CCAI 2020 大会特邀报告(II)

南加州大学ACM/AAAI Fellow：

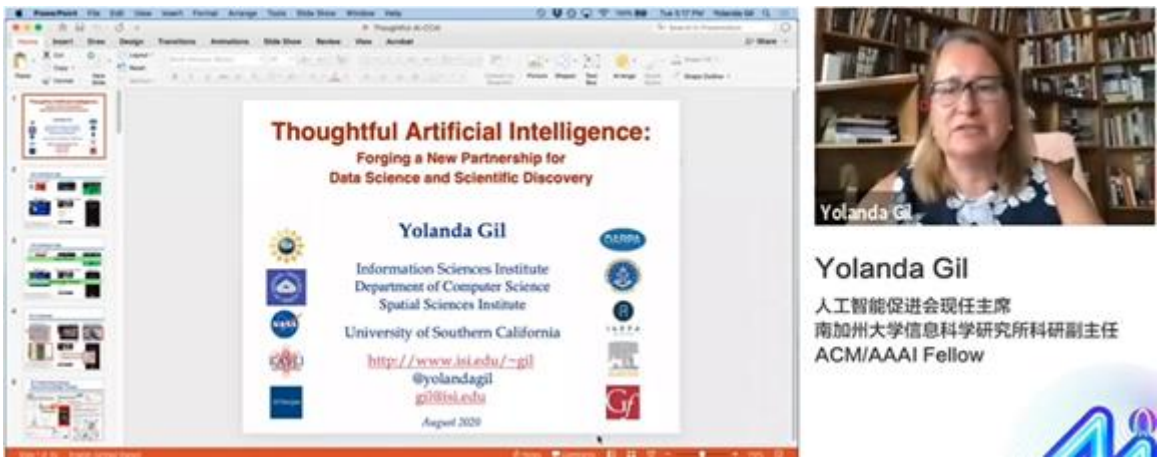
Thoughtful Artificial Intelligence：

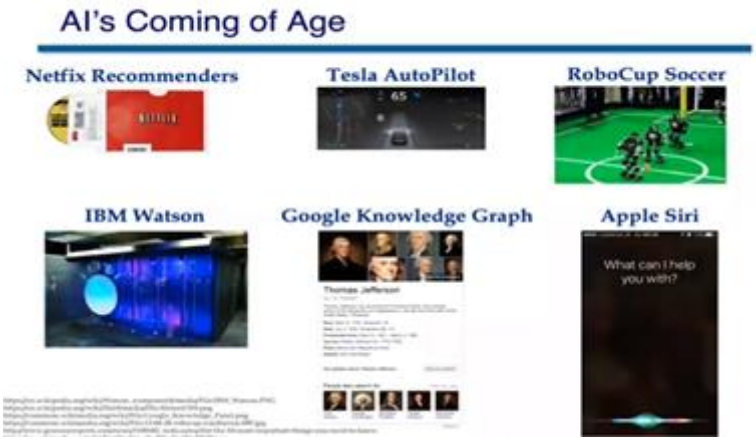
Forging a New Partnership for Data Science and Scientific Discovery

