ALLIANCE RESEARCH INTERNSHIP PROGRAM
AT COLUMBIA UNIVERSITY
ACADEMIC YEAR 2021-2022

DEADLINE TO APPLY: NOVEMBER 5, 2021
Created in 2002, the Alliance Program is an innovative joint-venture between Columbia University, the École Polytechnique, Sciences Po, and Paris 1 Panthéon-Sorbonne University. Every year, Columbia University offers a number of student internships in scientific disciplines, which are open to École Polytechnique students. The process for applying to these internships is outlined below.

I. Internship Description

- Students work with a faculty member, who acts as an academic advisor and supervises their research project. **Internships will start in March/April 2022.** The duration, objectives and tasks of the internship will be discussed with the supervisor at the host center or department.

- Internships are not paid. **If a stipend is offered, it will be specified in the internship offer.**

- Students are responsible for finding housing.

- All students are required to apply for a J1 Visa to conduct an internship in the United States.

II. Applications requirements

- Applicants must include: a CV, a cover letter (1 page), and a letter of recommendation.

- Students must send their application to the Alliance Program: alliance@columbia.edu

- All materials must be submitted in English.

**DEADLINE – November 5, 2021**

All applications must be sent to alliance@columbia.edu
1. **Faculty Sponsor:**

   Pamela Smith, Seth Low Professor of History; Director, Making and Knowing Project (https://www.makingandknowing.org/)

2. **Number of interns:**

   One (1) to three (3)

3. **Type of support available:**

   ✓ Access to campus services and facilities
   ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

   Innovative interfaces and infrastructure for historians of art and science: analyzing and visualizing a sixteenth-century artisanal, technical French manuscript

5. **Description:**

   The Making and Knowing Project is a research and pedagogical initiative in the Center for Science and Society at Columbia University that explores the intersections between artistic making and scientific knowing. Drawing on techniques from laboratory, historical, and digital development research, the Project crosses the science/humanities divide and explores the relationships between today’s labs and the craft workshops of the past. In the last six years, the Project’s focus has been the creation of a Digital Critical Edition of a sixteenth-century technical and artisanal manuscript, Secrets of Craft and Nature in Renaissance France. A Digital Critical Edition and English Translation of BnF Ms. Fr. 640 (launched in February 2020). To prepare the content of the Edition, the Project has modeled an innovative methodological approach, combining hands-on research and pedagogy with iterative, collaborative work and techniques from the digital humanities and computer science. This has informed the Project’s next phase of work, the publication of a Research and Teaching Companion.
6. **Skills:**

- Programming/scripting languages (Python, Ruby, Java, C++)
- Linux command line utilities
- Familiarity with web development (HTML5, CSS, JavaScript)
- Desirable experiences:
  - Cloud infrastructure tools (AWS, Google Cloud, Docker, etc.)
  - Version control systems, especially git
  - Third-party REST APIs, especially oAuth and the Google Drive API
  - Advanced text manipulation and transforms, including regular expressions
  - Natural language processing, text mining, machine learning

7. **Additional Information:**

The intern will join the active research of the Making and Knowing Project team to carry out original research, working closely with Project Digital Lead, Assistant Director, and student programming assistants. As part of an interdisciplinary and collaborative project, the intern will learn to manage discipline- and domain-specific priorities, knowledge, and working methods. The Project has recently published a digital critical edition of Ms. Fr. 640, a 16th-century artisanal manuscript, and is actively working on additional features as well as upcoming publications of a Research and Teaching Companion and an experimental online “Sandbox,” where the work of previous École Polytechnique interns has already been published. The intern will work with the team to develop an independent research program that considers the Project’s needs, user experience of the Edition and future Companion, and the intern’s own interests and experience. Particular focus will be on experimentation with and development of innovative, interactive, and analytic interfaces to Project data. These could include textual analysis and visualizations using tools such as Jupyter Notebook, Voyant, and GIS; and techniques such as natural language processing, text mining, and machine learning. As an active and evolving project in the digital humanities, the Project also explores methods and solutions not just for front-end interfaces but also for innovative and sustainable infrastructures. The intern will be tasked with considering the underlying support for such exploration or, alternatively, focus on the development or optimization of infrastructure. For example, projects may concentrate on data management, archiving, and data migration. In line with the Project’s larger philosophy and methodology, the intern’s research project will serve to provide both additional scholarly insights as well as pedagogical tools to engage users, students, and researchers from varied disciplines. The intern will gain experience in active research, in developing and overseeing an independent research program, in working with a wide range of disciplines, and in exploring and developing interfaces and infrastructures that serve both scholarly and pedagogical aim.

