

Artificial Intelligence's Grand Challenges: Past, Present, and Future

Ganesh Mani

■ *Innovative, bold initiatives that capture the imagination of researchers and system builders are often required to spur a field of science or technology forward. A vision for the future of artificial intelligence was laid out by Turing Award winner Raj Reddy in his 1988 Presidential address to the Association for the Advancement of Artificial Intelligence. It is time to provide an accounting of the progress that has been made in the field, over the last three decades, toward the challenge goals. While some tasks such as the world-champion chess machine were accomplished in short order, many others, such as self-replicating systems, require more focus and breakthroughs for completion. A new set of challenges for the current decade is also proposed, spanning the health, wealth, and wisdom spheres.*

Grand challenges are important as they act as compasses for researchers and practitioners alike — especially young professionals — who are pondering worthwhile problems to work on, testing the boundaries of what is possible! Challenge tasks also unleash the competitive spirit in participants as evidenced by the plethora of active participants in Kaggle competitions (and forum discussions therein). Prize money and research bragging rights also accrue to the winners. The Defense Advanced Research Projects Agency Grand Challenges¹ and X prizes² are some of the best-known successful programs that have helped make significant progress across many domains applying artificial intelligence (AI). As grand challenges are accomplished, other than the long-term benefits the solutions engender, the positive press they garner helps rally society behind the field. Trickle-down benefits include renewed respect for and trust in science and technology by citizens, as well as a desirable focus on science, technology, engineering, and mathematics education.

The challenge tasks laid out by Turing Award winner and Carnegie Mellon University professor Raj Reddy in his 1988 Association for the Advancement of Artificial Intelligence (AAAI) Presidential address and published in *AI Magazine* (Reddy 1988) touched upon everyday elements — spanning communication, transportation, and games — plus infrastructure requirements (on earth as well as for space explorations).

The Grand Challenges from 1988: A Retrospective

The challenges, as originally laid out, were for the subsequent thirty years and we now are just over that time period. A summary of the original tasks and their current status is presented in table 1.

World Champion Chess Machine

The achievement of winning the world champion chess machine challenge turned out to be a relatively easy one to accomplish. Within a decade of 1988, the Computer Chess Fredkin Prize, honoring the first program to beat a reigning human world champion, was awarded to the Deep Blue chess machine's designers for successfully defeating Garry Kasparov.³ Campbell et al. (2002) provide a good description of the key success factors: a single-chip chess search engine; massive parallelism for tree traversal; fast and slow evaluation functions; search extensions; and a Grandmaster game database.

Of related note is recent progress with two other games: *Go* and *Poker*. *Go* is a perfect-information game; however, the complexity is high, with 10^{170} possible board configurations. *AlphaGo* (Silver et al. 2016, 2017) was the start of a sequence of superhuman *Go* programs. It used dual deep neural nets: a value network to evaluate board positions, and a policy network to select moves. Citing the rise of AI,⁴ the human *Go* champion, Lee Sedol (who lost four games, but won one to AlphaGo in 2016) recently announced his retirement! *Poker* — an imperfect information game, as other players' cards are hidden — has also seen tremendous advances of late, with machines trouncing over humans (Brown and Sandholm 2019).

Mathematical Discovery

There have been two kinds of advances in the area of mathematical discovery:⁵ numerical explorations that hint at new facts and then are proven rigorously by human mathematicians; and an automated theorem prover (such as the HOList environment described in Bansal et al. 2019).

The sphere-packing problem embodied in the Kepler Conjecture was proven by Hales (2006) with the help of computer-aided techniques. Hales also pointed out that there is an open challenge to build an AI system that can win a gold medal in the International Mathematical Olympiad.⁶

Prizes for ongoing research have been awarded.⁷ While minor discoveries have been made so far in

the process of computer-aided experimental mathematics and theorem proving, discovery of a major result heretofore unknown to human mathematicians will be a significant step.

Translating Telephone

The translating telephone challenge can arguably be deemed complete. The speak-to-translate feature in the Google Translate⁸ app comes close to the intended goal. Using a smartphone's microphone, it allows two people to talk in real-time with the app acting as the interpreter. Google Assistant's⁹ interpreter mode also is a related feature, covering forty-four languages ranging from Arabic to Vietnamese. Microsoft and other companies also have products and services that can permit real-time translation in multiple languages. Facebook AI recently introduced and open-sourced M2M-100,¹⁰ a multilingual machine translation model that can translate between any pair of one-hundred languages without relying on English data.

The accuracy of the various translation offerings is quite reasonable; however, figures of speech (like metaphors) and highly technical content (such as a verbal treatment note from a physician) can still stymie the systems. Likewise, slang usages and acronyms that (especially, young) people use can also be problematic to chatbots. User experience can be another area of focus for future enhancement. On the research side, more attention should be paid to low-density and endangered languages, but otherwise this challenge is nearly complete.

Accident-Avoiding Car

There has been significant progress in this challenge, especially in the last decade, around mobility in general and specifically with intelligent software embodied in vehicles. A significant milestone was accomplished as early as the 1990s, when Carnegie Mellon's NavLab 5¹¹ completed the first coast-to-coast drive in the USA. This was a specially-rigged prototype vehicle, not amenable to facile mass production. An objective Defense Advanced Research Projects Agency Grand Challenge was held in 2004¹² for research teams to showcase autonomous driving and none of the teams finished the route and no winner was declared; however, the very next year (2005) saw five vehicles complete the off-road course spanning one-hundred and thirty-two miles and the first prize of \$2 million was awarded to the Stanford University Research Team for their vehicle Stanley (Carnegie Mellon's vehicles came in second and third). This was followed by an Urban Challenge in 2007,¹³ which involved the vehicles competing in a sixty-mile urban course, merging into and navigating other traffic, while obeying customary traffic rules. Carnegie Mellon's robotized Chevy Tahoe won first place and the \$2 million prize (Urmson et al. 2009).

