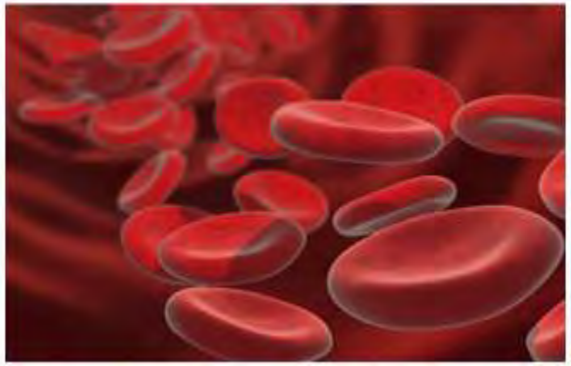
Welcome to...

# AI and Medical Research Opportunities At NSF

Sylvia Spengler (CISE/IIS/III); sspengle@nsf.gov


# NSF Organizational Structure

Office of Equity and Civil Rights (OECR)

Office of the General Counsel (OGC)

Office of Integrative Activities (OIA)

Office of International Science and Engineering (OISE)

Office of Legislative and Public Affairs (OLPA)

## Office of the Director (OD)

### Director

Deputy Director

Chief Operating Officer

Chief of Research Facilities

Chief Information Officer

Chief of Research Security, Strategy and Policy

## NSB National Science Board

Chair, Vice Chair,
NSB Executive Officer

## Office of the Inspector General (OIG)

Directorate for Biological Sciences (BIO)

Directorate for Computer and Information Science and Engineering (CISE)

Directorate for STEM Education (EDU)

Directorate for Engineering (ENG)

Directorate for Geosciences (GEO)

Directorate for Mathematical and Physical Sciences (MPS)

Directorate for Social, Behavioral and Economic Services (SBE)

Directorate for Technology, Innovation and Partnerships (TIP)

Office of Budget, Finance and Award Management (BFA)

Office of Information and Resource Management (OIRM)