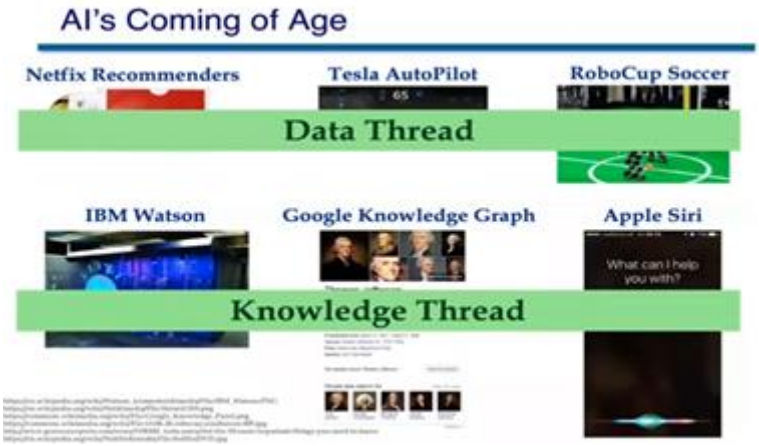
临菲信息技术港

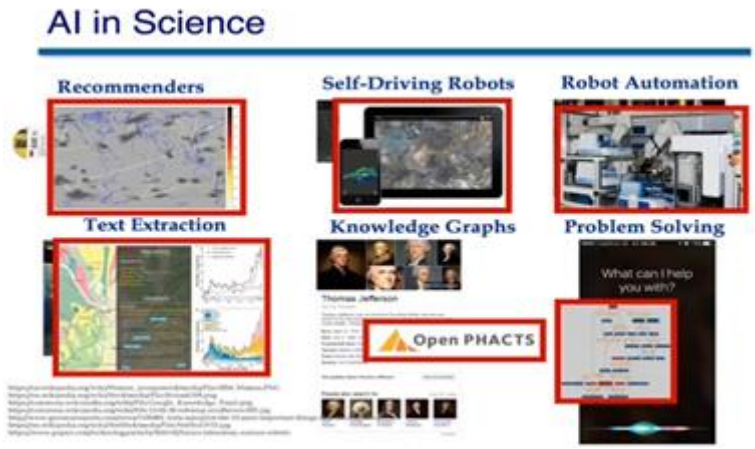
CCAI 2020 第二天大会特邀报告。ppt 根据视频摘录。

报告人：Yolanda Dil，人工智能促进会现任主席，南加州大学信息科学研究所科研副主任，ACM/AAAI Fellow。

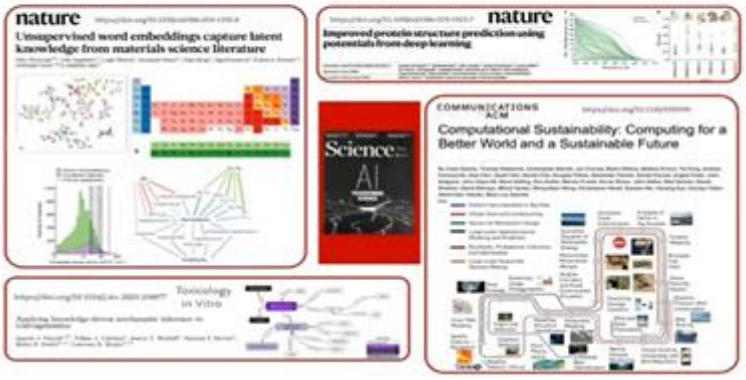



AI's Coming of Age
Netfix Recommenders
Tesla AutoPilot
RoboCup Soccer
IBM Watson
Google Knowledge Graph
Apple Siri
What can I help you with?


AI's Coming of Age
Netfix Recommenders
Tesla AutoPilot
RoboCup Soccer
Data Thread
IBM Watson
Google Knowledge Graph
Apple Siri
What can I help you with?
Knowledge Thread


AI in Science
Recommenders
Self-Driving Robots
Robot Automation
Text Extraction
Knowledge Graphs
Problem Solving
Open PHACTS
What can I help you with?

AI Transforming Science: Data and Knowledge Threads



A Tradition of AI and Scientific Discovery: Data and Knowledge Threads

- [Simon 1955]
- [Lenat 1976]
- [Lindsay, Buchanan, Feigenbaum & Lederberg 1980]
- [Langley & Simon 1981]
- [Simon et al 1983]
- [Falkenhainer 1985]
- [Langley et al 1987]
- [Kulkarni and Simon 1988]
- [Cheeseman et al 1989]
- [Zytkow et al 1990]
- [Valdes-Perez 1997]
- [Todorovski et al 2000]

1. Knowledge technologies are increasingly important

Thirst for Knowledge Technologies is Growing: 1) Industry-Scale Knowledge Bases

- Google's Knowledge Graph (1B) [Noy et al 2019]
- Microsoft's Satori (2B)
- Amazon's Product Graph (100M)
- Facebook's Social Graph (50M)
- eBay (100M)
- IBM (100M)
- Many others
  - Thompson Reuters
  - Springer Nature
  - …