The research prototype vehicles have paved the way for increasing amounts of automation to be

| Challenge | Status | Comment |
|--------------------------------|---|--|
| Explicit: | | |
| World champion chess machine | Completed. | Deep Blue (IBM, ex-Carnegie Mellon University) Team awarded Fredkin Prize in 1997. |
| Mathematical discovery | Minor discoveries completed. | A major discovery with real-world implications will get people's attention. Some ongoing research and foundational work was recognized with prizes. |
| Translating telephone | Mostly done. | Translation apps, tools (from Google and other vendors) are in widespread, everyday use. |
| Accident-avoiding car | More than half the journey is complete. | A pedestrian fatality in Arizona in an Uber car in 2018 and deaths in Tesla cars employing autopilot have been reported. No consensus yet on safety and ethical criteria. |
| Self-organizing systems | Moderate amounts of progress. | Broader interpretation: Swarm computation, Xenobot-based systems. |
| Self-replicating systems | Modicum of progress. | Needed for Mars colonization, back-up to Silicon Valley, financial exchanges, clearinghouses, and redundant hospital infrastructure (including electronic medical records). Some of the above is taken care of, via the cloud infrastructure, but needs richer capabilities. |
| Implicit: | | |
| Sharing knowledge and know-how | Efficient framework in place, but more features needed (for example, to help focus and to weed out misinformation). | Via Google and other web platforms. Speed of information generation is increasing, while average quality of information is decreasing. Human attention and curation cannot keep pace. |

Table 1. Current Status of the 1988 Grand Challenges.

built into vehicles over the last decade; although we are getting closer to the ultimate goal of fully autonomous driving, we are not quite there yet. The Society of Automotive Engineers,¹⁴ a standards-developing organization, has suggested a classification system ranging from level 0 (fully manual) to level 5 (full automation with the common human-driver controls, such as pedals and a steering wheel, eliminated completely). No mass-produced vehicle has attempted sustained level-5 driving yet.

Reddy had called for an eighty to ninety percent reduction in the automobile-accident fatality rate. According to the Insurance Institute for Highway Safety statistics covering all motor vehicle deaths, over the thirty years spanning 1988 to 2018, the fatality per 100,000 people came down from 15.4 to 11.2, a twenty-seven percent reduction; and, in terms of fatality per 100 million miles traveled, from 2.32 to 1.13, a fifty-one percent reduction. Advanced driver-assist features and electronic stability control are having a positive impact. It should be noted that a number of additional factors, such as the increase in airbags, seat-belt compliance, and fewer alcohol-related fatalities, have also contributed to the improved numbers.

There have also been recent setbacks in the field. For instance, the first pedestrian fatality by a

self-driving car is attributed to the Uber accident in Arizona, in March of 2018. Although various contributing factors ranging from the human overseer in the car being distracted, to improper programming that detected something in its pathway but failed to classify it as a (jaywalking) pedestrian, were involved,¹⁵ the consensus is that more technical or algorithmic improvements will be required to further strengthen the self-driving risk management protocols. Open tasks include programming of answers to moral dilemmas or trade-offs that an autonomous vehicle may face (for example, should it swerve onto the sidewalk with a couple of pedestrians to prevent harm to the car's occupants and perhaps any occupants in the stalled car, directly in front of it?). Awad et al. (2018) provide an analysis of some of the simulated dilemmas and summarize opinions crowdsourced from millions of global citizens.

In summary, the accident-avoiding car, or the intended goal of a responsible, ethical self-driving car remains a challenge, even though significant progress has been made toward it. We seem to have covered more than half the distance on this important journey affecting the future of mobility for much of society.

| | |
|--------|--|
| Health | Nursing home with ninety percent of the resident care being performed by robots and smart infrastructure. |
| | Assistant for patient with dementia (evaluate via performance threshold: example given, caregivers rating it at a ninety percent satisfaction rate or other objective measures). |
| | Wearable device providing reliable alerts (for clinical consult or auto-summoning ambulance/calling 911 based on implied criticality). Advanced versions may provide preliminary diagnosis. |
| Wealth | Thrift assistant that automatically goes through monthly payments (mortgage, auto insurance, and others) and e-negotiates lower payments (for same asset and coverage levels). |
| | Benefits assistant (covering, for example, US Social Security, any basic income promises, healthcare) ensuring quick credits to the end-user wallets (without fraud and overheads) even for people with limited digital infrastructure. Obviates paperwork; efficient push (to citizens) versus bureaucratic pull. |
| | Savings assistant (automatically saving toward certain consumption goals such as college education, retirement, wedding/honeymoon; and alerting, when not tracking desired trajectory). |
| Wisdom | Successfully arguing a case in front of a judge (related thought: Would defending be harder than being a plaintiff's AI counsel?). |
| | Winning the New Yorker Cartoon Caption Contest (multiple times and with explanation). |
| | Information checker (multimedia; with dialog and nuanced explanations). |
| | Explaining the reasoning behind AI system's decisions and arguing that it is being fair and ethical (and hence should be trusted). This could be considered a metachallenge. |

Table 2. New Grand Challenges (for the 2020s).

Self-Organizing Systems

The original goal called for acquiring significant capabilities via perception-mediated learning and discovery. For instance, reading from textbooks is a commonly used mode by which young humans all over the world acquire knowledge. People also learn by observation. Thus, some specific challenge-use cases that were suggested included machine reading of a first-year physics textbook, followed by successfully answering questions covering the material in the book chapters; and assembling an appliance after watching a human mechanic perform the task.

The Aristo project from the Allen Institute for AI (Clark 2019) reports a performance metric of over ninety percent in the New York Regents Eighth-Grade Science Exam. While the vocabulary comprehended is significant, we are still in the realm of non-diagram, multiple-choice questions for that test. Earlier attempts had side-stepped the natural-language processing task by hand-encoding the textual knowledge as well as the questions. Recent advances in language models (such as BERT [Bidirectional Encoder Representations from Transformers]; see Devlin et al. 2019) have continued to help in better organizing knowledge from a textbook, permitting reasoning toward more meaningful question-answering. Deep neural nets and large, pretrained transformer models have also helped with performance on the Winograd schema challenge, a somewhat related task. Kocijan et al. (2020) review the various approaches and benchmark datasets to the challenge, which principally involves pronoun disambiguation in a pair of tricky sentences differing by just one or two words. Similar prior work — on deciphering the harder

questions using commonsense reasoning — includes the advances showcased via the quiz show *Jeopardy!* in 2011, when IBM's computer Watson defeated the human champions Ken Jennings and Brad Rutter. Ferrucci et al. (2010) describe Watson's architecture and some of its algorithmic approaches.

Another important building block with respect to perception-mediated learning and reasoning is the novel object-captioning task. Hu et al. (2020) describe some recent results on a benchmark data set.

Self-organization can also be thought of as emergence of order and efficacy via peer-to-peer interactions, without external or central control. In nature, we see this prominently in ant colonies and bee swarms. Karaboga and Akay (2009) present a survey of algorithms based on the intelligence in bee swarms and their applications.

In a recent development, xenobots (Kriegman et al. 2020) — living machines assembled from cells, informed by suitable simulation on a supercomputer — are amenable to collectible behaviors. Simple group behaviors such as collision between two xenobots forming a temporary mechanical bond and orbiting about each other for several revolutions were observed, *in vivo*, by the authors. It has been suggested that xenobots can be applied to tasks ranging from drug delivery in humans to cleaning up plastics in oceans.

