CUTTER Business Technology Journal

Management, Innovation, Transformation

Vol. 34, No. 5, 2021

Navigating the Prospects and Perils of AI

Michael Eiden
Guest Editor

CUTTER Business Technology Journal

As business models for creating value continue to shift, new business strategies are constantly emerging and digital innovation has become an ongoing imperative. *Cutter Business Technology Journal* delivers a comprehensive treatment of these strategies to help your organization address and capitalize on the opportunities of this digital age.

Cutter Business Technology Journal is unlike academic journals. Each monthly issue, led by an expert Guest Editor, includes five to seven substantial articles, case studies, research findings, and/or experience-based opinion pieces that provide innovative ideas and solutions to the challenges business technology professionals face right now — and prepares them for those they might face tomorrow. *Cutter Business Technology Journal* doesn't water down or delay its content with lengthy peer reviews. Written by internationally known thought leaders, academics, and practitioners — you can be certain you're getting the uncensored perspectives of global experts.

You'll benefit from strategic insight on how the latest movements in digital innovation and transformation, artificial intelligence/machine learning, Internet of Things, blockchain, analytics, and cloud, to name a few, are changing the business landscape for both new and established organizations and how cutting-edge approaches in technology leadership, enterprise agility, software engineering, and business architecture can help your organization optimize its performance and transition to these new business models.

As a subscriber, you'll also receive the *Cutter Business Technology Advisor* — a weekly bulletin featuring industry updates delivered straight to your inbox. Armed with expert insight, data, and advice, you'll be able to leverage the latest business management thinking to achieve your organization's goals.

No other journal brings together so many thought leaders or lets them speak so bluntly — bringing you frank, honest accounts of what works, what doesn't, and why. Subscribers have even referred to *Cutter Business Technology Journal* as a consultancy in print and likened each month's issue to the impassioned discussions they participate in at the end of a day at a conference!

Get the best in thought leadership and keep pace with the technologies and business models that will give you a competitive edge — subscribe to *Cutter Business Technology Journal* today!

Founding Editor: Ed Yourdon
Publisher: Karen Fine Coburn
Group Publisher: Christine Generali
Production Manager: Linda Dias
Editors: Jennifer Flaxman, Tara K. Meads
Client Services: service@cutter.com

***Cutter Business Technology Journal®** is published monthly by Cutter Consortium, an Arthur D. Little company, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA • Tel: +1 781 648 8700 • Fax: +1 781 648 8707 • Email: cbtjeditorial@cutter.com • Website: www.cutter.com • Twitter: @cuttertweets • Facebook: Cutter Consortium. ISSN: 2475-3718 (print); 2475-3742 (online).*

Subscription rates are US $485 a year in North America, US $585 elsewhere, payable to Cutter Consortium. Reprints, bulk purchases, past issues, and multiple subscription and site license rates are available on request.

NOT FOR DISTRIBUTION
For authorized use, contact Cutter Consortium +1 781 648 8700 or service@cutter.com.

☐ Start my print subscription to *Cutter Business Technology Journal* ($485/year; US $585 outside North America).

Name _____ Title _____

Company Address _____

City _____ State/Province _____ ZIP/Postal Code _____

Email (Be sure to include for weekly *Cutter Business Technology Advisor*) _____

Fax to +1 781 648 8707, call +1 781 648 8700, or send email to service@cutter.com.
Mail to Cutter Consortium, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA.

Request Online License Subscription Rates

For subscription rates for online licenses, email or call: **sales@cutter.com** or **+1 781 648 8700.**

CUTTER CONSORTIUM
Access to the Experts

Opening Statement

by Michael Eiden, Guest Editor

Artificial intelligence (AI) as a concept has been with us for decades now, but how much of a reality is it? Over the past 10 years in particular, AI has been both promoted as the solution to a multitude of problems and decried as a threat to jobs and human potential. But what do we even mean by AI? Is it destined to be a universal panacea or the ultimate disruptor of society?

I have no doubt that AI is here to stay, especially given the sheer amount of progress we have made in research over the last decade. However, AI is also in danger of being reduced to a buzzword, and of being set up to fail in real-world business settings.

If we were to compare the AI revolution with the Industrial Revolution, then we're only a few years on from the invention of the steam engine. One major challenge we face is that there are few widely recognized standards or procedures for the development and lifecycle management of AI models. If AI is to be taken seriously — and be genuinely deemed useful — it needs to be developed using best-practice methodologies. This starts with the initial selection and quality control of the input data, followed by the selection of the appropriate algorithmic concept; the parametrization and architecture; the training, validation, and deployment; and, finally, the monitoring and safeguarding processes once put into production. Poorly developed AI applications that fail at any of these stages can infringe on privacy or contain unintended bias — and with AI being used in myriad sensitive scenarios as diverse as credit references and medical diagnosis, this can have very serious real-world consequences.

It's Still Early

Although tremendous progress has been made in the field of AI research, we are still at the beginning stages when it comes to bringing AI into real-world business applications. Deep learning has produced models with superhuman performance in certain fields of application, such as natural language processing and computer vision, but is it really the go-to approach to solve real-world business problems? It obviously depends on the circumstances, but in many business contexts, extremely large data sets aren't always available. On the contrary, data sets are often small in size and sparsely populated.

If AI is to be taken seriously — and be genuinely deemed useful — it needs to be developed using best-practice methodologies.

Even more importantly, data sets are quite often intrinsically incomplete (i.e., they only partially describe a more complex system in the real world that isn't fully observable). Deep learning approaches are clearly not applicable here. Instead, probabilistic machine learning (ML) coupled with reinforcement learning provides a more promising approach in such settings, particularly where transparency is also a key requirement.

Another concept generating more and more traction in dealing with complex real-world problems is coupling ML to graph representations. Graphs are highly scalable, fully transparent, and human-readable representations of systems of interest. This approach allows for easier human intervention and interrogation — and also facilitates an elegant expansion to different use cases as well as the straightforward integration of additional data sets.

More importantly, though, it also means that domain experts can be brought into the loop when models are actually being built. While rare at the moment, it is vital that this becomes a standard industry practice going forward, not only to improve the efficiency and accuracy of AI systems, but also to increase trust in such systems from the people who use them. For models to actually do the job in which they were intended, they must be subject to domain expert scrutiny before being

released, and they need to be regularly reassessed as new information emerges.

Going Forward

How will AI transform the jobs we do? It's difficult to say with any precision, but my vision is of a future where AI helps to make smart people smarter. It's not about taking agency away from experts; rather, it's about augmenting the decision-making process, particularly when those decisions require the consideration of multiple dimensions. As the COVID-19 crisis has amply demonstrated, even the cleverest humans aren't always good at thinking exponentially and identifying the right course of action when so many different factors are in play. In such scenarios, AI could help experts see the bigger picture and react to it more quickly.

> *Data science has to work in tandem with "data storytelling."*

The immediate future of AI isn't about building entirely autonomous systems, but about this type of augmentation. It also makes the adaptation process for the end user easier, as inevitably there will be differing levels of resistance to using AI in the workplace. Importantly, data science has to work in tandem with "data storytelling" (i.e., illustrating how data is used to generate AI models as a part of the overall solution approach so that end users clearly understand what a model can and can't do). True creative thinking, judgement, and the ownership of ideas are still very much in the human sphere of activity. AI is merely there to perform repetitive tasks and ensure that all possibilities are covered as decisions are made.

As a fast-evolving area, AI presents innumerable opportunities and applications that we haven't even imagined yet. In this issue of *Cutter Business Technology Journal* (*CBTJ*), we discuss the current factors and considerations surrounding AI today and take a look at where trends might be heading in the future.

In This Issue

We begin with Cutter Consortium Senior Consultant Paul Clermont diving straight into the three overarching issues related to AI. The first is *unintended consequences* like erosion of human skills and the scope expansion that takes us from reconnecting with old friends online to channels that broadcast "un-fact-checked 'news.'" The second is *unintended bias*, especially for systems that could have life-changing consequences (think business loans, hiring, college admissions, and bail-setting). The third is *privacy*, which AI's "hunger for data" takes to a new level as evidenced by deepfake photos used by trolls and recent misuse of facial recognition technology by Everalbum. Clermont offers no-nonsense advice for dealing with these issues, advocating for laws that make organizations responsible for the algorithms they use (whether bought or built) and prohibit unexplainable AI in applications that could harm people physically or affect their lives in significant ways.

AI has an important role to play in product development, in particular, says Michael Jastram in our second article. He outlines the four trends driving product complexity and explains how AI has the potential to help us overcome the limitations of current development approaches. Both systems engineering and Agile struggle to keep up with today's exponential growth in complexity. Model-based systems engineering (MBSE) was built to address complexity but requires a large up-front investment and frequently meets with cultural resistance. Jastram advocates for AI-based solutions that offer some of the benefits of MBSE without the need for long, expensive training processes. Regardless of the exact path, he's excited for the coming years, saying ready-to-use solutions like IBM's Watson barely scratch the surface of what's possible.

Upcoming Topics

Analytics Value Now! Proven Approaches for Immediate Business Value
Dave Cherry

Leadership Lessons: Keys to Thriving in the Post-Pandemic Business World
Noah Barsky

Why Do Technology Projects Fail?
Steve Andriole

Next up, Cutter Consortium Senior Consultant Claude Baudoin and Clayton Pummill give us the keys to achieving trust in AI. The first step is building cross-disciplinary teams, including psychologists, ethicists, sociologists, spiritual leaders, and legislators to develop solid AI policies. Then we must impart AI with emotional intelligence, which involves not only transparency, but also explainability and accountability. Eliminating bias and ensuring fairness must, of course, be in the mix. Baudoin and Pummill explore critical details, like what is the modern equivalent of Isaac Asimov's three laws of robotics[1] and should an AI system know how to lie (or should your car tell your insurance company how fast you were driving). The future is unlikely to be an AI utopia bringing unprecedented efficiencies nor a dystopia, write the authors, and right now, we have control over whether AI will be a trusted aide to humanity or a threat.

In our fourth article, Aswani Kumar Cherukuri, Annapurna Jonnalagadda, and Cutter Consortium Senior Consultant San Murugesan look at potential applications and impacts of AI on education. Although not initially embraced by the education sector, AI can help students receive personalized lessons, provide educators with deep insights into students' learning styles, revolutionize skills improvement for professionals, and lower the cost of education. The authors present the AI technologies being applied in education and then describe the platforms and applications now available in each of eight categories: adaptive and personalized learning; content preparation; proctoring and assessment; online learning and immersive learning through augmented reality/virtual reality; language learning; coding and robotics; tutoring and mentoring; and management and scheduling.

Finally, Jayashree Arunkumar outlines how five AI trends are being slotted into real-world use, including graph-accelerated ML (NASA uses it to extract knowledge from its Lessons Learned database[2]), generative AI (which helped Reuters create a fully automated sports newscast[3]), edge AI (like cameras used to control traffic and catch criminals), artificial general intelligence (OpenNARS and OpenCog are two examples), and coding (such as finding and fixing human errors). Arunkumar then examines how AI is helping the environment by accelerating the pace of delivering on the United Nation's Sustainability Development Goals[4] and how it might apply similar tactics to help improve world health. The article closes with four of the most recent AI developments, including data sets and an advanced recommendation system from Facebook, an interesting self-driving car development, and language algorithms that can write a coherent article from a text prompt.