METAWEB Freebase yago WIKIDATA

Thirst for Knowledge Technologies is Growing: 2) Natural Language

- Extracting common sense knowledge from text
  - [Forbes and Choi ACL 2017]
  - [Rashkin et al 2016]
- Adversarial examples
  - [Jia and Liang EMNLP 2017]

Article: Super Bowl 50
Paragraph: *"Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV."*
Question: *"What is the name of the quarterback who was 38 in Super Bowl XXXIII?"*
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

Figure 1: An example from the SQuAD dataset. The BiDAF Ensemble model originally gets the answer correct, but is fooled by the addition of an adversarial distracting sentence (in blue).

Thirst for Knowledge Technologies is Growing: 3) Computer Vision

Challenges on Content and Updating

-- Jamie Taylor, Google Knowledge Graph Group, ISWC'17

In this project, we attempt to study how many articles are needed to cover the sum of all human knowledge. As of 20 October 2019, English Wikipedia has 5,954,910 articles and Wikidata includes 64,116,277* items. This page, still in expansion, estimates that the total notable articles figure is over 104,701,020.

At current creation rate, 8,000 new items per day, Wikidata singularity will occur in the 2040s [...] It will be the first time in history that all human knowledge is stored in a machine-readable format and ready to be consumed, understood and used by computers or any device.

(* was 27,721,774 as of July 2017)

1. Knowledge technologies are increasingly important
2. AI offers systematic, correct, unbiased approaches and rigorous reporting

The Need for Knowledge Technologies in Scientific Discovery [Gil DSJ 2017]

Human limitations curb scientific progress:

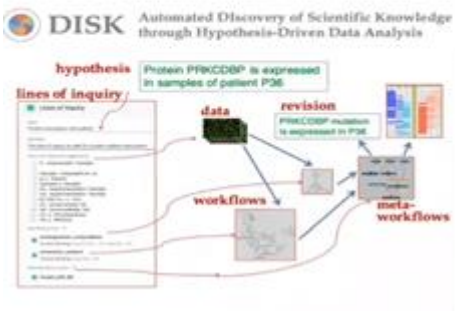
- Not systematic:
  - e.g., [Peters et al PLOS 2014]
- Errors:
  - e.g., [Herndon et al CJE 2013]
- Biases:
  - e.g., [Anderson et al ACS 2014]
- Poor reporting:
  - e.g., [Garijo et al PLOS 2013]

Artificial Intelligence offers systematic, correct, unbiased approaches and rigorous reporting

Automated vs Manual Hypothesis Testing

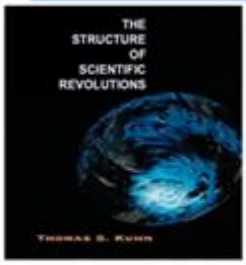
With Parag Malick, Ravali Adusumilli, Hunter Boyce (Stanford); Arunima Srivastava (OSU); Daniel Garijo, Varun Ratnakar, Rajiv Mayani (USC/ISI); Thomas Yu (Sage Bionetworks)

[Gil et al ACS 2016; Gil et al AAAI 2017; Garijo et al 2017; Srivastava et al PSB 2019; Mallick et al 2020]



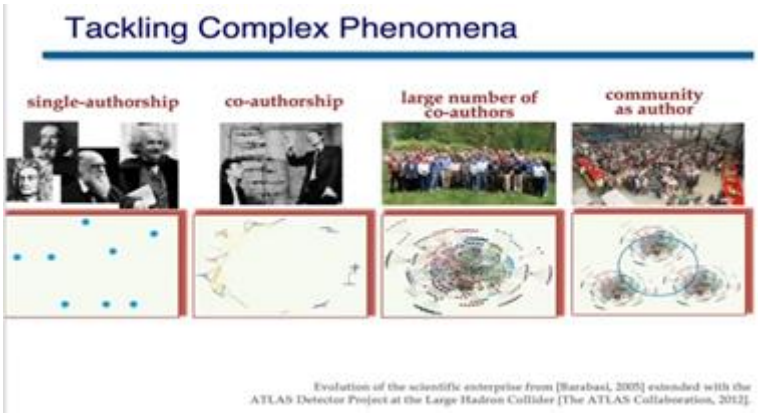
35% of protein identifications are not robust to changing just one analysis step, eg search engine