8. **Remote option if travel is not allowed:** YES
1. **Faculty Sponsor:**

Szabolcs Marka, Walter O. LeCroy, Jr. Professor of Physics

2. **Number of interns:**

Two (2)

3. **Type of support available:**

✓ Access to campus services and facilities
✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Multimessenger astrophysics with gravitational waves, neutrinos, and electromagnetic radiation

5. **Description:**

Discoveries of cosmic gravitational waves, the bonanza of black-hole merger detections, and multimessenger astrophysics coming to age revolutionized how we understand the Universe. To fully appreciate this leap we need to understand multiple factors, including instrumental advances, theoretical frameworks, data/computing breakthroughs, and science goals. We will explore how the interplay between the emerging interferometric gravitational wave detectors and other bands of astronomy play a critical role in discovery, enabling signals to be extracted from ripples in Einstein's space-time. We will develop and deploy multimessenger searches for elusive cosmic sources that utilize gravitational waves, high energy neutrinos and a broad spectrum of electromagnetic radiation from gamma-rays to the optical bands. Students will carry out an impactful research project, acquire communication skills, and prepare a polished paper draft.

6. **Skills:**

Curiosity, self-motivation, independence, statistics, Python

7. **Additional Information:**

The research will be carried out in a Team-based setting where the students are expected to communicate and collaborate with peers and other early career scientists.

8. **Remote option if travel is not allowed:** YES
1. **Faculty Sponsor:**

Andrew J. Millis, Professor of Physics

2. **Number of interns:**

One (1)

3. **Type of support available:**

- ✓ Stipend: Up to $2500/month will be provided for reimbursement for lodging and other expenses incurred by the visit.
- ✓ Access to campus services and facilities
- ✓ Immigration and visa assistance/spoon

4. **Internship Title:**

Dynamical mean field studies of novel quantum materials

5. **Description:**

Working with members of my group and with scientists at the Simons Foundation’s Flatiron Institute, the intern will learn forefront methods of many body physics and apply them to elucidate the physics of new classes of materials, especially the transition metal tri and di halides such as the Kitaev spin liquid material RuCl3

6. **Skills:**

Knowledge of quantum mechanics; Experience with scientific computation; ability to program in python

7. **Additional Information:**

None

8. **Remote option if travel is not allowed:** NO
1. **Faculty Sponsor:**

Kyle T. Mandli, Associate Professor of Applied Physics and Applied Mathematics

2. **Number of interns:**

Two (2)

3. **Type of support available:**

✓ Access to campus services and facilities
✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Development of Computational Mathematics for Coastal Flooding under the Influence of Climate Change

5. **Description:**

Internship involves addressing a number of topics regarding coastal flooding and climate change in the context of the mathematics and computational difficulties of the problem. Topics of interest include uncertainty quantification and adaptive algorithm development.

6. **Skills:**

Numerical analysis, computational science, programming ability, command line comfort.

7. **Additional Information:**

None

8. **Remote option if travel is not allowed:** YES
1. **Faculty Sponsor:**

Agostino Capponi, Associate Professor of Industrial Engineering and Operations Research.

2. **Number of interns:**

One (1)

3. **Type of support available:**

- ✓ Stipend: Up to $5,000 depending on the student profile
- ✓ Access to campus services and facilities
- ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Mathematics and Control with Applications to Financial Technology

5. **Description:**

Use control techniques and game theory to assess incentives behind blockchain adoption. Use stochastic dynamical systems to model the price evolution of cryptocurrencies.

6. **Skills:**


7. **Additional Information:**

None

8. **Remote option if travel is not allowed:** NO
1. **Faculty Sponsor:**

Elham Azizi, Assistant Professor of Biomedical Engineering and Herbert and Florence Irving Assistant Professor of Cancer Data Research (in the Herbert and Florence Institute for Cancer Dynamics and in the Herbert Irving Comprehensive Cancer Center)

2. **Number of interns:**

One (1) to two (2)

3. **Type of support available:**

- ✓ Stipend
- ✓ Access to campus services and facilities
- ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Deep generative models for analysis of high-dimensional genomic and imaging data

5. **Description:**

We have multiple research projects involving the development of novel machine learning techniques for characterizing the dynamics of diverse cell types within cancer tumors towards understanding cancer progression and response to treatments. One project models temporal dynamics and interactions between cells during response to immunotherapies using graph convolutional networks with variational inference. Another project focuses on developing a variational autoencoder to integrate genomic and imaging data.

6. **Skills:**

Background in statistics, computer science, or bioengineering; and strong programming skills in Python.

7. **Additional Information:**

Please visit our lab website [www.azizilab.com](http://www.azizilab.com) for further information about our research.

