To appear in AI Magazine, 2021.

AAAI Presidential Address

Will AI Write Scientific Papers in the Future?

Yolanda Gil, University of Southern California
gil@isi.edu

Delivered on February 9, 2020
at the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20) in New York, NY, USA

Abstract

In this presidential address, I would like to start with a personal reflection on the field and then share with you the research directions I am pursuing and my excitement about the future of AI. In my personal research to advance AI while advancing scientific discoveries, one question that I have been pondering for some years now is whether AI will write scientific papers in the future. I want to reflect on this question, and look back at the many accomplishments in our field that can make us very hopeful that the answer will be yes, and that it may happen sooner than we might expect.

A Personal Perspective on AI

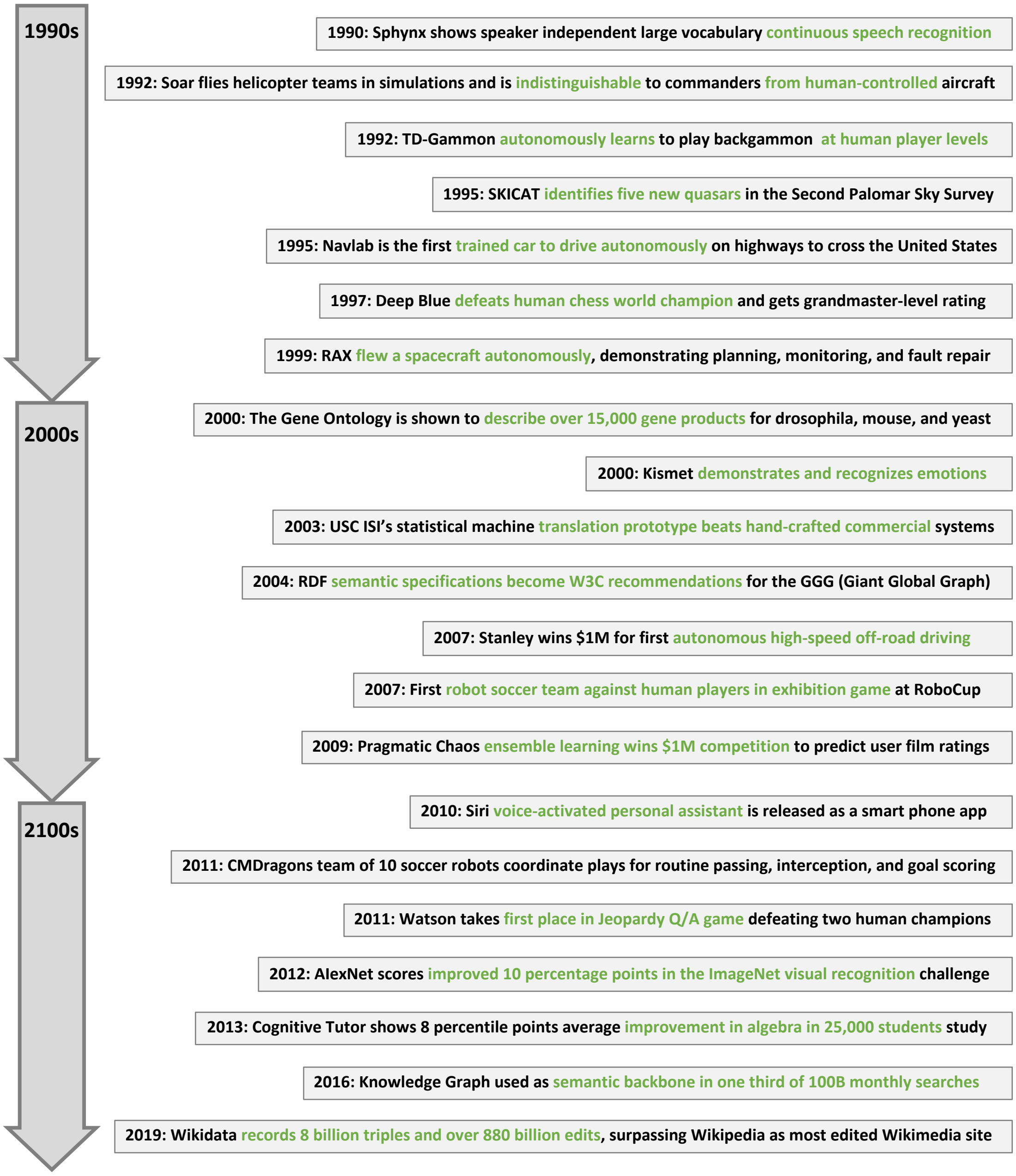
My first AAAI conference was in 1986. I had just landed in Pittsburgh from Spain, where I am originally from just two weeks earlier. I found AAAI to be a very inspiring community. In my view, AI researchers have been visionary, broad, inclusive, interdisciplinary, determined, and successful at challenging endeavors. It is a very vibrant field, and I think the AAAI conference is where I see that breadth, interdisciplinarity, and determination at its best. And I am very proud to be part of this AI community that has been tackling profound challenges, and I am often amazed by what it has accomplished as I look back over the years.

In the 80s, when I arrived at Carnegie Mellon University, there was a course on architectures for intelligence, where I learned so much about the breadth of ideas in approaching intelligence. Allen Newell was working with others on the SOAR architecture, adopting cognitive models of intelligence. Jaime Carbonell, who was my advisor, was doing research on engineering diverse intelligent capabilities for learning, reasoning, and meta-reasoning, all within the same Prodigy framework. Tom Mitchell was working on THEO, which was a sophisticated frame system to organize general knowledge and facts about the world. Geoff Hinton, who was at CMU at the time, was working on backpropagation as an alternative basis for intelligent architectures. In that course, we also learned about different architectures for intelligence that were being explored elsewhere. Roger Schank at Yale emphasized how intelligence is demonstrated in telling interesting stories that prompt others to respond with a good follow up question or a related and equally interesting story. Rod Brooks at MIT was investigating subsumption architectures, with basic capabilities controlling sensors and actuators, and then on those capabilities you build more elaborate ones that demonstrate higher intelligence. All this provided a very broad view on the approaches to study intelligence and on the field of AI. I hope that students today continue to seek these kinds of opportunities to appreciate the breadth of our field.

Over the decades, I have seen tremendous accomplishments from our community in all areas of AI. Figure 1 shows some of these highlights. In the early 90s, when I was finishing my Ph.D. I used the Sphynx voice recognition system to actually write my thesis. I thought it was incredible that it worked so well. After graduation, I moved to the Information Sciences Institute at the University of Southern California, where Paul Rosenbloom was collaborating with Allen Newell and John Laird in the SOAR architecture that I mentioned earlier, using it to fly helicopters in teams that a commander found to be indistinguishable from human pilots. I thought that was remarkable. Also during those years, the TD-Gammon system was learning to play backgammon with itself and performing at human levels. SKICAT, built at NASA's JPL, was identifying new quasars using data from Sky Surveys. CMU's Navlab was an autonomous vehicle with a neural network at its core that learned enough to drive itself across the US. Deep Blue was able to beat the World Chess Champion, a feat that Kasparov himself has said was eventually inevitable. In 1999, RAX flew in a spaceship and generated plans for the use of its instruments and becoming the first intelligent system to reach outer space. It was an amazing decade for AI.

In the 2000s, many more significant accomplishments would come from all areas of AI. I was excited to see ontologies become very popular, with the Gene Ontology starting to be used to describe thousands of gene products and becoming the core connector for most biology knowledge sources. Kismet demonstrated emotions embodied in an interactive robot. In 2003, a statistical machine translation system beat the performance of manually-coded commercial-grade systems that had been built over many years. RDF, a knowledge representation language that become a standard for the Giant Global Graph (the next generation of the World Wide Web), was the initial seed for today's widely-used knowledge graphs. Stanley won a million-dollar challenge for off-road autonomous driving. The first robot soccer exhibition game against humans took place, demonstrating teamwork in a highly dynamic environment. At the end of that decade, an ensemble learning system won a million dollars for predicting user movie ratings.

