

Columbia-IBM Center for Blockchain and Data Transparency Research Seed Funds Program

Funding Opportunity for Faculty and Research Staff

The digital transformation of the economy is creating a new natural resource: Digital data. Businesses transform the data about their customers, suppliers, or partners, into profitable products and services. It is natural to ask: What is the value of this data? How can individuals or organizations whose data is being collected participate in this data marketplace and exercise control over the access and use of their data? Extrapolating further, the value of a collection of individual datasets is often far greater than the values of each of the individual datasets; for example, fraud detection is likely to be more effective when banks combine their data, the success of precision medicine is predicated on merging records across multiple hospitals across the world, etc. What incentives can we provide for organizations to share their data for social good? What policies and mechanisms can be implemented to avoid biases in such collection processes? How can one provide guarantees of security and privacy?

Advances in technologies such as blockchain, homomorphic encryption, secure multi-party computation, zero knowledge proofs, and even secure hardware now make it possible for mutually distrusting organizations to share their data with each other in a secure, privacy-preserving, and tamper-proof manner. Some of these technologies do not rely on a (centralized) trusted third-party. These capabilities are already leading to novel applications in supply-chain management and food security, but the potential to impact all sectors, such as health, finance, energy, legal, media and entertainment, is enormous.

Columbia-IBM Center for Blockchain and Data Transparency aims to support research, educational, and entrepreneurial activities broadly in the area of data sharing, fair use of data, blockchains or related technologies.

Program Goal & Topics

- **Area 1 - Technology, systems and algorithms:** Advances in security, cryptography, game theory, network science, machine learning, and distributed systems that enable and incentivize trusted data sharing & secure distributed computations
- **Area 2 - Business model, services and applications:** Innovative applications of blockchain technology and data transparency technologies, including specific industry use cases
- **Area 3 - Policy, regulation, law and behavior:** Fundamental legal and policy frameworks for distributed data sharing networks, and business models to create, sustain, and evolve these networks and their applications

While specific topics of interest in each of these areas are given below, we welcome other proposals that are broadly in the area of data transparency and blockchains.

A list of potential IBM collaborators is available on the seed fund website
<https://datascience.columbia.edu/columbia-ibm-seed-funds-program/collaborators>

Area 1 - Technology, systems and algorithms

- Common features of problems where blockchains are likely to be useful
- Scalable security technologies for secure and privacy preserving data sharing
- Mechanism design for permissioned and permission-less data sharing
- Methods for valuation of networked data
- Cryptographic techniques for the secure, scalable, auditable access and computation on highly regulated data
- Collaborative/multi-party federated machine learning
- Model transparency in machine learning and AI
- Secure, privacy preserving, and fair Machine learning
- Hash functions for physical objects based on material properties for provenance verification

Area 2 - Business model, services and applications

- Data sharing applications in healthcare, finance, supply chain & logistics, energy, and media & entertainment.
- Practical approaches for handling private data in highly regulated domains, e.g. healthcare and financial services
- Digital tokens, in particular tokens representing physical assets, to create new marketplaces & financial products
- Blockchains for provisioning public goods and donations
- Business models for incentivizing data sharing

Area 3 - Policy, regulation, law and behavior

- Economic and regulatory issues in the design of smart contracts
- Valuation methodologies for data centric firms and industries
- The economics, policy and regulatory issues in data and model marketplaces
- Benefits of transparency and auditability of data and data analysis technologies in the context of public policy

The goal of this program is to support novel projects that are exploratory and not yet ready for funding from traditional sources. We anticipate successful projects will be able to secure long-term funding from external sources after completion of the seed projects.

Funding and Obligations

Funding is available for up to five projects, up to \$100,000 annually, with the opportunity to apply for renewal of support up to a maximum of two years. As a condition of funding, awardees will be required to submit quarterly financial reviews and biannual progress reports. Eligibility for continued funding for a second year will also require a progress report and justification of needs for continued research. All reports must include progress on external funding proposal submission(s). Furthermore, the awardees would be expected to participate in furthering the mission of the Center by acknowledging the support in all papers on the topics, and participating in workshops and demo days, as appropriate.

Proposal Process

The deadline for submission is November 1, 2018 by 3:00 pm. We will not accept incomplete or late submissions. We anticipate notifying award recipients by December 10, 2018.

Please submit the following materials via email, in **.doc** or **.pdf** format, to CU-IBM-SeedFund@columbia.edu, by the **November 1, 2018, 3:00pm deadline:**

- Application Cover Page
- Project Proposal (5-page maximum, single space, 12-font, Times New Roman)
- Budget in Excel ([template provided](#)) with a budget justification. Funds will be awarded as an Industry Grant (PG Project) with 35% IC to be included in your budget.
- CV's for Faculty/Collaborators ([2 page NSF style format](#))

The proposals will be reviewed by a technical committee consisting researchers drawn from relevant Departments and Schools across Columbia University and from IBM. The final selection will be done by the Center Steering Committee.

All the software and copyrightable material created in the performance of the projects funded by these seed grant must be made publicly available, either by publication or by open source license or open copyright license. Unless otherwise agreed upon, Columbia will grant a license for patents (if any) filed for inventions developed under support of this seed funding to IBM which will be worldwide, perpetual, irrevocable, non-exclusive, and non-transferable.

Review Criteria

All projects must be relevant to data transparency and blockchain, and the proposers should select the areas closest to their proposal in the section on Program Goals. The following criteria will be used in ranking and selecting submissions for seed funding:

1. Impact on advancing the state of the art of the proposed project.
2. Novelty of the proposed project.
3. Need of the seed funding for the success of this project.
4. The intended follow-up to apply for subsequent external funding
5. The review committee will rank projects within each area, and recommend a portfolio of projects across the areas for possible funding.
6. Impact on establishing legal systems or policies needed for broad applications of blockchain and data transparency technologies

