

# Roadmap of scalable analytic methods (SAM) and their applications to research

SAM refers to a broad collection of novel data-intensive methods, including artificial intelligence, machine learning, techniques for manipulating “big data,” and data science.

SAM approaches are promising but can be challenging to incorporate effectively into research.

Enter this map from one of three routes:

## Route 1

You know your research objective and want to identify relevant SAM methods.

## Route 2

You have a favorite method and want to see how it fits within this framework.

## Route 3

You want to see examples of how SAM methods were used in published RAND projects.

1

Research Objective



### Exploratory Analysis

**Purpose: Explore data and generate viable hypotheses**

This type of analysis helps make sense of what data you have and is well suited to unsupervised learning and factor analysis.

### Statistical Inference

**Purpose: Test a hypothesis or calculate an estimate and its precision**

This type of analysis is a mathematically formalized way of approximating reality. The approximation can be used to test relationships and estimate outcomes.

### Modeling & Simulation

**Purpose: Create and tune models that meaningfully represent real-world dynamics.**

This type of analysis is useful for observing dynamics that cannot or should not be directly observed in the real world.

2

Methods



Examples of traditional (non-SAM) approaches

- Text analysis
- Principal component analysis (PCA) & factor analysis
- Data visualization
- Network analysis

- Linear regression
- Decision trees
- Traditional causal inference

- Ordinary differential equation model
- Monte Carlo analysis

Examples of SAM approaches

- Clustering algorithms
- Synthetic network model
- Web scraping
- RAND Lex

- Neural network
- Causal inference supported by machine learning estimation

- Agent-based model (ABM)
- Advanced tuning of microsimulation models

3

Applications



### A Identifying dependencies and systemic risk in the broad economy

Used SAM to build more complete and detailed network maps of U.S. firms to identify dependencies between them. Uncovered new insights because of the maps' higher fidelity.

[www.rand.org/t/RR4185](http://www.rand.org/t/RR4185)

### B Hunting Russian trolls on Twitter and Reddit with AI

Tested whether machine learning (ML) can be used to identify Russian trolls across social media platforms. The ML model achieved 98% precision and almost 100% accuracy.

<https://technarrativelab.org/defending-the-2020-us-elections-and-beyond-hunting-russian-trolls-on-twitter-and-reddit-with-ai/>

### C Using ML to dramatically reduce the time needed to update systematic reviews

Compared standard methods for updating systematic reviews with ML-assisted updating. ML drastically reduced the effort required to update systematic reviews without sacrificing quality.

[www.rand.org/t/EP67251](http://www.rand.org/t/EP67251)

### D Testing a social network intervention to reduce substance use

Employed computer-supported social network visualizations with motivational interviewing to test the impact of social network interventions on high-risk behavior among individuals transitioning from unhoused to supportive housing.

[www.rand.org/t/EP67485](http://www.rand.org/t/EP67485)

### E Better representations of flu vaccination behaviors using reinforcement learning

Used reinforcement learning (RL) on agent-based models (ABMs) of flu vaccination behavior to test RL's impact. Found that RL effectively generates reward-seeking and maximizing behaviors and can outperform default ABMs.

<https://arxiv.org/abs/2006.05048>

### F Seeking COVID-19 exit strategies for reopening California

Calibrated a disease model using a SAM technique piloted at RAND (incremental mixture approximate Bayesian computation, or IMABC algorithm) that is better suited to high-dimensional parameter spaces and many calibration targets.

<https://projecteuclid.org/journals/annals-of-applied-statistics/volume-13/issue-4/Microsimulation-model-calibration-using-incremental-mixture-approximate-Bayesian-computation/10.1214/19-AOAS1279.short>

For inquiries and SAM expertise, please email us at [AskSAM@rand.org](mailto:AskSAM@rand.org).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

For more information on this publication, visit [www.rand.org/t/CPA1893-1](http://www.rand.org/t/CPA1893-1).  
[www.rand.org](http://www.rand.org) | © Copyright 2022 RAND Corporation