Historians may look back at this period in time as the point at which AI really started to have an impact across industry and society. AI may still be in a state of relative infancy, but one thing is for certain: it is an unstoppable wave, because anybody can access the tools to build AI systems and potentially produce world-changing applications. The downside of this is that, as we have already seen, bad actors can use AI to manipulate information that humans consume, which at its worst amounts to malign social engineering. However, AI can just as easily be used to counteract such misuse and build systems for good. It is a powerful form of technology-based democratization, the like of which has never been seen before. We hope this issue of *CBTJ* enlightens you on the state of AI today and helps you navigate the prospects and perils along your AI journey.

References

[1]See Wikipedia's "Three Laws of Robotics."

[2]Meza, David. "How NASA Finds Critical Data Through a Knowledge Graph." Neo4j, 17 May 2017.

[3]Chandler, Simon. "Reuters Uses AI to Prototype First Ever Automated Video Reports." *Forbes,* 7 February 2020.

[4]"Take Action for the Sustainable Development Goals." United Nations, accessed May 2021.

Michael Eiden is a Senior Consultant with Cutter Consortium's Data Analytics & Digital Technologies practice. Dr. Eiden, who serves as Head of AI at Arthur D. Little, is an expert in machine learning (ML) and artificial intelligence (AI) with more than 15 years' experience across different industrial sectors. He has designed, implemented, and productionized ML/AI solutions for applications in medical diagnostics, pharma, biodefense, and consumer electronics. Dr. Eiden brings along deep expertise in applying supervised, unsupervised, as well as reinforcement ML methodologies to a very diverse set of complex problem types. He has worked in various global technology hubs, such as Heidelberg (Germany), Cambridge (UK), and Silicon Valley (US), with clients ranging from small and medium-sized enterprises to globally active organizations. Dr. Eiden holds a doctorate in bioinformatics. He can be reached at experts@cutter.com.

PAY ATTENTION!

AI: Boon or Bane? (Hint: It Depends on Us)

by Paul Clermont

News stories and opinions about artificial intelligence (AI) are everywhere — from articles and podcasts to TED talks, think tank symposia, and philosophers' musings. Some enthusiastically tout AI's benefits for workers, enterprises, and society overall; others paint dystopian pictures of intrusive governments and employers surveilling and micromanaging our lives. Many foretell increased unemployment and greater disparities in income and wealth between educated elites and a proletariat consigned to miserable, low-paying jobs not yet taken by robots. Some envision AI morphing into artificial natural intelligence indistinguishable from our own but able to grow into a superintelligence that can determine our fate just like HAL 9000,[1] the computer in Stanley Kubrick's 1968 science fiction film, *2001: A Space Odyssey*. Scoffers — there are some — think AI is just the latest overhyped technology fad. All are a bit right and at least a little wrong.

> *There is no question whether AI applications will proliferate. They will. It's when we get down to just-because-we-can-doesn't-mean-we-should arguments that things get interesting.*

The purpose of this article is to find the space between the positive and the negative. Like any technology, AI can be used for good purposes and bad. It can be overused and misused, and it will be. The job at hand is to determine how specific dangers can be recognized and what might be required to avert them. Of course AI presents risks, but they are not entirely new. As computer applications have become dramatically more complex and interconnected, the risks posed to people by IT have grown exponentially over the years even without AI (see Table 1).

This article contends that the broadening scope of potential damage and the increasing speed with which it can happen mean innovations based on AI can no longer simply be what a corporation brings to the marketplace. Just as cities had to develop building codes to reduce fires, pestilences, and issues related to shoddy construction, so, too, must humans develop standards aimed at keeping IT a force for good (or at least not for ill). Too much technology has already been unleashed that's of dubious benefit or outright harm.

There is no question whether AI applications will proliferate. They will. It's when we get down to just-because-we-can-doesn't-mean-we-should arguments that things get interesting. Unlike simpler forms of IT, questions of what, why, and how around AI initiatives will not necessarily be well enough addressed by technologists and managers. There are roles for sociologists, psychologists, behavioral economists, ethicists, and even historians.

In this article, we address three overarching issues for AI aside from technology: unintended consequences, unintended bias, and privacy.

Overarching Issue #1: Unintended Consequences

In theory, there are good unintended consequences, but it's human nature to claim they were intentional if they happened, so not-so-good unintended consequences are of primary interest. These include the initiative itself, its (sometimes) logical extensions, and its implementation. The biggest challenge is to think far enough ahead to recognize their possibility. Here are some unintended consequences (and examples) to consider:

- **Collateral damage from an otherwise good idea**
  - The Internet plus social networks and blogging and publishing sites make it easy for ordinary people to make their voices heard. They also make it possible for mischief makers, cranks, and

Approximate Time Frame	Nature of IT	Scope of Potential Damage Outside System's Owner
Before ~1980	Batch processing in secure unconnected data centers	None
1980-2000	Online real-time computing with controlled proprietary access	Theft of private and proprietary data
2000+	Universal access via Internet	Expanded opportunity for data theft and hacking (e.g., distributed denial of service, malware, ransomware, Trojan horses)
2010+	Computer control of physical systems (e.g., smart homes, cars, generating plants, Internet of Things)	Above plus physical mayhem

Table 1 — Evolution of IT risk without AI.

conspiracy theorists to fill cyberspace with the misinformation, disinformation, and outright lies that have contributed greatly to political polarization.

- **Radical and questionable expansion of scope of a good idea once a base of users is established**
  - Social networks that made it easier for us to find and reconnect with old friends morphed into a channel for microtargeted advertising and broadcast of un-fact-checked "news."
- **Poor machine learning (ML) performance due to inadequate training**
  - When the scope of training cases was overly narrow, face recognition and skin lesion evaluation did poorly on dark skin.
  - If the range of possibilities is constricted to what has happened in the last 50 years, phenomena like 100-year floods will never be identified as possibilities.[2]
- **Overconfidence in sensors and logic**
  - A vehicle on "autopilot" is faked out by a white-painted truck trailer that essentially disappeared in bright sunlight, and the driver was killed in the crash.
  - A slightly modified 35-mph speed limit sign is misread as 85, and the car takes off without noting the context (a winding country road). Never mind that there may be no place in the US with an 85-mph limit!
- **Poorly designed human interface for dealing with emergencies**
  - Impaired sensors on two Boeing 737 MAX 8 planes led to crashes when the autopilot seized control of the plane and took inappropriate "corrective" action that cockpit crews were unable to override.
- **Culture-driven failure to think through what could go wrong**
  - An ethic of "moving fast and breaking things" does not foster hard critical thinking but succeeds admirably at breaking things.
  - A rah-rah, high-fiving culture spawns groupthink, sidelining the devil's advocates who ask the tough questions that in retrospect should have been asked and answered.
- **Erosion of human skills**
  - Marine navigation by charts, compass, and dead reckoning is becoming a lost skill as GPS proliferates. What if the electronic connection is lost?
  - Terrestrial navigation by map and landmark are deteriorating for the same reason.[3]
  - The intuitive sense based on experience that something is not quite right — *before* automatic alarms go off in a refinery.

A methodical approach to thinking through unintended consequences must include getting explicit not just about what we want to have happen, but what we do not want, and how success in the former could risk

including too much of the latter. It may also be useful to seek help from disciplines not usually associated with IT initiatives, such as behavioral economists and psychologists.

Overarching Issue #2: Unintended Bias

Biases leading to unethical (if not unlawful) discrimination in algorithmic decision making have arisen as a major concern in areas like loan applications, hiring, and criminal justice. In theory, an algorithm won't know the color of your skin without your picture, or your accent without a voice recording, so it will deal only in relevant facts with complete objectivity. If only it were that simple.

> *AI "training" approaches that use a phenomenon that reflects past human decisions as a model of excellence will perpetuate all the biases that underlay those decisions.*

Humans, including AI designers, have biases. We're products of the various people and cultures we've encountered, all of which have contributed biases. Thus, bias too easily creeps in despite our best efforts to eliminate it. As a very simple example, using past success patterns to train AI in evaluating job candidates bakes in all the biases that resulted in those patterns. In general, AI "training" approaches that use a phenomenon that reflects past human decisions as a model of excellence will perpetuate all the biases that underlay those decisions.

Some kind of oversight is needed way beyond the normal skills and perspectives of technologists and professional managers. A wise organization embarking on a major AI initiative will manage appearances as well as substance. An oversight board should include people with a variety of perspectives, including sociologists, historians, psychologists, lawyers, and maybe even community organizers.

It is very hard to define and recognize bias. Helpful criteria like the four-fifths rule[4] may "prove" lack of bias along a single dimension like race or gender but could miss intersectional disadvantage for non-white women.

The auditing model for financial reporting would make sense if there were a bias-recognizing equivalent of Generally Accepted Accounting Principles (GAAP)[5] to refer to, plus a certification process for practitioners. Absent that, a self-described "bias auditor" may not be reliable; the incentives are wrong versus those of a licensed third party like a CPA or a government agency with a charter like the US Securities and Exchange Commission (SEC). Some kind of GAAP equivalent will surely evolve to provide a template for bias "accounting" just as GAAP did for financial accounting. Like GAAP, it will provide a certain amount of latitude, but adherence to it, absent falsification of data, will provide some immunity against legal action. New careers will emerge to define and practice it and monitor adherence, just as with accountancy.

Even an algorithm reasonably believed to be unbiased should not be the final authority in every case, a notion that extends to potentially life-changing decisions for individuals and their families like home and business loans, hiring, college admissions, and elements of the criminal justice system like bail-setting, sentencing, and parole decisions. By definition, an algorithm yields a number that is compared to a boundary criterion. No numbering system that attempts to quantify things that are not precisely measurable is good enough to decide when the algorithmic result is close to the boundary. In that situation, basic fairness and common sense call for human intervention. Could that introduce bias? Yes, but a decision that could change the life of a person and family should not be made on the equivalent of a coin toss.[6] Human judgment should always have a place in such situations

Overarching Issue #3: Privacy

Privacy issues are not unique to AI, but AI's hunger for data for ML training has upped the threat. This was brought home by the recent enforcement actions taken by the US Federal Trade Commission (FTC) for misuse of facial recognition technology by Everalbum, which scraped millions of people's pictures and used them without the consent of the pictured for commercial purposes.[7] It will be interesting to see the implications of this on current and future development of AI technologies and applications: will it hinder innovation?

Shortcuts like Everalbum's that take advantage of loopholes and ambiguities in existing law will be prohibited eventually, some by legislation, some by judicial opinions. Innovation will be slowed but so is innovation in medical science slowed by laws governing ethically permissible protocols for human tests and experiments. As a general principle, ethics

trump expediency, but specific situations may call for more nuanced attention to risk and transparency.

Internet trolling is an increasing problem, made worse by AI's ability to create seemingly authentic deepfake photos of scenes that never happened or recordings of words never spoken. Their capacity for mischief requires no elaboration.

Dealing with the Issues

As the ability of digital and IT to create harm grows, so must the legal and regulatory infrastructure to prevent, or at least minimize, disasters and assign liability when they happen. The law always lags technology, and this is as it should be — though not by too much. The EU instituted its General Data Protection Requirements in 2018, largely copied by the US state of California in 2020; when two such sizable chunks of the world or the US market act, it becomes a de facto standard.[8] The EU also recently proposed new laws to govern AI, ensuring oversight and limitations in the interest of public safety and privacy (see sidebar). Although restrictions like this are new for IT, they're established practice elsewhere, including:

- Building codes
- Requirements for approval by licensed professional engineers
- Periodic auto safety inspections
- Underwriters Laboratories (UL) certification of electrical products
- US Federal Aviation Administration (FAA) and US Federal Drug Administration (FDA) certifications for safety of air travel, food, and pharmaceuticals

There should be standards for policies and practices when creating the software equivalent of guardrails, fire doors, containment vessels, and intrusion detectors. There need to be penalties when something goes wrong and those standards were not met. These may limit AI "creativity" (i.e., unorthodox, out-of-nowhere actions that just might work), but who knows? Such an action beat the *Go* champion, but *Go* is, well, just a game.

