Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery

Hiroaki Kitano

■ *This article proposes a new grand challenge for AI: to develop an AI system that can make major scientific discoveries in biomedical sciences and that is worthy of a Nobel Prize. There are a series of human cognitive limitations that prevent us from making accelerated scientific discoveries, particularity in biomedical sciences. As a result, scientific discoveries are left at the level of a cottage industry. AI systems can transform scientific discoveries into highly efficient practices, thereby enabling us to expand our knowledge in unprecedented ways. Such systems may outcompute all possible hypotheses and may redefine the nature of scientific intuition, hence the scientific discovery process.*

What is the single most significant capability that artificial intelligence can deliver? What pushes the human race forward? Our civilization has advanced largely by scientific discoveries and the application of such knowledge. Therefore, I propose the launch of a grand challenge to develop AI systems that can make significant scientific discoveries. As a field with great potential social impacts, and one that suffers particularly from information overflow, along with the limitations of human cognition, I believe that the initial focus of this challenge should be on biomedical sciences, but it can be applied to other areas later. The challenge is "to develop an AI system that can make major scientific discoveries in biomedical sciences and that is worthy of a Nobel Prize and far beyond." While recent progress in high-throughput "omics" measurement technologies has enabled us to generate vast quantities of data, scientific discoveries themselves still depend heavily upon individual intuition, while researchers are often overwhelmed by the sheer amount of data, as well as by the complexity of the biological phenomena they are seeking to understand. Even now, scientific discovery remains something akin to a cottage industry, but a great transformation seems to have begun. This is an ideal domain and the ideal timing for AI to make a difference. I anticipate that, in the near future, AI systems will make a succession of discoveries that have immediate medical implications, saving millions of lives, and totally changing the fate of the human race.

Grand Challenges as a Driving Force in AI Research

Throughout the history of research into artificial intelligence, a series of grand challenges have been significant driving factors. Advances in computer chess demonstrated that a computer can exhibit human-level intelligence in a specific domain. In 1997, IBM's chess computer Deep Blue defeated human world champion Gary Kasparov (Hsu 2004). Various search algorithms, parallel computing, and other computing techniques originating from computer chess research have been applied in other fields. IBM took on another challenge when it set the new goal of building a computer that could win the TV quiz show *Jeopardy!* In this task, which involved the real-time answering of open-domain questions (Ferrucci et al. 2010, Ferrucci et al. 2013), IBM's Watson computer outperformed human quiz champions. IBM is currently applying technology from Watson as part of its business in a range of industrial and medical fields. In an extension of prior work on computer chess, Japanese researchers have even managed to produce a machine capable of beating human grand masters of Shogi, a Japanese chess variant with a significantly larger number of possible moves.

RoboCup is a grand challenge founded in 1997 that traverses the fields of robotics and soccer. The aim of this initiative is to promote the development by the year 2050 of a team of fully autonomous humanoid robots that is able to beat the most recent winners of the FIFA World Cup (Kitano et al. 1997). This is a task that requires both an integrated, collective intelligence and exceptionally high levels of physical performance. Since the inaugural event, the scheme has already given birth to a series of technologies that have been deployed in the real world. For example, KIVA Systems, a technology company that was formed based largely on technologies from Cornell University's team for RoboCup's Small Size League, provided a highly automated warehouse management system that Amazon.com acquired in 2012. Various robots that were developed for the Rescue Robot League — a part of RoboCup focused on disaster rescue — have been deployed in real-world situations, including search and rescue operations at New York's World Trade Center in the aftermath of the 9/11 terror attacks, as well as for surveillance missions following the accident at the Fukushima Daiichi Nuclear Power Plant.

These grand challenges present a sharp contrast with the Turing test, aimed as they are at the development of superhuman capabilities as opposed to the Turing test's attempts to answer the question "Can machines think?" by creating a machine that can generate humanlike responses to natural language dialogues (Turing 1950). These differing approaches present different scientific challenges, and, while going forward we may expect some cross-fertilization between these processes, this article focuses on the grand challenge of building superhuman capabilities.

History provides many insights into changes over time in the technical approaches to these challenges. In the early days of AI research, it was widely accepted that a brute force approach would not work for chess, and that heuristic programming was essential for very large and complex problems (Feigenbaum and Feldman 1963). Actual events, however, confounded this expectation. Among the features critical for computer chess were the massive computing capability required to search millions of moves; vast memory to store a record of all past games; and a learning mechanism to evaluate the quality of each move and adjust search paths accordingly. Computing power, memory, and learning have proven to hold the winning formula, overcoming sophisticated heuristics. The 1990s saw a similar transformation of approach in speech recognition, where rule-based systems were outperformed by data- and computing-driven systems based on hidden Markov models (Lee 1988). Watson, the IBM computer that won the *Jeopardy!* quiz show, added new dimensions of massively parallel heterogeneous inference and real-time stochastic reasoning. Coordination of multiple different reasoning systems is also key when it comes to Shogi. Interestingly, similar technical features are also critical in bioinformatics problems (Hase et al. 2013; Hsin, Ghosh, and Kitano 2013). Elements currently seen as critical include massively parallel heterogeneous computing, real-time stochastic reasoning, limitless access to information throughout the network, and sophisticated multistrategy learning. Recent progress in computer GO added a combination of deep learning, reinforcement learning, and tree search to be the winning formula (Silver et al. 2016). Challenges such as those described have been highly effective in promoting AI research. By demonstrating the latest advances in AI, and creating high-impact industrial applications, they continue to contribute to the progress of AI and its applications.

The Scientific Discovery Grand Challenge

It is time to make an even greater stride, by imagining and initiating a new challenge that may change our very principles of intelligence and civilization. While scientific discovery is not the only driving force of our civilization, it has been one of the most critical factors. Creating AI systems with a very high capability for scientific discovery will have a profound impact, not only in the fields of AI and computer science, but also in the broader realms of science and technology. It is a commonly held perception that scientific discoveries take place after years of dedicated effort or at a moment of great serendipity. The process of scientific discovery as we know it today is considered unpredictable and ineffi-

cient and yet is blithely accepted. I would argue, however, that the practice of scientific discovery is stuck at a level akin to that of a cottage industry. I believe that the productivity and fundamental modalities of the scientific discovery process can be dramatically improved. The real challenge is to trigger a revolution in science equivalent to the industrial revolution.

It should be noted that machine discovery, or discovery informatics (Gil et al. 2014, Gil and Hirsh 2012), has long been a major topic for AI research. BEACON (Langley and Simon 1987), DENDRAL (Lindsay et al. 1993), AM, and EURISKO (Lenat and Brown 1984) are just some of the systems of this nature developed to date.

