MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

# Center for Computational Drug Discovery ($C^2D^2$)
## Whitepaper for Prospective Companies

### Introduction

The University of Michigan's Institute for Data Science (MIDAS) has been awarded a planning proposal by the National Science Foundation's Industry-University Cooperative Research Centers (NSF IUCRC) program to establish a research center for precompetitive research into computational drug discovery using novel machine learning and artificial intelligence techniques. We are seeking industry partners whose research efforts align with the proposed Center's focus and who will be active collaborators in shaping the next generation of drug discovery research. The motivation for the Center, its research areas, and your potential role within it are detailed below.

### Motivation

As many newly developed drugs fail in clinical trial or to yield a profit, pharmaceutical researchers use computational techniques to discover new clinical applications or drug-target interactions. Few machine learning techniques exist for discovering new drug/target candidates, while the fragmented nature of drug/target databases impede efforts to leverage this data for research. The proposed Center's machine learning approach to drug discovery using novel techniques for finding new drug/target interactions in integrated public and proprietary datasets can accelerate the rate of drug discovery by addressing industry data silos and reducing research and development costs.

### The Center for Computational Drug Discovery ($C^2D^2$)

$C^2D^2$ will focus on three main areas with the goal of significantly accelerating the pace of drug discovery while reducing research costs. These research areas represent the general capabilities that the Center will offer, but can be readily extended to meet the needs of participating industry partners:

a) *Development, testing, and validation of machine learning methods for drug discovery* – State of the art machine learning and artificial intelligence techniques will be brought to bear on discovering new drug-target interactions using public, academic, and industry partner-provided drug/target databases. These include advanced tensor algebraic techniques that were designed to systematically integrate auxiliary information currently ignored or under-utilized by extant methods. Additionally, novel machine learning-based methodologies developed for molecular representation and other chemoinformatic tasks will also be further developed.

b) Provide an industry-wide and vendor-agnostic Secure Data Hub for pharmaceutical and patient data with third-party private search capabilities – By applying advanced algebraic and statistical methods, the Center will create large, encrypted, and integrated databases of drug targets for which the participating companies need only share full encryptions of their proprietary databases. Such databases enable third-party private search, allowing partners to search securely and anonymously for new drugs targets or other data.

c) Enable federated machine learning for drug repositioning over encrypted databases – Center investigators have adapted a number of machine learning algorithms for use over encrypted data. The Center will further integrate these algorithms into a federated learning framework. Such technologies will enable machine learning over databases that cannot be integrated into the Secure Data Hub.

MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

**About the NSF IUCRC Program and Your Potential Collaboration**

The proposed Center will leverage the existing community of data science and domain knowledge experts of MIDAS and the University of Michigan. MIDAS and its affiliated faculty will actively engage with industry partners to construct and implement a computational platform using novel algorithmic approaches to extract, validate, and integrate information from different drug/target datasets in a secure and vendor-agnostic manner.

The NSF IUCRC program has been enabling long-term research partnerships between industry, academia, and government for over 40 years. The IUCRC program facilitates the establishment of a thematic center focused on pre-competitive research projects, in which participating members have the ability to solicit and select proposals for development. The NSF funds a single center at the national level for any one research area and has indicated its support for our proposal to be ***the*** NSF-funded center fulfilling the national need for computational drug discovery.

The creation of a new center is a multi-phase process. We are currently in the planning phase, during which we will determine the Center's organizational structure, operating procedures, intellectual property policies, and initial experimental plan.

During this planning phase, financial commitment letters will be required for all companies who wish to become members of the proposed Center. Membership entails royalty-free use of intellectual properties developed through the Center's research efforts, along with voting rights on the Industry Advisory Board (IAB), with each member receiving a pro-rated number of votes based on their membership status. We anticipate the following annual membership fee structure: *Full Membership* at $80,000 for one vote, $120,000 for one and one-half votes, or $160,000 for two votes; and *Associate Membership* at $40,000 for one-half vote. Members exercise their voting rights through the Center's IAB, choosing which pre-competitive research projects to fund. The NSF provides a minimum of $150,000 annually in the first 5 years of the Center's operations, with additional matching funds depending upon industry participation. The NSF IUCRC program stipulates that 90% of all membership and NSF monies directly fund research projects.

Meetings were held with prospective Center members on August 17-19, during which attendees had an opportunity to learn more about the Center and review multiple pre-competitive research project proposals for potential inclusion in the initial set of proposals to be evaluated by the Industry Advisory Board (meeting recordings are available on request). Prospective members and MIDAS will now work together to finalize the proposal through mid-September, at which time financial commitment letters are due. Upon completion of the planning phase, the full proposal will be sent to the NSF for consideration by December 8, 2021. We anticipate the NSF proposal review process to take six to nine months, with a potential Center start date of September 2022.

To learn more about the Center for Computational Drug Discovery, please contact Dr. Jonathan Gryak (gryakj@med.umich.edu), Senior Scientist for MIDAS.