Self-Replicating Systems

Space manufacturing was cited as the motivation for this challenge. Instead of transporting a whole factory, the goal would be to generate almost all the parts needed for the factory using locally available raw materials by simply transporting a minimal

viable set of tools including perhaps some seeding robots. The parts would then be assembled in place to instantiate the comprehensive factory and presumably this process can be repeated at other remote sites.

The US National Aeronautics and Space Administration has announced a Space Robotics Challenge¹⁶ to help develop technologies and architectures toward a lunar in-situ resource utilization mission. The current phase of the challenge is to develop software that will aid a virtual team of robots to navigate the simulated lunar landscape, locate resources and extract, for instance, water (ice), methane, and ammonia. Winners are expected to be announced in late 2021; progress in this avenue is ongoing, albeit slowly.

Sharing Knowledge and Know-how (Implicit Challenge)

The Internet has enabled facile indexing and fast retrieval with widespread sharing of information. News organizations post digital content in real-time and there is a plethora of user-generated content being added every second on social media platforms. This also has introduced new challenges: how to discern the veracity and source authority of a news story, separating facts from opinions, summarizing news stories, and highlighting any unique details a particular news article may provide.

Reddy in his Heidelberg Laureate Lecture in 2019¹⁷ termed the unfinished business in this milieu to be threefold: summarizing media content (such as that from books, talks as well as movies and music); creating an encyclopedia on demand; and providing the right information to the right person at the right time in the right language. Filtering out information that is wrong — or deliberately circulated to mislead — is a related problem that has recently become more critical.

Other Related Accomplishments of Note

A deep learning model was recently used to discover an antibiotic, Halicin, by performing predictions on multiple chemical libraries (Stokes et al. 2020). In the process, the algorithm found that a molecule — structurally different from existing antibiotics — from the Drug Repurposing Hub¹⁸ could potentially exhibit strong activity against a broad range of pathogens. Halicin was tested in vitro and then in vivo in mice, confirming the AI system's prediction.

BenevolentAI,¹⁹ a UK-based company, armed with domain knowledge about 2019-nCoV, searched for previously approved drugs that could help block the viral infection mechanism and suggested baricitinib — a rheumatoid arthritis drug — as having the potential to reduce the virus' ability to infect lung cells (Richardson et al., 2020). Doctors familiar with the drug found it to be a novel, yet reasonable suggestion, and initiated steps toward a formal clinical trial.

Based on all the aforementioned summaries, a reasonable question to ask is why all the challenge tasks from 1988 have not yet been fully accomplished, despite the three-decade span, novel algorithms, and the exponential increase in computing power? One

possibility is the focus on narrow AI — well-defined tasks in a specific domain that are easier to make progress on — as opposed to broader accomplishments spanning multiple domains and exhibiting what humans would term *common sense*. Stone et al. (2016) come to a similar conclusion while describing progress in eight domains ranging from transportation to entertainment, and argue that human-aware AI that enriches life and society in creative ways is the next frontier. Fairness and bias-free implementations are important embedded themes. Rahwan et al. (2019) argue that the interdisciplinary and systematic study of machine behavior can inform better human-machine teaming (which is one immediate approach to overcoming the limitations of narrow AI).

I invited half a dozen thought-leaders with varying vantage points — involved in different aspects of AI, including influencing funding toward the field — to opine and suggest Grand Challenges; their commentaries are featured in the sidebars. Francesca Rossi proposes an AI ethics switch and also astutely observes that many grand challenges are interconnected. Frank Chen and Steve Cross address the theme of human-machine teaming — partly congruent with (Grosz 2012) — while Ken Stanley describes open-endedness as a metachallenge. Tom Kalil emphasizes the need for reskilling and workforce training at scale, as well as healthcare cost-cutting. Vanathi Gopalakrishnan, via her wish list, describes two agents: one parent-like, to help with timely reminders for children; and another for dynamic budgeting in a business setting. Their design and satisfactory development could be considered significant challenges.

I also introduce a new set of potential challenges spanning the health, wealth, and wisdom spheres; progress toward them will require technical accomplishments as well as deliberations around policy implications and societal impact.

AI Grand Challenges for the 2020s

Keeping in mind some of the lessons from the set of incomplete challenges in the previous decades, I propose the following new challenges for the current decade (see table 2 for a summary). Instead of the original challenges slated for 30 years, a shorter time frame is in order given the higher velocity of innovation as well as faster, networked computers aided by the cloud infrastructure. Multiple sources of data and advances in Quantum Computing may also serve as additional catalysts in actualizing some of these challenges sooner than later.

Grand Challenges in the Health Milieu

Old age is a challenge across the world, including in many developed countries; skilled assistance for seniors in their golden years, when they are not able to be fully independent, is in short supply. Seniors will have care needs spanning multiple areas:

Teaming

The AI community has historically fetishized beating or replacing humans. We design AI systems to beat *Go* grandmasters, *Starcraft* teams, and *Texas hold 'em* players. We challenge ourselves to build systems that can replace radiologists, website designers and real-time translators.

While some of these goals seem like the right ones (self-driving cars are the only path I know to get to a zero car-accident fatality future), I would like to propose a set of new AI Grand Challenges with a different design center: namely, making AI + humans = better together. These challenges would shift our design focus from *surpass or replace humans* to a *better together* focus. In other words, how can we best blend machine systems that can consider massive data sets, make accurate predictions, and avoid repeatable cognitive biases (such as preferring people who look or talk like us) with humans who can be creative, empathic, wise, loving, encouraging, and inspiring? To that end:

Education: Humans and AI teachers improve K-12 educational outcomes more than teacher alone or AI alone.

Creativity: Humans and an AI team create an original music video more popular than a human alone or AI alone.

Healthcare: Humans and AI primary care teams deliver better health outcomes along with a more empathic bedside manner than a human doctor alone or an AI system alone.

Justice: Humans and AI judge teams render a set of fairer, less-biased set of judgments, considering the most relevant precedents, than human judges alone or AI judges alone.

– Frank Chen

functional (such as dressing or eating), behavioral (such as modulating actions or moods), cognitive (such as assistance with memory), medical (such as help with catheters or other medical devices), and social (such as interactions with other residents, or with video-calling relatives).

Given the importance of needs in the senior-care sphere, I propose two new challenges covering that domain. The first is a nursing home environment where roughly ninety percent of the care is being performed by robots and devices with smart software, to take care of seniors who are functionally independent and do not have behavioral or cognitive impairment. Specialized medical care (for example, helping with catheters) may require human help or supervision and would constitute the remaining ten percent of the care. At-home care can be considered a special case of this broader challenge.