The past decade has also seen many exciting AI accomplishments, and I will just mention them briefly since they are more present in our minds. In 2010, Siri was released as a smart phone app and soon became an integral part in the lives of many mobile phone users. Robot soccer teams were now passing and intercepting across ten robots coordinating together. Watson won the popular Jeopardy question-answering contest against the best human players. Alexnet raised the bar tremendously on image recognition and contributed to the revival of neural networks that are flourishing today. Cognitive tutors were demonstrated to improve learning for thousands and thousands of students. Knowledge graphs were used to improve a third of the hundred billion searches conducted in a month. Wikidata recorded a billion triples, or logic assertions, and became the largest crowdsourced knowledge base with more edits than its older sister Wikipedia.



1990s

1990: Sphynx shows speaker independent large vocabulary continuous speech recognition

1992: Soar flies helicopter teams in simulations and is indistinguishable to commanders from human-controlled aircraft

1992: TD-Gammon autonomously learns to play backgammon at human player levels

1995: SKICAT identifies five new quasars in the Second Palomar Sky Survey

1995: Navlab is the first trained car to drive autonomously on highways to cross the United States

1997: Deep Blue defeats human chess world champion and gets grandmaster-level rating

1999: RAX flew a spacecraft autonomously, demonstrating planning, monitoring, and fault repair

2000s

2000: The Gene Ontology is shown to describe over 15,000 gene products for drosophila, mouse, and yeast

2000: Kismet demonstrates and recognizes emotions

2003: USC ISI's statistical machine translation prototype beats hand-crafted commercial systems

2004: RDF semantic specifications become W3C recommendations for the GGG (Giant Global Graph)

2007: Stanley wins $1M for first autonomous high-speed off-road driving

2007: First robot soccer team against human players in exhibition game at RoboCup

2009: Pragmatic Chaos ensemble learning wins $1M competition to predict user film ratings

2100s

2010: Siri voice-activated personal assistant is released as a smart phone app

2011: CMDragons team of 10 soccer robots coordinate plays for routine passing, interception, and goal scoring

2011: Watson takes first place in Jeopardy Q/A game defeating two human champions

2012: AlexNet scores improved 10 percentage points in the ImageNet visual recognition challenge

2013: Cognitive Tutor shows 8 percentile points average improvement in algebra in 25,000 students study

2016: Knowledge Graph used as semantic backbone in one third of 100B monthly searches

2019: Wikidata records 8 billion triples and over 880 billion edits, surpassing Wikipedia as most edited Wikimedia site

Figure 1. Highlights of significant AI accomplishments over the last few decades, spanning cognitive systems, machine learning, multi-agent systems, knowledge representation, search, planning, robotics, and natural language.

These are all incredible things that our community has accomplished, and there are many more that I have not covered here. But they suffice to illustrate that these accomplishments cut across the spectrum of all areas of AI: interaction, collaboration, search, constraint reasoning, sensing, perception, planning, robotics, knowledge systems, learning, and so on. All across the board we have been able to make tremendous progress.

This message of diversity and breadth is very important for the AI community: that we have accomplished things that the rest of the world may or may not have noticed but that we know place us in a solid footing to tackle future problems. All this progress gives us hope that AI is at the core of new approaches to tackle humanity's crucial challenges, from science to health to innovation to education to policy.

The Imperative for AI in Science

Accelerating scientific advances is an important grand challenge of our time. In the case of science in particular, I believe that it is not just useful to have AI. AI has really become an imperative for science.

There is a long history of AI in science. Back in the 1950s, Herb Simon already started to look at cognitive aspects of how scientific discovery occurs, and over the years he had many collaborators demonstrating and replicating scientific discoveries. Ed Feigenbaum, who was his student, went off to Stanford to collaborate with Joshua Lederberg and Bruce Buchanan on molecular biology and molecular discovery. There is a long tradition in AI in science. In many cases, the work has been naturally focused on machine learning. But there is also a lot of analysis on how scientists approach new discoveries in terms of mechanism design, causality, and paradigm shifts.

At the same time, it is important to recognize the human limitations that actually curb scientific progress [Gil 2017]. So when we write papers and when scientists look at the world, sometimes there are errors and biases, there is poor reporting in papers, and sometimes the work is not systematic or complete. I will mention, for example, work by Liz Bradley and colleagues on analyzing paleoclimate data where their AI system generated a range of hypotheses and some had not appeared in published papers. Scientists considered them to be valid hypotheses, but they chose not to mention them. And so important hypotheses are often left not discussed or explored. There are many cases of errors because scientists are human and make mistakes, and so we read about published papers being retracted after some scrutiny about the results. I use the example of a graduate student trying to reproduce the results of a paper by two influential economists, and finding out that they had made an error by omitting some of the data in their tables. Another important problem is that reproducing the results of a published paper is often very challenging, because authors do not include enough information that is crucial to understand how their methods actually worked. These are all important issues that illustrate how human limitations can curtail scientific progress and that I believe AI is in a great position to address.

At the same time, there is very exciting work that is more recent that I wanted to share with you on AI in science. The computational sustainability community is developing new approaches in multi-agent systems, constraint reasoning, machine learning, and optimization to address a range of environmental problems (e.g., [Gomes et al 2019]). There is also very important work on materials discovery, where from automated text extraction from published articles they are able to identify particular molecules, recreate the periodic table, and actually predict discoveries that have occurred in the past just from looking at the trends in the literature (e.g., [Tshitoyan et al 2019]). There is also increasing value for knowledge-based biomedical data science on how knowledge is used to make progress in life sciences (e.g., [Callahan et al 2020]). There are also discoveries through machine learning, one of note is on protein folding prediction which incorporated physical and geometrical constraints to outperform any other algorithms and even years of work invested by some labs dedicated to specific proteins. (e.g., [Senior et al 2020]). It is very exciting to see AI research leading to these advances. And I want to point out that these advances come from a diversity of areas in AI. Given the formidable science challenges that our generation faces — understanding the brain, preserving our planet, deciphering the origins of the universe — I believe we need the diversity and breadth of research areas in AI to make strides on these frontiers.

Capturing Scientific Knowledge

My recent research has focused on capturing scientific knowledge and the new AI approaches that can use this knowledge to provide novel capabilities. I will show our work on how we use scientific knowledge in two major ways: to do systematic data analysis and to enable interdisciplinary research.

When you think of scientific knowledge, you think of physics, mathematics, biological processes, and so on. We take a step back and focus instead on the compositionality and modularity of scientific artifacts that already capture that knowledge. We take models in science as something that we can use as building blocks. For example, a hydrology model can use sophisticated physics equations to represent where rain water goes in a complex ecosystem. We think that AI systems do not necessarily need to understand exactly how a model works, but instead it needs knowledge about how to use that model to make estimates and predictions. So we think of models and other scientific artifacts as modular computational objects — that is a very important concept for us. We focus on the knowledge needed to use them as modular objects: we want to represent about the input data their require, the physical variables that they model, the constraints for their use, when their assumptions are appropriate for a particular problem, the parameters that we would like to adjust, and the interventions or changes that we want to do on a situation in order to improve outcomes.

We have worked on a range of projects over the years that touch on different aspects of capturing or representing knowledge. I will discuss here a few of them.