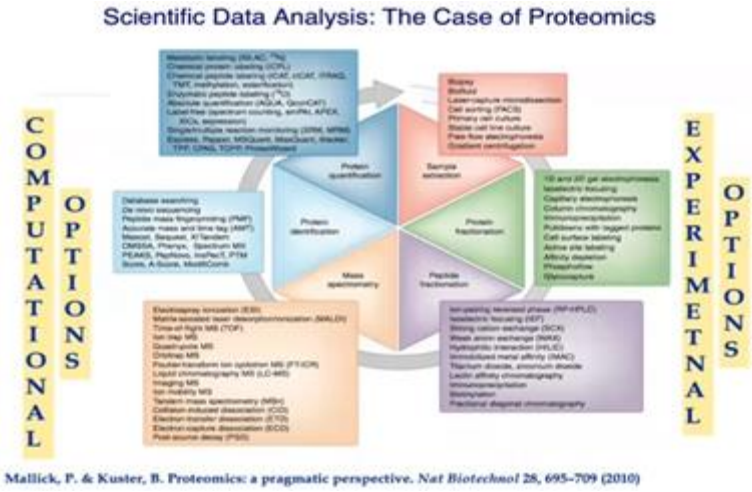
Analyzing the Scientific Discovery Process

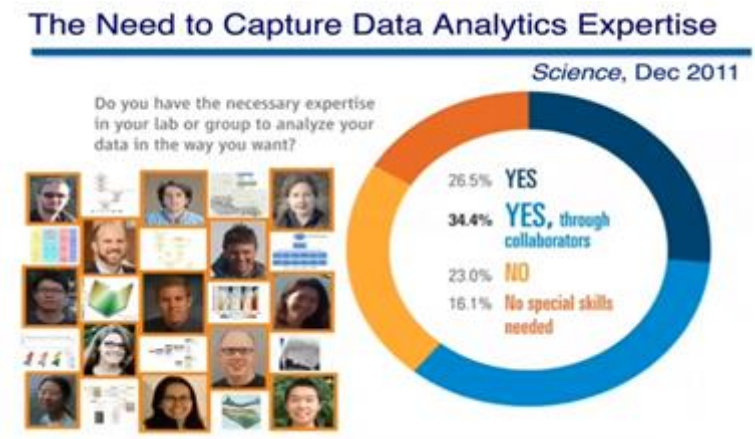


1. Knowledge technologies are increasingly important
2. AI offers systematic, correct, unbiased approaches and rigorous reporting

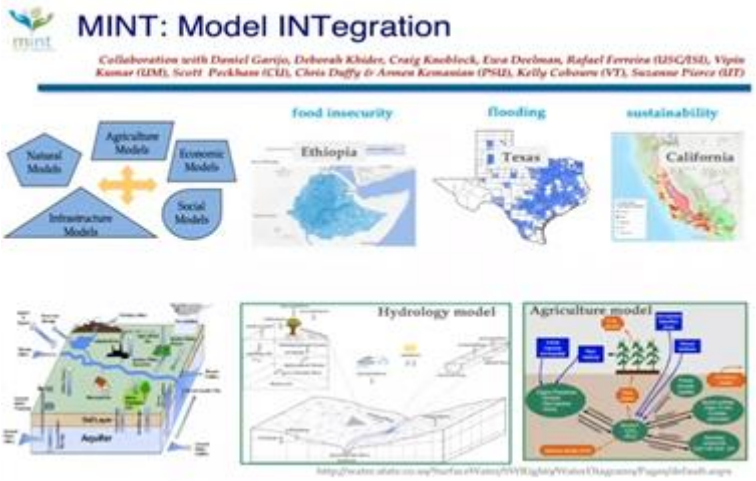
3. AI will excel at assembling fragmented knowledge about complex systems and pursue interdisciplinary frontiers

