Content.

Area: **Entertainment**
Category: **Project internship**

In this project, you will be working with Entertainment team members at PFN. You will focus on the design and implementation of applications that combine human creativity and AI technology.
You will develop new features for existing services or new applications. (Please refer to the information below for existing services.)

- Crypko https://crypko.ai/
- Memes https://memes.crypko.ai/

Communication language

English/Japanese

Must requirements

One of the following skills or practical experience:

- Application & web development, system design
- Product design, UI & UX design

Want requirements

- Frontend (JavaScript, Vue, React, etc.)
- Backend (Python, Django, etc.)
- Cloud computing
- UI design tools (such as Figma)
- Data analysis
- Experience with deep learning frameworks, such as PyTorch
- Interest in anime, games, etc.

↑ Back to list of themes

27
Application of deep learning techniques to creation

Area: **Entertainment**
Category: **Research internship**

In this project, you will be working with Entertainment team members at PFN. You will take part in the research of deep learning techniques for creativity.

Communication language

English/Japanese

Must requirements

- Familiarity with Python

Want requirements

- Experience in using deep learning frameworks such as PyTorch
- Publications in machine learning domain
- Domain knowledge in Computer Vision, Computer Graphics, Generative Model
- Skills in software development
- Like anime/game

↑ Back to list of themes

36
Autonomous driving system for trucks. Development of perception modules such as 3D object detection.

Area: **Autonomous driving for trucks** [1]
Category: **Project internship**

Target tasks include development of 3D object detection methods using LiDAR data and images, reproduction of existing methods and their incorporation into real vehicles, as well as implementation of automatic scene extraction pipelines for subsequent annotation.

Communication language

English

Must requirements

- Python, C++ programming
- Experience with deep learning frameworks such as PyTorch, TensorFlow etc.

Want requirements

- Experience working on 2D or 3D object detection (papers/conference presentations are desirable)
- Experience with TensorRT
- Ability to communicate in English
- Experience developing practical tools, web applications, etc

↑ Back to list of themes

[1] Regarding the "Autonomous driving for trucks" area, a position for part-time engineer is also being offered by T2 inc., which was established by Preferred Networks and Mitsui & Co. for the practical application of self-driving trucks. If you are interested, please visit the T2 website (https://t2.auto/recruit/Part-time_Engineer.html). You may apply for both positions.

40
Developing machine learning methods for financial applications

Area: **Finance**
Category: **Project internship**

You will work on developing machine learning methods for financial applications (FinanceML). Your main task would be (i) developing python library for FinanceML (e.g., pfhedge) or (ii) (re-)implementation of state-of-the-art FinanceML methods. Example topics: deep hedging, deep-learning-based time-series prediction and generation, and machine-learning-based trading strategy.

Communication language

English/Japanese

Must requirements

- Python programming
- Experience in deep learning frameworks (We recommend PyTorch)
- Basic knowledge and development experience in machine learning and deep learning

Want requirements

- Interest in quantitative finance, financial engineering, economics and related fields (We welcome in-depth knowledge of the fields)
- In-depth knowledge in one or more areas of computer science, statistics, optimization etc.

↑ Back to list of themes

41
Machine learning research for financial applications

Area: **Finance**
Category: **Research internship**

PFN Quantitative Finance Team invites research internship applications on the topic of machine learning techniques for financial applications, including (1) deep hedging, (2) deep-learning-based time-series prediction and generation, and (3) machine-learning-based trading strategy. We welcome self-driven students who can independently and proactively execute research projects.

Communication language

English/Japanese

Must requirements

- Python programming
- In-depth knowledge in one or more areas of computer science, statistics, optimization etc.
- Basic knowledge of finance and related fields
- Experience in writing and publishing research papers

Want requirements

- Experience in publishing a paper of finance AI

↑ Back to list of themes

42
Research and development on the estimation/prediction of various weather conditions using the advanced high-spatio-temporal resolution 3D weather data

Area: **Remote sensing**
Category: **Project internship**

We are collaborating with NICT to analyze the advanced high-spatio-temporal resolution 3D weather radar, which NICT developed recently. Interns will be expected to engage research tasks that can take leverage of this rich and massive radar data. Possible tasks are: quantitative precipitation estimation, precipitation particle classification, lightening localization and prediction, 3D wind estimation, and others.

Communication language

English/Japanese

Must requirements

- Expertise in Deep Learning-based machine learning
- Good experiences in developing machine learning models in PyTorch
- Capability of propelling your R&D projects proactively and in a self-motivated manner: identify problems, analyze causes, devise solutions

Want requirements

- Experiences in writing a technical paper
- Knowledgeable about physics of whether and atmosphere
- Mathematical skill to formulate and approximate physical phenomenon into statistical models
- R&D experiences in a multi-person team with Git, Github
- R&D experiences with virtual container/environment such as Docker, and Kubernetes
- Development experiences on GPU clusters (private or public)
- Experiences dealing with GIS data and related libraries, softwares
- Accustomed to command-line interfaces in Linux and/or Mac

- Eager to start R&D activities from collecting and organizing your own dataset
- Team working and communications skills

↑ Back to list of themes

43
Research and development of advanced segmentation techniques for SAR image data

Area: **Remote sensing**
Category: **Research internship**

SAR images are actively sensed images using radar, which makes analysis of the surface possible from satellites and airborne sensors even during night and bad weather conditions. Interns will be expected to research Domain Adaptation and Weakly Supervised Learning using large scale datasets from a variety of remote sensing sensors for computer vision tasks such as semantic segmentation in aerial SAR images.