Legal principles need to be established, including:

- **Unexplainable AI does not belong in any application that could harm people physically or affect their lives in significant ways.** The same applies to black-box algorithms sold by vendors that refuse to explain their "proprietary" workings. Buyers must receive explanation, acknowledge receipt, and possibly sign nondisclosure agreements.
- **Organizations are responsible for algorithms they use, whether bought or built.** If bought, relief would require that algorithms don't behave as advertised, similar to how an auto accident caused by a design or factory defect can relieve the driver of some responsibility.

Laws or regulations will also be required to:

- Ensure the spirit of principles like those established in the US Bill of Rights is applied to government activities (e.g., requirements for warrants).
- Limit the use of algorithms for close calls on potentially life-changing decisions such as home loans, business startup loans, hiring, and college admission.
- Establish what constitutes a "legitimate interest" for third parties and governments to access data we have no choice but to generate about ourselves.
- Obtain a better balance between freedom of speech and freedom from malicious misinformation and disinformation. AI can be highly useful here to

The EU Gets into the Act

As this article was being written, the EU put forth a framework for regulating AI that could become a de facto standard. As with the GDPR, it fills a gap that was unlikely to be filled in the near term by others. It recognizes that different forms of AI pose different levels of risk:[1]

- Low-risk AI only requires transparency. For example, a deepfake photograph would have to be labeled as such. Otherwise, it would be essentially unregulated.
- High-risk AI, which includes most of the examples cited in this article, would require not just transparency but thorough risk assessment, user education, and adequate human control before being brought to market.
- Unacceptable-risk AI, such as applications to manipulate public opinion or develop social credit scores (as in China), would be prohibited.

[1]Marcia, Valeria, and Kevin C. Desouza. "The EU Path Towards Regulation on Artificial Intelligence." The Brookings Institution, 26 April 2021.

improve fact-checking. Even more important, AI can be used to broaden people's perspectives on events rather than how it is now used to narrow and intensify them, fostering political polarization.

- Recognize that Internet communications services are not the same as voice telephone services because they can monitor content, and experience is showing they need to. The viral spread of absurd conspiracy theories and proliferation of lies and trolling are giving free speech a bad name.

Major emphasis on cybersecurity is not a matter of choice. There are more algorithms to steal, or worse, hack. There is more data to steal or pollute. Risks really grow with the Internet of Things:

- We must carefully limit our trust in hackable AI to safely govern infrastructure like power grids, nuclear plants, hydroelectric dams, air and vehicle traffic control, and drinking water supplies.
- We may need better firewalls to separate remote access *control* from the provision of information. For example, there's a huge difference between downloading potentially hackable engine management software to a moving vehicle via the Internet versus downloading GPS information the driver needs in real time.

In short, just because we can do something that's "cool" doesn't mean we should.[9]

Picturing an AI-Enabled Future

The future can look quite good if governments and technology providers deal effectively with the overarching issues. Beyond that, the most tangible changes for many people will be in the workplace:

- Use of robots and robotic process automation for repetitive and unambiguously describable tasks will increase wherever it's economical. Where it's not, people's jobs will be made as robotized as possible by industrial engineers, as in Amazon fulfillment centers. (Taylor lives on![10]) In the physical realm, some processes may be amenable to "cobots": robots that work directly with a person to do parts of a job that are too finicky, strenuous, or dangerous for a human. In effect, the robot amplifies the person's capability. In the office realm, we're already seeing a trend toward customer service phone systems that use AI to determine from the customer's words where to route the call if it can't be addressed by the computer alone; in effect, the person is amplifying the capability of the machine.
- Paraprofessional or lower-level professional tasks entailing a lot of search and pattern recognition will be taken over by AI, amplifying the person's ability and productivity.
- New jobs in AI design, monitoring, and ML training will offset some lost jobs.
- A high degree of worker surveillance will be irresistible to employers unless it's regulated.
- We may at long last see questioning of the continuing appropriateness of the 40-hour workweek, the standard for almost a century after a century of rapid decline from 72 or 84 hours.

On the more sobering side, workplace prospects for the poorly educated continue to dim. Failure of governments to address "left behind" people and regions will lead to increasing political unrest with predictably negative consequences.

Conclusion

The risks to society in general from AI are real and important enough right now to demand attention from a lot of very talented, busy people. That said, we need to put the power of AI in perspective. As sophisticated and brilliant as some AI applications are, they do not come near artificial general intelligence (AGI). Consider three milestones.

Championship-level checkers playing by a computer dates back to 1962.[11] It was based, like today's AI, on ML (i.e., playing against itself a huge number of times, gradually getting better by learning through trial and error). It took 34 years for computers to advance to the level of grandmaster chess and another 20 to conquer *Go*. At their core, these feats, however brilliant, were the same: pursuit of a clearly defined goal with strict rules governing the moves you can make to get there. In short, AI applications to date are idiot savants, able to do a limited range of tasks with superhuman brilliance … and nothing else.

When we apply our human intelligence to addressing AI's "right now" challenges, we can safely ignore

the prospect of imminent AGI that could morph into superintelligence, whatever that is. Time and energy spent on risks that certainly will not materialize in the next decade or two (if ever) is not just a waste of time, it's a counterproductive diversion.

> Sufficient unto the day is the evil thereof.
>
> — Sermon on the Mount, Matthew 6:34[12]

References

[1]See Wikipedia's "HAL 9000."

[2]Farnum Street. "The Lucretius Problem: How History Blinds Us." Farnam Street Media, accessed May 2021.

[3]I've seen myself becoming increasingly dependent on GPS directions in the car, and not just for the first time or two I go there.

[4]Mondragon, Nathan. "What Is Adverse Impact? And Why Measuring It Matters." HireVue, 25 March 2018.

[5]"Generally Accepted Accounting Principles (GAAP)." Investopedia, 22 February 2021.

[6]This stricture does not apply to non-life-changing decisions such as auto loans. If the algorithm thinks you can't afford the Cadillac, a Chevrolet will still get you where you need to go.

[7]Art photographs often include random people. Usage is permitted for art but not for purely commercial purposes like advertising.

[8]Whatever its imperfections, it filled a void. Only the US had comparable void-filling power, but its political structure at the time had a strong bias toward inaction.

[9]Does my clothes dryer really need an IP address?

[10]Frederick Winslow Taylor was an efficiency expert from the early 20th century who turned time and motion study into a science.

[11]In the 1960s, the checkers problem was addressed by "brute force" techniques, evaluating a tree of all potential moves multiple generations in the future and choosing the one with the highest probability of a good outcome. By the 1980s, this had evolved to pattern recognition. All of these approaches were hindered by lack of sufficient computer processing power.

[12]See Wikipedia's "Sufficient unto the day is the evil thereof."

Paul Clermont is a Senior Consultant with Cutter Consortium's Business Technology & Digital Transformation Strategies practice. He has been a consultant in IT strategy, governance, and management for 40 years and is a founding member of Prometheus Endeavor, an informal group of veteran consultants in that field. His clients have been primarily in the financial and manufacturing industries, as well as the US government. Mr. Clermont takes a clear, practical view of how IT can transform organizations and what it takes to direct both business people and technicians toward that end. His major practice areas include directing, managing, and organizing IT; reengineering business processes to take full advantage of technology; and developing economic models and business plans. Mr. Clermont is known for successfully communicating IT issues to general managers in a comprehensible, jargon-free way that frames decisions and describes their consequences in business terms. In his consulting engagements, he follows a pragmatic approach to the specific situation and players at hand and is not wedded to particular models, methodologies, or textbook solutions.

Before going into individual practice, Mr. Clermont was a Principal with Nolan, Norton & Co., a boutique consultancy that became part of KPMG. Before joining Nolan, Norton & Co., he directed IT strategy at a major Boston bank and launched its IT executive steering committee. Mr. Clermont has spoken and written about the challenges of getting significant and predictable value from IT investments and has taught executive MBA courses on the topic. His undergraduate and graduate education at MIT's Sloan School of Management was heavily oriented toward operations research. He can be reached at experts@cutter.com.

AI's Role in Accelerating Product Development

by Michael Jastram

Humans have centuries of product development experience, but the recent exponential rise in product complexity means current development approaches are reaching their limit. Fortunately, artificial intelligence (AI) has the potential to help us overcome those limitations.

Product complexity is currently driven by four trends. First, products include a growing amount of embedded software. We can roughly measure product complexity by looking at number of lines of code.

Second, modern products are increasingly connected, leading to a "system of systems," a trend also known as the Internet of Things (IoT). Such systems of systems exhibit emergent behavior — properties the individual systems do not have.

Third, complexity is compounded by larger teams, which are often distributed. This results in growing communications overhead. To cite just one example, automotive manufacturers frequently share requirements specifications containing tens of thousands of requirements.

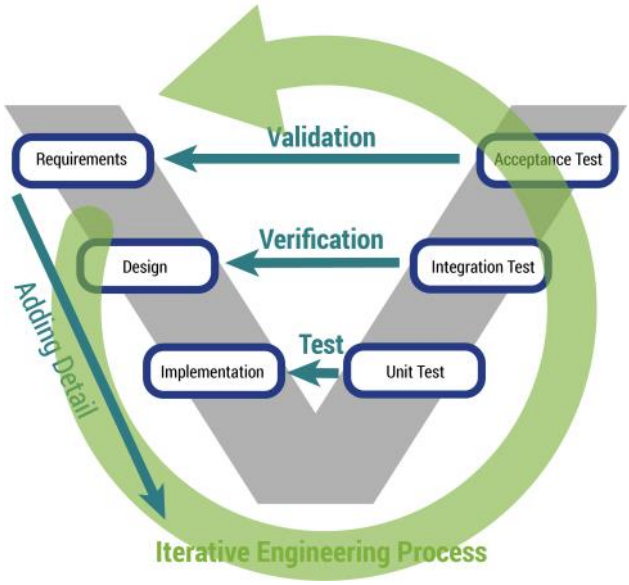

Figure 1 — The V-Model.

Fourth, regulatory compliance creates overhead. With more IoT devices entering our lives, manufacturers must demonstrate their safety. A failed audit can result in a delayed product launch.

To understand why complex products require new approaches, let's first briefly review the existing approaches.

Systems Engineering

In the 1960s, systems engineering became an interdisciplinary field of engineering for developing complex products. It was driven primarily by the space industry, as well as some high-profile engineering disasters. A key concept of systems engineering is the V-Model, which is primarily known as a development process, but it also represents the various artifacts of product development. The V-Model puts requirements, implementation, and validation and verification (V&V) into context, a key concept that is highly relevant to the rest of this article (see Figure 1).

Interpreted as a process model, the V-Model starts at the top left: defining the product requirements. These requirements are refined and decomposed, eventually leading to an implementation. The right arm indicates that on each level, V&V activities ensure the right product is being built, and that it's being correctly built. All elements must function as designed (verification), and the result must be what the customer desired (validation). Testing is the most common approach for V&V, but there are others. Similarly, the left arm of the V often consists of more levels than the three shown.