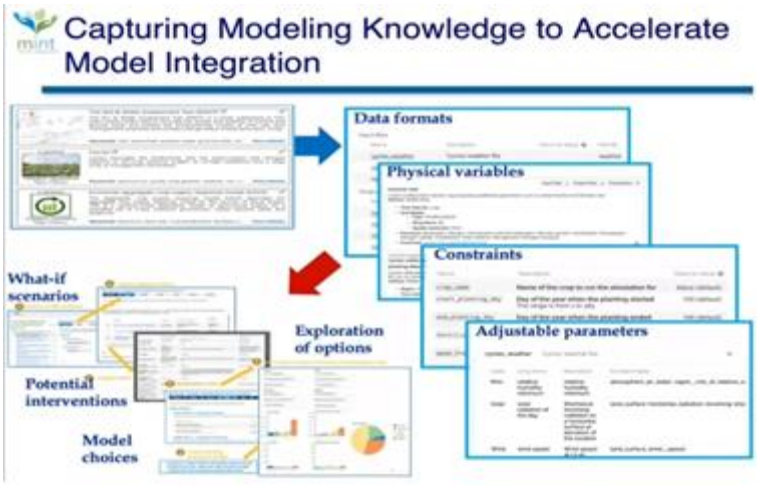

Tackling Complex Phenomena
single-authorship
co-authorship
large number of co-authors
community as author
Evolution of the scientific enterprise from [Barabasi, 2005] extended with the ATLAS Detector Project at the Large Hadron Collider [The ATLAS Collaboration, 2012].


Scientific Data Analysis: The Case of Proteomics
COMPUTATIONAL OPTIONS
EXPERIMETNAL OPTIONS
Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. Nat Biotechnol 28, 695-709 (2010)


The Need to Capture Data Analytics Expertise
Science, Dec 2011
Do you have the necessary expertise in your lab or group to analyze your data in the way you want?
26.5% YES
34.4% YES, through collaborators
23.0% NO
16.1% No special skills needed


Artificial Intelligence systems should be partners that excel at assembling fragmented knowledge about complex systems and pursue interdisciplinary frontiers


MINT: Model INTegration
Collaboration with Daniel Garijo, Deborah Khider, Craig Knoblock, Ewa Deelman, Rafael Ferreira (USC/ISI), Vipin Kumar (UM), Scott Peckham (CU), Chris Duffy & Arnoen Kemanian (PSU), Kelly Cobourn (VT), Suzanne Pierce (UT)
food insecurity
flooding
sustainability
Agriculture Models
Natural Models
Economic Models
Social Models
Infrastructure Models
Ethiopia
Texas
California
Hydrology model
Agriculture model



1. Knowledge technologies are increasingly important
2. AI offers systematic, correct, unbiased approaches and rigorous reporting

3. AI will excel at assembling fragmented knowledge about complex systems and pursue interdisciplinary frontiers

4. *Thoughtful AI* will exploit knowledge technologies for effective human-AI partnerships