The first one I will describe is on crowdsourcing vocabularies to describe scientific data and create metadata annotation standards [Gil et al 2017b]. We collaborated with Julien Emile-Geay and Deborah Khider of the University of Southern California, and with Nick McKay of Northern Arizona University. This is very interesting work, because in a lot of disciplines it takes a lot of effort to agree to standard vocabularies that can be used to describe scientific data. This is the case with paleoclimate, where scientists study past climates in the last few hundred or thousand years by studying very diverse data. They drill cores in different locations on Earth, just to see what happened many years ago through what is buried in the ground. Some scientists drill in the ice, and study the size of the trapped air bubbles. Others look at marine cores, and study the remains of coral that appear and how it grew based on the climate at that time. Others drill on lakes to study sediments. So it is very challenging to come up with a standard way to represent all this data. And as a result, it takes them years to analyze all this diverse data in a consistent way to create a model of past climate at planetary scale.

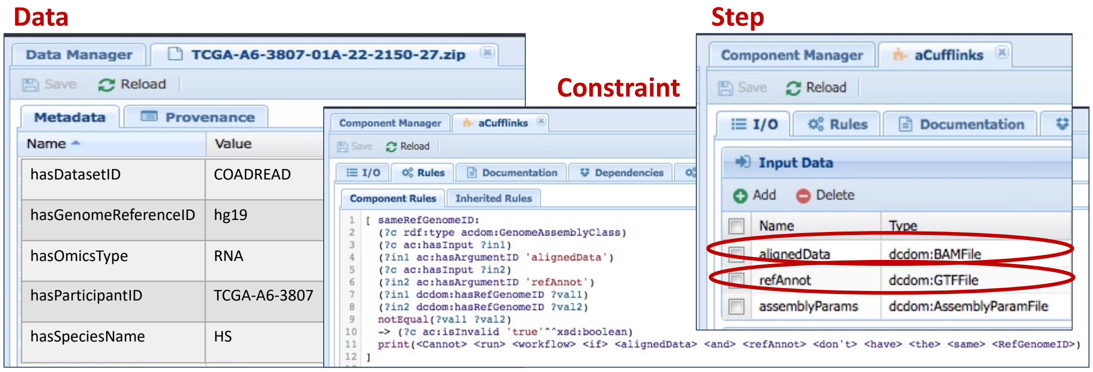
We developed a new approach that we call *controlled crowdsourcing*. Scientists describe the datasets that they are using as they do their work, and they are each asked to propose terms that they would like use for their own data. They can choose to adopt the terms proposed by others, and pretty soon you have convergence at least for some of the terms. And we place an editorial process on top, very much following the footsteps of Wikipedia, but focused on deciding on terms that are worthy of more general adoption by the community and eventually turning them into ontologies. So users will describe their data as they go, continuously adding what we call *crowd properties* to extend a very solid set of core ontologies that we either reused or created.

We had to address the challenges of living with an evolving ontology. We would start with an ontology, scientists would start annotating the data and proposing new properties, and eventually we had a new version of the ontology. But we could not just adopt the new ontology, because we had many datasets already annotated using the prior version. We were able to address this with AI techniques from ontology development, and principles for non-monotonic and monotonic reasoning.

The resulting standard for describing paleoclimate data emerged over a period of two years, and there was great convergence on how to describe the different datasets as well as a few basic terms that were adopted by all [Khider et al 2019]. And I should mention it was accomplished with zero face to face meetings, just a single meeting was held at the very beginning to agree to the overall process. Thanks to AI we were able to enable this standard.

It is important to realize that this is, in effect, an approach that uses AI techniques to synthesize new scientific knowledge in the form of those vocabulary standards which did not exist before.

Another project that I would like to discuss is capturing knowledge about data analysis processes as *semantic workflows* in our WINGS workflow framework [Gil et al 2011]. Workflows represent multi-step computational methods in science that are repeated and are re-used often. We view them as plans and study them with AI techniques to reason about goals and effects, execution monitoring and replanning, failure detection and recovery, and abstractions. This provides a very powerful



Name	Value
hasDatasetID	COADREAD
hasGenomeReferenceID	hg19
hasOmicsType	RNA
hasParticipantID	TCGA-A6-3807
hasSpeciesName	HS

```
[ sameRefGenomeID:
  (?c rdf:type acdom:GenomeAssemblyClass)
  (?c ac:hasInput ?in1)
  (?in1 ac:hasArgumentID 'alignedData')
  (?c ac:hasInput ?in2)
  (?in2 ac:hasArgumentID 'refAnnot')
  (?in1 dcdom:hasRefGenomeID ?val1)
  (?in2 dcdom:hasRefGenomeID ?val2)
  notEqual(?val1 ?val2)
  -> (?c ac:isInvalid 'true'^^xsd:boolean)
  print(<Cannot> <run> <workflow> <if> <alignedData> <and> <refAnnot> <don't> <have> <the> <same> <RefGenomeID>)
]
```

Name	Type
alignedData	dcdom:BAMFile
refAnnot	dcdom:GTFFile
assemblyParams	dcdom:AssemblyParamFile

Figure 2. Semantic workflows in WINGS capture constraints about science methods and datasets, enabling the automatic elaboration of high-level methods to customize them to the characteristics of the given data.

framework for many science processes and data analysis methods described in the methods sections of scientific papers.

We do not just treat workflows as computations, instead we treat them as objects of science that have meaning and purpose and we attach to them semantic annotations and constraints. Every constituent of the workflow, whether data or computational step, has an identifier and we can make assertions and express constraints about them. We can assert that a certain type of input data has a property, and that having that property makes it compatible with some analysis step being done downstream in the workflow. And we can attach to a workflow a lot of such constraints. These constraints allow us to reason about how to set parameters of specific method steps so they are customized to the data, to reason about generating metadata for the workflow outputs, to reason about how to choose an implementation among many available for a given workflow step, and how to validate that the overall workflow is appropriate for the data at hand.

Figure 2 shows an example workflow constraint. The left side shows the metadata that we have for any data set, and the right shows the step using that type of data and other data, and then a constraint that indicates that if two different steps are used together, in this case alignment and assembly, they have to use the same reference genome.

Semantic workflows also use abstraction, so a given step could be executed with different algorithms and implementations. We can very easily reason about abstract methods in science versus specific implementations. Through AI, we are able to do very powerful reasoning for composition and exploration of these workflows. We can take a very simple high-level workflow and elaborate it to add many sub-steps. It is basically what some of us in AI would recognize as skeletal planning. Starting with a skeletal plan with high-level steps, we can specialize each step to the current input data based on the constraints that we have about each option for implementing the step.

We capture many scientific methods as semantic workflows. For example, we created a library of workflows for population genomics, and we were able to reproduce papers where we could get a hold of the original data. We got the same significant results that the papers did, just by reusing workflows from that library. Our workflows used open source software, while some of the papers used proprietary software. Our workflows could use more modern algorithms, while some of these papers used very old algorithms that were known at the time. But we were able to get the same results so we are using very powerful AI techniques to capture sophisticated scientific methods.

Semantic workflows also enable us to use machine learning to detect common workflow fragments that scientists use with different data, and we demonstrated this with a large collection of neuroimaging workflows [Garijo et al 2014]. This is work with Daniel Garijo and Oscar Corcho of the Polytechnic University of Madrid. We were able to access hundreds of workflows created manually by scientists and found common workflow fragments that scientists use about how the warping of brain images is done, and general ways in which they approach neuroimaging analysis. To do this, we extended process mining techniques to exploit the semantic annotations in the workflows, essentially treating them like labeled graphs where we could then map steps across them and create generalizations of any specific workflow fragment. We are using AI to synthesize a new form of scientific knowledge as commonly used abstract workflow fragments that had not been detected before and can now be reused for future neuroimaging analyses.