A key takeaway is the triangular relationship between requirements, implementation, and test, which provides robustness. A failed test might indicate a problem with the implementation, but it could also point to a faulty requirement or a problem with verification.

Initially, systems engineering was paper-based; every iteration carried significant overhead because long

documents had to be read, reread, checked, and aligned. For complex products like rockets, planes, and cars, iterations typically took six months to two years.

Many organizations still practice traditional development without many changes (substituting Word and Excel files for paper), even though more sophisticated tools are available. This does not necessarily result in unsafe systems, but the overhead for each iteration is increasing because of rising complexity. Furthermore, long iterations lead to a slow response to customer feedback and changes in market and technology, creating a competitive disadvantage.

Since the advent of computers, software has been used for making development more efficient, like CAD and CAE systems in the 1980s. However, these solutions helped engineers solve a specific problem, rather than addressing the overall engineering process.

In the 1990s, we saw the emergence of software tools known as requirements management tools, which cover large parts of the development process. Rather than dealing with documents, they manage fine-grained items and allow organizations to define a data model that makes dealing with changes and traceability more effective. Key market players include:

- IBM Dynamic Object-Oriented Requirements System (DOORS) ("classical" DOORS)
- IBM Rational DOORS Next Generation (the codebase is not related to classical DOORS)
- Jama Software Jama Connect
- Siemens Polarion
- Intland Software codebeamer
- PTC Requirements Management and Validation (formally PTC Integrity)

These tools break development artifacts into smaller chunks, but most of the data is still text, written by people for people. There may be formulas, figures, and schematics, but the actual content cannot be processed by machines — they only help with content management.

Some of these tools offer AI add-ons. These typically analyze individual requirements to provide some quality feedback. Although this can be useful, it only scratches the surface of what is possible.

Another category of software tools addresses product lifecycle management, but these focus on manufacturing and bill-of-materials management and so are not central to the product development process.

Agile Product Development

In the early 2000s, Agile methods became popular in software development, and their application has now spread to product development and beyond.

Agile practices involve rapid iteration cycles and close collaboration with all stakeholders, including customers, in cross-functional teams. Agile addresses the problem of slow development cycles and communication problems. Agile practices also tend to embrace new technologies that take friction out of development (e.g., through automation).

A key takeaway is the triangular relationship between requirements, implementation, and test, which provides robustness.

Applying Agile development in the context of developing complex products doesn't change that much, however. This is particularly true if the system under development involves safety. Safety regulation embraces systems engineering and the idea behind the V-Model.

There are specialized tools for Agile development, but most of them have a strong software focus (e.g., application lifecycle management). For developing complex products, all vendors of the requirements management tools mentioned above support frameworks for working in a more Agile way.

Model-Based Systems Engineering

Model-based systems engineering (MBSE) takes the idea behind the V-Model to an extreme, abandoning documents and replacing them with a formal systems model. Documents only exist in the form of specialized views of the model.

MBSE relies on a modeling language instead of human language. Prominent examples include Unified

Modeling Language (UML), Systems Modeling Language (SysML), Business Process Modeling Notation (BPMN), and Event-B. SysML is by far the most visible systems modeling language, while UML is prevalent for pure software systems.

Better-known MBSE tools include Sparx Systems Enterprise Architect, IBM Rational Rhapsody, and Dassault Systèmes Cameo Systems Modeler. Introducing an MBSE tool into an organization almost always requires training, coaching, and consulting to make it successful; the tools and methods are too complicated to use without expert help.

Modeling languages can be visual or textual, open or proprietary. But they all differ from traditional approaches in their ability to allow machines to directly reason. This opens the door to a number of useful capabilities, like consistency checks, test generation, and simulation, all of which can be automated.

MBSE has been around for 20 years and is widely used in avionics, defense, rail, and automotive. However, MBSE requires a large up-front investment with the promise of a positive ROI somewhere down the road; this is particularly true for organizations that build tailored, complex systems and therefore must deal with variants (see Figure 2).[1] To date, only a few small organizations have been willing to risk this type of large investment.

There are also cultural challenges, including the fact that many stakeholders are unwilling or unable to learn a new modeling language and the corresponding methods. This severely limits MBSE adoption.

Applying AI to Product Development

Systems engineering and Agile both struggle to keep up with today's exponential growth in complexity. MBSE was built to address complexity, but it requires a large up-front investment and often meets with cultural resistance. Those are two very different challenges, but both can be addressed with AI — at least in principle.

AI is already used to support specific traditional development activities like quality checking and detecting similarities. Similarly, AI is being employed to extract models from MBSE text.

However, AI in product development is in its infancy. This is partly because the market for product development is small compared to areas where AI is already adding value, such as marketing.

The most visible player right now is IBM, which has a portfolio of tools for product development that includes DOORS Next for requirements engineering and Rhapsody for modeling. IBM also has a prominent AI engine: Watson. As expected, IBM adapted Watson as an add-on for DOORS Next. Currently, this add-on analyzes individual requirements to rate their quality.

Several smaller players are also active, trying to climb their way toward wider recognition. No clear leader has emerged, but it's early days; we'll see a lot of activity in this area over the next five years.

Finally, it's important to recognize that there are many AI subfields, each at a different development level, which we explore more fully next.

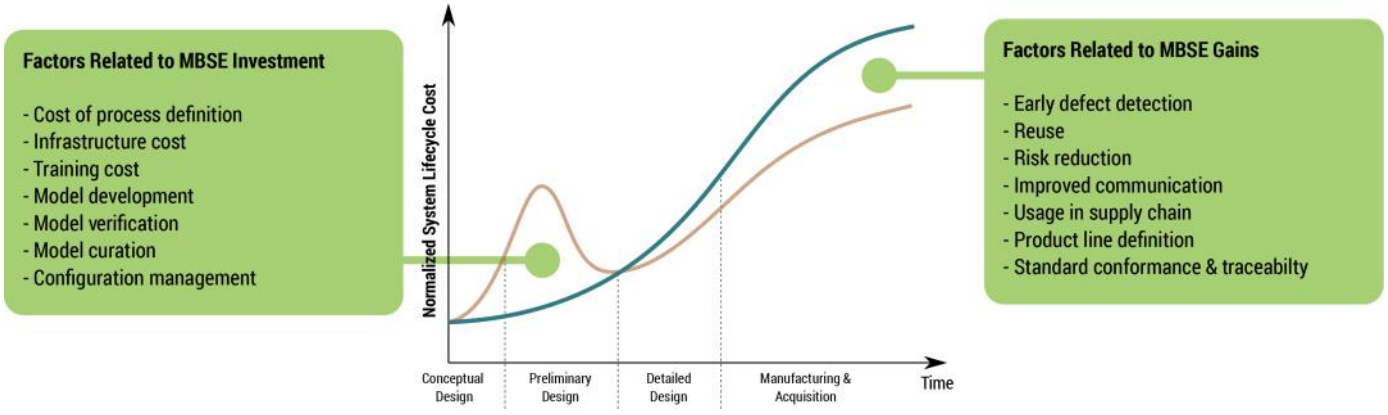

Figure 2 — MBSE factors related to investments and gains. (Adapted from: Madni and Purohit.)

Natural Language Processing

By far, the most activity in AI related to product development is in natural language processing (NLP). This is not surprising since most of the artifacts in product development are still text: requirements, risk items, test cases, and so on. This AI subfield is well suited to both traditional and Agile product development.

By far the most common application of NLP in requirements engineering is quality scoring. Two big players, IBM (DOORS) and Siemens (Polarion), have solutions for this.

Although these solutions work and provide value to practitioners, they do not address the problem of rising product development complexity. These solutions only analyze one requirement at a time, making it impossible to spot contradictions between two requirements.

There is a lot of research aimed at complete product descriptions, rather than focusing on individual items. Analyzing complete product descriptions opens the door to a number of interesting applications:

- Identifying issues that can be spotted only in context, such as gaps, contradictions, and ambiguities
- Improving traceability (traceability analysis is possible only when all items are analyzed together)
- Extracting MBSE models from natural language specifications

The last point in particular could be transformative. It would provide a bridge from traditional product development to MBSE, making MBSE's benefits accessible to a much wider audience.

These ideas have not yet been realized in commercial products, but both academic and industry researchers are working toward that goal.

Machine Learning

Machine learning (ML) employs algorithms that use data to improve automatically through experience. In product development, a simple example is the categorization of items. For instance, you could assign a technical category like software, hardware, or electronics to an item. This is well understood but of limited value, as it does not solve the problem of rising complexity.

AI is already applied in product support — providing first-level support or analyzing the customer's mood. A more relevant activity involves analysis and clustering of conversations to find critical or highly visible bugs or identify important feature requests or unmet needs. These can feed into the product roadmap to make future products more successful. Similar information can be extracted from marketing data or product telemetrics.

By far, the most activity in AI related to product development is in NLP. This is not surprising since most of the artifacts in product development are still text.

A more complex problem is the creation and analysis of traceability, a key aspect of systems engineering. Practitioners create and maintain the traceability in both traditional product development and Agile development. The company Relatics follows this approach by analyzing arbitrary data sources for traceability management. More players will certainly follow, but commercial solutions are currently scarce.

Some industries rely more on large-scale traceability systems than others. For example, the automotive industry has strict regulations regarding traceability and a large volume of requirements. It's not unusual for manufacturers to provide suppliers with specifications consisting of tens of thousands of requirements. The scale alone makes this use case interesting for ML.

In MBSE, traceability tends to follow clearly defined rules, limiting the use of ML. However, even MBSE starts with textual requirements, and ML could support the allocation of those requirements. Once again, there are few commercially available solutions.

Interactive AI

Human-computer interaction is already used extensively in customer service and support, most visibly in the form of chatbots and voice-based applications, but there is unexplored potential for such systems in product development. Today's problems are solved primarily by hiring outside experts. The goal is to raise

the organization's maturity to make it more effective with its existing resources.

Interactive AI has the potential to replace experts, at least in some areas. This could take the form of interactive wizards that help set up a development project or interactively improve the system architecture in a dialog format.

If such systems already exist, they're most likely for internal use at large organizations. It is less challenging to build such a system in a domain-specific context. For instance, a German manufacturer of lighting systems built a system to guide users in capturing the system behavior in a domain-specific language. Using the stakeholder input, the system makes suggestions for translating the input into a machine-readable format. Building a universal system is much more challenging, primarily due to missing training data.

There's little question we'll soon have an army of consultants, integrators, and service providers that promise to find areas of applications of AI technologies for their customers, in product development and beyond.

Training Data

Many AI systems rely on a large body of training data. This is particularly challenging in product development and is likely the reason we haven't seen much progress compared to other sectors. Detailed product descriptions are often considered core intellectual property; few organizations are willing to share the descriptions, even with an NDA in place.

There are workarounds. For instance, user manuals, data sheets, and maintenance handbooks can be substitutes, especially for high-level requirements. But this is not enough for a holistic body of knowledge on product development and therefore only of limited use.

This is the reason large multinationals are active in this area: they have enough training data to make AI systems useful, at least internally. It will be interesting to see what these organizations publish in the coming years.

AI Platforms

Until off-the-shelf AI product development systems are available, organizations must build their own. There are many players in this space, including established companies like IBM, Google, and Microsoft, and newcomers like Dataiku and SambaNova Systems.