We must aim high. What distinguishes the proposed challenge from past efforts is its focus on biomedical sciences in the context of dramatic increases in the amount of information and data available, along with levels of interconnection of experimental devices that were unavailable in the past. It is also set apart by the focus on research, with the extremely ambitious goal of facilitating major scientific discoveries in the biomedical sciences that may go on to earn the Nobel Prize in Physiology or Medicine, or achieve even more profound results. This is the moonshot in AI. Just as the Apollo project's goal went beyond the moon (Kennedy 1961, 1962), the goals of this project go far beyond the Nobel Prize. The goal is to promote a revolution in scientific discovery and to enable the fastest-possible expansion in the knowledge base of mankind. The development of AI systems with such a level of intelligence would have a profound impact on the future of humanity.

Human Cognitive Limitations in Biomedical Sciences

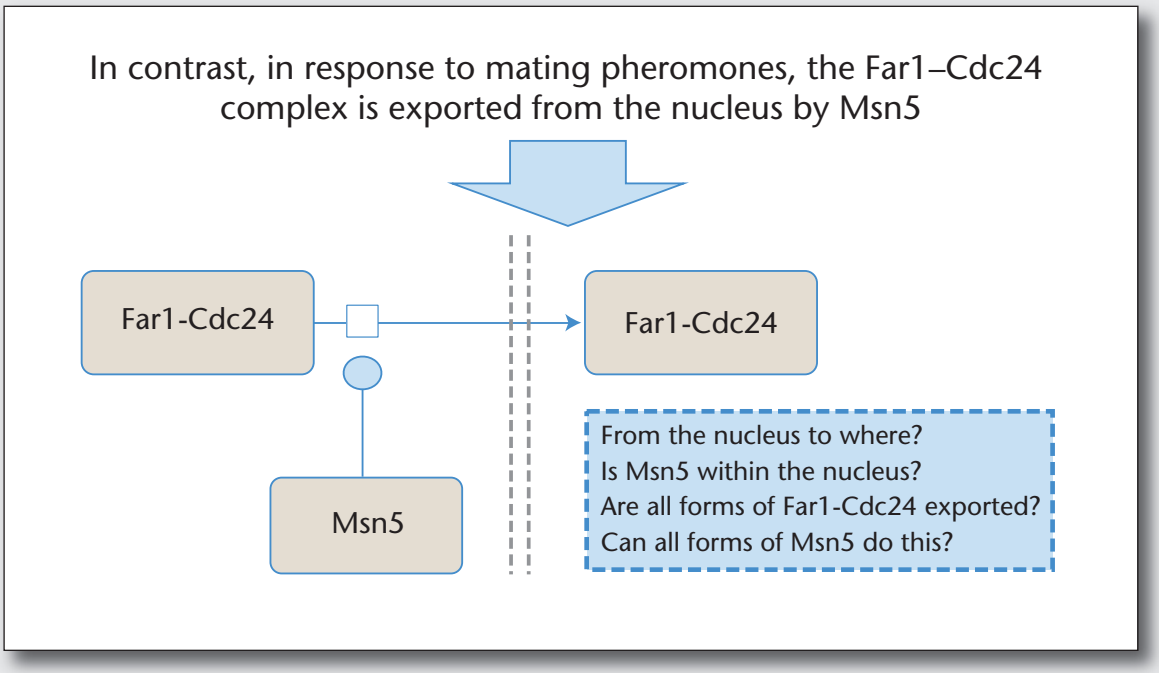
There are fundamental difficulties in biomedical research that overwhelm the cognitive capabilities of humans. This problem became even more pronounced with the emergence of systems biology (Kitano 2002a, 2002b). Some of the key problems are outlined below.

First, there is the information horizon problem. Biomedical research is flooded with data and publications at a rate of production that goes far beyond human information-processing capabilities. Over 1 million papers are published each year, and this rate is increasing rapidly. Researchers are already overwhelmed by the flood of papers and data, some of which may be contradictory, inaccurate, or misused. It is simply not possible for any researcher to read, let alone comprehend, such a deluge of information in order to maintain consistent and up-to-date knowledge. The amount of experimental data is exploding at an even faster pace, with widespread use of high-throughput measurement systems. Just as the rapidly expanding universe creates a cosmic event horizon that prevents even light emitted in the distant past from reaching us, thus rendering it unobservable, the never-ending abundance of publications and data creates an information horizon that prevents us from observing a whole picture of what we have discovered and what data we have gathered. It is my hope that, with the progress instigated by the challenge I am proposing, AI systems will be able to compile a vast body of intelligence in order to mitigate this problem (Gil et al. 2014).

Second, there is also the problem of an information gap. Papers are written in language that frequently involves ambiguity, inaccuracy, and missing information. Efforts to develop a large-scale comprehensive map of molecular interactions (Caron et al. 2010, Matsuoka et al. 2013, Oda and Kitano 2006, Oda et al. 2005) or any kind of biological knowledge base of any form will encounter this problem (see sidebar). Our interpretation, hence human-based knowledge extraction, largely depends on subjectively filling in the gaps using the reader's own knowledge or representation of knowledge with missing details, results in an arbitral interpretation of knowledge in the text.

Obviously, solving this is far beyond the capacity to convey information of the language of a given text (Li, Liakata, and Rebholz-Schuhmann 2014). It also involves actively searching for missing information to discern what is missing and how to find it. It is important to capture details of the interactions within a process rather than merely an abstracted overview, because researchers are well aware of overall interaction and expect such a knowledge base, or maps, to provide consistent and comprehensive yet in-depth description of each interaction. Similar issues exist when it comes to understanding images from experiments. They include: how to interpret images, checking consistency with the sum of past data, identifying differences and the reasons for these, and recovering missing information on experimental conditions and protocol.

Third, there is a problem of phenotyping inaccuracy. The word *phenotyping* refers to representation and categorization of biological anomalies such as disease, effects of genetics mutations, and developmental defects. Phenotyping is generally performed based on subjective interpretation and consensus of medical practitioners and biologists, described using terms that are relatively easy to understand. This practice itself is tightly linked with human cognitive limitations. Biomedical sciences have to deal with complex biological systems that are highly nonlinear, multidimensional systems. Naïve delineation of observation into coarse categorization can create significant inaccuracies and lead to misdiagnosis and inaccurate understanding of biological phenomena (figure 1a). This is a practical clinical problem as signified in some rare disease cases that took decades for patients to be diagnosed and had almost 40 percent



An Example of Missing Information in a Biological Statement

Biomedical science is knowledge-intensive and empirical science. Currently, knowledge is embedded in the text and images in publications. The figure exemplifies a case of missing information implicit in biomedical papers. Take the example of the following typical sentence from a biology paper: "In contrast, in response to mating pheromones, the Far1-Cdc24 complex is exported from the nucleus by Msn5" (taken from the abstract by Shimada, Gulli, and Peter [2000]). We can extract knowledge on a specific molecular interaction involving the Far1-Cdc24 complex and Msn5 and represent this graphically. The sentence itself does not, however, describe where the Far-Cdc24 complex is exported to, and where Msn5 is located. In such cases, researchers can fill in the conceptual gaps from their own biological knowledge. However, it is not clear if all forms of the Far1-Cdc24 complex will become the subject of this interaction, nor if all forms of Msn5 can conduct this export process. In this case, the general biological knowledge of researchers will generally prove insufficient to fill in such gaps, thereby necessitating either the inclusion of a specific clarifying statement elsewhere in the paper, or the need to search other papers and databases to fill this gap.