The last project that I will mention in terms of capturing scientific knowledge is recording provenance. Provenance represents how scientific workflows, as any plans do, once executed leave a trail of the steps that were carried out and the results that they generate. If you have a new dataset that results from an analysis, its provenance record has a very similar underlying structure to the provenance of a dataset collected through a sensor. You can also see that all of the steps that were followed to collect or to prepare data has a lot of similarities to the way that humans actually put together any other digital resource. It is also similar to the way that we describe how a piece of art is generated by a painter or an artist, then years later may go to a curator that cleans it up, and then eventually appear for sale at a gallery. So there is a provenance trail for datasets just like there is for pieces of art. A provenance record refers to agents, plans, objects, actions, successful and failed executions, and many other abstractions that we have been studying in AI for decades and designing sophisticated representations and abstractions for them. We worked closely with dozens of people representing diverse disciplines to create a general representation and ontology for provenance, which became a World Wide Web Consortium standard in 2013, and we are very proud that it has been widely adopted [Moreau et al 2013].

Let us look at scientific papers for a moment and the question that I posed earlier: Will AI write scientific papers in the future? When we look at scientific papers, we realize that they can be written better. I would like to convince all of us scientists to write better papers, because if we do that then the AI systems for machine reading and text extraction will work better. If we do that, AI systems (and other scientists) can actually be in a better position to understand our papers and reproduce the results.

What would a scientific paper of the future look like? How would we describe best the science findings and their provenance? We started with a group of visionary early career researchers to define the geoscience paper of the future [Gil et al 2016], other groups defined the geophysics paper of the future [Broggini et al 2017] and the neuroscience paper of the future [Poldrack et al 2017]. More recently, with Odd Erik Gundersen of Norwegian University of Science and Technology and David Aha of the Naval Research Laboratory, we looked at what the AI paper of the future would look like and proposed core principles for reproducibility of AI publications [Gundersen et al 2018]. I believe it is very important that as a field we formalize more how to capture knowledge properly in our publications. For the benefit of other researchers, and for the benefit of AI systems that will read them and eventually build on our work and extend our ideas.

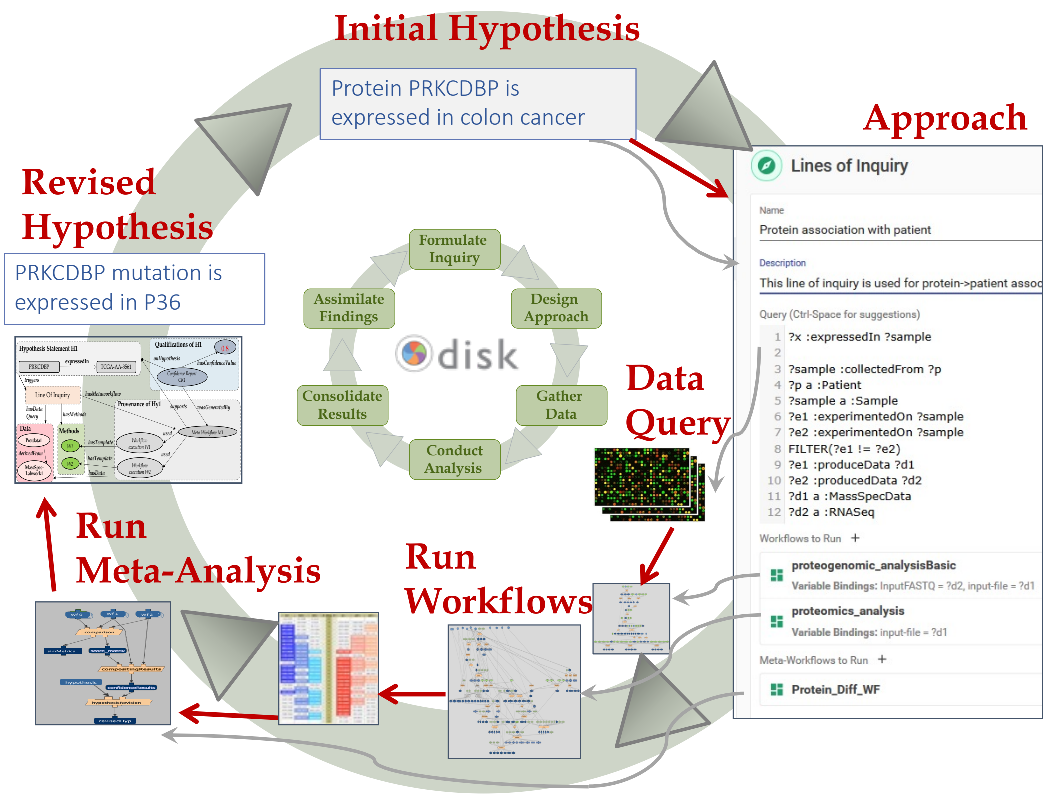
That is kind of a quick tour over our work on capturing scientific knowledge, and I want to remark that a lot of what we capture does not really exist beforehand. I described our use of AI to synthesize new forms of scientific knowledge as metadata standards for paleoclimate. I described how we represent scientific methods as semantic workflows, using AI techniques to synthesize new forms of scientific knowledge as reusable workflow fragments. I mentioned how we use general concepts from AI to create generalized representations of provenance for very diverse types of results, so our publications really capture the provenance of new findings which AI systems can read and eventually extend.

Once we capture all this knowledge, how do we use it for science? I will describe two key aspects of how we use this knowledge: systematic analyses and interdisciplinary research.

AI for Systematic Scientific Data Analysis

Our first major use of the scientific knowledge we capture is to do systematic scientific data analysis.

We are developing a framework, called DISK, to make hypothesis testing and data analysis more systematic [Gil et al 2017a]. We look at the discovery cycle, starting with formulating new hypotheses, figuring out what type of data and method can be used to test it (we call this a *line of inquiry*), retrieving the data from a shared repository, analyzing the data, and then revising the hypothesis. We have automated this hypothesis-driven discovery cycle in our DISK project, a collaboration with Parag Mallick and his group at Stanford. We work in particular in cancer omics, and there is extensive data in shared repositories that is continuously growing so we do not need to collect data ourselves.



Lines of Inquiry

Name
Protein association with patient

Description
This line of inquiry is used for protein->patient assoc

Query (Ctrl-Space for suggestions)

```
?x :expressedIn ?sample

?sample :collectedFrom ?p
?p a :Patient
?sample a :Sample
?e1 :experimentedOn ?sample
?e2 :experimentedOn ?sample
FILTER(?e1 != ?e2)
?e1 :produceData ?d1
?e2 :producedData ?d2
?d1 a :MassSpecData
?d2 a :RNASeq
```

Workflows to Run +

proteogenomic_analysisBasic
Variable Bindings: InputFASTQ = ?d2, input-file = ?d1

proteomics_analysis
Variable Bindings: input-file = ?d1

Meta-Workflows to Run +

Protein_Diff_WF

Figure 3. Overview of DISK, an autonomous system for hypothesis-driven discovery that relies on lines of inquiry specified by scientists about the datasets and methods they seek when they pursue specific kinds of hypotheses or questions.

Figure 3 illustrates how our approach works. A scientist has a hypothesis such as whether a protein is expressed in a certain type of cancer. The line of inquiry indicates that both proteomic and genomic data would be useful for this kind of hypothesis, so DISK executes a query to TCGA (The Cancer Genome Atlas) and to the CPTAC (Clinical Proteomics Tumor Analysis Consortium) data repositories to get data from patients who have that type of cancer. The line of inquiry then indicates that a proteomics workflow should be run with the proteomics data, a genomics workflow with the genomics data, and a proteogenomic workflow with both types of data together. Then the line of inquiry suggests a meta-workflow to combine the results. After executing the query, the workflows, and the meta-workflow, DISK notifies the scientist that it found evidence that the gene was expressed – but suggested a revised hypothesis that the gene that was expressed was a mutant form of the original hypothesized protein.