The Importance of Process

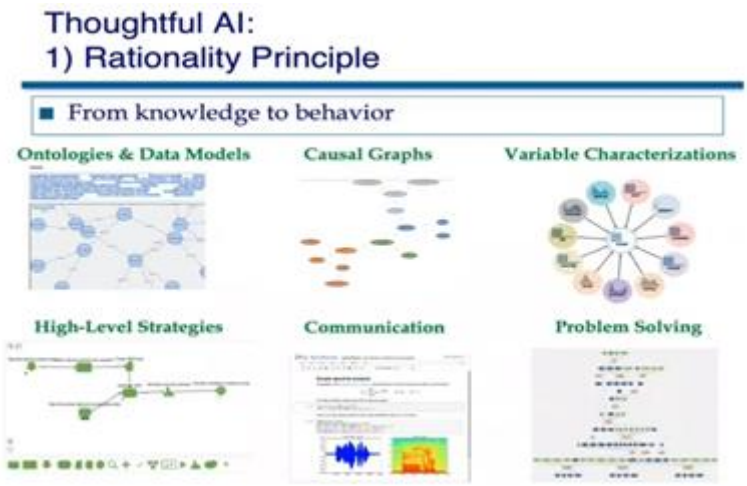
Freestyle Chess Champion Anson Williams

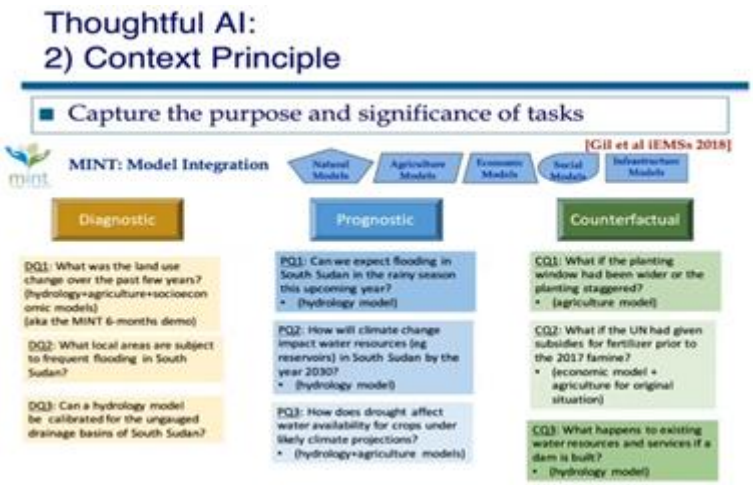
"The winner was revealed to be not a grandmaster with a state-of-the-art PC but a pair of amateur American chess players using three computers at the same time. Their skill at manipulating and "coaching" their computers to look very deeply into positions effectively counteracted the superior chess understanding of their grandmaster opponents and the greater computational power of other participants. **Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process.**"

– Gary Kasparov, 2010

Thoughtful Artificial Intelligence [Gil DSJ'17]

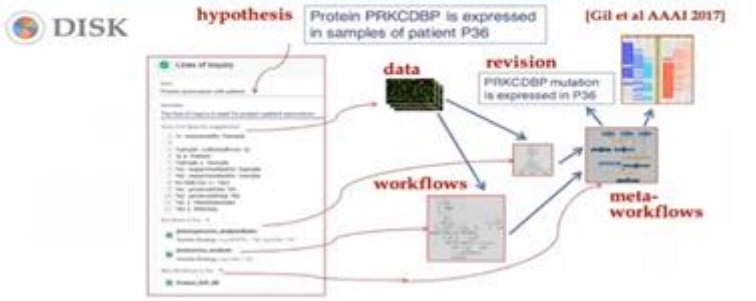
1. Rationality principle: knowledge to behavior
2. Context principle: purpose and significance
3. Initiative principle: self-driven learning
4. Network principle: seek more resources
5. Articulation principle: respond + ask
6. Ethical principle: uncertainty + unknowns
7. Systems principle: compositionality


Thoughtful AI:
1) Rationality Principle
From knowledge to behavior
Ontologies & Data Models
Causal Graphs
Variable Characterizations
High-Level Strategies
Communication
Problem Solving


Thoughtful AI:
2) Context Principle
Capture the purpose and significance of tasks
[Gil et al iEMSs 2018]
MINT: Model Integration
Diagnostic
Prognostic
Counterfactual
DQ1: What was the land use change over the past few years? (hydrology+agriculture+socioeconomic models) (aka the MINT 6-months demo)
DQ2: What local areas are subject to frequent flooding in South Sudan?
DQ3: Can a hydrology model be calibrated for the ungauged drainage basins of South Sudan?
PQ1: Can we expect flooding in South Sudan in the rainy season this upcoming year?
• (hydrology model)
PQ2: How will climate change impact water resources (eg reservoirs) in South Sudan by the year 2030?
• (hydrology model)
PQ3: How does drought affect water availability for crops under likely climate projections?
• (hydrology+agriculture models)
CQ1: What if the planting window had been wider or the planting staggered?
• (agriculture model)
CQ2: What if the UN had given subsidies for fertilizer prior to the 2017 famine?
• (economic model + agriculture for original situation)
CQ3: What happens to existing water resources and services if a dam is built?
• (hydrology model)