of initial misdiagnosis rate (EURORDIS 2007). Clinical diagnosis is a process of observation, categorization of observed results, and hypothesis generation on a patient's disease status. Misdiagnosis leads to inappropriate therapeutic interventions. Identification of proper feature combinations for each axis, the proper dimension for the representation of space, and the proper granularity for categorization shall significantly improve diagnosis, hence therapeutic efficacy (figure 1b). Extremely complex feature combinations for each axis, extreme high-dimensional representation of space, and extremely fine-grained categorization that can be termed as *extreme classification* shall dramatically improve accuracy of diagnosis. Since many diseases are constellations of very large numbers of subtypes of diseases, such an extreme classification shall enable us to properly identify specific patient subgroups that may not be identified as an isolate group at present and lead to specific therapeutic options. However, an emerging problem would be that humans may not be able to comprehend what exactly each category means in

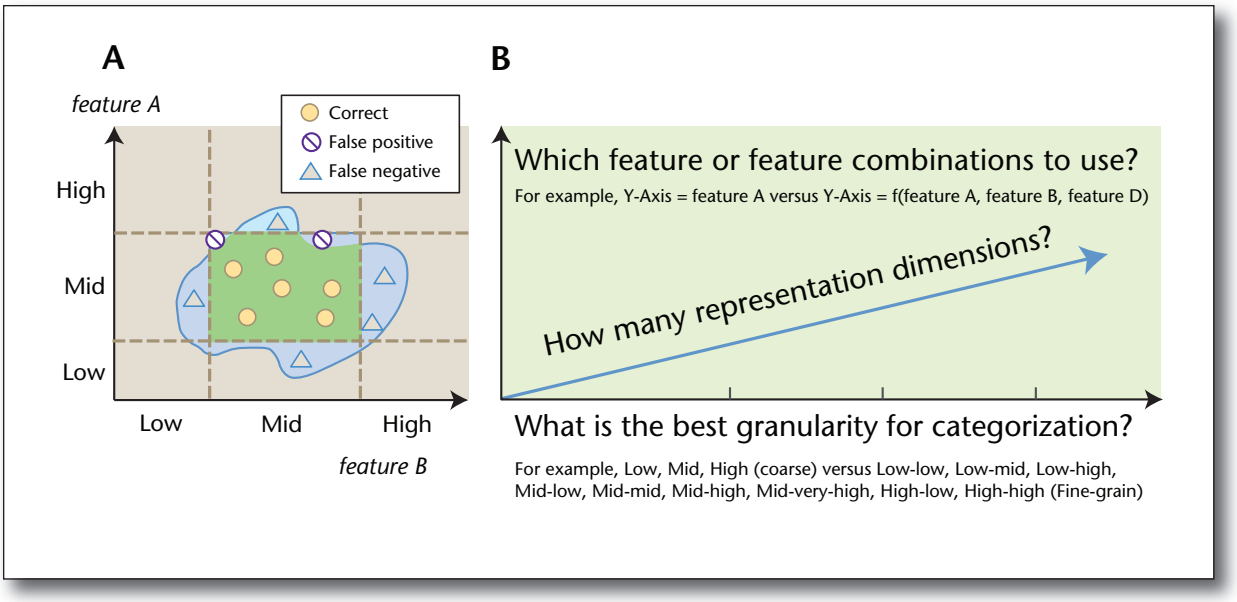


Figure 1. Problems in the Representation and Categorization of Biological Objects and Processes.

Left figure modified based on Kitano (1993). Figure 1a is an example of an attempt to represent a nonlinear boundary object, assumed to be a simplification of a phenotype, with a simple two-feature dimensional space with coarse categorization such as Low, Mid, and High. The object can be most covered with "feature A = Mid and feature B = Mid condition." However, it inevitably results in inaccuracy (false-positives and false-negatives). Improving accuracy of nonlinear object coverage requires the proper choice of the feature complex for each axis, the proper dimension of representational space, and the proper choice of categorization granularity (figure 1b).

relation to their own biomedical knowledge, which was developed based on the current coarse and low-dimensional categorization.

Another closely related problem is that of cognitive bias. Due to the unavoidable use of language and symbols in our process of reasoning and communication, our thought processes are inevitably biased. As discussed previously, natural language does not properly represent biological reality. Alfred Kozybski's statement that "the map is not the territory" (Korzybski 1933) is especially true in biomedical sciences (figure 2). Vast knowledge of the field comes in the form of papers that are full of such biases. Our ability to ignore inaccuracies and ambiguity facilitates our daily communication, yet poses serious limitations on scientific inquiry.

Then there is the minority report problem. Biology is an empirical science, meaning knowledge is accumulated based on experimental findings. Due to the complexity of biological systems, diversity of individuals, uncertainty of experimental conditions, and other factors, there are substantial deviations and errors in research outcomes. While consensus among a majority of reports can be considered to portray the most probable reality regarding a specific aspect of biological systems, reports exist that are not consistent with this majority (figure 3).

Whether such minority reports can be discarded as errors or false reports is debatable. While some will naturally fall into this category, others may be correct, and may even report unexpected biological findings that could lead to a major discovery. How can we distinguish between such erroneous reports and those with the potential to facilitate major discoveries?

Are We Ready to Embark on This Challenge?

I have described some of the human cognitive limitations that act as obstacles to efficient biomedical research, and that AI systems may be able to resolve during the course of the challenge I am proposing. Interestingly, there are a few precedents that may provide a useful starting point. Of the early efforts to mitigate the information horizon problem, research using IBM's Watson computer is currently focused on the medical domain. The intention is to compile the vast available literature and present it in a coherent manner, in contrast to human medical practitioners and researchers who cannot read and digest the entire available corpus of information. Watson was used in a collaboration between IBM, Baylor College of Medicine, and MD Anderson Cancer Center that led to the identification of novel modification sites of p53, an important protein for cancer suppression (Spangler et al. 2014). A recent DARPA Big Mechanism Project (BMP) aimed at automated extraction of

