# 2025 Summer Internship

## List of Courses and Themes

March 26, 2025

Preferred Networks

# List of Courses

Please refer to the recruitment instructions on our website for more details on the whole internship.

| Course | 10-days Short Dev. | 7-weeks R&D |
|---|---|---|
| Period | Aug. 18th (Mon.)<br>- Aug. 29th (Fri.)<br><br>Sep. 5th (Fri.)<br>- Sep. 19th (Fri.) | Aug. 7th (Thu.)<br>- Sep. 26th (Fri.) |
| | Must participate for the entire duration in principle | Temporary absence allowed |
| Theme Openings | This course requires Japanese language proficiency.<br>Please check the Japanese application guidelines. | 15 themes<br>RD01～RD38 |

There are two categories in the 7-week R&D course according to its orientation. Please use the following descriptions as a guide when selecting your preferred theme.

| Project Internship | Research Internship |
|---|---|
| Engage in projects focused on practical problem-solving, conducting research and development of core technologies, and contributing to the development of products and services. | Engage in pioneering and challenging research activities with the goal of publishing papers in top-tier international conferences. |

The course and theme of your choice will be selected when you submit the screening task that we will guide you through after your application. The actual theme and the job content that you will work on will be decided in consultation with our team members during the selection process.

# List of Themes

* IDs are non-continuous because the themes held only in Japanese are omitted from the list. If you are interested in these themes, please see the list of themes in Japanese.

## 7-weeks Research & Development Course

| ID | Theme | Area | Category |
|---|---|---|---|
| RD01 | Research and/or Development of evaluation system for large scale LLM | R&D of the Foundation Model | Project |
| RD02 | Maintenance of LLM training dataset | R&D of the Foundation Model | Project |
| RD03 | R&D on LLM post-training | R&D of the Foundation Model | Project |
| RD05 | R&D of Vision Language Model (VLM) | R&D of the Foundation Model | Research |
| RD06 | LLM inference optimization | LLM Fundamental Technology Development | Project |
| RD09 | Development of MN-Core Compiler | MN-Core/Computing Infrastructure | Project |
| RD10 | Development ML/HPC Application on MN-Core | MN-Core/Computing Infrastructure | Project |
| RD11 | Development of Optuna | OSS Development | Project |
| RD12 | Development of Optuna Dashboard | OSS Development | Project |
| RD14 | Applied research and development of machine learning and atomic simulation on materials | Material Science (R&D) | Research |
| RD15 | Evaluation of material discovery methods on Matlantis or the improvement of the software development process using LLM | Material Science (Core Technology Development) | Project |
| RD22 | Reconstruction, editing, and generation of 3D models and free-viewpoint videos (Project) | PFN 3D Scan (Product/Service Development) | Project |
| RD23 | Reconstruction, editing, and generation of 3D models and free-viewpoint videos (Research) | PFN 3D Scan (Core Technology Development) | Research |
| RD24 | Video generation model application service | Entertainment (Product/Service Development) | Project |
| RD38 | Research on machine learning methods for medical imaging | Health Care | Research |

# RD01
# Research and/or Development of evaluation system for large scale LLM

**Area: R&D of the Foundation Model**

**Course: 7-weeks Research & Development Course**

We want you to help us with the development and enhancement of our in-house evaluation system for large scale LLM

## Communication Language

English/Japanese

## Must Requirements

- Basic knowledge of computer science
- Experience in Python
- Experience using deep learning frameworks such as PyTorch

## Preferred Requirements

- Knowledge or experience in LLM benchmark

↑ Back to list of themes

# RD02
# Maintenance of LLM training dataset

**Area: R&D of the Foundation Model**

**Course: 7-weeks Research & Development Course**

We will work on maintaining training data for Large Language Models (LLM), e.g., improving existing large-scale datasets.

## Communication Language

English/Japanese

## Must Requirements

- Basic knowledge of computer science
- Coding ability using Python
- Experience in developing applications using LLMs (any scale)

## Preferred Requirements

- Experience in training LLMs
- Experience in large-scale distributed parallel data processing

↑ Back to list of themes

# RD03
# R&D on LLM post-training

**Area: R&D of the Foundation Model**

**Course: 7-weeks Research & Development Course**

We conduct research and development on post-training of Large Language Models (LLMs).
Specific themes include:

- Evaluation of data quality using LLMs and similar technologies (AI feedback, reward modeling)
- Generation of artificial data using LLMs
- Other developments to improve model performance (long-context, translation, etc.)