These workflows actually integrate knowledge across disciplines. For example, we use workflows that combine transcriptomics, genomics, proteomics. Typically this knowledge is spread across different labs and often different institutions. In addition, there is some analysis needed make sure that the combination of all of these methods is coherent. This type of knowledge to combine and validate analytic processes is very difficult to obtain and it takes a lot of effort. But once we have captured all this knowledge, it can be reused very broadly just by re-running these workflows. A great challenge for science today is the compartmentalization of all that knowledge: scientists that collect patient data from genomics and proteomics instruments do not necessarily have expertise in both. A lot of data just sits there waiting to be studied for lack of in-house expertise or collaborators. This presents a tremendous opportunity for AI systems to analyze tirelessly and thoroughly to extract as many findings as possible.

DISK represents a hypothesis statement as a triple that expresses a relationship between two objects, for example that a protein (first object) is associated with (the relationship) a cancer type (second object). Each hypothesis statement has a qualifier that indicates the confidence level on the hypothesis, and an evidence trail of the analysis details (that is, the executed workflow) that led to that confidence level. When more data becomes available, it will run a new analysis and revise both the confidence level and the hypothesis. In this case, it will re-examine not just the association of the protein with colon cancer, but it will look for a more specialized and refined hypothesis regarding a subtype of colon cancer. DISK has a hypothesis ontology to express hypothesis statements, qualifiers of confidence, analytical evidence, and the evolution of the statement and its confidence over time.

The way we do science today is that a paper is published with a certain finding and is considered final. But as more data becomes available, one wonders if that finding may be different than what was originally discovered, for example if the confidence is greater or the evidence more diverse or the hypothesis more precise. We envision AI systems like DISK that will continuously update findings, which would revolutionize our current approach to publications.

We used DISK to reproduce a seminal cancer omics paper by Bing Zhang and colleagues as part of the CPTAC collaboration [Zhang et al 2014]. It took two years to do this large-scale proteogenomic study. The paper's major result was the discovery of a new, fifth subtype of colon cancer based on patient proteomes. All the data is available, and the paper has many figures that show intermediate results and a long appendix with many details. It took us quite a long time to really understand what was done and reproduce the results.

Using DISK, we were able to reproduce the results and find the same new fifth subtype of colon cancer. But we wanted to quantify how close our results were to the original findings – not just at the highest level, but throughout the extensive hierarchy of sub-hypotheses that support the highest level, major hypothesis. It is often the case with complex hypotheses that they are based upon hundreds to thousands of lower-level hypotheses.

What we found is that even though DISK was able to independently discover a fifth subtype and confirm the main finding, it got very different results for the sub-hypotheses. At the lowest level of the hypothesis tree are the peptides (aminoacid chains that combine to form proteins) present in each patient sample. These then feed up to hypotheses about which proteins are present in the sample. DISK got very different results for the peptides and proteins that were present in each patient's tumors. It found a 10% difference, which is quite large. After discussion with the authors, we discovered they had used multiple thresholds and complex filtering in their analysis. In particular, they used a stringent filter at one step, and a less stringent filter in a subsequent step. None of that was mentioned in the paper. Incorporating these filters cut the difference in the results in half, which is still quite a large difference but it is more acceptable. But this points to how sensitive lower-level hypotheses may be to subtle variations in the methods. The overall finding still stands, but it is quite alarming that the detailed analytic results are so different.

We then used DISK to try a few alternative methods, for example different approaches to do peptide search. This is very easy to do using semantic workflows, since DISK can easily explore all the methods that fit a step and find the best performing ones. We found that there was a 35% difference in the protein identifications, and that key proteins were missed by some of the methods. This is alarming, because proteogenomics publications typically use only one search method. So many proteins are likely to be missed unless we use an AI system like DISK to do a more systematic exploration of alternative methods.

DISK also showed a marked difference when using the same method with different parameters, or different reference databases. The protein and peptide identifications also varied substantially. We also looked at other measures, like single aminoacid variants and variant peptides, and found the differences to be quite substantial. So there is a great need to use AI to systematically explore all these method parameters, and to understand their differences.

I mentioned earlier that another significant opportunity for AI is to continually analyze data and update findings as new data becomes available. We took the dataset from the original [Zhang et al 2014] paper, and updated the analysis with the new samples that have appeared in the data repositories since the publication of that original paper. When the first few new samples are added, they fall nicely in the five subtypes. The next few samples do as well, as do the next ones. When we added the last few samples, the clustering algorithm still generated five subtypes but it grouped the samples a bit differently so it is unclear that they are the same five subtypes as before. This makes the analysis more challenging, and one of the things we had to do to revise hypotheses over time is to develop new workflows with batch effect correction for the new samples.

Finally, using AI to do systematic analyses makes it easy to apply methods to new datasets. We applied the same workflows for datasets from colorectal cancer (CRC95), NCI60 (9 types of cancer, multiple tissues), ovarian cancer (OV), and breast cancer (BRCA). DISK found, for example, that colorectal cancer has a lot less variance than other types of cancer. This is well known, but we were able to find that very easily very quickly.

We think DISK can be used as a reference benchmarking framework, and we demonstrated this for the National Cancer Institute CPTAC DREAM Proteogenomics Challenge [Srivastava et al 2019]. For that challenge, the same datasets were analyzed by dozens of teams to identify proteins. It is very hard to pinpoint why a specific team did better, because their methods are quite similar and only differ in very small ways. The teams follow a typical analysis structure, where they first align the gene sequences, then the quantify and normalize the data, and then they predict the protein levels. We created generic semantic workflows that DISK could reason about to specialize them into the kinds of methods adopted by individual teams. With this framework, we can analyze the solutions proposed by different teams, and compare and contrast what specific differences made a method significantly better. What we found is that gene-specific models give significantly better results, while the other steps are only responsible for minor improvements. And by exploring all the possibilities systematically, DISK found the optimal combination of methods.

All this work illustrates how AI can make a different in making scientific research more systematic. Every year omics datasets are increasing in size, diversity, and complexity. The expertise is very fragmented. And there is more and more analytic complexity. At some point sooner or later, biomedical researchers will not be able to keep up. Imagine if AI systems like DISK were used to analyze proteogenomic data systematically in every cancer study. This would accelerate discoveries for new problems and easily update earlier findings when new data is available. AI will become a game changer for omics research, and many other areas of science where there is abundant data available.

AI for Interdisciplinary Science Frontiers

The second way in which we use scientific knowledge is in exploring interdisciplinary science frontiers.

We have a project called MINT, for Model INTegration, where we are looking at how climate is affecting water availability, agriculture and food production, socioeconomic factors, with extremes of flooding and drought cutting across all those aspects [Gil et al 2021]. To study these types of problems, we have to integrate climate models, models for hydrology, agriculture models, economics models, and social models. Each of these types of models is developed by different disciplines each with their own methods and approaches. For example, hydrology models use physics equations to simulate where rain water goes in an ecosystem. In contrast, social models often rely on agent-based simulations of social behaviors. It takes months or years to assemble integrated models for a certain region, since it is really a craft. There is increasing demand to understand these cross-domain regional phenomena. We have a growing community of collaborators in Ethiopia, Texas, and California.