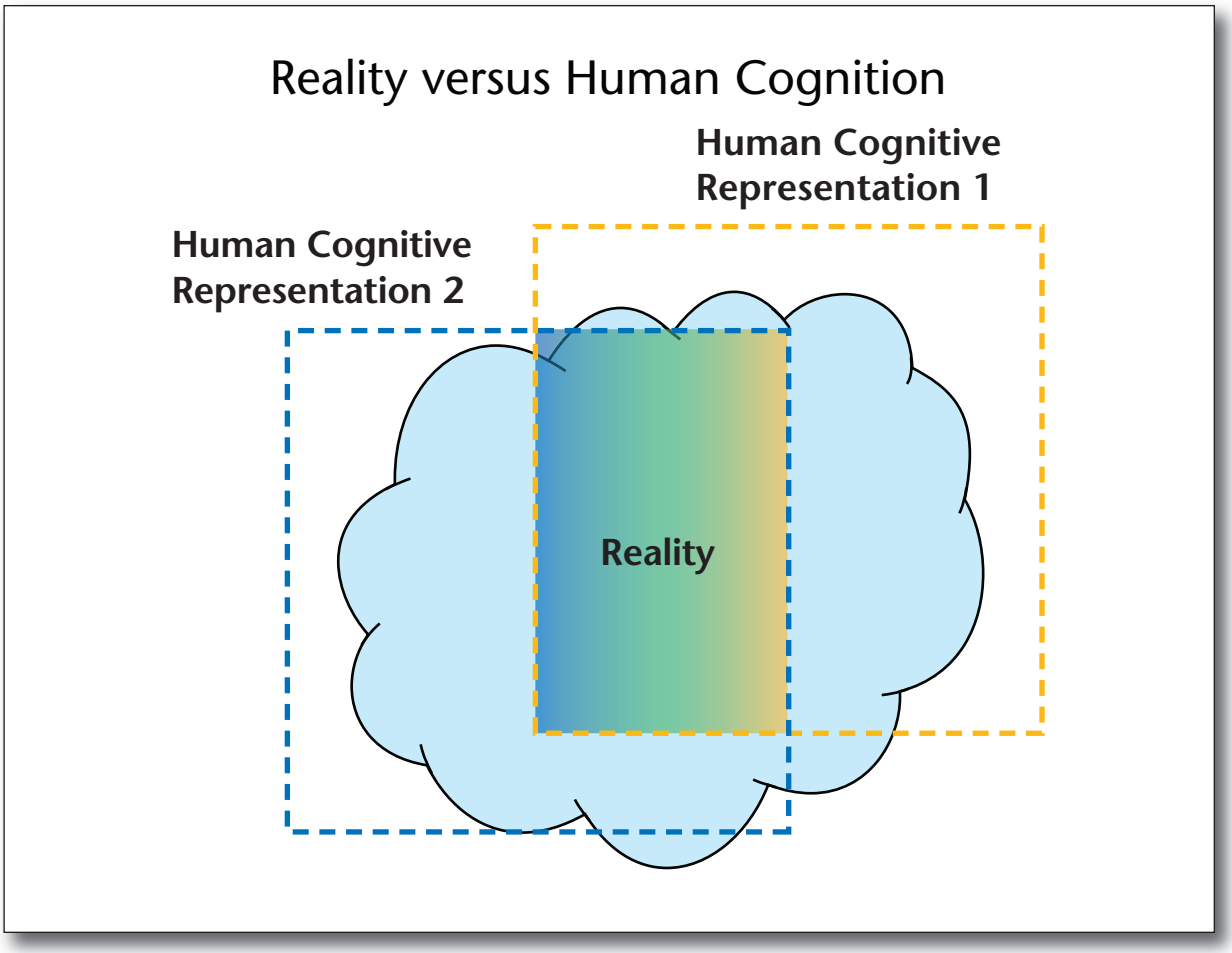


Figure 2. The Same Reality Can Be Expressed Differently, or the Same Linguistic Expressions May Represent Different Realities.

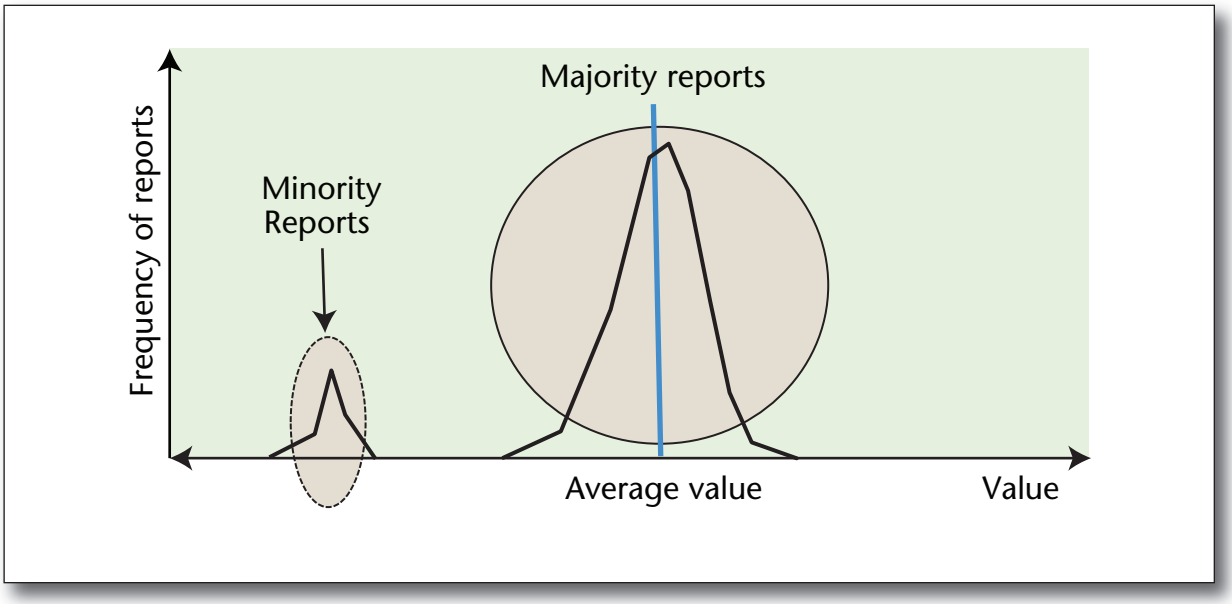


Figure 3. Should Minority Reports Be Discarded? Or Might They Open Up Major Discoveries?

large-scale molecular interactions related to cancer (Cohen 2014).

With regard to problems of phenotyping inaccuracy, progress in machine learning as exemplified in deep learning may enable us to resolve some cognitive issues. There are particular hopes that computers may learn to acquire proper features for representing complex objects (Bengio 2009; Bengio, Courville, and Vincent 2013; Hinton 2011). Deep phenotyping is an attempt to develop much finer-grained and in-depth phenotyping than current practice provides to establish highly accurate diagnosis, patient classification, and precision clinical decisions (Frey, Lenert, and Lopez-Campos 2014; Robinson 2012), and some of pioneering researchers are using deep learning (Che et al. 2015). Combining deep phenotyping and personal genomics as well as other comprehensive measurements leads to dramatically improved accurate diagnosis and effective therapeutic interventions, as well as improving drug discovery efficiency.

For generating hypotheses and verifying them, Ross King and his colleagues have developed a systematic robot scientist that can infer possible biological hypotheses and design simple experiments using a defined-protocol automated system to analyze orphan genes in budding yeast (King et al. 2009a, 2009b; King et al. 2004). While this brought only a moderate level of discovery within the defined context of budding yeast genes, the study represented an integration of bioinformatics-driven hypothesis generation and automated experimental processes. Such an automatic experimental system has great potential for expansion and could become a driving force for research in the future.