The second proposed challenge in the senior-care sphere is an assistant for an individual with dementia to help with quotidian activities. This may include reminders for nourishment and nutrition, exercise, personal hygiene, resting, recreation, and communication. The assistant may have varied form factors (one embodiment is a series of audio-video devices in the house) but allows the user to communicate naturally as they would, with a live-in human caregiver. The auto-assistant can

escalate confusing situations to a remote human, who may first attempt to resolve tricky situations via a video call and feasible remote operations. The remote overseer can then, depending on the escalated need, call for medical help or schedule an in-person caregiver visit. Dementia is usually associated with old age, but early onset is possible and a solution for senior care should also be potentially portable for the benefit of the younger cohort. Evaluation of successful completion of these challenges can be tricky but can be based on lack of adverse events as well as skilled, human caregivers scoring the AI assistant above a certain threshold on each of a plurality of task dimensions. Solving this challenge will help scale the scarce expertise of human clinicians and caregivers, as well as improve the quality and trust of overall care.

The third proposed grand challenge in health is a wearable device with reliable alerts. This could be akin to the warning or check lights on an automobile dashboard, primarily meant for the individual to take some action, such as eat a snack with carbohydrates or sugar for a low blood sugar alert, tele-consult a physician, or schedule a face-to-face appointment in the near term. The alerting bot or agent should be able to discern the criticality, auto-dialing an emergency call to 911 or 999 or calling an ambulance, as warranted.

A New Turing Test — The Reddy Test

Although Raj Reddy described why grand challenges were crucial for advancing the field of AI, I believe the research community has shown little enthusiasm for them. Funding agencies often talk about grand challenges, but they have evolved into sponsoring single-investigator, low-risk research. If AI is to advance, as envisioned in programs such as the Defense Advanced Research Projects Agency's AI Next,²⁰ then a new focus on grand challenges is required.

Perhaps the first AI grand challenge was the Imitation Game proposed by Alan Turing.²¹ In this game, two participants, a human and a machine, would be interrogated by an unseen person via a teletype. The objective was to determine which of the pair was human and which was machine. Turing said the test would be passed if the average interrogator would not have "... more than seventy-percent chance of making the right identification after five minutes of questioning."

Although it is a subject of ongoing spirited discussion, we have systems today that are close to or have passed the Turing Test. For example, Jill Watson²² (the AI-based teaching assistant used in the Georgia Tech online Master of Science in computer science program) fooled most of the students in a course who thought it was a human. I see a future, not too far distant, where it is difficult, if not impossible, to distinguish between the AI and the human. Thus, a new test is suggested — the Reddy Test.

Consider how this might work with teams. A high-performance team is one where the team members have trust in each other's abilities, there is shared understanding of both goals and intent, and communication patterns are unambiguous and effective; teams and their members adapt to changing situations, and overall team performance improves with experience. Teams are vital to us in just about every aspect of life. For example, the care team of doctors, nurses, dieticians, and counselors who support a loved one undergoing cancer treatment; the team of investment advisors and staff who manage one's retirement funds; the pilots and air traffic controllers who ensure safe transport; and the government and non-governmental agencies counted upon to help during a crisis such as the recent forest fires in California. We just assume or hope these are high-performance teams. With automated team members that pass the Turing Test, such teams will have a better chance of being high performance!

So, suppose these teams have human and AI-based members. For brevity, I will refer to the latter as AIs. It is suggested that AIs are the secret sauce for ensuring teams are high performance. I see a future where the AIs are not only indistinguishable from humans as suggested by the Imitation Game, but they are, in fact valued for their insights. They would derive these insights via rapid analysis of huge amounts of data in real-time and their uncanny ability to anticipate the need for deep analysis, and then explain the significance of these insights to other team members. In short, AI team members come up with options and insights not conceivable by human team members.

So, I boldly suggest a new kind of Turing Test — the Reddy Test for Teams. One objective is that a given team is assessed to be "high performance" using whatever criteria for high performance seems appropriate in a given domain (for example, pilots and air traffic personnel are able to address an unprecedented situation).²³ The second, and more interesting objective, is not to determine which team member is human or machine, but to identify which team members are AIs! The AI is distinguished not because of its non-human behavior, but because of its superior intelligence.

– Steve Cross

Grand Challenges in the Wealth Milieu

The challenges proposed in this domain have the common theme of money efficiency, behind the scenes, recognizing the inherent tradeoff between time and money. Reducing transaction friction is another goal. For instance, the first proposal of a Thrift Assistant that automatically suggests refinancing of a mortgage or switching to a different auto insurance carrier assumes that the workflow associated with it (such as sending personalized information, getting updated quotes and e-negotiating, or submitting additional documents) will be minimally obtrusive to the human principal. It is an example of a set of tasks that could be done manually every few months by monitoring interest rates and setting alerts for insurance rate changes. However, the time consumed in these tasks may reduce the effective savings. By doing it in the background in an automated fashion, it can be done more frequently, and greater savings may be accrued due to the finer-grained monitoring for rate changes. Event-based triggers and responses usually add value over a calendar-based workflow.

The second wealth-related challenge addresses a pressing need for the population that may not be as digitally savvy as the rest of us. A specific use case is that of a senior drawing US Social Security payments — ensuring that the payment reaches the end-user digital wallet or bank account, without fees and obviating any waste and fraud. It could also apply to basic income promises or gig economy workers, where the AI agent helps ensure that the right amount of monies due has been credited to the beneficiary's account. The agent may elicit relevant information from the user (on the subject of number of hours worked or change in hourly rates, for example) to make the workflow accurate. This can be thought of as a Benefits Assistant.

Personal savings rates in many parts of the world, including the USA, are low. To counter the instant-gratification phenomenon and save for a future need like retirement or a child's education, behavioral economists have suggested automatic mechanisms (such as payroll deduction as a default option). Extending this concept with additional features is what I am proposing as the final challenge in the wealth category. Setting up goals for big-ticket purchases (such as upgrading kitchen appliances) and other large consumption-centric life milestones (for example, weddings and honeymoons) would be enabled as this challenge is addressed using a Savings Assistant. The system will suggest contribution amounts toward each savings bucket (for example, \$x goes toward retirement, \$y toward a bucket-list vacation goal) based on the income and expense profile of the family or individual. Contribution amounts may be overridden, but smart alerts will be provided when not tracking desired savings trajectory to reach the goal with a high probability within the target timeframe.

It is also worth considering combining all three of the aforementioned assistants (thrift, benefits, and savings) into an all-purpose Financial Smart Agent, that can also handle purchases and payments. The agent should be able to comprehend conversational-style input via voice or text (including making sense of any e-mails that may be forwarded to it).