Our goal is to reduce the time to assemble cross-domain models from months or years down to days. To do this, we are using AI to do the mediation between models

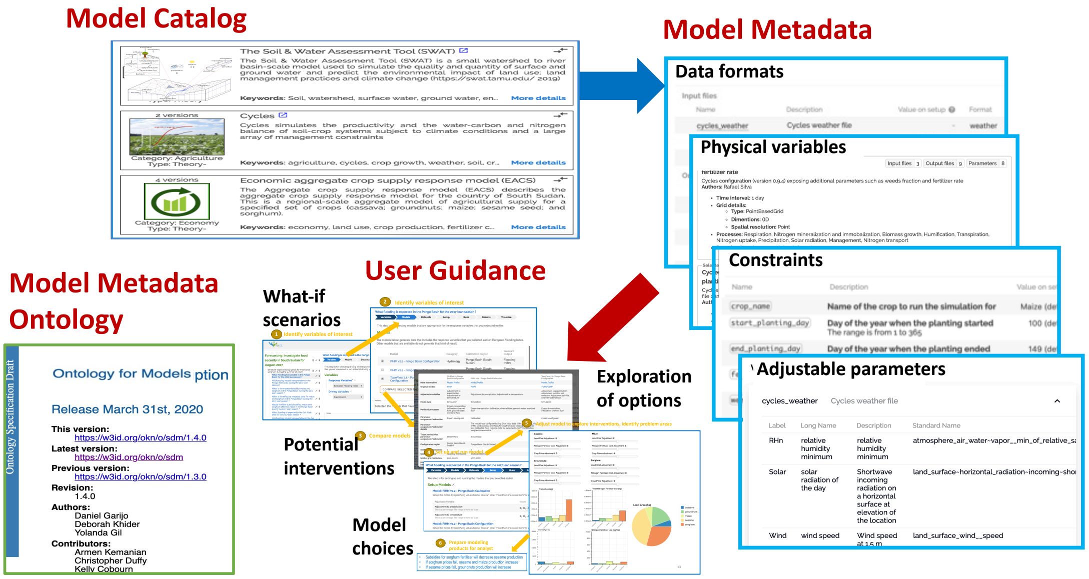


Figure 4. MINT captures metadata and constraints about models and datasets and uses them to guide users to develop and integrate models from different disciplines more efficiently.

at all levels: at the variable level, at the dataset level, at the process level, and so on. At the variable level, we use an ontology conceived by Scott Peckham at the University of Colorado that is very extensive and has thousands of terms. It has not been built manually, instead it was designed using principles for how to describe physical variables in a consistent, uniform way. With this ontology, we can be very precise about what a particular variable in a model is measuring. We think of a model as a computational object, where we need to represent the data formats, the physical variables, the constraints on how to use the model, and so on.

We also use AI to mediate models and datasets from different sources that have varying quality, and this is also a big challenge. We represent what those datasets contain, what variables they characterize, and the structure of those files. Then, we use AI planning techniques to automatically transform the datasets from their native format to the format that a model requires. Craig Knoblock and Jay Pujara at USC/ISI have been leading this research. There has been a lot of work in this particular topic over the years, and we are building on that to work with much more complex data.

In some regions, there is simply not enough data available for modeling. This is the case with regions like Ethiopia, where there are not enough river gauges. We do not have enough measurements of water levels over the last 30 years to calibrate models. So in the case in Ethiopia, if we model the Awash or Baro rivers, the models are incredibly coarse and the flooding indicators appear very pixelated. In contrast, in work that we are doing with Suzanne Pierce of the University of Texas at Austin we can show the extent of the flooding building by building in Travis County. And the models actually generate very accurate predictions of the flooding that correspond well to manual flood maps. In order to improve our models for Ethiopia, we use machine learning to extract from satellite data useful information about water level in rivers, the degree of flooding, what kinds of crops were typically planted over the years and so on. This is work by Vipin Kumar and his group at the University of Minnesota. There are many groups exploring how to combine physics-based models with machine learning models in a variety of ways, and this is a very promising direction to use AI to improve the accuracy of models in geosciences.

As far as the models themselves, we also use AI to represent abstractions and connect the models to decisions. For this I am going to use examples from an agriculture model, developed by Armen Kemanian at Penn State University. These kinds of models have a lot of detailed parameters, such as nitrogen stress or solar radiation levels. But those are things that somebody making decisions about what crops to plant does not really need to know. So we focus on the variables that a decision maker would want to adjust. How much corn should be planted? How much sorghum should be planted? How much irrigation should be used? Those are the kinds of factors that a decision maker would actually want to change, and then see how different weather predictions affect the outcome of interest, which is the crop yield.

This allows us to use a model in a very compositional way. It is still very challenging, but we have very rich descriptions of what each model can be used for, as illustrated in Figure 4. This is work with Daniel Garijo and Deborah Khider at USC/ISI. Each model has many different variants as we adapt it for different regions, and for each variant we capture the data formats, the physical variables, the use constraints, and the adjustable parameters. We have developed ontologies for describing scientific software, and extended them to describe models. These descriptions of the models can be used to drive their integration. We can see the areas that will be flooded (from a hydrology model) and how that affects other things: the crops that will be lost if not harvested earlier (from an agriculture model), the roads that will not be trafficable to bring the food where it is needed (from a transportation model), and the towns that people will flee to go to dry grounds.

This work illustrates how AI can integrate diverse knowledge from different disciplines and integrate it to create unified models to understand complex phenomena. AI provides a crucial capability for many problems where we need to characterize complex systems where natural, human, and engineered processes interact, such as environmental and geosciences research.

A Perspective on the Future

I have described how we capture scientific knowledge, and illustrated how we use all of this knowledge to do systematic data analysis and to do interdisciplinary science. Now I would like to take a step back, and revisit the question that I posed at the beginning: "Will AI write scientific papers in the future?" Because if we had AI systems with enough knowledge to follow the methods of science, we have to wonder if they could generate new findings on their own and report those findings by writing their own papers. And of course that would liberate scientists from doing more routine work and spend more time coming up with big ideas.

Scientific research is becoming more and more complex. Many significant discoveries are the product of collaborations of hundreds or thousands of scientists. The Human Genome Project was an ambitious endeavor that involved hundreds of scientists. The discovery of the Higgs boson in 2012 was reported in an article with more than four thousand authors [Aad et al 2021]. The science questions that we want to ask are more and more complex, and these kinds of discoveries take several years and involve many people. These are very unique kinds of efforts in science today, but I wonder what would it take for AI to help make scientific advances of this caliber more commonplace.

Since we have Gary Kasparov as a speaker at the conference, which I am really thrilled about, as well as the team that played him with Deep Blue, I wanted to share an interesting observation he made about freestyle chess. Here, a player can be a combination of any number of humans and any number and type of machines. He remarked that they were surprised that a weak human plus a machine plus a better process was superior to a strong computer alone, and superior to a strong human and a machine and a less ideal process. This highlights the importance of organizing and distributing work appropriately, that defining a good process is crucial when combining the skills of humans and machines together.

What would it take for machines to collaborate with scientists in similarly powerful teams? In the case of chess, an AI system does not need any special skills to become a team member since the only requirement is to communicate to others the recommended next move. But for an AI system to be a collaborator in science I believe that it needs to have very sophisticated skills. I recently proposed a set of principles for the design of *thoughtful AI systems* that will be good partners for a scientist [Gil 2017]. Today, a researcher gives an AI system some data, an algorithm, and the function to optimize. There is a lot of research that we need to do so an AI system can be a partner. Table 1 summarizes the principles that I have proposed for thoughtful AI systems, namely that they govern their behavior with knowledge structures, understand the context of their tasks, seek knowledge when needed, access new knowledge by asking others or through the web, justify and engage in argumentation to defend their beliefs, integrate modularly with other systems or humans, and finally, understand how to convey their limitations.