There's little question we'll soon have an army of consultants, integrators, and service providers that promise to find areas of applications of AI technologies for their customers, in product development and beyond. Considering how specific product development is, retaining such experts will be a good idea for many organizations. As always, the key to success will be providing clear strategic direction and strong oversight.

Confidentiality is a major challenge with these platforms, as they typically operate in the cloud. On-premise systems are available but typically have limitations. For instance, they cannot access cloud-based analysis systems by third-party providers that continuously evolve their AI engines.

Conclusion

Organizations must respond to the pressure of rising complexity. Thus far, we've seen companies respond by adding either manpower (work harder) or expertise through consulting, training, and tooling (work smarter).

Adding more and better tools has been practiced successfully since the 1990s, starting with CAD systems, requirements management systems, and simulation tools. In the 2000s, we saw a certain amount of "tool-fatigue," with practitioners noticing that simply adding a tool did not always solve the problem. This gave a huge boost to Agile methods, which require training more than new tools. However, tools continued to provide efficiency. It is surprising how much product development work is still done using Word and Excel.

Organizations that reach their limits due to complexity growth typically turn to MBSE, but today's MBSE tools are far too complicated to be a good starting point. Successful MBSE involves a strategic, long-term objective and clear, achievable milestones.

We'll continue to see MBSE deployment over the next five years or so, but growth will be severely limited by

the number of available experts. This opens up a huge opportunity for AI-based solutions that provide some of the benefits of MBSE without requiring a long, expensive training process.

We currently see attempts to address product complexity by using AI to bridge the gap between MBSE and the language of the stakeholders. Most of this work is still in the research stage. Ready-to-use solutions like IBM's Watson barely scratch the surface of what's possible.

In the coming years, more powerful AI systems for product development will emerge, starting with written language. Initially, these systems will improve traditional or Agile development, but they will soon focus on helping with MBSE. This will reduce (but not remove) the need for companies to pay outside experts in order to benefit from MBSE.

Such systems will guide practitioners through the development process by creating and maintaining architecture, testing traceability, identifying quality issues, and more. All these help prevent waste, reduce risk, and speed product development.

Reference

[1]Madni, Azad M., and Shatad Purohit. "Economic Analysis of Model-Based Systems Engineering." *Systems*, Vol. 7, No. 1, 2019.

Michael Jastram is a systems engineer with more than 20 years' professional experience, including 10 years in the US, and has worked as a software engineer and architect for various startups. With expertise in requirements modeling, his current focus is in developing Semiant, an AI-based virtual quality assistant for product development and organizational knowledge management. Dr. Jastram is active in open sourcing as the founder and project lead of the Eclipse Requirements Modeling Framework and coauthor of the Requirements Interchange Format (ReqIF), an international open standard. As an advocate for openness, he shares his knowledge via books, articles, talks, and even as organizer, as well as via his weekly blog System Engineering Trends *(in German) and his monthly* Formal Mind Blog. *Dr. Jastram earned a master's degree from the Massachusetts Institute of Technology (MIT). He can be reached at michael.jastram@formalmind.com.*

THE ROAD AHEAD

Bridging the AI Trust Gap

by Claude Baudoin and Clayton Pummill

Artificial intelligence (AI) in general, and specific fields within AI such as image recognition, machine learning (ML), and natural language processing (NLP), are going through a rapid revival that powers both consumer-oriented and industrial applications. These capabilities long ago left the confines of university research labs and are making real decisions that impact our lives and safety.

So what's the best way forward as we strive for trustworthy AI? AI ethics and trustworthiness are trending topics in the field, but the power of AI is already in play, so we have to progress accordingly. In this article, we examine some of the pathways to greater AI maturity, trustworthiness, and acceptance.

The history of technology revolutions is full of examples of reluctance, resistance, and sometimes violent opposition.

The AI Trust Gap

Why is there an issue with trusting AI? After all, humans have a long history of rushing to adapt new, often unproven, remedies. However, the history of technology revolutions is full of examples of reluctance, resistance, and sometimes violent opposition. When the steam locomotive was invented, people argued that the human body could not survive being moved at speeds of 50 mph. When the first cars hit the roads, there were some short-lived attempts to compel drivers to follow a man on foot waving a warning flag. Cultural anthropologist Genevieve Bell attributes these reactions to "the 'moral panic' that a society experiences when particularly revelatory technological advances show up — specifically, ones which interfere with or alter our relationships with time, space, and each other."[1]

It is not clear that AI falls within that very specific description, although it certainly challenges our understanding of what intelligence is, not to mention our belief in the inherent superiority of human thinking over machines. But we think the causes of mistrust go beyond these existential considerations and include the following:

- AI is mysterious. The vast majority of society does not understand how it works, and deep neural networks in particular can produce results that we cannot readily explain. People generally fear what they don't understand.
- AI is often equated to robots (which in reality may or may not be powered by AI), and there have been decades of sci-fi literature about rogue robots wreaking havoc. Think about Isaac Asimov's novels involving his three laws of robotics[2] and HAL 9000 in Arthur Clarke's *2001: A Space Odyssey*.[3]
- AI is seen as having a huge potential to eliminate jobs. A typical example is call centers: voice recognition, and the interpretation of a caller's requests or commands, reduces the need for operator interaction and can sometimes result in processing queries entirely without human intervention.
- The dramatic increase in hate speech or grave threats in social media requires the use of AI to recognize them among billions of daily posts. Such NLP applications "may be exposed to text that contains falsehoods or lies, or hatred, or is meant to incite trouble," notes Stanford Engineering Professor Russ Altman.[4] Many users have now experienced false negatives (hate speech that was not blocked) or false positives (their account was suspended for some innocuous remarks that confused the AI-based filter).
- Self-driving cars have become the poster child for high-risk AI applications and an example of a technology where a bad decision could have dramatic consequences (whether we are riding in autonomous cars or just sharing the road with them). There is credible evidence that human drivers make more dangerous mistakes than self-driving cars, but it is the occasional opposite case that gets the headlines. In application domains other than self-driving, there have been calls for a "human in the loop" approach,

even though it may negate the benefits of the technology in situations that call for a very short reaction time (e.g., shutting down an oil well as soon as a gas leak is detected by a sensor).

- Healthcare presents a similar challenge: we may not be ready to entrust our lives or well-being to an AI application, preferring to trust a resident who may be nearing the end of a 30-hour shift.

- Although this is not well-known to the general public, technology-savvy people now understand that ML is only as good as the training data sets, which can impart bias. A typical example is the higher rate of error in facial recognition of people of color and women.[5]

- An even more arcane risk is the malicious injection of data to fool AI models into reaching the wrong conclusion. This has been an active area of research over the last decade, and the news overall is not good.[6]

- Finally, a concern shared by ethicists is about AI making unexplained (or unexplainable) life-or-death decisions. We're back to the famous tramway dilemma: which track do you choose for the runaway tramway car, given that whatever the decision, someone will get killed, and you have to decide based on the number, gender, age, and other characteristics of the potential victims. The equivalent dilemma in these pandemic times could be how do we feel about AI deciding which patient to take off of a respirator in an overloaded ICU? There were unfounded alarms in the US about "death panels" during the 2009 debates preceding the adoption of the Affordable Care Act; what if this became a reality, except that the death panel is a bunch of computers rather than doctors in white lab coats?

The result of all this is that the phrase "AI trust gap" became very popular in the last two years (at the time of this writing, Google reports 44.9 million results on that phrase), and the phrase "AI ethics" is even more popular (263 million results), while the lesser but still notable popularity of the phrase "AI ethics oxymoron" (1.37 million hits) should concern us.

It is noteworthy that not everyone agrees these issues are really impeding the adoption of AI, but those dissenting voices are somewhat biased. For example, after the European Commission published a white paper on AI[7] with a section entitled "An Ecosystem of Trust" that contains certain regulatory recommendations, the Federation of German Industries (BDI) issued a response that includes this:

> ... the Commission should critically examine whether lack of trust is really a main factor holding back a broader uptake on AI…. Analogous to the so-called privacy paradox, a discrepancy between concerns about the trust-worthiness of an application and the actual user behavior is also apparent in the case of AI. Thus, even a minimal additional benefit or cost saving could be sufficient to induce consumers to use less trustworthy AI applications.[8]

However, this and other points made by BDI were clearly aimed at rejecting the need to enact additional legislation, redefine liability for AI systems, single out AI as a more problematic technology than others, or label AI-based systems as such. In essence, they were arguing for industry self-regulation.

But enough bad news. Let's turn to what we should do about this.

The Keys to Achieving Trust in AI

If the ultimate goal of AI is to efficiently replicate and exceed human thought for the good of humanity, then building trust requires that AI incorporate the multitude of sound human decision capabilities. Just reading this statement makes it clear that the journey toward AI ethics is no easy road. Even the assumption that AI should be used for the good of humanity is wishful thinking; in reality, AI is mostly developed for profit or for political or economic gain, along with many other motives (consider the fact that the same image classification techniques can be used to detect breast cancer and perform racial profiling).

We must therefore consider that in aiming for trusted AI, the world is a very diverse and complex stage (to paraphrase Shakespeare). Agreement on, and adoption of, global AI trust policies will require navigating that global political and economic theater. Alliances will be forged, agreements made, and oversight and enforcement will be necessary. This will require more than computer and data scientists. We should build cross-disciplinary teams that include psychologists, ethicists, sociologists, spiritual leaders, and even lawyers and legislators, in order to develop solid and accepted AI policies. On the technology side, a major evolution that needs to take place is to impart AI with emotional intelligence.

What policies, procedures, practices, technologies, and training should be implemented to ensure the integrity, transparency, and accountability of algorithms, AI-enabled applications, or automated processes? And how do we transform the reality of trustworthy AI (assuming we achieve it) into a correct perception by users, so that unsubstantiated fears do not linger, preventing us from achieving the technology's benefits?

Transparency, Explainability, Accountability

> Perhaps blackbox models cannot be avoided because the problems we want to solve are complex and non-linear. If simpler models were used, there [would] be a tradeoff in performance (for explainability), which we don't want.
>
> — Prajwal Paudyal, ML/AI scientist[9]

Among the reasons listed earlier for the trust gap, the mystery of how an AI model arrives at a decision is a recurrent theme. Life-or-death situations aside, we don't want to be denied a loan, ask the banker why, and receive this answer: "Er … I'm not quite sure, but the computer said so." Nor will our legal system permit such an answer.

The desire for AI transparency is particularly acute when neural networks are involved, since by design the way the network generates its output is complex. No one explicitly programmed the weights and thresholds applied at each node to pass as inputs to the next layer; those were arrived at by training the model (hence the phrase "machine learning"). In fact, there is an argument that the more powerful an AI model is, the less explainable it is and vice versa (this is, if you will, the AI version of Heisenberg's uncertainty principle in physics).

This has been a subject of serious research over the last couple of years. A fundamental paper from the US Defense Advanced Research Projects Agency (DARPA) proposed the term "XAI" ("eXplainable AI") and presented a model of the inverse relationship between performance and explainability (see Figure 1).[10]

The DARPA paper goes on to propose a model for an explanation process, in which an explanation of an AI decision can be subjected to "goodness" criteria and tests of satisfaction, leading to the user replacing mistrust with "appropriate trust and reliance."

Some authors point out that explainability isn't something you add after the fact. An AI model or algorithm needs to be designed with explainability in mind, so that appropriate intermediate data that contributes to the ultimate answer is captured during execution.