Grand Challenges in the Wisdom Milieu

Three challenges and a metachallenge are proposed under this category, where, broadly, the AI system is playing the role of a knowledge agent and exhibiting what many would call *wise behavior*. The first is a potential legal role, where the task is to advocate for a plaintiff in front of a judge. Acting as counsel for a defendant is a related challenge. Legal reasoning can involve complex interpretation of laws, precedent, and context, including societal expectations. Many of these elements need to be tied to available facts and evidence, in the process of reasoning and constructing persuasive arguments. Often arguments about what the language — of a contract or law — means or should mean is central to a case. Apps like DoNotPay²⁴ (that can help, for instance, with airline flight compensation and disputing parking tickets) are early steps in the direction of legal process automation.

Winning, especially more than once, the *New Yorker* Cartoon Caption Contest²⁵ is a second challenge that is proposed. On being queried, the system should be able to elaborate why the catchphrase is apt and funny, much like a human would explain to a child or colleague from a different culture (who does not fully understand the joke immediately). Humor is considered difficult to precisely describe, quantify, and systematize and so, while subjective, this could be one of the tasks that showcases the breadth and creativity of AI systems in the coming decade.

Today's world, especially our digital environment, is awash in information of questionable quality; misinformation, sometimes propagated by malicious agents, is on the rise. It is getting harder to access reliable guidance to aid even in quotidian tasks, let alone occasional knowledge-intensive problem solving for important issues or crises. Solving the proposed Information Checker challenge will help quickly and robustly ascertain the source authority, vintage, and other attributes of a document or video. It should also permit further interaction based on the initial information nugget, such as follow-up queries or a dialog that can elicit nuanced explanations, guidance, and related media. Good teachers and mentors are a scarce resource, especially in developing economies, where educating youngsters is or should be one of the highest national priorities. The information checker can assist many people who may not have easy access to a guru with ready answers to a nuanced query.

Useful Agents

A common question that I am asked is whether I consider AI safe for our human race. Our AI community must find ways to communicate the state of our technology truthfully and aligned with reality. Humans are yet to agree on a definition for commonly used words such as *intelligence* and therefore, I first offer my definition and then discuss our capability to develop general AI. I define *intelligence* from an agent point of view as: *Intelligence is clear thinking aligned with natural laws, using multidimensional, multimodal perception that is transformed into decisions of how and when to act.* Clear thinking employs reasoning that is unbiased, and critically examines underlying assumptions and human emotions or beliefs. By defining intelligence thus, I posit that unless uncertainties regarding knowledge about natural laws can be encoded, along with their validity within contextual applications, it is unlikely that we can develop general AI agents without a human in the loop. Below, I list two AI agents that we could develop, test, and use — these constitute grand challenges, as they require integration of different abilities to achieve their goals.

Madre would be a parentlike AI agent. Children, especially at a young age, rely on their parents or caregivers to keep track of their must-do's for each day and to remind them of the same in a timely fashion. Many of these agenda items are day-to-day tasks, and *Madre*, the Parent-like AI Agent, will need to learn the personal calendars of every child, recognize them by voice or otherwise monitor them via sensor feeds, and issue timely reminders of major action items. For example, a child may need to be reminded to brush their teeth at bedtime every day. The child may have to be present at a soccer game every Tuesday during the spring season. *Madre* should automatically monitor the local weather report and provide advice regarding whether, for example, the kids should check with their coaches to find out if the game is still on. There can be many special variations of *Madre* to include cultural preferences for communicating, planning meals, helping choose outfits, and similar tasks. *Madre* can be evaluated by parents and children using survey tools. Evaluation measures to rate *Madre* for successfully performing tasks that result in kids accomplishing parts of their to-do lists over certain time periods such as a week can be compared against parents doing the same, from various households, which would be used as control data. Consistency and efficiency achieved by *Madre* or similar parent-like agents can be used to measure success in AI's abilities to achieve vision or sensor-based monitoring, effective use of real-time information, and natural language communication. (Nothing should be made of the *Madre* name; it could be *Padre* or have a gender-agnostic label; the focus should be on the functionality.)

Diya is an AI agent for dynamic scenario and budget forecast planning. I strongly believe that it is time for static budgeting that happens each fiscal year to be evaluated and modified due to its undesirable influence on any unit's spending habits, especially when sufficient levels of financial stability exist within the higher-level organization. The focus should be on policy related to financial matters, and how the guidelines can be implemented in a dynamic, ongoing fashion. Hard budgets can lead to undesirable spending and creation of wants that are not necessarily aligned with our needs related to business, family, or social projects. Moreover, emergencies such as the ongoing novel coronavirus pandemic, demonstrate the need for flexible and efficient budget reallocation to handle and monitor unanticipated spending. The development of *Diya*, an AI agent for dynamic scenario and budget forecast planning to continuously monitor expenditure reports using fuzzy rules that encode policies, should provide anytime support to businesses, non-profits, and corporations to better use their resources instead of spending significant amounts of time each year for planning and replanning. *Diya's* evaluation can be based on the number of human hours saved and how well it calibrates itself via dynamic reallocation to yield reasonably accurate budgeting functions across various levels of an organization. Integration between secure financial systems such as payroll processing and billing offices within the organization will need to be accomplished. *Diya* could aggregate financial information needed for planning and budgeting offices via use of dashboards. The human-machine interactions needed to successfully develop and test AI agents such as *Diya*, would draw upon and inform foundational research in user interfaces, cybersecurity applications, financial operations, law, policy, and strategic planning across various levels within an organization.

– Vanathi Gopalakrishnan

AI Ethics

Grand challenges can be very inspirational for researchers and practitioners. Often the path to the result is more important than the result itself. Even before the challenge is achieved, many new techniques, methodologies, and general lessons can be derived; and these can be reused or adapted in other contexts, leading to advancements toward other challenges as well. So, I am definitely in favor of AI grand challenges, and I would like to define one in the area of AI ethics.

AI ethics is a multidisciplinary field of study that identifies issues in the pervasive deployment of AI in our life that could lead to undesired and negative outcomes, and defines technical as well as non-technical solutions for such issues. Examples of AI ethics issues are those relating to fairness, transparency, explainability, privacy, accountability, human dignity, and agency, as well as impact on jobs and society. Technical solutions can be novel algorithms to detect and mitigate bias; to derive explanations from an AI model; and be toolkits to help developers revise their AI pipeline to include new processes addressing AI ethics. Non-technical solutions can be guidelines, principles, policies, standards, certifications, incentives, and laws.

Many AI researchers have devised techniques to make an AI system compliant to some ethics directive (such as not passing a threshold in testing for a certain notion of bias). However, this check is usually done by humans, and during the development phase of an AI system. Once the system is deployed, its behavior can possibly evolve as new data are ingested. We can only recheck it by employing the same testing procedure we used during development.