Most experimental devices these days are highly automated and connected to networks. In the near future, it is likely that many will be supplemented by high-precision robotics systems, enabling AI systems not only to access digital information but also to design and execute experiments. That would mean that every detail of experimental results, including incomplete or erroneous data, could be stored and made accessible. Such progress would have a dramatic impact on the issues of long-tail distribution and dark data in science (Heidorn 2008).

Crowdsourcing of science, or citizen science, offers many interesting opportunities, and great potential for integration with AI systems. The protein-folding game FoldIt, released in 2008, demonstrated that with proper redefinition of a scientific problem, ordinary citizens can contribute to the process of scientific discovery (Khatib et al. 2011). Patient-powered research network Patientslikeme is another example of how motivated ordinary people can contribute to science (Wicks et al. 2015, Wicks et al. 2011). While successful deployment of community-based science requires carefully designed missions, clear definition of problems, and the implementation of appropriate user interfaces (Kitano, Ghosh, and Matsuoka 2011), crowdsourcing may offer an interesting opportunity for AI-based scientific discovery. This is because, with proper redefinition of a problem, a system may also help to facilitate the best use of human intelligence.

There are efforts to develop platforms that can connect a broad range of software systems, devices, databases, and other necessary resources. The Garuda platform is an effort to develop an open application programming interface (API) platform aimed at attaining a high-level of interoperability among biomedical and bioinformatics analysis tools, databases, devices, and others (Ghosh et al. 2011). The Pegasus and Wings system is another example that focuses on sharing the workflow of scientific activities (Gil et al. 2007). A large-scale collection of workflow from the scientific community that may direct possible sequences of analysis and experiments used and reformulated by AI systems would be a powerful knowledge asset. With globally interconnected high-performance computing systems such as InfiniCortex Michalewicz, et al. 2015), we are now getting ready to undertake this new and formidable challenge. Such research could form the partial basis of this challenge. At the same time, we still require a clear game plan, or at the very least an initial hypothesis.

Scientific Discovery as a Search Problem: Deep Exploration of Knowledge Space

What is the essence of discovery? To rephrase the question, what could be the engine for scientific discovery? Consistent and broad-ranging knowledge is essential, but does not automatically lead to new discoveries. When I talk about this initiative, many scientists ask whether AI can be equipped with the necessary intuition for discovery. In other words, can AI systems be designed to ask the "right" questions that may lead to major scientific discoveries? While this certainly appears to be a valid question, let us think more deeply here. Why is asking the right question important? It may be due to resource constraints (such as the time for which researchers can remain active in their professional careers), budget, competition, and other limitations. Efficiency is, therefore, the critical factor to the success of this challenge. When time and resources are abundant, the importance of asking the right questions is reduced. One might arrive at important findings after detours, so the route is not of particular significance. At the same time, science has long relied to a certain extent on serendipity, where researchers made a major discovery by accident. Thinking about such observations, it is possible to arrive at a hypothesis that infers that the critical aspect of scientific discovery is how many hypotheses can be generated and tested, including examples that may seem highly unlikely.

This indicates the potential to scientific discovery

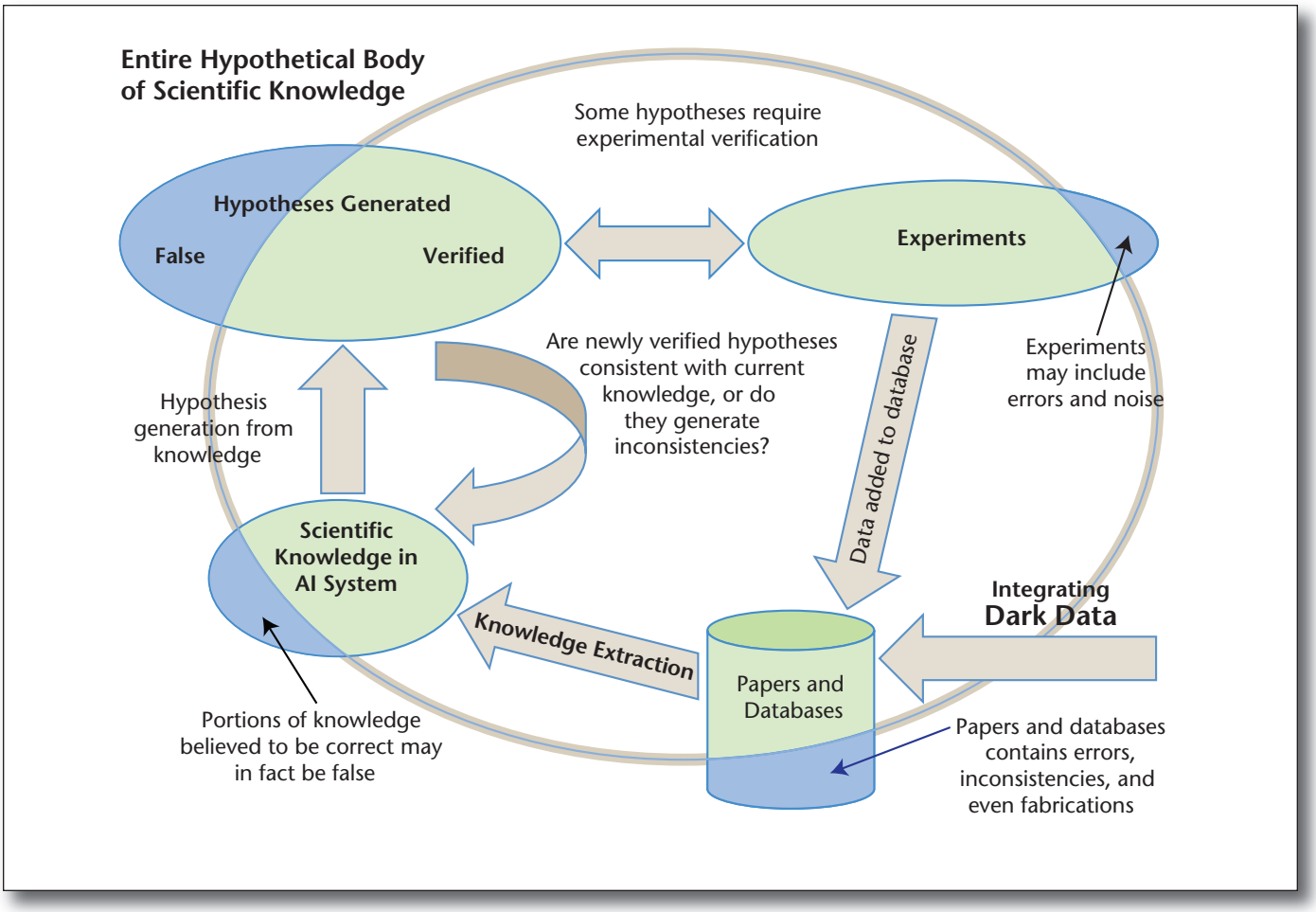


Figure 4. Bootstrapping of Scientific Discovery and Knowledge Accumulation.