Thoughtful AI: 3) Initiative Principle

- Self-driven learning
  - From the scientific record
  - Through independent inquiry



Thoughtful AI: 4) Network Principle

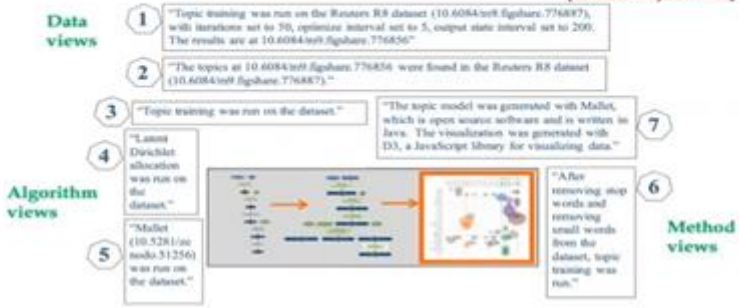
- Seek resources from a knowledge web of the scientific record

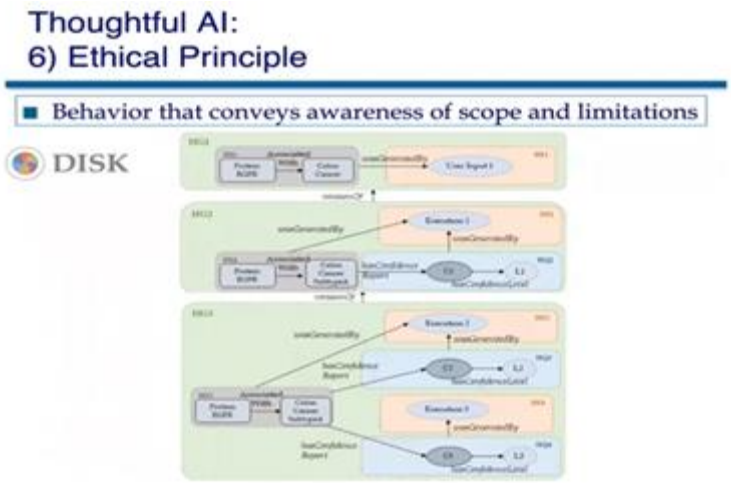
[Gil et al ISWC 2017; Garijo et al FGCS 2017; Gil et al eScience 2015; Gil et al IEEE IS 2011]

Thoughtful AI: 5) Articulation Principle

- Communicating findings to different audiences
- Respond, but also ask and discuss!

[Gil and Garijo IUI 2017]




Thoughtful AI:
6) Ethical Principle
Behavior that conveys awareness of scope and limitations
DISK


Thoughtful AI:
7) Systems Principle
Compositionality, abstraction, connectivity

Thoughtful Artificial Intelligence [Gil DSJ'17]

1. Rationality principle: knowledge to behavior
2. Context principle: purpose and significance
3. Initiative principle: self-driven learning
4. Network principle: seek more resources
5. Articulation principle: respond + ask
6. Ethical principle: uncertainty + unknowns
7. Systems principle: compositionality

Thoughtful Artificial Intelligence:

Forging a New Partnership for Data Science and Scientific Discovery

1. Knowledge technologies are increasingly important
2. AI offers systematic, correct, unbiased approaches and rigorous reporting

3. AI will excel at assembling fragmented knowledge about complex systems and pursue interdisciplinary frontiers

4. *Thoughtful AI* will exploit knowledge technologies for effective human-AI partnerships

相关阅读：**CCAI 2020 大会特邀报告**

临菲信息技术港

临菲信息技术港公众号

临菲学堂