# Directorate for Computer and Information Science and Engineering

## Office of the Assistant Director

**Gregory Hager**
Assistant Director

**Joydip Kundu**
Deputy Assistant Director

## CISE Advisory Committee

## Division of Computer and Network Systems (CNS)

**Ellen Zegura**
Division Director

**Behrooz Shirazi**
Deputy Division Director

## Division of Computing and Communication Foundations (CCF)

**Dilma DaSilva**
Division Director

**Irina Dolinskaya**
Deputy Division Director

## Division of Information and Intelligent Systems (IIS)

**Michael Littman**
Division Director

**Wendy Nilsen**
Deputy Division Director

## Office of Advanced Cyberinfrastructure (OAC)

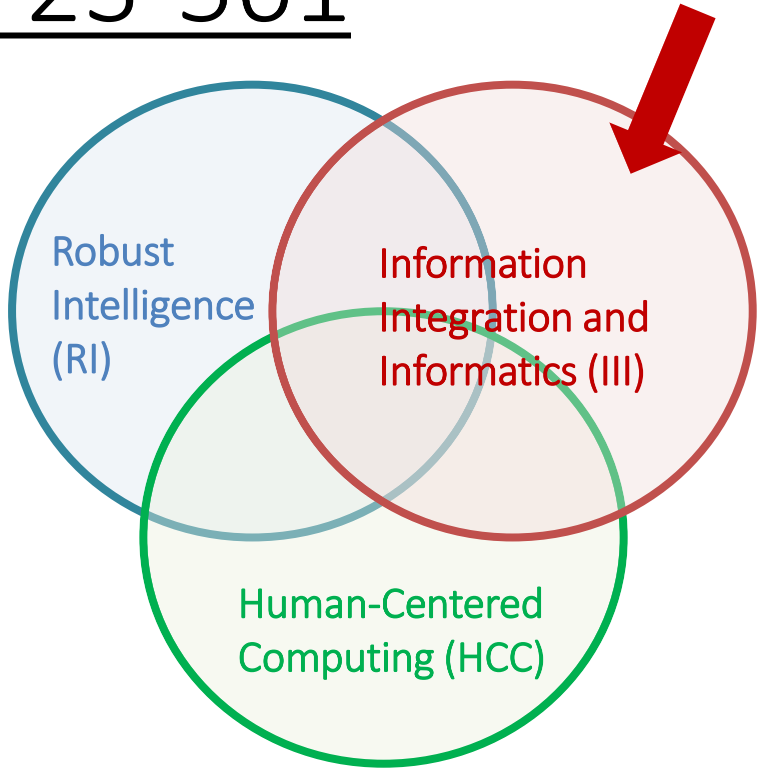
**Katie Antypas**
Office Director

**Amy Walton**
Deputy Office Director

# Information and Intelligent Systems CORE Solicitation NSF 23-561

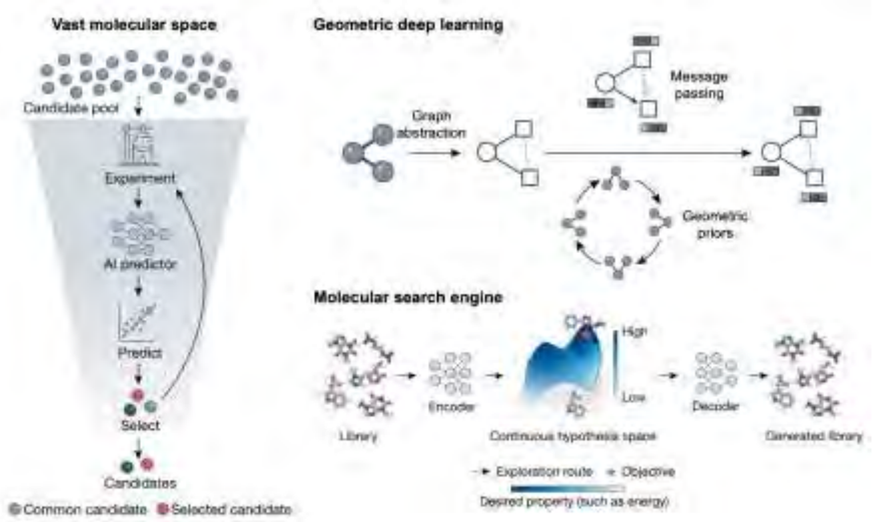
IIS supports research in the inter-related roles of people, computers, and information.

IIS includes three core programs:
Human-Centered Computing (HCC), Information Integration and Informatics (III), and Robust Intelligence (RI).
+ many interdisciplinary crosscutting programs.



**CAREER: Geometric Deep Learning to Facilitate Algorithmic and Scientific Advances in Therapeutics (2339524)**

Institution: Harvard University
PI: Marinka Zitnik




**Summary:** This project aims to reshape therapeutic science through machine learning algorithms for graph-structured datasets, focusing on the vast and complex world of drug-like molecules and proteins. It leverages these algorithms to create **molecular search engines** that find useful molecules for a given application considering dozens of criteria at once with the potential to explore fundamentally larger molecular spaces that would never be accessible to experimental screening at a fraction of the cost.

**Motivation:**

- Drug-like molecular space is estimated at $10^{60}$ diverse molecules
- Multimodal data, including sequences, geometric structures, and biological activity, contain an enormous amount of information about molecules with likely therapeutically useful effects
- Critical gap is how to navigate the vast molecular space to find therapeutically useful molecules
- Unlocking this information requires generalizable and multimodal geometric deep learning algorithms adept at a broad array of tasks

**Broader Impacts**: To address the medical needs of billions worldwide, this project ambitiously strives to help compress traditional 13-15 year and $2-3 billion drug development process through novel geometric deep learning algorithms trained on large molecular interaction datasets.