Correct and incorrect knowledge, data, and experimental results are involved throughout this process, though some may be ambiguous. Scientific discovery requires an iterative cycle aimed at expanding our knowledge on this fragile ground. The aim is to compute, verify, and integrate every possible hypothesis, thereby building a consistent body of knowledge.

of a brute-force approach in which AI systems generate and verify as many hypotheses as possible. Such an approach may differ from the way in which scientists traditionally conduct their research, but could become a computational alternative to the provision of scientific insights. It should be stressed that while the goal of the grand challenge is to make major scientific discoveries, this does not necessarily mean those discoveries should be made as if by human scientists.

The brute-force approach empowered by machine learning and heterogeneous inference has already provided the basis of success for a number of grand challenges to date. As long as a hypothesis can be verified, scientific discovery can also incorporate computing to search for probable correct hypotheses from among the full range of possible ones. The fundamental thrust should be toward massive combinatorial hypothesis generation, the maintenance of a consistent repository of global knowledge, and perhaps a number of other fundamental principles that we may not be aware of at present. Thus, by using computing to generate and verify as quickly as possible the full range of logically possible hypotheses, it would mitigate resource constraint issues and enable us to examine even unexpected or seemingly far-fetched ideas. Such an approach would significantly reduce the need to ask the right questions, thereby rendering scientific intuition obsolete, and perhaps even enabling us to explore computational serendipity.

The engine of discovery should be a closed-loop system of hypothesis generation and verification, knowledge maintenance, knowledge integration, and so on (figure 4) and should integrate a range of technologies (figure 5). Fundamentally speaking, hypotheses, along with constraints imposed on

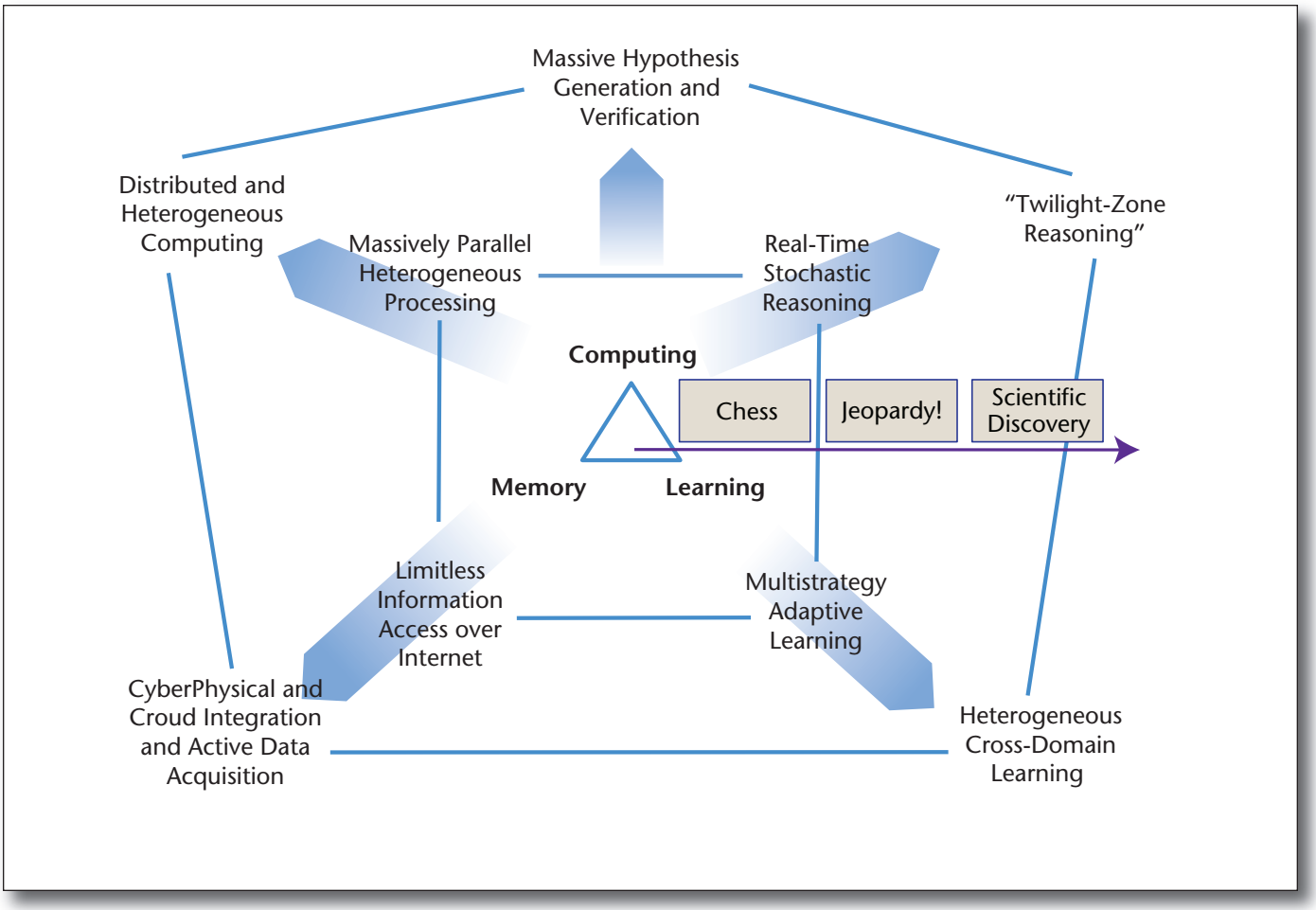


Figure 5. Evolution of Key Elements in Grand Challenges and Possible Elements of the Scientific Discovery Grand Challenge.

Computing, memory, and learning have long been key elements in computer chess. Further techniques have originated from the application of computers to the quiz show *Jeopardy!* To facilitate scientific discovery, an even more complex and sophisticated range of functions is required. The term *twilight-zone reasoning* refers to the parsing of data and publications that may be highly ambiguous, error-prone, or faulty. The elements introduced here represent general ideas on how to approach the scientific discovery grand challenge, rather than originating from precise technical analysis of the necessary functionalities.

hypothesis generation and the initial validation process, would be derived from the vast body of knowledge to be extracted from publications, databases, and automatically executed experiments. Successfully verified hypotheses would be added to the body of knowledge, enabling the bootstrapping process to continue. It is crucial to recognize that not all papers and data to emerge from the scientific community are correct or reliable; they contain substantial errors, missing information, and even fabrications. It may be extremely difficult to reproduce the published experimental results, and some may prove impossible to re-create (Prinz, Schlange, and Asadullah 2011). At the same time, major progress is continually being made in the field of biomedical science. How can this be possible if such a high proportion of papers present results that are false or not reproducible? While individual reports may contain a range of problems, collective knowledge has the potential to uncover truths from even an error-prone scientific process. This is a twilight zone of scientific discovery, and AI systems need to be able to reason in the twilight zone. The proposed challenge would shed light on this conundrum.

Advanced Intelligence

What is certain is that such a system would substantially reinforce the intellectual capabilities of humans in a manner that is entirely without precedent and that holds the potential to change fundamentally the way science is conducted.