Eliminating Bias, Ensuring Fairness

Another cause of mistrust is the suspicion — and too often the evidence — that AI can be just as biased as humans. There is even such a thing as "confirmation bias" (the tendency to perceive what you already believed, not the reality you see) in AI. Ranjay Krishna, a PhD student at Stanford, recently demonstrated how an AI model trained to recognize scenes from photographs can be given an image of a person sitting on a fire hydrant and "recognize" a nonexistent chair in the picture, simply because in all the training data containing a person sitting on something, that something was a chair.[11]

How does one test for bias? A key proposition is that a small change in the inputs to the AI model should

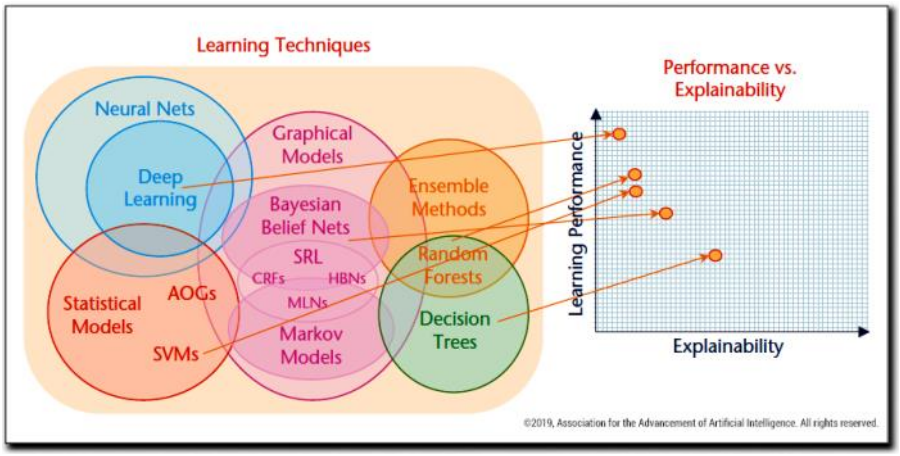

Figure 1 — Learning performance vs. explainability tradeoff for several categories of learning techniques. (Source: Gunning and Aha.)

generally produce a small change in the output — whatever "small change" means, and in the absence of a clear metric, human interpretation may be required. For example:

- If a loan application evaluation program processes two loan applications that are identical in all respects, except for the ethnicity of the applicant, and recommends to approve the loan to a white applicant and deny it to a person of color, that's a fairly damning sign that bias has been built in either in the program or the training data. In the latter case, the training data (history of loans manually granted or refused by loan officers) simply carries along the bias that underlay past human decisions and teaches the model to be equally biased.
- If adding some random noise, invisible to the human eye, to a picture of a dog causes an image classification program to "recognize" a cat instead, this may not be a sign of bias per se, but it at least means the algorithm is brittle and can be easily manipulated.

There is some relevant work worth considering here. In 2013, the Object Management Group (OMG) established the Structured Assurance Case Metamodel (SACM™):

> An Assurance Case is a set of auditable claims, arguments, and evidence created to support the claim that a defined system/service will satisfy its particular requirements. An Assurance Case is a document that facilitates information exchange between various system stakeholders such as suppliers and acquirers, and between the operator and regulator, where the knowledge related to the safety and security of the system is communicated in a clear and defendable way.[12]

SACM provides a formal basis to reason about claims, arguments, supporting evidence, and the artifacts they involve, building up to an assurance case. While SACM is often used to analyze the claims to safety of a cyber-physical system, the applicability of the model to assurances of AI fairness is clear. It thus provides a framework to answer questions such as:

- What is being claimed about the system's fairness?
- What arguments support that claim?
- What tests have been run to try to disprove the proposition, and where is the evidence that those tests have not discovered a problem?

Along similar lines, Professor Roberto Zicari and his team at Goethe University Frankfurt's Big Data Lab propose a method called "Z-inspection," which can be used either as part of an "AI ethics by design" process or, if the AI has already been designed and deployed, to perform an "AI ethical sanity check."[13] The method has been tested on a use case in healthcare, a novel AI-based method for assessing the risk of coronary heart disease.

Bias is not easy to define … it depends on the goal of the system and who the stakeholders of the decision process are.

AI systems can also be tested periodically for fairness. The sensitivity of a model to an adversarial input (e.g., changing the race of a person) can be established and judged against predetermined goals. An AI tester could search for biases by running scenarios against a model representing a diverse user base, ensuring consistent and as-intended results. These scores should be communicated to end users, encouraging model designers to continually improve.

We should note that bias is not easy to define in the first place: it depends on the goal of the system and who the stakeholders of the decision process are. On the one hand, systems that help in medical diagnosis are often less precise for minorities and women because most training data sets have historically underrepresented those populations. On the other hand, you wouldn't ask for equal gender representation in a data set of chest X-rays if the goal is to detect breast cancer. At minimum, developers of AI and those who train the models should consciously consider what their goals are and whether the algorithms and data they use were developed in a way that is consistent with those goals.

Data Protection

AI thrives on data — the more, the better, both at training time and execution time. How do we ensure the data collected by AI is not interpreted and delivered in a context that violates our privacy?

A prominent aspect of this problem is the collection and use of people's images by law enforcement. It is said that if you stroll for hours in London, your entire journey can be retraced through CCTV footage: the coverage of the city is practically 100%. Today, with AI-based facial recognition, the system could not only trace someone's journey, but also know his or her race, gender, and approximate age — and perhaps even the

exact identity of that person. Absent even the possibility of informed consent to this recognition and tracking, cities and regions in democratic countries are increasingly banning facial recognition. Recently, in fact, several companies announced they will no longer develop this technology.[14] Ultimately, we need a "privacy by design" methodology that is specifically tailored for the unique ethical and legal challenges created by AI algorithms.

Another interesting approach is the use of generative adversarial networks (GANs) combined with differential privacy. The combination of these technologies is showing promise in creating data sets that so mimic real data as to still provide utility for training purposes while protecting personally identifiable information (PII).[15]

What will make the legal landscape of AI even more complex than past situations is the issue of responsibility and liability.

AI and the Law

Ultimately, especially in a litigious country such as the US, many issues end up being decided by the courts. Show some lawyers an ambiguous document, and they see money. What will make the legal landscape of AI even more complex than past situations is the issue of responsibility and liability.

Let's go back to our favorite example, self-driving cars. When an accident happens, even if the exact sequence and timing of events is established, who is responsible?

- The programmer who developed the software and forgot a safeguard?
- The person who selected the training data set?
- The car manufacturer, who didn't test the car in similar conditions to the ones that caused the accident?
- The passenger (if there was one at the time) who could conceivably have taken over and avoided the accident but wasn't paying attention since the car was in autonomous mode?

Clearly, a new legal framework will need to be developed, and it will take years, if not decades, to arrive at a consensus that can be applied consistently and fairly. We're likely to see the equivalent of the infamous "Caution: coffee may be hot" warning, and this will not reassure users. There will be dozens of pages of disclaimers included with AI-enabled products. There is already a controversy as to whether users can reasonably be expected to read all the fine print before using a product; this will get worse before it possibly gets better.

AI Ethics

Here the discussion easily borders on philosophy, so we can only scratch the surface. Some considerations that will drive trust (or mistrust) include the following questions, to which there are currently no definite answers:

- **Should an AI product inform its users that it is one?** In other words, while the Turing test may be meaningful *as a test*, a deployed AI system should perhaps inform its users that this is what they're interacting with. This is what the proposed EU regulations are getting at with the mention of labeling. Of course, if ultimately AI is going to pervade everything from our cars to our thermostats to our pet's water bowl, such labels become meaningless.
- **What is the modern equivalent of Asimov's three laws of robotics?** If all AI systems reliably and demonstrably followed at least the first law (do not harm humans), this would go a long way to build trust. But law enforcement, the military, and the governments of authoritarian countries will never agree to this.
- **Is having a human in the loop for higher-risk applications a good idea?** In the short term, most people probably think it would be a good thing, but we know that in many circumstances, humans make more mistakes than machines. Most fatal airplane crashes are due to pilot error, including many where the autopilot was disengaged. When people suggest human supervision as a solution, is this similar to having a flagperson walking in front of a car?
- **Should an AI system know when to lie?** Humans have a sense that truth can be harmful in certain circumstances. Should we impart the same emotional subtleties to AI, and would this actually decrease

trust? If you have an accident, should your car tell the police or your insurance company how fast you were driving?

- **Who is an AI application accountable to: its owner/user, a larger community, society at large?** Consider a self-driving car coming unexpectedly upon a group of children crossing the street. The car can brake but still hit the children, or it can veer off, crash into a wall, and perhaps kill you. Would you ride in a car that came with a warning that "if circumstances require, in the sole interpretation of the onboard AI software, this car may deliberately cause injury or death to its occupants"? Or should society accept that the car will prioritize the life of a single occupant over that of several children?

- **Is it ethical to develop technologies that may eliminate some jobs through automation without a comprehensive plan to offset the economic impact?** AI could be the key to a 20-hour work week, or it could be the trigger of extreme poverty, especially in countries where many people are employed at the jobs that would be the first to be displaced (e.g., call center operators). Yes, we've gone through several industrial revolutions that raised the same fears, and each time new types of jobs were created and the world didn't end. But AI could be different because *its very goal* is to replace humans in what seemed to be their ultimate irreplaceable capability: reasoning and intelligent decision making.

The Role of Standards

Finally, what is the current and future role of standards in codifying the above solutions? Here we draw on our experience as cochairs and members of the OMG's AI Platform Task Force, where we have been examining that very issue.

The word "standard" is used to cover a wide range of recommendations, from very technical ones (e.g., definitions of units of measure) to very general and often buzzword-laden frameworks or guidelines. Other so-called standards are simply glossaries that aim to reduce confusion around the use of words that may or may not be synonyms.

With this in mind, Table 1 lists some (but certainly not all) efforts that are relevant to solving the AI trust gap. Their diversity and varying level of precision and obligation are clear indications that this domain is not mature and that much work remains to arrive at usable norms.

The Fork in the Road

"A little learning is a dangerous thing," wrote Alexander Pope. We think we're at that point in society's understanding of AI, and therefore its ability to trust it or its inclination to reject it.

The utopian view is that if we make the right decisions, AI will permeate our environment and bring unprecedented efficiencies, security, life-saving assistance to the sick and the elderly, and personalized top-notch education while freeing us from so many routine tasks that we can maintain or improve prosperity while working less.

> *We do have control — for now — over whether AI will be a trusted aide to humanity or a threat.*

The dystopian view, abundant in sci-fi literature, is that AI will be uncontrolled and used for nefarious purposes by corporations and governments alike: humans will lose control, and HAL will never open the pod bay doors. Alternatively, a new generation of Luddites will emerge and reject or destroy the machines. If today's conspiracy theorists believe that COVID-19 vaccines are used to implant us with microchips, imagine their reactions if they understood what deep learning can do!

We've been there before, notably with our ability to split and merge atoms. Although we have mastered the use of nuclear fission to produce energy, not just to destroy enemies, most people still do not trust nuclear power, and the rare accidents create a disproportionate reaction to the actual damage.