**Outreach and Broader Impacts Plan**:

- Interdisciplinary collaborations through Therapeutics Data Commons platform with AI models, tasks, and benchmarks
- Tutorials and workshops on AI4Science at conferences
- Graduate course in biomedical AI and undergraduate mentored research program
- Education materials for AI4Science jobs
- AI in Medicine workshops for middle and high school students at Alexander Twilight Academy
- Education about responsible use of AI in scientific research

**Intellectual Merit**: Establish a suite of geometric deep learning models that will be evaluated towards a vast array of therapeutic tasks. Foundation models, especially large language models, are transforming the paradigm of deep learning: instead of training many task-specific models, we can adapt a single pretrained foundation model to many tasks via few-shot prompting or fine-tuning. However, current foundation models apply to sequence data but not to graph data, which present unique challenges due to large disparity among different domains (e.g., knowledge graphs vs. molecules).

This project addresses 3 challenges open in molecular machine learning that must be solved to establish foundation graph models that generate general graph representations adept at diverse tasks and domains:

- **Generalizable graph models of molecular interactions**
- **Context-aware and multi-level graph models**
- **Multimodal graph pretraining models**

**Marinka Zitnik,** https://zitniklab.hms.harvard.edu, marinka@hms.harvard.edu

## Smart Health and Biomedical Research in the Age of Artificial Intelligence and Advanced Data Science (SCH)

- A cross-cutting program supported by NSF and NIH
- Most recent solicitation **NSF 23-614**
- The program funds:
  - Integrative Projects with ≤ $1,200,000 total cost
  - $300,000/year up to 4 years
  - Interdisciplinary teams
  - Contribution to fundamental science from at least two NSF sciences

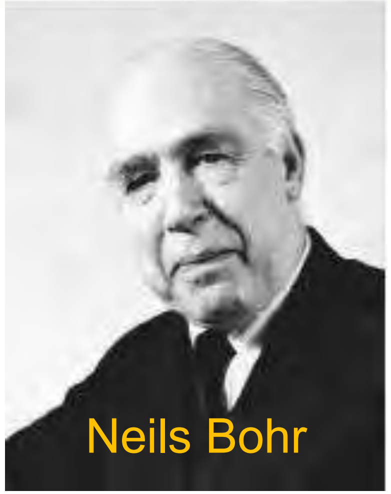
- ***SCH is a case of use-inspired basic research***

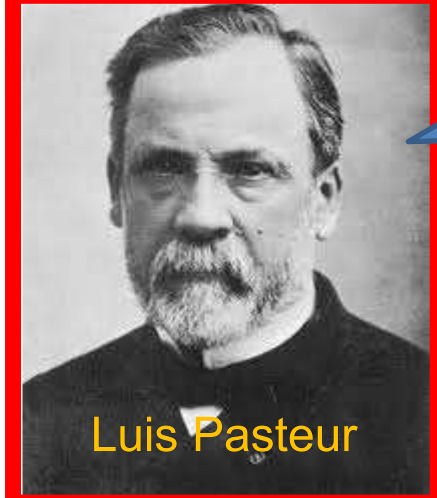
# Pasteur's Quadrant

This is the SCH quadrant: use-inspired basic research

Quest for Basic Understanding

High




Low


Low High

Application Inspired: Consideration of Use

Donald E. Stokes, *Pasteur's Quadrant – Basic Science and Technological Innovation, Brookings Institution Press, 1997*

# TWO DISTINCT SCIENTIFIC CONTRIBUTIONS

- Computer and Information Science, e.g., artificial intelligence, machine learning, informatics, computer architecture, databases, natural language technology, networks, robotics, etc.
- Engineering, e.g., multimodal sensor technology, signal processing, optimization, operations research, closed loop systems, dynamic systems analysis, materials, etc.
- Mathematics and Statistics, e.g., stochastic modeling, analysis, interpretation and characterizing uncertainty, etc.
- Social, Behavioral and Cognitive Sciences, e.g., perception, social psychology, cognition, emotion, economics, ethics, linguistics etc.