The first-ever defeat of a chess grand master by an AI system was followed by the emergence of a new style of chess known as advanced chess, in which human and computer work together as a team, to

take on similarly equipped competitors. This partnership may be considered a form of human-computer symbiosis in intelligent activities. Similarly, we can foresee that in the future sophisticated AI systems and human researchers will work together to make major scientific discoveries. Such an approach can be considered "advanced intelligence."

Advanced intelligence as applied to scientific discovery would go beyond existing combinations of AI and human experts. Just as most competitive biomedical research institutions are now equipped with high-throughput experimental systems, I believe that AI systems will become a fundamental part of the infrastructure for top-level research institutions in the future. This may involve a substantial level of crowd intelligence, utilizing the contributions of both qualified researchers and ordinary people to contribute, each for different tasks, thereby forming a collaborative form of intelligence that could be ably and efficiently orchestrated by AI systems. Drawing this idea out to its extreme, it may be possible to place AI systems at the center of a network of intelligent agents — comprising both other AI systems and humans — to coordinate large-scale intellectual activities. Whether this path would ultimately make our civilization more robust (by facilitating a series of major scientific discoveries) or more fragile (due to extensive and excessive dependence on AI systems) is yet to be seen. However, just as Thomas Newcomen's atmospheric engine was turned into a modern form of steam engine by James Watt to become the driving force of the industrial revolution, AI scientific discovery systems have the potential to drive a new revolution that leads to new frontiers of civilization.

References

Bengio, Y. 2009. Learning Deep Architecture for AI. *Foundations and Trends in Machine Learning* 2(1): 1–127. dx.doi.org/10.1561/2200000006

Bengio, Y.; Courville, A.; and Vincent, P. 2013. Representation Learning: A Review and New Prespectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8): 1798–1828. dx.doi.org/10.1109/TPAMI.2013.50

Caron, E.; Ghosh, S.; Matsuoka, Y.; Ashton-Beaucage, D.; Therrien, M.; Lemieux, S.; Perreault, C.; Roux, P.; and Kitano, H. 2010. A Comprehensive Map of the mTOR Signaling Network. *Molecular Systems Biology* 6, 453. dx.doi.org/10.1038/msb.2010.108

Che, Z.; Kale, D.; Li, W.; Bahadori, M. T.; and Liu, Y. 2015. Deep Computational Phenotyping. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery. dx.doi.org/10.1145/2783258.2783365

Cohen, P. 2014. Big Mechanism [Project Announcement]. Arlington, VA: Defense Advanced Research Projects Agency.

EURORDIS. 2007. Survey of the Delay in Diagnosis for 8 Rare Diseases in Europe (EurordisCare2). Brussels, Belgium: EURODIS Rare Diseases Europe.

Feigenbaum, E., and Feldman, J. 1963. Computers and Thought. New York: McGraw-Hill Book Company.

Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A; Lally, A.; Murdock, J. W.; Nyberg, E.; Prager, J.; Schlaefer, N.; Welty, C. 2010. Building Watson: An Overview of the DeepQA Project. *AI Magazine* 31(3): 59–79.

Ferrucci, D.; Levas, A; Bagchi, S.; Gondek, D.; and Mueller, E. 2013. Watson: Beyond Jeopardy! *Artificial Intelligence* 199–200: (June–July): 93–105. dx.doi.org/10.1016/j.artint.2012.06.009

Frey, L. J.; Lenert, L.; and Lopez-Campos, G. 2014. EHR Big Data Deep Phenotyping. Contribution of the IMIA Genomic Medicine Working Group. *Yearbook of Medical Informatics* 9: 206–211. dx.doi.org/10.15265/IY-2014-0006

Ghosh, S.; Matsuoka, Y.; Asai, Y.; Hsin, K. Y.; and Kitano, H. 2011. Software for Systems Biology: From Tools to Integrated Platforms. *Nature Reviews Genetics* 12(12): 821–832. dx.doi.org/10.1038/nrg3096

Gil, Y.; Greaves, M.; Hendler, J.; and Hirsh, H. 2014. Artificial Intelligence. Amplify Scientific Discovery with Artificial Intelligence. *Science* 346(6206): 171–172. dx.doi.org/10.1126/science.1259439

Gil, Y., and Hirsh, H. 2012. Discovery Informatics: AI Opportunities in Scientific Discovery. In *Discovery Informatics: The Role of AI Research in Innovating Scientific Processes*: Papers from the AAAI Fall Symposium, 1–6. Technical Report FS-12-03. Palo Alto, CA: AAAI Press.

Gil, Y.; Ratnakar, V.; Deelman, E.; Mehta, G.; and Kim, J. 2007. Wings for Pegasus: Creating Large-Scale Scientific Applications Using Sematic Representations of Computational Workflows. In *Proceedings of the 19th Innovative Applications of Artificial Intelligence* (IAAI-07). Palo Alto, CA: AAAI Press.

Hase, T.; Ghosh, S.; Yamanaka, R.; and Kitano, H. 2013. Harnessing Diversity Towards the Reconstructing of Large Scale Gene Regulatory Networks. *PLoS Computational Biology* 9(11): e1003361. dx.doi.org/10.1371/journal.pcbi.1003361

Heidorn, P. B. 2008. Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends* 57(2): 280–299. dx.doi.org/10.1353/lib.0.0036

Hinton, G. 2011. A Better Way to Learn Features. *Communications of the ACM* 54(10). dx.doi.org/10.1145/2001269.2001294

Hsin, K. Y.; Ghosh, S.; and Kitano, H. 2013. Combining Machine Learning Systems and Multiple Docking Simulation Packages to Improve Docking Prediction Reliability for Network Pharmacology. *PLoS One* 8(12): e83922. dx.doi.org/10.1371/journal.pone .0083922

Hsu, F.-H. 2004. *Behind Deep Blue: Buidling the Computer That Defeated the World Chess Champion*. Princeton, NJ: Princeton University Press.

Kennedy, J. F. 1961. Special Message to Congress on Urgent National Needs, 25 May 1961. Papers of John F. Kennedy. Presidential Papers. President's Office Files. JFKPOF-034-030 John F. Kennedy Presidential Library, Boston, MA.

Kennedy, J. F. 1962. Address at Rice University on the Nation's Space Effort 12 September 1962. Accession Number USG:15 reel 29. John F. Kennedy Presidential Library, Boston, MA.