We do have control — for now — over whether AI will be a trusted aide to humanity or a threat. But we need a number of convergent approaches to be able to choose the right path. We need an ethical framework, we need to understand our own biases and avoid conveying them to the machines, we need certain laws and regulations to address liability issues, and we need some standards to codify our commitment. Ultimately, this amounts to AI governance.[16]

Organization	Product	Description
International Organization for Standardization (ISO)	**ISO/IEC AWI 25059: Quality Model for AI-Based Systems** (see: https://www.iso.org/standard/80655.html)	This is an Approved Work Item (AWI), under development as of this writing, of the JTC 1/SC 42 Subcommittee on AI.
	ISO/IEC DTR 24027: Bias in AI Systems and AI-Aided Decision Making (see: https://www.iso.org/standard/77607.html)	Draft technical report under review.
	ISO/IEC TR 24028:2020: Overview of Trustworthiness in Artificial Intelligence (see: https://www.iso.org/standard/77608.html)	Includes approaches to establish trust in AI systems through transparency, explainability, controllability, etc.; engineering pitfalls and typical associated threats and risks to AI systems, along with possible mitigation techniques and methods; and approaches to assess and achieve availability, resiliency, reliability, accuracy, safety, security, and privacy of AI systems.
	ISO/IEC AWI TR 24368: Overview of Ethical and Societal Concerns (see: https://www.iso.org/standard/78507.html)	AWI under development.
Institute of Electrical and Electronics Engineers (IEEE)	**IEEE P7000™: Model Process for Addressing Ethical Concerns During System Design** (see: https://ethicsinaction.ieee.org/p7000/)	"Addresses ethical considerations at each stage of development to help avoid negative unintended consequences while increasing innovation."
	IEEE P7001™: Transparency of Autonomous Systems (see: https://ethicsinaction.ieee.org/p7000/)	"A standard for developing autonomous technologies that can assess their own actions and help users understand why a technology makes certain decisions in different situations."
	IEEE P7003™: Algorithmic Bias Considerations (see: https://ethicsinaction.ieee.org/p7000/)	"The project offers specific steps to take for eliminating issues of negative bias in the creation of algorithms."
Association for the Advancement of Artificial Intelligence (AAAI)	**AAAI Code of Professional Ethics and Conduct** (see: https://aaai.org/Conferences/code-of-ethics-and-conduct.php)	"The Code is designed to inspire and guide the ethical conduct of all AI professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses AI technology in an impactful way, and is particularly intended to act as a standard of ethical and professional conduct for all AAAI members."
National Institute of Standards and Technology (NIST)	**Draft NISTIR 8312: Four Principles of Explainable Artificial Intelligence** (see: https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf)	The four principles are Explanation, Meaningful, Explanation Accuracy, and Knowledge Limits.
Object Management Group (OMG)	**Structured Assurance Case Metamodel™** (see: https://www.omg.org/spec/SACM/2.1/About-SACM/)	See section on SACM in main article.
	AI Platform Task Force (see: https://www.omg.org/ai)	Created in late 2019, this OMG subgroup is developing a roadmap and considering the feasibility of issuing requests for proposals for various specifications, including on explainability and bias testing.
American Council for Technology and Industry Advisory Council (ACT-IAC)	**Ethical Application of Artificial Intelligence Framework** (see: https://www.actiac.org/act-iac-white-paper-ethical-application-ai-framework)	A white paper proposing four components to ethical application of AI: bias, fairness, transparency, responsibility, and interpretation.
European Commission	**Ethics Guidelines for Trustworthy AI** (see: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419)	Document prepared by the High-Level Expert Group on Artificial Intelligence (AI HLEG), an independent expert group set up by the European Commission in June 2018.

Table 1 — Efforts to solve the AI trust gap.

There are multiple initiatives to define the elements of such governance: AAAI, IEEE, ISO, the European Commission, and many other organizations are conducting parallel efforts as seen in the above standards table. To move "from slideware to software," a global consortium of industry leaders could focus on providing open source technologies, tools, and resources to make AI more secure and explainable, and therefore more trusted. These resources could aggregate and create AI impact assessments to ensure the least intrusive and most explainable models are being used, that the data driving models is updated and secure, that bias audits are conducted, attestations that the model is only being used for its intended purpose are generated, and that privacy-friendly technologies are being utilized where available. However, we've seen a deep sense of mistrust emerge in the past few years about the AI work of the private sector (e.g., Facebook and Cambridge Analytics), and it is unlikely that self-regulation, similar to what has been put in place for Payment Card Industry Data Security Standard (PCI DSS) compliance, will be sufficient to ensure trust.

We may also want to rethink the appropriateness or speed of developing artificial general intelligence (AGI). Similar to the trend in privacy to not collect or use more personal information than is necessary to perform the intended function, AI might, for now, be as limited in scope as possible to perform the intended result. Such narrower applications of AI will help create success stories while limiting risks, helping generate society's understanding of what AI can and cannot do and therefore enabling us to trust the technology and reap its benefits.

References

[1]Rooney, Ben. "Women and Children First: Technology and Moral Panic." *The Wall Street Journal,* 11 July 2011.

[2]See Wikipedia's "Three Laws of Robotics."

[3]See Wikipedia's "HAL 9000."

[4]Stanford University School of Engineering. "Dan Jurafsky: How AI Is Changing Our Understanding of Language." YouTube, 15 March 2021.

[5]Grother, Patrick, Mei Ngan, and Kayee Hanaoka. "NISTIR 8280: Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects." National Institute of Standards and Technology (NIST), December 2019.

[6]See Wikipedia's "Adversarial Machine Learning."

[7]"On Artificial Intelligence: A European Approach to Excellence and Trust." White paper, European Commission, 19 February 2020.

[8]"Statement on Section 5 'An Ecosystem of Trust' Within the White Paper on AI of the European Commission." Federation of German Industries (BDI), 18 June 2020.

[9]Paudyal, Prajwal. "Should AI Explain Itself? Or Should We Design Explainable AI So That It Doesn't Have To?" Towards Data Science, 4 March 2019.

[10]Gunning, David, and David W. Aha. "DARPA's Explainable Artificial Intelligence Program." Association for the Advancement of Artificial Intelligence (AAAI), *AI Magazine,* 24 June 2019, pp. 44-58.

[11]Krishna, Ranjay. "Compositionality in Computer Vision." *Compositionality in Computer Vision*, virtual workshop, 15 June 2020.

[12]"Structured Assurance Case Metamodel Specification, Version 2.1." Object Management Group (OMG), April 2020.

[13]Zicari, Roberto, et al. "Z-Inspection: Towards a Process to Assess Ethical AI." Cognitive Systems Institute Group (CSIG), 21 October 2019.

[14]Urban, Holly A. "Reputable Companies Rejecting Facial Recognition Due to Racial Bias." Lexology, 5 April 2021.

[15]Xie, Liyang, et al. "Differentially Private Generative Adversarial Network." Cornell University, February 2018.

[16]Durmus, Murat. "A Collection of Recommendable Papers and Articles on AI Governance." LinkedIn, 19 October 2020.

Claude Baudoin is a Senior Consultant with Cutter Consortium's Business & Enterprise Architecture and Data Analytics & Digital Technologies practices and a member of Arthur D. Little's AMP open consulting network. He has been Owner and Principal Consultant of cébé IT and Knowledge Management, a boutique consulting practice, since 2009. Prior to that, Mr. Baudoin spent 35 years in IT and software management roles in industry, including 26 years at Schlumberger, the global oilfield services company, in France and the US. Since 2015, he has served as an energy domain advisor to the Object Management Group (OMG) and the Industrial Internet Consortium (IIC) and cochairs OMG's AI Platform Task Force. Mr. Baudoin has authored two books on software engineering and numerous papers and conference presentations. He holds two patents related to IT infrastructure and security. Mr. Baudoin's undergraduate engineering degree is from École Polytechnique in Paris, and he holds a master of science degree in computer science from Stanford University. He can be reached at experts@cutter.com.

Clayton Pummill is a Strategic Account Manager for Sepio Systems and a member of the board of advisors for Tesseract Ventures. Mr. Pummill is a licensed attorney, focusing on the areas of machine learning, cybersecurity, and data privacy. He is active in the technology startup space, having previously developed patented technology, cofounded organizations, and brought them through to successful exits. Mr. Pummill also serves on the Object Management Group's (OMG's) AI Platform Task Force and on the advisory board of the Security Advisor Alliance. He also holds International Association of Administrative Professionals (IAAP) CIPP-US/EU/C certifications. Mr. Pummill earned a law degree from the University of Missouri School of Law and an MBA from Washington University's Executive MBA Program. He can be reached at fishtree48@protonmail.com.

A WORLD OF POSSIBILITY

AI in Education: Applications & Impact

by Aswani Kumar Cherukuri, Annapurna Jonnalagadda, and San Murugesan

Artificial intelligence (AI) and machine learning (ML) technologies have progressed significantly and rapidly in the past few years. Almost every sector, including education and training, is being influenced by these technologies.

Until about 10 years ago, the education sector had not embraced AI because it lacked the required level of digitalization. Today, the education sector is witnessing a massive transformation from "chalk-and-talk" classrooms to "click-and-learn" digital environments. As education and learning become predominantly digital, there are opportunities to enhance them through AI and other technologies, such as augmented reality (AR), virtual reality (VR), and cloud computing. According to a report from Research and Markets, the AI market in the education sector is estimated to reach US $3.68 billion by 2023 and $5.8 billion by 2025. The report predicts the education sector to grow at a CAGR of nearly 48% during the period 2018-2022.[1]

AI is helping educators create and curate customizable digital content that lets grade school and higher ed students learn in a more personalized way using skill mapping and microlearning.

In this article, we discuss how AI impacts the teaching and learning experience and the quality of education. We discuss an AI technology stack that can be applied to various educational processes and provide a summary of AI applications and platforms, along with their key features. We also look at why a cautious approach is best when adopting AI technology for education.

Impact of AI in Education

AI-based learning platforms are being used to personalize curriculum, content, online classes, and learning assessments. This is having a positive effect on students, teachers, educational institutions, and business professionals in various ways:

- Students receive adaptive, personalized lessons that help them more easily understand complex concepts, along with customized assessments that identify their competencies and learning gaps.
- Educators gain deeper insights into students' learning styles, helping them analyze student performance and knowledge gaps and follow data-driven teaching methods.
- Teachers improve their efficiency by automating managerial activities like assessments.
- Institutions can offer higher-quality, more affordable education and achieve better student retention, thanks to the scale of online education.
- Business professionals can improve their skills and explore new knowledge areas.

AI is helping educators create and curate customizable digital content that lets grade school and higher ed students learn in a more personalized way using skill mapping and microlearning. Using AI-powered learning paths, platforms like Coursera and edX offer courses aimed at both students and professionals worldwide. With on-demand digital education, students and professionals gain knowledge and skills at their convenience and on their own timeline.

As students access digital content, educators get real-time feedback on each student's knowledge gaps and areas of difficulty. AI systems also help educators create personalized learning paths for students. AI-fueled analyses provide detailed information about where students are failing and, more importantly, why they are failing.

Intelligent assistants and autonomous chatbots provide assistance to students, reinforcing the concepts being taught, and they can challenge students with micro-tasks. At the higher education and research level, AI-based graphics processing units like those from Nvidia

provide cost-effective, efficient hardware and software stacks for high-performance computing. This infrastructure equips graduate students and researchers to accelerate AI-based research in various scientific fields.

AI is helping educators reduce the amount of time they spend planning, scheduling, and doing managerial tasks. AI facilitates automation in student admissions, monitoring and alerting about student absenteeism, budgeting, HR management, and parent interactions. In general, it enhances learning outcomes, productivity, and employee engagement in corporate learning and training. AI-powered personalized training, digital learning assistants, data-driven insights, and feedback are proving effective for both employees and organizations. Students tend to dislike learning through reinforcement (learning from failure); using gamification, AI-powered educational platforms allow and even encourage students to explore and learn from failures.