8. **Remote option if travel is not allowed:** YES
IRVING INSTITUTE FOR CANCER DYNAMICS (IICD)

1. **Faculty Sponsor:**
   Professor Simon Tavaré (Director, Irving Institute for Cancer Dynamics [IICD]; Professor of Statistics and of Biological Sciences)

2. **Number of interns:**
   Two (2)

3. **Type of support available:**
   - ✓ Stipend
   - ✓ Access to campus services and facilities
   - ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**
   Irving Institute for Cancer Dynamics Internship

5. **Description:**
   The Irving Institute for Cancer Dynamics has opportunities for internships for students with interests in the mathematical sciences and their applications in cancer research. We have opportunities in areas such as analysis of single-cell DNA sequencing data, stochastic models of tumor growth, mathematical immunology, geometric and topological data analysis, mathematical population genetics, and phylogenetics. Internships can be crafted to suit the applicant.

6. **Skills:**
   Background in mathematics, statistics, computer science or biological sciences with a strong quantitative component. Prior experience in cancer research would be useful, but is not required.

7. **Additional Information:**
   The IICD is an interdisciplinary institute located on the Morningside Heights campus of Columbia University and focused on the interplay between the mathematical sciences and cancer research, collaborating across disciplinary boundaries to develop tools and methods that can improve our understanding of cancer biology, origins, treatment and prevention. Our website, at cancerdynamics.columbia.edu, gives an overview of our research teams and our current projects.

8. **Remote option if travel is not allowed:** NO
1. Faculty Sponsor:

Roxana Geambasu, Associate Professor of Computer Science

2. Number of interns:

One (1)

3. Type of support available:

✓ Stipend
✓ Access to campus services and facilities
✓ Immigration and visa assistance/sponsor

4. Internship Title:

Privacy Budget Scheduling in Machine Learning Systems

5. Description:

5.1. Project Background

Machine learning (ML) models trained on personal data have been shown to leak information about users [1]. Differential privacy (DP) enables model training with a guaranteed bound on this leakage [2]. Each new model trained with DP increases the bound on data leakage and can be seen as consuming part of a global privacy budget that should not be exceeded. This budget is a scarce resource that must be carefully managed to maximize the number of successfully trained models.

We extending Kubernetes, a popular datacenter orchestrator, to add privacy budget as a new type of resource to be managed alongside other traditional compute resources, such as CPU, GPU, and memory. DP ML jobs request privacy resource in addition to traditional resources when they want to compute something on user data. Our Kubernetes extension determines whether, and when, a privacy request should be granted in order to maximize the number of successfully trained models and minimize training delays. Because the privacy resource behaves differently from traditional resources, we require new algorithms to efficiently schedule privacy. Recently, we developed Dominant Private Block Fairness (DPF) – a variant of the popular Dominant Resource Fairness (DRF) algorithm – that is geared toward the non-replenishable privacy resource but enjoys similar theoretical properties as DRF, namely max-min fairness [3]. Unfortunately, like DRF, DPF does not guarantee to maximize the number of successfully trained models, and in fact can make very poor scheduling decisions in some cases. Thus, new algorithms are needed to optimize efficiency of the scheduling decisions and maximize the number of models successfully trained over time.
5.2. Internship Objectives

We are looking for an intern to work on one of the following related topics:

**Privacy scheduling**
- Review the literature for online scheduling algorithms (e.g. online multidimensional knapsack problem) and how to formalize privacy as a resource with differential privacy and in particular, with Rényi differential privacy, an efficient form of differential privacy [4].
- Develop online scheduling algorithms tailored to the privacy resource.
- Implement the algorithms on a distributed system, such as Kubernetes with an ML platform.
- Evaluate how the system and the algorithms perform in realistic settings.

**Privacy attacks**
- Review the main techniques for membership inference and database extraction attacks. Understand how differential privacy can protect individual models against attacks, and how composition works across different models.
- Design privacy attacks to extract sensitive information across different models. Compare different settings (types of models, data, training techniques) to find a realistic use-case that is vulnerable to attacks.
- Implement the attacks (potentially as an open-source package) and evaluate them against public datasets.

5.3. References


6. **Skills:**

   Required:
   - Good mathematical and algorithmic foundations.
   - Coding experience (which language matters less, but most of our project is Python and Go).
   - Interest in implementing and evaluating systems.

   Desirable:
   - Interest in theoretical analysis of algorithms.
   - Experience with a machine learning framework such as PyTorch.
   - Experience with distributed systems.

7. **Additional Information:**

   The intern will work most closely with Prof. Roxana Geambasu ([https://roxanageambasu.github.io/](https://roxanageambasu.github.io/)) and her Ph.D. student, Pierre Tholoniat ([https://tholoniat.me/](https://tholoniat.me/)), on one of the proposed projects.