# Data Challenges and Opportunities

New analytic methods should be:

- Efficient
- Standardized
- Personalizable
- Scalable
- Domain-Sensitive
- Explanatory
- Able to deal with Uncertainty
- Visually Representable

# Data Challenges and Opportunities

- Methods to identify batch effects, “missingness” and non-representativeness
- Simulate missing data based on redundancies in health data sets
- Methods that would allow real-time assessment and information on problematic data
- Dealing with biases

## Data Validity

- Does the data represent what we think it does?
- Is the variance important – signal or noise?

# Smart Health Research Themes

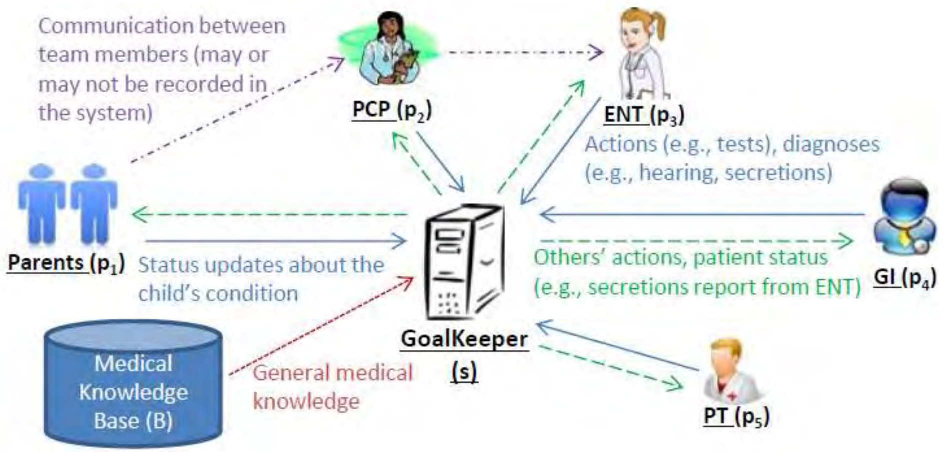
- Robotics
- Transformative Analytics in Biomedical and Behavioral Research
- Fairness and Trustworthiness
- **Your Idea**
- Cyber-Physical Systems
- Next Generation Multimodal and Reconfigurable Sensing Systems
- Biomedical Image Interpretation
- Unpacking Health Disparities and Health Equity

# Effective Research is a relay between basic and applied science

# SCH: INT: COLLABORATIVE RESEARCH INTELLIGENT INFORMATION SHARING: ADVANCING TEAMWORK IN COMPLEX CARE

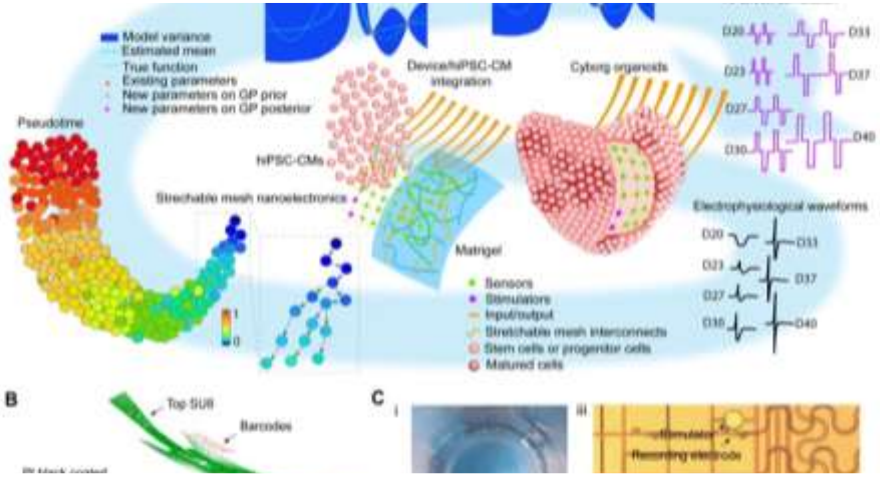
Barbara Grosz,
Harvard University IIS-1521820/Lee Sanders,
Stanford University IIS-1521832
NIH R01-CA204585