## Communication Language

English/Japanese

## Must Requirements

- Basic knowledge of machine learning
- Proficiency in Python programming
- Experience using deep learning frameworks such as PyTorch
- Familiarity with and experience using LLMs

## Preferred Requirements

- Experience in research and development using LLMs

↑ Back to list of themes

# RD05
# R&D of Vision Language Model (VLM)

**Area: R&D of the Foundation Model**

**Course: 7-weeks Research & Development Course**

Interns will be involved in cutting-edge research and development of the Vision Language Model (VLM).
Example tasks include SotA model implementation, dataset curation, analysis or resolving open research problems.
We welcome students who can independently carry out research or development projects.

## Communication Language

English/Japanese

## Must Requirements

- Communication in English
- Knowledge of LLM
- Knowledge of image processing and natural language processing
- Understanding of machine learning and deep learning
- Advanced coding skills using Python

## Preferred Requirements

- Knowledge of VLM (including typical VLM architectures and basic concepts of multimodal learning)
- Experience in R&D of VLM
- Experience in R&D as a team
- Experience implementing using a deep learning framework
- Experience of having a domestic/international paper accepted as the first author
- Experience of contributing to OSS or giving a demonstration

↑ Back to list of themes

# RD06
# LLM inference optimization

**Area: LLM Fundamental Technology Development**

**Course: 7-weeks Research & Development Course**

In this internship, you will have the opportunity to work on cutting-edge performance optimization for LLM inference. At PFN, we use an open source inference engine to deploy our in-house LLM in production. Together, we will choose an open issue that is both interesting to you and useful for PFN. You will spend your time researching the state of the art of the specific issue and implementing a feature to address it. The focus will be on improving throughput, latency or memory utilization, for either generic or specific workload scenarios.

## Communication Language

English/Japanese

## Must Requirements

- Coding skills in Python
- Experience in PyTorch
- Basic knowledge of the Transformer architecture

## Preferred Requirements

- Basic knowledge of inference optimization
- Knowledge of the Mamba architecture
- Experience with using open source LLM inference engines (e.g. vLLM, llama.cpp, …)
    - Strong bonus: Contributions to these engines
- Coding experience in CUDA/Triton

↑ Back to list of themes

# RD09
# Development of MN-Core Compiler

**Area: MN-Core/Computing Infrastructure**

**Course: 7-weeks Research & Development Course**

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

- Improvement of MN-Core compiler algorithm
- Development of toolchain for MN-Core
- Development of verification tools for MN-Core design

## Communication Language

English/Japanese

## Must Requirements

- Basic knowledge of computer science
- Experience of Python, C++

## Preferred Requirements

- Experience of low level performance optimization
- Knowledge and experience of optimization algorithms
- Knowledge and experience of deep learning compiler stack

↑ Back to list of themes

# RD10
# Development ML/HPC Application on MN-Core

**Area: MN-Core/Computing Infrastructure**

**Course: 7-weeks Research & Development Course**

You will work on ML/HPC application porting for MN-Core accelerator. Your main task would be development for :

- Library for HPC (e.g. LAPACK)
- HPC Application (e.g. Atmospheric Model, Molecular Dynamics or RNA secondary structure prediction)
- ML workload

You can collaborate with compiler-core developers if you need it.

## Communication Language

English/Japanese

## Must Requirements

- Basic knowledge of computer science
- Experience of Python, C++

## Preferred Requirements

- Experience of low level performance optimization
- Experience on application development/porting for accelerator-based HPC system

↑ Back to list of themes

# RD11
# Development of Optuna

**Area: OSS Development**

**Course: 7-weeks Research & Development Course**

You will work on
(1) implementing new algorithms and
(2) implementing visualization and analysis functions
for the hyperparameter optimization framework, Optuna. You will focus on real-world applications and extend existing methods, such as Bayesian optimization and evolutionary computation, while also investigating, prototyping, and validating the effectiveness of new approaches based on deep learning, large language models, and diffusion models.
Reference:https://www.preferred.jp/ja/projects/optuna/

## Communication Language

English/Japanese

## Must Requirements

- Coding ability using Python
- Knowledge of black-box optimization (such as Bayesian optimization and evolutionary computation) or experience using Optuna

## Preferred Requirements

- Experience in OSS development
- Knowledge and development experience in machine learning
- Knowledge and development experience in deep learning
- Knowledge and development experience in large language models
- Knowledge and development experience in exploratory landscape analysis