I would like to see AI systems that can recognize when their behavior goes outside certain AI ethics boundaries defined in the design stage; and, if that occurs, they alert humans or switch themselves off. Many parts of this challenge statement are still not clear and thus require research work to be clarified and resolved. For example, how to define the ethical boundaries in a clear but flexible way, so it can be adapted depending on the context? Also, how to provide AI systems with the introspection capability to recognize that it is likely going out of this boundary, either through the current action or through a sequences of actions starting with the current one? And finally, how to embed such an AI ethics switch module in an AI system so that it cannot be tampered with, by the system itself or by others?

This challenge also covers the case of AI systems that work in collaboration with, or in support of human beings, and not in isolated autonomy. In this case, the human-machine team should be considered as a whole, and the AI system should be able to evaluate not just its own behavior but also the behavior of the other human members on its team. Thus, the AI ethics switch should activate when some member of the team, or a group of them, leads the whole team outside the ethics boundaries. Moreover, in this scenario, the AI boundary itself could evolve over time, because the human beings could decide to modify their normative and ethics constraints.

By achieving this challenge, we will be able to trust that the AI systems we use behave within the agreed-upon AI ethics limitations and help humans comply as well. While working toward this challenge, I expect that many other metachallenges will need to be addressed, such as how to significantly advance AI's capability to learn from data; reason with knowledge; understand causality; be able to generalize and abstract; and robustly adapt to new environments.

Grand challenges are not isolated from each other. Working on one will bring new insights for many other ones!

– Francesca Rossi

Economic Outcomes

I have previously described in other forums a number of ideas for maximizing the economic and social benefits of AI (AI for Good).²⁶ Two significant challenges are excerpted and highlighted here: AI for the workforce, and AI in healthcare.

AI and Workforce: The goal is to increase the wages of non-college-educated workers (or unemployed or under-employed veterans) by \$10,000 in six months or less, by enabling them to master a skill that is a ticket to a middle-class job. Of course, if advanced training technologies dramatically increased the supply of workers with a given skill, it might reduce the wages for workers with this skill; therefore, the target areas need to be chosen carefully. For example, the Defense Advanced Research Projects Agency's Education Dominance program created an AI-based digital tutor that is allowing new Navy recruits with a high school degree to outperform Navy information-technology technicians with seven to ten years of experience, using both written exams and the ability to solve real-world trouble tickets.

The impact of AI for accelerated training would be increased with other types of innovation. For example, firms could collaborate to identify critical skills, sponsor the development of competency-based assessments that are accurate predictors of on-the-job-performance, share these assessments with training providers, and embrace hiring based on skills as opposed to credentials. Training providers could offer a pay-for-success model, where they are paid based on an increase in future earnings of a worker.

AI and Healthcare: Reduce error rates in medical diagnostics by eighty percent.

Identify X medical conditions, where it is possible to both improve health outcomes and lower costs by at least five to ten billion dollars each.

– Thomas Kalil

A high level of emotional intelligence is attributed to many of us who are able to look in the mirror, accurately rating and judging ourselves regarding our strengths and (especially) weaknesses. Such self-reflection or self-realization is posited as a challenge to cap off this category — a system that can reason about and rate itself on whether it is being unbiased, ethical, and exhibiting good judgment. A bonus, certainly, if it can persuade humans in these important dimensions!

Raj Reddy, while commenting on the challenges, pointed out the need for the resulting solutions to result in tangible value to the common person, especially somebody at the bottom of the pyramid. He evoked the Hindi phrase of “Roti, Kapada aur Makaan,” which translates to bread, clothing, and housing, spanning basic needs. With many policy-makers, across the world discussing universal basic income recently, an interesting goal would be for AI to help efficiently satisfy this triad of needs within the budget envelope of the basic income amount!

One important point to note is that the new, proposed grand challenges strike a balance between being specific enough for researchers and system developers to tackle via pointed interim milestones

and final goals, and being broadly useful for a wide variety of other tasks. For instance, a smarter wearable device for health monitoring can be used with modifications to track and ensure the safety of a young child. Likewise, developing a system advocating legal arguments will also help design a system to provide business strategy advice (ingesting a different set of data and facts).

Of related note is the set of grand challenges proposed for AI and education by Woolf et al. (2013) in *AI Magazine*. Mentors for every learner, instilling twenty-first century skills (such as critical thinking, presentation training, and active listening), interaction data to support learning, access to global classrooms, and lifelong learning, were mentioned as the challenge goals. The authors provided a vision and brief research agenda for each goal.

A special issue of *AI Magazine* published in 2016 discussed specific tests and some of the challenges inherent in constructing robust, valid, and reliable tests for advancing the field of AI (Marcus et al. 2016). Shieber (2016) has also suggested holding competitions when reasonable entrants exist and has laid out criteria for inducement prizes. These include reasonable but absolute (as opposed to relative) milestones

Open-Endedness

Even with all the progress in AI, there is still something enigmatic about human intelligence that seems to defy mechanization. Indeed, whenever intelligence is successfully formalized for a particular purpose, like beating humans at a popular game, it still seems mechanical compared with the natural fluidity of human intellect. Admittedly this impression is subjective, but I think I know why this mechanical feeling continues to prevail: The real power of our intellect is not in our ability to learn, but rather to continually imagine something new. We could call it *creativity*, but that word does not do justice to the profoundness of this generative capacity. The entire history of civilization, all of human invention, all of science, and all of art is the product of our intelligence. This elusive gift within our species is not the ability to invent one thing or to solve one problem, but to invent everything. Even more than that, our gift is to retain perpetually the potential for this process to keep climbing higher with ever more complex inventions, ideas, and creations. To crystallize this notion in a single term, the fundamental property that vastly separates our intelligence from every attempt so far is that our intelligence is *open-ended*. Open-endedness is the ability to invent without end — not only to solve problems, but to invent the problems too. It is the never-ending algorithm, a process of creation that, once sparked, explodes forever. It is worth noting that there is one other example of open-endedness in nature that unfolded without human input — the evolution of all the forms of life in nature — an unfolding process of more than a billion years that happened also to produce humans. In this way, interestingly, not only do we ourselves exhibit open-endedness in our cognition, but we are also a product of it.

In my view, there is one grand challenge for AI that sits above all others, which is to achieve open-endedness. We need to learn how to write algorithms that become more interesting the longer they run, whose discoveries increase in complexity and sophistication as they go, that diverge into innumerable fascinating concepts rather than converging to a solution, and that invent not only solutions to problems, but also the problems themselves. We need these algorithms for two complementary reasons: open-endedness may be the only viable path to discovering architectures that manifest human-level intelligence (as it was in nature); and open-endedness is perhaps the most fundamental aspect of the human mind that gives it its unique unbounded character. The longer we are preoccupied with problems and benchmarks, the longer we will be distracted from this true prize. Who will write the first never-ending algorithm that would be worth running for thousands or even billions of years?