- The research aims to provide foundations for intelligent interactive systems that assist care teams in using integrated care plans in the context of shared decision-making to improve the care of children with complex chronic conditions.
- Develop new multi-agent systems algorithms that are able to reason about plans with only partial information, in conjunction with human-computer interaction methods to support use of integrated care plans.
- The resulting GoalKeeper system supports care plan use through the development of a novel representation, mutual influence potential networks.



Harvard University
PI: Jia Liu

## AI-embedded cyborg system for accelerating human stem cell-derived cardiac tissue maturation




## Summary:

This proposal seeks to establish an AI-driven "cyborg tissue" platform that integrates flexible electronic sensors and actuators with developing human stem cell-derived cardiac tissues for monitoring and controlling tissue maturation. The closed-loop system will collect multimodal data from cells, analyze the data using statistical models, and generate control policies through reinforcement learning (RL) to promote tissue maturation via distributed electrical actuations.

## Motivation:

- Align with NSF mission to advance biomedical research by artificial Intelligence and data science
- Driven by the need to integrate AI with bioelectronics for biological and biomedical studies
- Critical gap to be addressed is how to use AI to interpret multimodal biological data for closed-loop control

## Broader Impacts:

- Technical innovations will bring broader impacts to AI, bioelectronics, and bioengineering communities
- Developed hardware and software can be applied to virtually any current biological systems
- Research and education plans will further merge the field of AI, bioelectronics, and regenerative medicine

## Outreach and Broader Impacts Plan:

- Engage graduate and undergraduate students with AI, bioelectronics, and regenerative medicine.
- Develop interdisciplinary curricula.
- Support underrepresented minority students in AI, bioelectronics, and bioengineering.
- Expand K-12, local, and global community outreach.

## Intellectual Merit:

- Integrate multifunctional bioelectronics with cardiac organoids, enabling real-time monitoring and control.
- Develop machine learning models and tools to identify cell maturation states based on multimodal recording.
- Develop data-efficient and scalable RL methods to determine optimized stimulation policy to promote cardiac maturation

## Research Plan:

Aim1. Multimodal bioelectronics for cardiac organoids
Aim2. AI-enabled online learning from multi-modal data
Aim3. Bio-inspired RL for cardiac maturation
Evaluation: 1) Electronics performance; 2) Statistical model accuracy; 3) Numerical test of RL model; 4) Tissue maturation

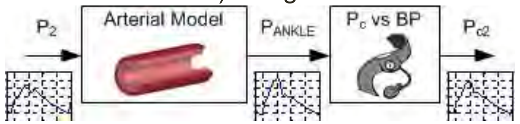
**Contact: Jia Liu, https://liulab.seas.harvard.edu/**
**jia_liu@seas.harvard.edu**

# A Low-Cost and Non-Invasive Measurement Method for Personalized Cardiovascular Risk Predictors

Jin-Oh Hahn, University of Maryland
Rama Mukkamala, Michigan State University
IIS-1404436/IIS-1403004

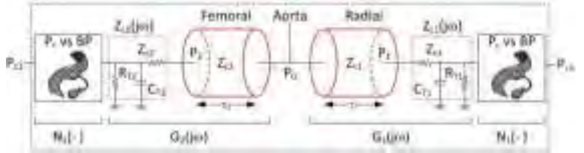
**Goal:** To improve cardiovascular risk stratification by non-invasive and affordable measurement of cardiovascular risk predictors