Khatib, F.; DiMaio, F.; Foldit Contenders Group; Foldit Void Crushers Group; Cooper, S.; Kazmierczyk, M.; Gilski, M.; Krzywda, S.; Zabranska, H.; Pichova, I.; Thompson, J.; Popovi, Z.; Jaskolski, M.; Baker, D. 2011. Crystal Structure of a Monomeric Retroviral Protease Solved by Protein Folding Game Players. *Natural Structural and Molecular Biology*

18(10): 1175–1177. dx.doi.org/10.1038/ nsmb.2119

King, R. D.; Rowland, J.; Oliver, S. G.; Young, M.; Aubrey, W.; Byrne, E.; Liakata, M.; Markham, M.; Pir, P.; Soldatova, L. N.; Sparkes, A.; Whelan, K. E.; Clare, A. 2009a. The Automation of Science. *Science* 324(5923): 85–89. dx.doi.org/10.1126/science.1165620

King, R. D.; Rowland, J.; Oliver, S. G.; Young, M.; Aubrey, W.; Byrne, E.; Liakata, M.; Markham, M.; Pir, P.; Soldatova, L. N.; Sparkes, A.; Whelan, K. E.; Clare, A. 2009b. Make Way for Robot Scientists. *Science* 325(5943), 945. dx.doi.org/10.1126/science. 325_945a

King, R. D.; Whelan, K. E.; Jones, F. M.; Reiser, P. G.; Bryant, C. H.; Muggleton, S. H.; Kell, D. B.; Oliver, S. G. 2004. Functional Genomic Hypothesis Generation and Experimentation by a Robot Scientist. *Nature* 427(6971): 247–252. dx.doi.org/10. 1038/nature02236

Kitano, H. 1993. Challenges of Massive Parallelism. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence,* 813–834. San Mateo, CA: Morgan Kaufmann Publishers.

Kitano, H. 2002a. Computational Systems Biology. *Nature* 420(6912): 206–210. dx.doi. org/10.1038/nature01254

Kitano, H. 2002b. Systems Biology: A Brief Overview. *Science* 295(5560): 1662–1664. dx.doi.org/10.1126/science. 1069492

Kitano, H.; Asada, M.; Kuniyoshi, Y.; Noda, I.; Osawa, E.; and Matsubara, H. 1997. RoboCup: A Challenge Problem for AI. *AI Magazine* 18(1): 73–85. dx.doi.org/10.1145/267658. 267738

Kitano, H.; Ghosh, S.; and Matsuoka, Y. 2011. Social Engineering for Virtual 'Big Science' in Systems Biology. *Nature Chemical Biology* 7(6): 323–326. dx.doi.org/10.1038/nchembio.574

Korzybski, A. 1933. S*cience and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Chicago: Institute of General Semantics.

Langley, P., and Simon, H. 1987. *Scientific Discovery: Computational Exploration of the Creative Processes*. Cambridge, MA: The MIT Press.

Lee, K. F. 1988. *Automatic Speech Recognition: The Development of the SPHINX System*. New York: Springer.

Lenat, D., and Brown, J. 1984. Why AM and EURISKO Appear to Work. *Artificial Intelligence* 23(3): 269–294. dx.doi.org/10.1016/0004-3702(84)90016-X

Li, C.; Liakata, M.; and Rebholz-Schuhmann, D. 2014. Biological Network Extraction from Scientific Literature: State of the Art and Challenges. *Brief Bioinform* 15(5): 856–877. dx.doi.org/10.1093/bib/bbt006

Lindsay, R.; Buchanan, B.; Feigenbaum, E.; and Lederberg, J. 1993. DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation. *Artificial Intelligence* 61(2): 209–261. dx.doi.org/10.1016/ 0004-3702(93)90068-M

Matsuoka, Y.; Matsumae, H.; Katoh, M.; Eisfeld, A. J.; Neumann, G.; Hase, T.; Ghosh, S.; Shoemaker, J. E.; Lopes, T.; Watanabe, T.; Watanabe, S.; Fukuyama, S.; Kitano, H.; Kawaoka, Y. 2013. A Comprehensive Map of the Influenza: A Virus Replication Cycle. *BMC Systems Biology* 7: 97(2 October). dx.doi.org/10.1186/1752-0509-7-97

Michalewicz, M.; Poppe, Y.; Wee, T.; and Deng, Y. 2015. InfiniCortex: A Path To Reach Exascale Concurrent Supercomputing Across the Globe Utilising Trans-Continental Infiniband and Galaxy Of Supercomputers. Position Paper Presented at the Third Big Data and Extreme-Scale Computing Workshop (BDEC), Barcelona, Spain, 29–30 January.

Oda, K., and Kitano, H. 2006. A Comprehensive Map of the Toll-Like Receptor Signaling Network. *Molecular Systems Biology* 2: 2006 0015. dx.doi.org/10.1038/msb4100057

Oda, K.; Matsuoka, Y.; Funahashi, A.,; and Kitano, H. 2005. A Comprehensive Pathway Map of Epidermal Growth Factor Receptor Signaling. *Molecular Systems Biology* 1 2005 0010. dx.doi.org/10.1038/msb4100014

Prinz, F.; Schlange, T.; and Asadullah, K. 2011. Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets? *Nature Reviews Drug Discovery* 10(9): 712. dx.doi.org/10.1038/nrd3439-c1

Robinson, P. N. 2012. Deep Phenotyping for Precision Medicine. *Human Mutation* 33(5), 777–780. dx.doi.org/10.1002/humu.22080

Shimada, Y.; Gulli, M. P.; and Peter, M. (2000). Nuclear Sequestration of the Exchange Factor Cdc24 by Far1 Regulates Cell Polarity During Yeast Mating. *Nature Cell Biology* 2(2): 117–124. dx.doi.org/10.1038/35000073

Silver, D.; Huang, A.; Maddison, C.; Guez, A.; Sifre, L.; den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D. 2016. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529(7587): 484-489. dx.doi.org/10.1038/nature16961

Spangler, S.; Wilkins, A.; Bachman, B.; Nagarajan, M.; Dayaram, T.; Haas, P.; Regenbogen, S.; Pickering, C. R.; Corner, A.; Myers, J. N.; Stanoi, I.; Kato, L.; Lelescu, A.; Labire, J. J.; Parikh, N.; Lisewski, A. M.; Donehower, L.; Chen, Y.; Lichtarge, O. 2014. Automated Hypothesis Generation Based on Mining Scientific Literature. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery. dx.doi.org/10.1145/2623330.2623667

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59(236): 433–460. dx.doi.org/10.1093/mind/LIX.236.433

Wicks, P.; Lowe, M.; Gabriel, S.; Sikirica, S.; Sasane, R.; and Arcona, S. 2015). Increasing Patient Participation in Drug Development. *Nature Biotechnology* 33(2): 134–135. dx.doi.org/10.1038/nbt.3145

Wicks, P.; Vaughan, T. E.; Massagli, M. P.; and Heywood, J. 2011. Accelerated Clinical Discovery Using Self-Reported Patient Data Collected Online and a Patient-Matching Algorithm. *Nature Biotechnology* 29(5): 411–414. dx.doi.org/10.1038/nbt.1837

Hiroaki Kitano is director, Sony Computer Science Laboratories, Inc., president of the Systems Biology Insitute, a professor at Okinawa Institute of Science and Technology, and a group director for Laboratory of Disease Systems Modeling at Integrative Medical Sciences Center at RIKEN. Kitano is a founder of RoboCup, received the Computers and Thoughts Award in 1993, and Nature Award for Creative Mentoring in Science in 2009. His current research focuses on systems biology, artifiical intelligence for biomedical scientific discovery, and their applications.