Will AI write scientific papers in the future? I am certainly working on many projects in that direction, and many others are working in relevant areas. And there is a lot of good work on re-imagining the future of science and research. We have worked with the geosciences community to envision a future with AI [Gil et al 2019], and I will highlight also Hiroaki Kitano's recent article in *AI Magazine* on AI to win the Nobel Prize [Kitano 2016]. I think that is quite an ambitious goal, but also very exciting. There are a lot of great possible outcomes along the way: AI to reproduce published articles, AI as a research assistant, AI as a partner and co-author.

Table 1: Principles for Thoughtful AI Systems that will partner with scientists to accelerate discoveries.

Principles for Thoughtful AI Systems	
Rationality principle	Behavior is governed by knowledge
Context principle	Seek to understand the purpose and significance of tasks
Initiative principle	Proactively new learn knowledge relevant to a task
Network principle	Access external sources of knowledge and capabilities
Articulation principle	Respond with persuasive justifications and arguments
Systems principle	Facilitate integration and compositionality with other systems
Ethical principle	Behavior that conveys limitations, uncertainty, and unknowns

So will AI write scientific papers in the future? I will go back to the observation that I did at the very beginning of this talk: the AI community has always been visionary, broad, inclusive, interdisciplinary, determined, and successful at challenging endeavors. That is all standing in one side of the equation. On the other side, I also showed that human scientists are not systematic, they make errors, they have biases, and they do poor reporting. So this might tilt the balance for AI, that maybe AI will be a much better approach and will end up writing of all of our papers. But at the same time, I reflect on the wonders of human ingenuity. Think that penicillin came out of a human error, where Alexander Fleming accidentally left the lid open in a culture plate and it ended up contaminated with Penicillium mold that killed the bacteria.

In the end, if we find imaginative ways to combine humans and machines I believe we will see really remarkable outcomes in science. My hope is that this perspective will inspire us all to pursue significant research goals, that if we read an article or work on a problem where we think to ourselves "I could write an AI for that" then we will be less excited to pursue that line of work. But if we read an article and have a sense that only a human could come up with that kind of idea, I am willing to bet that in many cases we will hear things like "Well, I just was stubborn, "I had this firm belief in this possibility", or even "I made a mistake and that led me to see things differently." So I do not know what the future will look like as scientists partner with AI to do significant discoveries, but I am very excited to design that future together with all of you.

In this talk, I have focused on science but the kinds of AI research that I discussed are universally applicable to other important areas where AI could be a game changer: wellbeing and quality of life, education, social justice, innovation, and many others. You can find very compelling use cases and interesting research directions in a recent community roadmap for the next 20 years of AI research [Gil and Selman 2019]. Please continue to push for the breadth and diversity of AI. All of your ideas, every single one, are important for a future where AI changes the way that humanity approaches the most formidable challenges.

Acknowledgements

I would like to thank my collaborators over the years, especially (and in no particular order) Varun Ratnakar, Daniel Garijo, Deborah Khider, Maximiliano Osorio, Hernan Vargas, Jihie Kim, Paul Groth, Ewa Deelman, Karan Vahi; Rafael Ferreira, Rajiv Mayani, Craig Knoblock, Jay Pujara, Kevin Knight, Chris Mattmann, Paul Ramirez, Dan Crichton, Ralph Bergmann, William Cheung, Oscar Corcho, Pedro Gonzalez, Ricky Sethi, Carole Goble, Natalia Villanueva, Deanna Pennington, Julien Emile-Geay, Nick McKay, Felix Michel, Matheus Hauder, Paul Hanson, Jordan Read, Scott Peckham, Maria Stoica, Chris Duffy, Armen Kemanian, Vipin Kumar, Neda Jahanshad, Phil Bourne, Chris Mason, Joel Saltz, Tahsin Kurk, Jill Mesirov, Michael Reich, Shannon McWeeney, Christina Zhang, Parag Mallick, Ravali Adusumilli, Hunter Boyce, Tom Harmon, Sandra Villamizar, Suzanne Pierce, Cédric David, Ibrahim Demir, Jonathan Goodall, Bakinam Essawy, Robinson Fulweiler, Allen Pope, Mimi Tzeng, Odd Erik Gundersen, and so many others. I would also like to thank my thesis advisor Jaime Carbonell for his support from the beginning of my career, and Tom Mitchell, Nils Nilsson, and Herb Simon for being such an inspiration as I was learning to do research. My thanks also go to Kevin Knight for coming all the way to hear me talk for the millionth time, and to Victoria Knight for patiently transcribing the original talk. I am very grateful for the sponsorship of my work over the years from the National Science Foundation, the Defense Advanced Research Projects Agency, the National Institutes of Health, the Intelligence Advanced Research Projects Activity, the National Aeronautics and Space Administration, the Air Force Office of Scientific Research, the Office of Naval Research, Elsevier, the Kavli Foundation, the Bill and Melinda Gates Foundation, and J.P. Morgan Chase.

References

[Aad et al 2012] "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. " Georges Aad et al. Physics Letters B, 716(1), 2012. https://doi.org/10.1016/j.physletb.2012.08.020

[Broggini et al 2017] "Reproducible research: Geophysics papers of the future — Introduction." Filippo Broggini, Joseph Dellinger, Sergey Fomel, and Yang Liu. Geophysics, 82(6), 2017. https://doi.org/10.1190/geo2017-0918-spseintro.1

[Callahan et al 2020] "Knowledge-Based Biomedical Data Science." Tiffany J. Callahan, Ignacio J. Tripodi, Harrison Pielke-Lombardo, and Lawrence E.

Hunter. Annual Review of Biomedical Data Science, 3(1), 2020. https://doi.org/10.1146/annurev-biodatasci-010820-091627

[Garijo et al 2014] "FragFlow: Automated Fragment Detection in Scientific Workflows." Daniel Garijo, Oscar Corcho, Yolanda Gil, Boris A. Gutman, Ivo D. Dinov, Paul M. Thompson, and Arthur W. Toga. Proceedings of the IEEE Conference on e-Science, Guarujua, Brazil, 2014. https://doi.org/10.1109/eScience.2014.32

[Gil et al 2011] "A Semantic Framework for Automatic Generation of Computational Workflows Using Distributed Data and Component Catalogs." Yolanda Gil, Pedro Antonio Gonzalez-Calero, Jihie Kim, Joshua Moody, and Varun Ratnakar. Journal of Experimental and Theoretical Artificial Intelligence, 23(4), 2011. https://doi.org/10.1080/0952813X.2010.490962

[Gil et al 2016] "Towards the Geoscience Paper of the Future: Best Practices for Documenting and Sharing Research from Data to Software to Provenance." Yolanda Gil, Cédric H. David, Ibrahim Demir, Bakinam T. Essawy, Robinson W. Fulweiler, Jonathan L. Goodall, Leif Karlstrom, Huikyo Lee, Heath J. Mills, Ji-Hyun Oh, Suzanne A Pierce, Allen Pope, Mimi W. Tzeng, Sandra R. Villamizar, and Xuan Yu. Earth and Space Science, 3(10), 2016. http://dx.doi.org/10.1002/2015EA000136

[Gil 2017] "Thoughtful Artificial Intelligence: Forging A New Partnership for Data Science and Scientific Discovery." Gil, Yolanda. Data Science, 1(1), 2017. http://doi.org/10.3233/DS-170011

[Gil et al 2017a] "Towards Continuous Scientific Data Analysis and Hypothesis Evolution." Yolanda Gil, Daniel Garijo, Varun Ratnakar, Rajiv Mayani, Ravali Adusumilli, Hunter Boyce, Arunima Srivastava, and Parag Mallick. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, 2017. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/11157