1) Derive blood pressure waves from non-invasive blood volume waves (e.g., oscillometric cuff oscillations) using a mathematical model

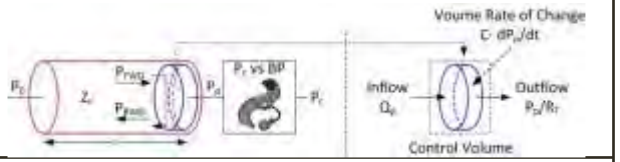



2) Derive cardiovascular risk predictors from blood pressure and volume waves using adaptive signal processing

Using 2 Cuff Signals (Arm and Ankle)




Using 1 Cuff Signal (Arm or Ankle)




## Motivation:

•This project is relevant to the SCH program since it may accelerate the development of novel methods for cardiovascular risk predictor measurement, thus improving the quality of life and well being.

•This research is motivated by the fact that conventional methods to measure cardiovascular risk predictors are expensive, inconvenient and non-personalized.

•The critical gap to be addressed is the investigation of low-cost, non-invasive methods for measurement of cardiovascular risk predictors.

## Transformative:

•This project is transformative since it will improve the standard of care and stimulate the development of related medical devices.

•This project vertically advances the field because it may advance mathematical modeling and signal processing relevant to physiologic systems modeling and health monitoring.

## Broader Impacts:

•Advancement of personalized healthcare

•Stimulation of medical device industry and economy

•Research and Education are integrated by developing course modules on physiologic systems modeling and analysis, and by disseminating the knowledge to K-12, women and minority students

## Technical Approach:

•We use two blood volume waveform signals to derive cardiovascular risk predictors. Specifically, we use two oscillometric cuffs to measure cuff oscillations at an arm and an ankle.

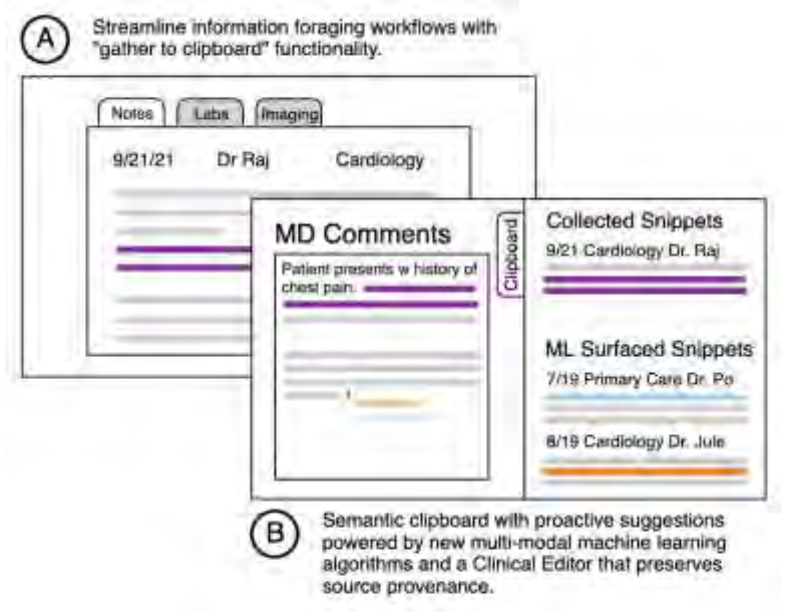
•These oscillations are analyzed by an adaptive signal processing algorithm incorporating the mathematical model relating blood volume to blood pressure waveforms in order to derive personalized cardiovascular risk predictors.