Because the field of open-endedness is not yet well known within AI, I coauthored an introductory article for people interested in learning more.³⁰

– Ken Stanley

and flexibility in the interpretation of the rules. Transparency and replicability are also good goals that can enable robust progress building on prior achievements. Shieber's framework can perhaps be used to elaborate some of the aforementioned Grand Challenges along with a partner organization that may want to sponsor a prize.

Projects and efforts from other organizations complement some of the discussion herein. For instance, AI Impacts²⁷ depicts the trajectory of certain AI achievements and helps readers understand the effects of human-level AI. The Leverhulme Centre for the Future of Intelligence researchers

have assembled several resources²⁸ at the intersection of intelligent technologies and policy issues. Organizations like MLPerf.org and Benchmarks.ai are creating different sets of hardware, software, and dataset benchmarks to objectively measure AI performance along certain dimensions. The AI Index report²⁹ has statistics and information relating to research publication volumes (broken down, for example, by country and gender), patent volumes, technical performance (such as for tasks ranging from visual-question answering to specialist-level detection of diabetic retinopathy), and economic impact (such as industry adoption or startup investment activity).

Conclusion

In closing, conjuring up good problems whose solutions will extend the frontiers of knowledge and aid society is an important step in any field. Challenge problems need to be relatively unrestricted so as not to stifle creativity and innovation. As initial research gets completed, replicability becomes an important dimension. Then, as the research gets translated and deployed in society — via products and services supporting, for instance, learning, health, and wellness; work and play — human factors and ethics come into focus. Society should be cognizant of the broader implications of technologies, especially ones that can have surprising and indelible effects on individuals or our environment. *AI is fire* is perhaps the apt metaphor. Along with the promise of inordinate benefits, come the dangers of unintended consequences and misuse with malicious intent.

Post-Pandemic Notes

Many of the ideas discussed herein had their origins in 2019 and the frame of reference was the pre-Coronavirus-19 world. In revisiting the proposed challenges with a new lens before finalizing this article, a few observations are in order.

Surprisingly, the challenges outlined broadly seem to be even more relevant in light of the pandemic. Solutions to the challenges would have helped partly cope with the crisis; for instance, having robots in nursing homes would probably have reduced infection rates and brought down the fatality numbers. The unfinished business that Reddy alluded to — of providing the right information at the right granularity to the right person, at the right time — would be immensely valuable to a clinician dealing with a new disease such as Coronavirus-19 (under fog-of-war conditions). We are witnessing the rapid publication of papers and revisions of treatment protocols; and physicians are struggling to keep up with the information flow, especially under resource and infrastructure constraints. The information-checker challenge task under the wisdom category should address this need (Mani and Hope 2020). A solution to the benefits assistant challenge in the wealth category, for instance, would have helped disburse the government stimulus payments quicker and more efficiently.

The algorithms and building blocks that constitute the solutions to the outlined challenges will help prepare us for the next strategic surprise as well as make progress toward the sustainable development goals recommended by the United Nations.

Finally, the end goal is to make AI more impact-aware and human-centric — the ability for systems to work in the background, independently to make life easier in society; and to team with people and other machines, especially in exceptional and unusual scenarios (to reduce risk). Steve Jobs, decades ago, waxed

eloquent on the amplification of human ability by referring to the efficiency of man with a bicycle³¹ in the context of personal computers; it is particularly relevant today in the context of human-AI teaming and collective intelligence. The mantra should be, of the people, by the people with machines, for the people!

Acknowledgments

I am grateful for my initial conversations with Raj Reddy and his pointed input and guidance. Ashok Goel's enthusiasm was key to the project and its unique format. The critical reviewers suggested additional pointers and their comments helped hone the manuscript. I am also grateful for the sidebar contributions of Francesca Rossi, Frank Chen, Steve Cross, Ken Stanley, Thomas Kalil, and Vanathi Gopalakrishnan. I dedicate this article to the memory of AI pioneer Jaime Carbonell, one of my mentors and long-term collaborators.

Notes

1. en.wikipedia.org/wiki/DARPA_Grand_Challenge
2. www.xprize.org
3. www.scs.cmu.edu/link/then-and-now-2
4. inews.co.uk/news/go-champion-lee-sedol-retires-admitting-ai-cannot-be-defeated
5. mathscholar.org/2019/04/google-ai-system-proves-over-1200-mathematical-theorems
6. github.com/IMO-grand-challenge/IMO-grand-challenge.github.io
7. www.ams.org/profession/prizes-awards/ams-supported/atp-prizes
8. translate.google.com
9. assistant.google.com
10. github.com/pytorch/fairseq/tree/master/examples/m2m_100
11. www.cs.cmu.edu/~tjochem/nhaa/navlab5_details.html
12. www.darpa.mil/news-events/2014-03-13
13. www.grandchallenge.org
14. SAE International, www.sae.org
15. www.nts.gov/investigations/AccidentReports/Reports/HAR1903.pdf
16. www.nasa.gov/directorates/spacetech/centennial_challenges/space_robotics/index.html
17. www.heidelberg-laureate-forum.org/video/lecture-grand-challenges-in-ai-unfinished-agenda.html
18. www.broadinstitute.org/drug-repurposing-hub
19. www.benevolent.com
20. www.darpa.mil
21. plato.stanford.edu/entries/turing-test
22. www.youtube.com/watch?v=WbCguICyFTA
23. See McMillan, P. 2001 *The Performance Factor: Unlocking the Secrets of Teamwork*. Nashville, TN: Broadman & Holman. The book presents a fascinating scenario on how a team handled an unprecedented aircraft emergency.
24. The World's First Robot Lawyer: DoNotPay, Donotpay.com

25. The New Yorker, www.newyorker.com/cartoons/contest
26. AI for Good, <https://cra.org/crc/wp-content/uploads/sites/2/2017/04/aiforgood-032917.pdf>
27. aiimpacts.org/about
28. www.lcfi.ac.uk/projects
29. hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf
30. www.oreilly.com/radar/open-endedness-the-last-grand-challenge-youve-never-heard-of
31. www.youtube.com/watch?v=0lvMgMrNDIlg&feature=emb_logo (relevant excerpt starts at 5:20).