Communication language

English/Japanese

Must requirements

- Expertise in deep Learning-based machine learning
- Good experiences in developing machine learning models in PyTorch
- Capability of propelling your R&D projects proactively and in a self-motivated manner: identify problems, analyze causes, devise solutions

Want requirements

- Experiences in writing a technical paper
- Knowledgeable about, and/or having R&D experiences on computer vision, in particular, semantic segmentation
- Knowledgeable about, and/or having R&D experiences on satellite image processing
- Knowledgeable about, and/or having R&D experiences on domain adaptation and/or weakly supervised learning
- R&D experiences in a multi-person team with Git, Github
- R&D experiences with virtual container/environment such as Docker, and Kubernetes
- Development experiences on GPU clusters (private or public)
- Experiences dealing with GIS data and related libraries, softwares
- Accustomed to command-line interfaces in Linux and/or Mac

- Eager to start R&D activities from collecting and organizing your own dataset
- Team working and communications skills

↑ Back to list of themes

45
Basic research to introduce structural induction bias into diffusion models

Area: **Deep learning**
Category: **Research internship**

Recently, diffusion models were proven to be very powerful in image generation; however, its application to different modalities still have much room for improvements and theoretical investigation. We are looking for a courageous applicant to join us in our research in developing a new framework to introduce structural prior knowledge of the dataset to the basic framework of diffusion type model to extent its applications to various types of dataset.

Communication language

English/Japanese

Must requirements

- Certain level of mathematical knowledge, such as those related to stochastic process and algebra
- Coding Skills in Python and Deep learning

Want requirements

- Ph.D Candidateship
- Experience with theoretical research

↑ Back to list of themes

46
Symmetry-based 3D representation learning and its computer vision applications

Area: **Deep learning**
Category: **Research internship**

We are looking for a research-minded applicants to join us in developing a new framework of utilizing geometric inductive bias to improve the sample efficiency and out-of-distribution generalization capability of the neural-network model. Although we expecting to focus much on the analysis of 3D image, we are open to explore a new direction of application.

Communication language

English/Japanese

Must requirements

- Experiences with Deep learning frameworks (preferably PyTorch)
- Deep knowledge of Machine Learning and Computer Vision

Want requirements

- Ph.D. candidacy
- Record of publication in top-tier conference/journal as a corresponding author
- capability of propelling your R&D projects proactively and in a self-motivated manner: identify problems, analyze causes, devise solutions.

↑ Back to list of themes

47
Interactive System for 3D/4D reconstruction

Area: **HCI**
Category: **Research internship**

We invite you to join our HCI research into building interactive systems in 3D/4D reconstruction. We are interested in utilizing human knowledge to reconstruct and control 3D/4D spaces captured by cameras. More specifically, we will research supporting frameworks to capture real-world environments and edit reconstructed environments.

Communication language

English/Japanese

Must requirements

- Familiarity with the Python programming
- Experience in development of user interface or computer graphics (any platform: Javascript, Unity, etc.)

Want requirements

- Experiences in writing a technical paper
- Experience in deep learning frameworks (e.g., PyTorch)
- Experience in 3D reconstruction technologies

↑ Back to list of themes

48
Desktop VR

Area: **HCI**
Category: **Research internship**

Research and develop novel interactive experiences for desktop VR

Communication language

English/Japanese

Must requirements

- Experience developing XR applications

Want requirements

- Proven track record of XR application development using Unity, e.g. projects on Github, apps, games, research prototypes etc.
- Publications at HCI/XR conferences or journals

↑ Back to list of themes

49

Object manipulation using robotic arms with a multitude of sensors

Area: **Robotics**
Category: **Research internship**

We will research object manipulation using a Franka Emika panda arm with multiple sensors, such as a tactile sensor and RGB-D camera. The robot will recognize and grasp transparent objects, such as test tubes, and place them. The research topics are the recognition of transparent objects, high-precision manipulation under uncertain, (and the development of a robot hand).

Communication language

English/Japanese

Must requirements

- Minimum level of English to be able to communicate and read research articles (Interviews will be in English.)
- Robot system development experience, using C++ and Python with ROS Noectic (Knowledge of ros_control is required.)
- Machine Learning (Computer Vision/Reinforcement Learning/Self-Supervised Learning. Experience in Robot Perception/Control methods is a plus)

Knowledge and experience* in the above three and at least one of the following:
* Required level is the ability to reproduce and implement a research article, or submit an article.

- Object manipulation with a robotic arm (Experience in working with novel hardware and sensors is a plus. R&D experience at Franka Emika panda arm is a plus.)
- Mechanical engineering (Experience in designing, planning, and deploying robotic systems at scale is a plus)
- Physics Simulators (Experience in using MuJoCo/PyBullet/Isaac Sim to set up novel environments)

Notes: Requirements are updated at Apr. 14

↑ Back to list of themes