## Contacts:

- PI: Jin-Oh Hahn (University of Maryland)
- Co-PI: Rama Mukkamala (Michigan State University)
- Collaborator: Hao-Min Cheng (Taipei VGH)
- www.terpconnect.umd.edu/~jhahn12

## *Machine Learning Driven User Interfaces for Information Gathering and Synthesis from Medical Records*

Institutions: MIT and Beth Israel Deaconess Medical Center
PIs: David Sontag, David Karger, Steven Horng




**Motivation:**

- Relates to *Smart Health* as we aim to leverage AI and HCI to streamline clinical workflows, specifically for clinical documentation, by surfacing and reasoning using robust knowledge representations of multi-modal clinical data
- Research is motivated by easing burden in interfacing with EHRs, as clinicians currently spend more time in the EHR than with patients
- Critical gap to be addressed is seamless integration of robust machine learning algorithms where prior labelled data does not exist
- Vertically advances the field because implicitly learning information needs in an end-to-end fashion and creates cleaner data sources for future work
- Transformative because better information retrieval in EHRs can make clinicians more efficient and decrease medical errors

**Objective**:

- To automatically, dynamically, surface relevant past clinical history and assist in note-writing process
- To seamlessly capture data on what information needs clinicians have in the documentation process and provide automatic data provenance

**Technical Approach**:

We develop and deploy a foraging user interface in the EHR that seamlessly captures implicit usage-based information needs, to be used as training data. Next, we develop machine learning algorithms that use the collected data to proactively suggest context-specific and user-tailored clinical information. Finally, we develop and pilot a novel interface for integrating ML suggestions into the note writing process.

**Outreach and Broader Impacts Plan**:

- Decrease the time spent in the EHR, currently posited to be a leading cause of physician burnout.
- Provide an illustrative example of seamless, unobtrusive integration of decision support in clinical workflows
- Research and Education are integrated by training cross-disciplinary scientists (teaching clinicians ML, and CS researchers medicine) in a diverse research group of all stages.

**Prior Results, Deliverables:**

- Result: ML and user interface for deployed context-specific autocomplete.
- Result: Auto-surfaced information based on weakly supervised signals.
- Deliverable: Efficient algorithms for prediction of information needs from multi-modal sources. Design & implementation of 'semantic clipboard'.
- Deliverable: Open-source framework for integration of suggestions into note writing interface.

**Schedule:**

We expect to develop the initial foraging interface in the 1st year and start deployment in the 2nd year. We will simultaneously start development of algorithms in the 1st year, and evaluate and evolve continuously as more data becomes available in years 2-4. An integrated note-writing suggestion interface will be developed in years 2 and 3, and piloted in year 4.

**Contact: David Sontag,**
dsontag@csail.mit.edu http://clinicalml.org

# Assessing fit and prepping to write

- Do your homework
  - NEW PAPPG Effective MAY 2024, NSF 24-1
  - Read the solicitation (NSF-23-614)carefully multiple times
  - Check for webinars, FAQs, other program resources
- Talk to NSF Program Officers about your ideas
  - Prep a 1-2 page writeup in advance
  - Key questions, contributions, activities

# Help the Reviewers

- You are writing to reviewers and program officers.
- Make what they are looking for easy to find, using the language of the review criteria and headings to highlight the elements of the project description.
- Don't assume that all reviewers will know the jargon of your community or commonly used acronyms; reviewers may not be in your subspecialty.
- Make sure the most important things – the contributions, the plans – receive the most space.

# Common proposal gotchas

Importance

- Problem not scientifically/societally important
- Weak, vague, or no connection to STEM content
- Relevant literatures not cited

Methods

- Inadequate or inappropriate research design
- Vague or inappropriate data collection & analyses
- Evaluation missing/unclear
- Appropriate expertise not represented on team
- Unclear generalizablity

# Other gotchas

- Ignoring requirements stated in the solicitation or PAPPG
- The "Trust Me" approach. Provide citations or evidence for critical assertions made, and details of work.
- The "Oversell" of yourself or your project. Persuade with the quality of the proposal, not glitzy/sales language.
- General, vague, or rambling narrative; low precision/details
- Too much rationale and prior work, not enough method and details of what will actually be implemented.

# Thank you!

- Questions or Comments?

Sspengle@nsf.gov