FAIR, Workflow, and AI

Stuart Feldman
Schmidt Futures
12 September 2019

“AI”

- Broad usage in this talk
- Includes many concepts and techniques
  - Pattern recognition
  - Machine Learning
    - Not just CNNs
  - Image Analysis
  - Bayesian optimization
  - Robotics
  - Autonomy
  - Natural Language Processing

FAIR and AI

- A huge topic, needing a separate full-scale study
- AI is key to producing new data (image analysis etc., robotic labs, decisions)
- AI techniques will be applied to extant data, so systems will be dependent on FAIR principles for success
- AI applications will themselves be important providers of metadata – what techniques were used, what was the result of learning, what were details of the inputs

FAIR and Workflow

- Automated workflows will be a major producer of truly FAIR data
- By formalizing the activities, abundant metadata can be defined and captured far better than manual intervention
  - experimental state (all available measurements)
  - human and mechanical operations (detailed)
  - intermediate results
  - decisions
- But workflow doesn't guarantee openness or FAIRness – it only enables it

Workflow

- Computer-assisted process management is an old topic, almost as old as computers
  - A spectrum of rigor, from *ad hoc* to *formal* systems to meet user demands and output requirements
  - A spectrum of degree of automation from *advice* to *full control*
- The (business) Workflow Management Coalition was founded in 1993.
- Computer scientists love writing workflow engines – a perfect fit to the software mentality ☺
- Researchers usually don't enjoy formalizing their work (serendipity!) or being controlled by a system
- Workflow engines are numerous (1000s), workflow languages almost as common
  - Cynically, more creators than users of these systems

Workflow – What is Changing

- Usage has been spotty among researchers
  - Obligatory in certain regulated areas
    - e.g., pharma with FDA requirements
  - Easily available open systems (e.g., Kepler) but slowly growing user base
  - Not just US – a major aspect of e-science and EU science tech and policy
- But a likely renaissance driven by
  - Scientific Productivity opportunities –more science faster and cheaper
  - Replicability demands
    - Openness and full records
  - Notebook use
  - New generation of researchers comfortable with automation and computers in the loop

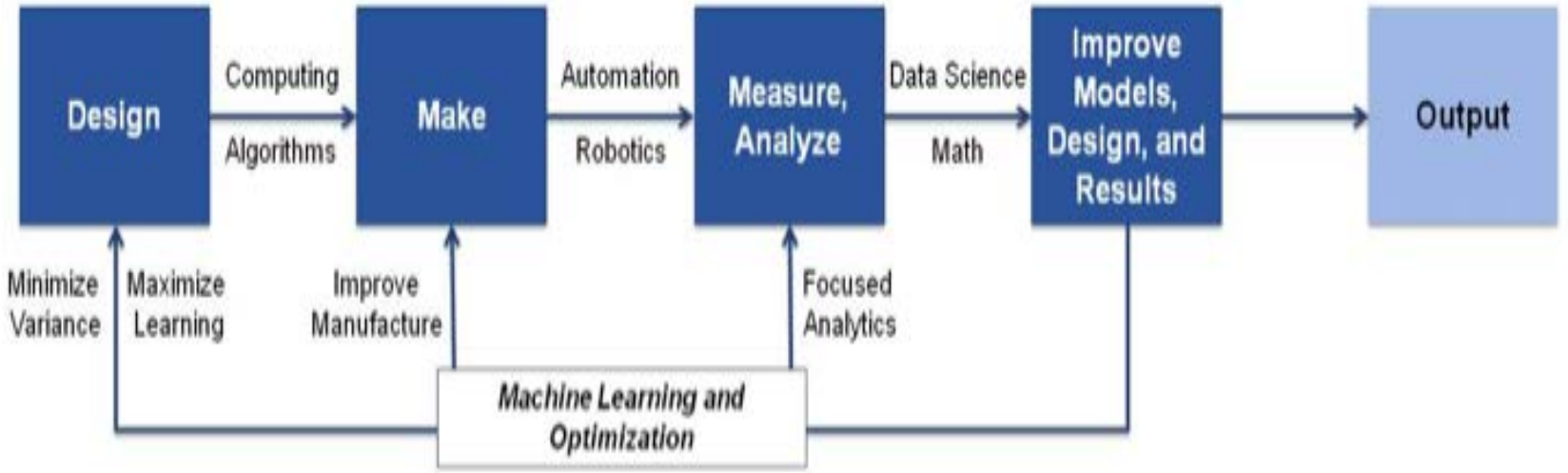
Scientific Research and AI

- AI as a tool for research
  - Image Analysis
  - Pattern Matching
  - Lab Automation
  - Natural Language analysis
  - Etc. etc.
- AI as a driver of research
  - AI is not just a tool, it's a mindset
  - How to approach problems in the large
  - Planning to generate and analyze significant amounts of complex data
  - Using objective validated measures, not just small-scale intuition
  - With automation and mathematics, a university-scale lab could out-produce a traditional industrial lab
- AI as a creator of knowledge
  - To be continued

Workflow and AI

- Use AI as a tool in many of the steps
- Use AI to (help) determine the steps to be taken
- Humans remain part of the whole process

AI Driving the Outer Loop: AI-Accelerated Science



Schmidt Futures
AI-Accelerated Science Projects

- Astronomy: Spectrographic Observations
- Biology: New Antibiotic Discovery
- Chemistry: New Catalysts and Nanostructures
- Earth Science: Earth Systems Climate Model
- Neuroscience: Development of Neural System
- Social Science: Liberating Government Data
- More coming…

National Academies Study

- BRDI is sponsoring a study on this subject
  - Approved by the Academies
  - Schmidt Futures is financial sponsor (only)
- Daniel Atkins is the chair of a distinguished panel representing many viewpoints
- First steering board meeting August 2019
- A workshop will be held in DC

Study: Realizing Opportunities for Advanced and Automated Workflows in Scientific Research

An ad hoc committee will conduct a study that examines current efforts to develop advanced and automated workflows for scientific research. The study will also identify promising research approaches to accelerating progress in the effectiveness and utilization of workflow systems and tools.

The committee's primary information gathering will consist of a workshop that examines the status of research workflows in several example fields, key barriers and enablers, and emerging opportunities.

The workshop will explore the role of open science, in the form of broad access to research articles, data, and analytical code, and other enabling factors.

Based on insights from the workshop, a review of the literature, and other inputs, the committee will produce a consensus report that identified research needs and priorities in the use of advanced and automated workflows in scientific research.

Summary

- New generations of researchers are embracing the techniques due to convenience of technology and opportunities for research output
- Modern workflow management is a huge opportunity for academic research.
  - AI will be a growing part of the practice
- Automation is a great lever for supporting the best open practices
- Automated workflows will be a major user and provider of FAIR data and a promoter of the cause