[Gil et al 2017b] "A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations." Yolanda Gil, Daniel Garijo, Varun Ratnakar, Deborah Khider, Julien Emile-Geay, Nicholas McKay. Proceedings of the Sixteenth International Semantic Web Conference (ISWC), Vienna, Austria, 2017. http://doi.org/10.1007/978-3-319-68204-4_24

[Gil and Selman 2019]. "A 20-Year Community Roadmap for Artificial Intelligence Research in the US." Yolanda Gil and Bart Selman (Editors). Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI) report, August 2019. Published in arXiv, https://arxiv.org/abs/1908.02624v1

[Gil et al 2019] "Intelligent Systems for Geosciences: An Essential Research Agenda". Yolanda Gil, Suzanne A Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron

Ridley, Hanan Samet, Shashi Shekhar, Katie Skinner, Padhraic Smyth, Basil Tikoff, Lynn Yarmey, and Jia Zhang. Communications of the ACM, 62(1), 2019. http://doi.org/10.1145/3192335

[Gil et al 2021] "Artificial Intelligence for Modeling Complex Systems: Taming Expert Models to Improve Decision Making." Yolanda Gil, Daniel Garijo, Deborah Khider, Craig A. Knoblock, Varun Ratnakar, Maximiliano Osorio, Hernán Vargas, Minh Pham, Jay Pujara, Basel Shbita, Binh Vu, Yao-Yi Chiang, Dan Feldman, Yijun Lin, Hayley Song, Vipin Kumar, Ankush Khandelwal, Michael Steinbach, Kshitij Tayal, Shaoming Xu, Suzanne A. Pierce, Lissa Pearson, Daniel Hardesty-Lewis, Ewa Deelman, Rafael Ferreira da Silva, Rajiv Mayani, Armen R. Kemanian, Yuning Shi, Lorne Leonard, Scott Peckham, Maria Stoica, Kelly Cobourn, Zeya Zhang, Christopher Duffy and Lele Shu. ACM Transactions on Interactive Intelligent Systems (TiiS), 2021.

[Gomes et al 2019] "Computational sustainability: computing for a better world and a sustainable future." Carla Gomes, Thomas Dietterich, Christopher Barrett, Jon Conrad, Bistra Dilkina, Stefano Ermon, Fei Fang, Andrew Farnsworth, Alan Fern, Xiaoli Fern, Daniel Fink, Douglas Fisher, Alexander Flecker, Daniel Freund, Angela Fuller, John Gregoire, John Hopcroft, Steve Kelling, Zico Kolter, Warren Powell, Nicole Sintov, John Selker, Bart Selman, Daniel Sheldon, David Shmoys, Milind Tambe, Weng-Keen Wong, Christopher Wood, Xiaojian Wu, Yexiang Xue, Amulya Yadav, Abdul-Aziz Yakubu, Mary Lou Zeeman. Communications of the ACM, 62(9), 2019. https://doi.org/10.1145/3339399

[Gundersen et al 2018] "On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications." Odd Erik Gundersen, Yolanda Gil, and David W. Aha. AI Magazine, (39)3, 2018. http://doi.org/10.1609/aimag.v39i3.2816

[Khider et al 2019] "PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data." D. Khider, J. Emile-Geay, N. P. McKay, Y. Gil, D. Garijo, V. Ratnakar, M. Alonso-Garcia, S. Bertrand, O. Bothe, P. Brewer, A. Bunn, M. Chevalier, L. Comas-Bru, A. Csank, E. Dassié, K. DeLong, T. Felis, P. Francus, A. Frappier, W. Gray, S. Goring, L. Jonkers, M. Kahle, D. Kaufman, N. M. Kehrwald, B. Martrat, H. McGregor, J. Richey, A. Schmittner, N. Scroxton, E. Sutherland, K. Thirumalai, K. Allen, F. Arnaud, Y. Axford, T. Barrows, L. Bazin, S. E. Pilaar Birch, E. Bradley, J. Bregy, E. Capron, O. Cartapanis, H.-W. Chiang, K. M. Cobb, M. Debret, R. Dommain, J. Du, K. Dyez, S. Emerick, M. P. Erb, G. Falster, W. Finsinger, D. Fortier, N. Gauthier, S. George, E. Grimm, J. Hertzberg, F. Hibbert, A. Hillman, W. Hobbs, M. Huber, A. L. C. Hughes, S. Jaccard, J. Ruan, M. Kienast, B. Konecky, G. Le Roux, V. Lyubchich, V. F. Novello, L. Olaka, J. W. Partinv C. Pearce, S. J. Phipps, C. Pignol, N. Piotrowska, M.-S. Poli, A. Prokopenko, F. Schwanck, C. Stepanek, G. E. A. Swann, R. Telford, E. Thomas, Z. Thomas, S. Truebe, L. von Gunten, A. Waite, N. Weitzel, B. Wilhelm, J. Williams, M. Winstrup, N. Zhao, and Y. Zhou. Paleoceanography and Paleoclimatology, 34(10), 2019. http://doi.org/10.1029/2019PA003632

[Kitano 2016] “Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery.” Hiroaki Kitano. AI Magazine. 37(1), 2016. https://doi.org/10.1609/aimag.v37i1.2642

[Moreau et al 2013]. “PROV-DM: The PROV Data Model.” Luc Moreau, Paolo Missier, Khalid Belhajjame, Reza B'Far, James Cheney, Sam Coppens, Stephen Cresswell, Yolanda Gil, Paul Groth, Graham Klyne, Timothy Lebo, Jim McCusker, Simon Miles, James Myers, Satya Sahoo, Curt Tilmes. Published as a W3C Recommendation on 30 April 2013. Available from http://www.w3.org/TR/prov-dm/

[Poldrack et al 2017] “Scanning the horizon: towards transparent and reproducible neuroimaging research.” Russell A. Poldrack, Chris I. Baker, Joke Durnez, Krzysztof J. Gorgolewski, Paul M. Matthews, Marcus R. Munafò, Thomas E. Nichols, Jean-Baptiste Poline, Edward Vul, and Tal Yarkoni. Nature Reviews Neuroscience,18(2), 2017. https://doi.org/10.1038/nrn.2016.167

[Senior et al 2020] “Improved protein structure prediction using potentials from deep learning.” Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu & Demis Hassabis. Nature, 577, 2020. https://doi.org/10.1038/s41586-019-1923-7

[Srivastava et al 2019] “Semantic Workflows for Benchmark Challenges: Enhancing Comparability, Reusability and Reproducibility.” Arunima Srivastava, Ravali Adusumilli, Hunter Boyce, Daniel Garijo, Varun Ratnakar, Rajiv Mayani, Thomas Yu, Raghu Machiraju, Yolanda Gil, and Parag Mallick. Proceedings of the Pacific Symposium on Biocomputing (PSB), Waimea, HI, 2019. https://doi.org/10.1142/9789813279827_0019

[Tshitoyan et al 2019] “Unsupervised word embeddings capture latent knowledge from materials science literature.” Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder, and Anubhav Jain. Nature 571, 2019. https://doi.org/10.1038/s41586-019-1335-8

[Zhang et al 2014] “Proteogenomic characterization of human colon and rectal cancer.” Bing Zhang, Jing Wang, Xiaojing Wang, Jing Zhu, Qi Liu, Zhiao Shi, Matthew C. Chambers, Lisa J. Zimmerman, Kent F. Shaddox, Sangtae Kim, Sherri R. Davies, Sean Wang, Pei Wang, Christopher R. Kinsinger, Robert C. Rivers, Henry Rodriguez, R. Reid Townsend, Matthew J. C. Ellis, Steven A. Carr, David L. Tabb, Robert J. Coffey, Robbert J. C. Slebos, Daniel C. Liebler & the NCI CPTAC. Nature, 513, 2014. https://doi.org/10.1038/nature13438