   There is a larger team of professors and students involved in this project and the interns will have the opportunity to closely interact with this larger group. You can find a bit more information about the project on our website: [https://systems.cs.columbia.edu/PrivateKube/](https://systems.cs.columbia.edu/PrivateKube/)

   Finally, we are part of a vibrant group of software systems researchers ([https://systems.cs.columbia.edu/](https://systems.cs.columbia.edu/)) and hold weekly reading groups during the semester, which interns will be welcome to attend and interact with others.

   Both Roxana and Pierre would be happy to provide further information about this opportunity. Please email us at [roxana@cs.columbia.edu](mailto:roxana@cs.columbia.edu) or [pierre@cs.columbia.edu](mailto:pierre@cs.columbia.edu), respectively.

8. **Remote option if travel is not allowed:** YES
COLUMBIA UNIVERSITY RESEARCH GROUPS

Please find below a list of some Columbia University Research Groups.

➢ *Please note this list is not exhaustive and only includes some Departments that you might be interested in.*

➢ If you need to find the address email of a faculty/contact at Columbia University, please go to [http://www.columbia.edu/content/university-news.html](http://www.columbia.edu/content/university-news.html)

On the top of the page, please type the name of the person and you will access his exact department and contact information.
<table>
<thead>
<tr>
<th>COLUMBIA ENGINEERING</th>
<th>Research Involvement Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPLIED PHYSICS AND APPLIED MATHEMATICS</td>
<td>Research Groups</td>
</tr>
<tr>
<td>Applied Physics</td>
<td>Research Groups &amp; Centers</td>
</tr>
<tr>
<td>Applied Mathematics</td>
<td>Research Groups</td>
</tr>
<tr>
<td>Materials Science and Engineering</td>
<td></td>
</tr>
<tr>
<td>BIOMEDICAL ENGINEERING</td>
<td>Research Groups</td>
</tr>
<tr>
<td>Cell and Tissue</td>
<td></td>
</tr>
<tr>
<td>Biosignals and Biomedical Imaging</td>
<td></td>
</tr>
<tr>
<td>Biomechanics</td>
<td></td>
</tr>
<tr>
<td>CHEMICAL ENGINEERING</td>
<td>Research Groups</td>
</tr>
<tr>
<td>Biotechnology and Biological Engineering</td>
<td></td>
</tr>
<tr>
<td>Energy and the Environment</td>
<td></td>
</tr>
<tr>
<td>Soft Materials</td>
<td></td>
</tr>
<tr>
<td>Interdisciplinary Centers</td>
<td></td>
</tr>
<tr>
<td>CIVIL ENGINEERING AND ENGINEERING MECHANICS</td>
<td>Research Groups</td>
</tr>
<tr>
<td>Transportation engineering</td>
<td></td>
</tr>
<tr>
<td>Multi-Hazard Risk Assessment and Mitigation</td>
<td></td>
</tr>
<tr>
<td>Advanced Materials</td>
<td></td>
</tr>
<tr>
<td>Fluid mechanics</td>
<td></td>
</tr>
<tr>
<td>Geotechnical Engineering</td>
<td></td>
</tr>
<tr>
<td>Safety and Reliability</td>
<td></td>
</tr>
<tr>
<td>COMPUTER ENGINEERING</td>
<td>Research Groups</td>
</tr>
<tr>
<td>Computer Science</td>
<td></td>
</tr>
<tr>
<td>Electrical Engineering</td>
<td></td>
</tr>
</tbody>
</table>

Page 15 of 16
## EARTH AND ENVIRONMENTAL ENGINEERING

<table>
<thead>
<tr>
<th>Research Groups</th>
<th>Environmental Health Engineering</th>
<th>Sustainable Water</th>
<th>Sustainable Energy</th>
</tr>
</thead>
</table>

## BIOLOGICAL SCIENCES

<table>
<thead>
<tr>
<th>Research Groups</th>
<th>Dept. of Biology</th>
</tr>
</thead>
</table>

## CHEMISTRY

<table>
<thead>
<tr>
<th>Research Groups</th>
<th>Dept. of Chemistry</th>
</tr>
</thead>
</table>

## PHYSICS

<table>
<thead>
<tr>
<th>Research Groups</th>
<th>Dept. of Physics</th>
</tr>
</thead>
</table>

## INDUSTRIAL ENGINEERING & OPERATIONS RESEARCH

| Research Centers    | Optimization.  
|---------------------| Machine learning and Analytics  
|                      | Stochastic modeling and Simulation  
|                      | Financial engineering and Risk Management  
|                      | Operations                        |

## NEUROSCIENCE

| Contacts           | Dept. of Neuroscience  
|--------------------| Centers and Institute  
|                    | Contacts               |