↑ Back to list of themes

# RD12
# Development of Optuna Dashboard

**Area: OSS Development**

**Course: 7-weeks Research & Development Course**

Optuna Dashboard is a web dashboard for Optuna, enabling users to check the optimization history in graphs and tables. Currently, Optuna Dashboard doesn't have capabilities to collect and monitor its internal metrics, making it difficult to provide the reliable service to users. In this theme, you will enhance the observability of the Optuna Dashboard by introducing OpenTelemetry.
Reference:https://www.preferred.jp/ja/projects/optuna/

## Communication Language

English/Japanese

## Must Requirements

- Coding ability using Python and TypeScript
- Development experience of the server-side application
- Experience of the web development using React

## Preferred Requirements

- Knowledge and experience in OpenTelemetry
- Experience in OSS development

↑ Back to list of themes

# RD14
# Applied research and development of machine learning and atomic simulation on materials

**Area: Material Science (R&D)**

**Course: 7-weeks Research & Development Course**

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

## Communication Language

English/Japanese

## Must Requirements

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

## Preferred 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

# RD15
# Evaluation of material discovery methods on Matlantis or the improvement of the software development process using LLM

**Area: Material Science (Core Technology Development)**

**Course: 7-weeks Research & Development Course**

One of the following two topics is the main topic for this theme.

a) Implementing and evaluating computational chemistry for physical property using Matlantis.
b) Improve the software development process of Matlantis using LLM (Large Language Model).

In addition to this theme, there are also themes that focus more on research (RD14), service provision (RD16, 17,18). We can discuss which theme to join after application.
Reference: Matlantis: https://matlantis.com/

## Communication Language

English/Japanese

## Must Requirements

- Coding skills in Python
- Basic knowledge of computer science and software development
- Interest in physics, chemistry, or material science

## Preferred Requirements

- Motivation to bring value to customers and connect academic research to actual products
- College-level knowledge in physics, chemistry, or materials science

- Computational chemistry experience. For example, knowledge or usage of simulation techniques such as quantum chemical calculations, molecular dynamics methods, etc.
- Experience of the development process improvement using LLM
- Research or publication experience

↑ Back to list of themes

# RD22
# Reconstruction, editing, and generation of 3D models and free-viewpoint videos (Project)

**Area: PFN 3D Scan (Product/Service Development)**

**Course: 7-weeks Research & Development Course**

We are working on improving and expanding the technologies related to 3D/4D reconstruction, generation, and editing. Specifically, we anticipate you will be involved in developing new features for 3D-related products and services, as well as researching, implementing, and validating cutting-edge technologies. Reference: https://pfn3d.com/overview

## Communication Language

English/Japanese

## Must Requirements

- Proficiency in coding using Python.
- Experience with deep learning frameworks such as PyTorch.
- Knowledge in computer vision and 3D-related topics.

## Preferred Requirements

- Knowledge and research experience in 3D reconstruction, generation, free viewpoint image generation, CG, AR/VR, video production, and camera-related topics.
- Experience with 3DCG production software.

↑ Back to list of themes

# RD23
# Reconstruction, editing, and generation of 3D models and free-viewpoint videos (Research)

**Area: PFN 3D Scan (Core Technology Development)**

**Course: 7-weeks Research & Development Course**

We are working on improving and expanding the technologies related to 3D/4D reconstruction, generation, and editing. Specifically, we anticipate you will implement state-of-the-art 3D research and conduct fundamental research aimed at publication. We highly encourage students who have the ability to independently conduct research projects. Reference: https://pfn3d.com/overview

## Communication Language

English/Japanese

## Must Requirements

- Proficiency in coding using Python.
- Experience with deep learning frameworks such as PyTorch.
- Knowledge in computer vision and 3D-related topics.

## Preferred Requirements

- Track record of presenting research at international conferences and/or publishing in academic journals.
- Knowledge and research experience in 3D reconstruction, generation, free viewpoint image generation, CG, AR/VR, video production, and camera-related topics.
- Experience with 3DCG production software.

↑ Back to list of themes

# RD24
# Video generation model application service

**Area: Entertainment (Product/Service Development)**

**Course: 7-weeks Research & Development Course**

We welcome flexible ideas for solving problems in practical application of video generation models, interface improvements, and application proposals.

## Communication Language

English/Japanese

## Must Requirements

- Proficiency in coding using Python.
- Experience with deep learning frameworks such as PyTorch.
- Knowledge in 2D image generation

## Preferred Requirements

- Knowledge and research experience in 2D image or video generation / detection

↑ Back to list of themes

# RD38
# Research on machine learning methods for medical imaging

**Area: Health Care**

**Course: 7-weeks Research & Development Course**

We will develop machine learning methods to solve problems in the medical imaging domain (e.g. applications of foundation models, dealing with noisy labels, survey of novel 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

## Preferred 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