References

- Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.-F.; and Rahwan, I. 2018. The Moral Machine Experiment. *Nature* 563(7729): 59–64. doi.org/10.1038/s41586-018-0637-6
- Bansal, K.; Loos, S.; Rabe, M.; Szegedy, C.; and Wilcox, S. 2019. HOList: An Environment for Machine Learning of Higher Order Logic Theorem Proving. *Proceedings of Machine Learning Research* 97: 454–63.
- Brown, N., and Sandholm, T. 2019. Superhuman AI for Multiplayer Poker. *Science* 365(6456): 885–90. doi.org/10.1126/science.aay2400
- Campbell, M.; Hoane, A. J. Jr.; and Hsu, F. H. 2002. Deep Blue. *Artificial Intelligence* 134(1–2): 57–83. doi.org/10.1016/S0004-3702(01)00129-1
- Clark, P. 2019. Project Aristo: Towards Machines that Capture and Reason with Science Knowledge. In *10th International Conference on Knowledge Capture (K-CAP 2019)*. New York: Association for Computing Machinery.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics.
- Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A. A.; Lally, A.; Murdock, J. W.; Nyberg, E.; Prager, J.; Schlaefer, N.; and Welty, C. 2010. Building Watson: An Overview of the DeepQA Project. *AI Magazine* 31(3): 59–79. doi.org/10.1609/aimag.v31i3.2303
- Grosz, B. 2012. What Question Would Turing Pose Today? *AI Magazine* 33(4): 73. doi.org/10.1609/aimag.v33i4.2441
- Hales, T. C. 2006. Historical Overview of the Kepler Conjecture. *Discrete & Computational Geometry* 36(1): 5–20. doi.org/10.1007/s00454-005-1210-2
- Hu, X.; Yin, X.; Lin, K.; Wang, L.; Zhang, L.; Gao, J.; and Liu, Z. 2020. VIVO: Surpassing Human Performance in Novel Object Captioning with Visual Vocabulary Pre-Training. arxiv preprint. arXiv:2009.13682 [cs.CV]. Ithaca, NY: Cornell University Library.
- Karaboga, D., and Akay, B. 2009. A Survey: Algorithms Simulating Bee Swarm Intelligence. *Artificial Intelligence Review* 31(1–4): 61–85. doi.org/10.1007/s10462-009-9127-4
- Kocijan, V.; Lukasiewicz, T.; Davis, E.; Marcus, G.; and Morgenstern, L. 2020. A Review of Winograd Schema Challenge Datasets and Approaches. arxiv preprint. arXiv:2004.13831[cs.CL]. Ithaca, NY: Cornell University Library.
- Kriegman, S.; Blackiston, D.; Levin, M.; and Bongard, J. 2020. A Scalable Pipeline for Designing Reconfigurable Organisms. *Proceedings of the National Academy of Sciences of the United States of America* 117(4): 1853–9. doi.org/10.1073/pnas.1910837117
- Mani, G., and Hope, T. 2020. Viral Science: Masks, Speed Bumps, and Guard Rails. *Patterns* 1(6): 100101. doi.org/10.1016/j.patter.2020.100101
- Marcus, G.; Rossi, F.; and Veloso, M. 2016. Beyond the Turing Test. *AI Magazine* 37(1): 3–4. doi.org/10.1609/aimag.v37i1.2650
- Rahwan, I.; Cebrian, M.; Obradovich, N.; Bongard, J.; Bonnefon, J. F.; Breazeal, C.; Crandall, J. W.; Christakis, N. A.; Couzin, I. D.; Jackson, M. O.; Jennings, N. R.; Kamar, E.; Kloumann, I. M.; Larochelle, H.; Lazer, D.; McElreath, R.; Mislove, A.; Parker, D. C.; Pentland, A.; Roberts, M. E.; Shariff, A.; Tenenbaum, J. B.; and Wellman, M. 2019. Machine Behaviour. *Nature* 568(7753): 477–86. doi.org/10.1038/s41586-019-1138-y
- Reddy, R. 1988. Foundations and Grand Challenges of Artificial Intelligence: AAAI Presidential Address. *AI Magazine* 9(4): 9. doi.org/10.1609/aimag.v9i4.950
- Richardson, P.; Griffin, I.; Tucker, C.; Smith, D.; Oechsle, O.; Phelan, A.; Rawling, M.; Savory, E.; and Stebbing, J. 2020. Baricitinib as Potential Treatment for 2019-nCoV Acute Respiratory Disease. *The Lancet, Correspondence*, 395(10223): E30–E31. doi.org/10.1016/S0140-6736(20)30304-4
- Shieber, S. M. 2016. Principles for Designing an AI Competition, or Why the Turing Test Fails as an Inducement Prize. *AI Magazine* 37(1): 91–6. doi.org/10.1609/aimag.v37i1.2646
- Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, L.; Kavukcuoglu, K.; Graepel, T.; and Hassabis, D. 2016. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529(484–489). doi.org/10.1038/nature16961
- Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; Chen, Y.; Lillicrap, T.; Hui, F.; Sifre, L.; van den Driessche, G.; Graepel, T.; and Hassabis, D. 2017. Mastering the Game of Go without Human Knowledge. *Nature* 550(7676): 354–9. doi.org/10.1038/nature24270
- Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; and Collins, J. J. 2020. A Deep Learning Approach to Antibiotic Discovery. *Cell* 180(4): 688–702.e13.
- Stone, P.; Brooks, R.; Brynjolfsson, E.; Calo, R.; Etzioni, O.; Hager, G.; and Leyton-Brown, K. 2016. *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel*. Stanford, CA: Stanford University. ai100.stanford.edu/2016-report
- Urmson, C.; Baker, C.; Dolan, J.; Rybski, P.; Salesky, B.; Whittaker, W.; Ferguson, D.; and Darms, M. 2009. Autonomous Driving in Traffic: Boss and the Urban Challenge. *AI Magazine* 30(2): 17. doi.org/10.1609/aimag.v30i2.2238
- Woolf, B. P.; Lane, H. C.; Chaudhri, V. K.; and Kolodner, J. L. 2013. AI Grand Challenges for Education. *AI Magazine* 34(4): 66–84. doi.org/10.1609/aimag.v34i4.2490

Ganesh Mani is an adjunct faculty member at Carnegie Mellon University. He cofounded one of the earliest firms

454 × 510 pt

(acquired by SSgA) that applied machine learning to investment management. He has been involved in numerous other academic and industry innovation projects utilizing AI. He is past-president and current board member of TiE.org's Pittsburgh chapter.

Frank Chen contributed a commentary to this article. He is an operating partner at the venture capital firm Andreessen Horowitz.

Steve Cross contributed a commentary to this article. He is an independent consultant and a retired Georgia Institute of Technology faculty member.

Vanathi Gopalakrishnan contributed a commentary to this article. She is an associate professor and director of the Intelligent Systems Program at the University of Pittsburgh.

Thomas Kalil contributed a commentary to this article. He is chief innovation officer at Schmidt Futures.

Francesca Rossi contributed a commentary to this article. She is president-elect of the Association for the Advancement of Artificial Intelligence, and an IBM fellow.

Ken Stanley contributed a commentary to this article. He is a research manager at OpenAI.