Here are the key benefits that AI-based education offers over traditional instruction modes:

- Adaptive and gamified learning
- Automated content preparation and curation
- Efficient proctoring and automated assessment
- Analytics-driven student performance analysis
- Adaptive team formation and collaboration
- Personalized skills development and competency building for professionals
- Skills enhancement, including language, comprehension, critical thinking, and problem solving
- Tutoring and mentoring
- Immersive learning

A number of AI technologies are being applied in education to achieve these advances. Table 1 provides a brief summary of the technology stack.

AI-Powered Educational Platforms

Many AI-powered educational platforms and applications are now available to help educators leverage AI. We've grouped these platforms based on their features to create Tables 2-9, of which applications, listed in no particular order, enable these functions:

- Adaptive and personalized learning
- Content preparation
- Proctoring and assessment
- Online learning and immersive learning through AR/VR
- Language learning
- Coding and robotics
- Tutoring and mentoring
- Management and scheduling

AI is helping educators reduce the amount of time they spend planning, scheduling, and doing managerial tasks.

The platforms in Table 2 use AI to analyze the knowledge gaps in learning and offer personalized learning paths to the students. Some of them provide insights into why these learning gaps arise and create recommendations for educators.

Creating lesson content and study materials for the students is an important task. AI is helping educators create and curate content from multiple sources. Table 3 briefly describes a few AI-powered educational platforms designed to assist with content preparation.

Learning assessments and grading are essential educational processes. AI-based tools are helping educators conduct learning assessments and gain insights into student learning gaps. Table 4 highlights a few such platforms.

Online learning platforms leverage AI technology to personalize learning paths and understand gaps in learning and teaching. AR/VR-assisted learning platforms provide an immersive learning experience that can lead to a richer understanding of complex concepts. These AI-powered platforms allow educators to understand and analyze students' cognitive responses during the learning process. Table 5 presents some of these tools.

Technology Stack	Application
Natural language processing (NLP)	NLP can be used to automate essay grading. AI-powered speech-to-text applications can reduce language barriers. Real-time language translation and online lecture scripts enhance knowledge transfer.
Big data and analytics	Big data algorithms can collect billions of data points based on student interaction with digital educational platforms. AI-powered analytics can use that data to diagnose student performance and customize course content.
Visualization	Visualization can be used to gain deeper insights into student performance and knowledge gaps. Visualization models help educators more thoroughly explain complex concepts.
Neuroscience and cognitive models	AI models are created by linking cognitive neuroscience with educational technology to build personalized learning applications.
Chatbots and digital twins	Chatbots, digital twins, and avatars can assist students, serving as personal tutors. Chatbots can answer common queries; digital twins can assist in problem solving. Voice assistants and chatbots can help students learn in real-life environments and offer human-like interaction.
Recommendation models	Recommendation models can be coupled with AI techniques to help students select suitable courses, pursue relevant learning paths, and track actionable items.
Gamification	AI-enabled gamification can motivate learners to actively participate. AI can identify learning traits and habits while gamification improves retention and makes learning fun.
Audio and video processing	Facial recognition, speech recognition, and audio/video processing techniques can be applied to automated proctoring systems designed to conduct assessments.
Augmented reality (AR) and virtual reality (VR)	AR/VR, combined with AI, can provide immersive learning experiences in which students gain insights into subjects through interaction with virtual objects representing physical concepts.

Table 1 — Technology application in education.

Educational Platforms	Features
Alta (Knewton)	Alta is an integrated, adaptive learning technology providing a personalized learning experience. With each interaction, the platform identifies the competency levels of the students and their knowledge gaps and adapts to address those concerns. It also assists instructors in teaching mathematics, science, statistics, and economics courses at various levels.
Cognii	Cognii provides AI-based education technology solutions for K-12, higher ed, and corporate training. Students receive personalized lessons and intelligent tutoring. Teachers receive data and analytics about students' learning habits. Short essay auto-grading reduces teacher grading burdens.
CENTURY	CENTURY is an educational technology platform based on AI, learning science, and neuroscience. It provides personalized learning solutions for English, math, and science at various grade levels. Using the AI-powered engine, the platform can identify knowledge gaps. It also identifies and addresses student misconceptions during lessons. CENTURY adapts to individual learning habits and is suitable for both classroom and online learning. It helps teachers identify the students who require additional input, and its recommendation engine suggests suitable micro-lessons as "nuggets." Analytics-driven dashboards for students, parents, teachers, and administrators provide insights into pupil performance.
Carnegie Learning	Carnegie Learning uses AI to help teachers understand students' learning behaviors and habits, particularly at K-12 level. AI-enabled solutions reduce learning loss, accelerate learning, and help teachers monitor student performance. The platform offers student communities and allows educators to connect and share ideas/ resources.
Sana Labs	Sana Labs offers personalized learning using ML-powered engines for assisted learning, course creation, and learning analytics.
Fishtree	Fishtree provides scalable, personalized learning. The platform offers AI- and analytics-powered learning solutions, such as dynamic authoring, automatic resource generation, and real-time data analytics to K-12, higher ed, and corporate trainers.
Realizeit	Realizeit offers personalized corporate training solutions at scale for workforce development. Courses and content are customized to close participant knowledge gaps. For academic institutions, the platform offers an adaptive learning experience to students as well as tools to reduce educators' administration burden.
DreamBox	DreamBox provides differentiated learning and interactive sessions for mathematics. It helps teachers track student progress, identify proficiency gaps, and create small, virtual study groups.
Volley	Volley is focused on AI-based customized talent recruitment and training.
ALEKS	ALEKS is an AI-based learning and assessment system that reviews student knowledge and recommends learning paths. It uses "knowledge space theory" principles and identifies student boundaries, offering learning topics based on those boundaries.

Note: Descriptions of platform features come from each vendor's website.

Table 2 — AI-powered educational platforms for adaptive and personalized learning.

Educational Platforms	Features
Dragon Speech Recognition (Nuance)	Dragon Speech Recognition transcribes speech at rates of up to 160 words per minute to help teachers prepare lesson plans and other materials automatically using dictation. It also assists students experiencing writing and pronunciation challenges. Students can orally express themselves and eliminate the typing barrier.
Content Technologies	Content Technologies provides AI-powered instructional design and content application solutions.
Rev	The Rev tool provides automated audio-to-text transcription that is useful in turning lecture transcripts into study materials.

Note: Descriptions of platform features come from each vendor's website.

Table 3 — AI-powered educational platforms for content preparation.

Educational Platforms	Features
Querium	Querium offers an AI-based virtual tutor that works on smartphones and systems, delivering short, personalized lessons with tutorial assistance. The platform focuses on STEM topics and student performance evaluation. As part of analyzing answers, the engine considers several parameters to understand depth of student learning, including types of mistakes made by student while answering, number of hints requested by student, type of hints provided, amount of time spent by student on each step in a solution and on each solution in total, skills employed in solving the problem, and overall accuracy in problem solving.
MAP Suite (NWEA)	MAP Suite is a set of assessment and learning tools that provides insights on student progression over time as well as knowledge gaps. It allows educators to adjust instructions according to student need.
Gradescope	Gradescope helps teachers grade assignments, homework, quizzes, programming problems, and even paper-based exams. It can "read" paragraphs, theorems proofs, and more. It also helps educators manage regrade requests from students and uses rubric-level statistics to gain insights into student achievements.
Kidaptive	Kidaptive is an AI-powered, cloud-based adaptive learning solution that can be integrated with any digital learning platform. Educators can track, analyze, and predict student performance using AI. It can collate student learning data from multiple sources, such as learning activities, assessments, educator feedback, and parent input, and create a holistic map of a student's learning patterns.
HireMee ProEx	HireMee ProEx is an AI-powered online examination system for entrance exams, certifications, and final exams.
neo Exam (iamneo.ai)	neo Exam offers an AI-driven assessment platform to securely conduct online examinations and automatically grade them. An AI-enabled proctoring feature helps students complete online tests without human interaction. The platform also supports corporate recruitment and upskilling.

Note: Descriptions of platform features come from each vendor's website.

Table 4 — AI-powered educational platforms for proctoring and assessment.

Educational Platforms	Features
Coursera	Coursera leverages AI to identify gaps in lectures and lesson materials based on student performance.
BYJU'S	The BYJU'S platform provides online learning programs for students from pre-K-12. It provides personalized tutoring services using visualization techniques and AR models to provide conceptual clarity.
WizKlub	WizKlub offers programs to enhance the critical thinking, comprehension, and complex problem-solving abilities of children aged six to 14. The technology is powered with AI and cognitive learning to offer a personalized approach, structured curriculum, and competitive challenges.
LinkedIn Learning	LinkedIn offers personalized learning for competencies building and skills enhancement. Analytics-driven recommendations on the courses are based on the learner's profile and those of similar learners.
Blippar	Blippar provides AR-based solutions that help educators increase interactivity and enhance student engagement through visualizing complex concepts with AR and gamification.
3rdFLIX	3rdFLIX uses AR for experiential learning and VR for immersive learning, mixed reality, and 3D model simulations. It provides an assessment platform and analytics.

Note: Descriptions of platform features come from each vendor's website.

Table 5 — AI-powered educational platforms for online learning and immersive learning through AR/VR.

Educational Platforms	Features
Grammarly	Grammarly provides NLP and AI-powered writing assistance. It offers an intuitive interface to identify and rectify language-related communications issues.
WritetoLearn (Pearson)	WritetoLearn provides a Web-based platform for enhancing writing, reading speed, and reading comprehension.
VIPKid	VIPKid is a customized solution to help educators/students with English learning.
Babbel	Babbel provides a platform for learning languages based on cognitive techniques, speech recognition, and analytics.
Duolingo	Duolingo is a gamified, personalized application for learning languages. Content is tailored toward an individual's learning style; teachers can use it to enhance lessons.

Note: Descriptions of platform features come from each vendor's website.

Table 6 — AI-powered educational platforms for language learning.

Educational Platform	Features
Tynker	Tynker is an interactive, gamified, story-based learning platform to help students ages five to 18 learn programming, critical thinking, and problem solving.
CodeHS	CodeHS is a comprehensive, data-driven learning platform for teaching and learning computer science and programming. It supports educators and managers with professional development.
neo Colab (iamneo.ai)	neo Colab is an AI and analytics-driven online programming platform that enables students to learn programming languages. It helps educators efficiently mentor students and monitor their progress.

Note: Descriptions of platform features come from each vendor's website.

Table 7 — AI-powered educational platforms for coding and robotics.

Educational Platform	Features
Thinkster Math	Thinkster Math provides an intelligent tutoring system that helps educators and parents understand why students are having trouble solving math problems.
Quizlet	Quizlet assists students in learning languages, science, and math using flashcards and gamification and breaking down topics in a customized manner.
Squirrel Ai Learning	Squirrel Ai Learning offers personalized after-school tutoring services based on AI-powered adaptive learning engines.
Classcraft	Classcraft is a gamification- and analytics-driven solution to understand student behavior and provide support through personalized strategies.

Note: Descriptions of platform features come from each vendor's website.

Table 8 — AI-powered educational platforms for tutoring and mentoring.

Educational Platform	Features
Abl Master Scheduler	Abl Master Scheduler is a data-driven platform that enables school managers/ administrators to plan, schedule, and manage resources to improve the student experience.
Creatrix Campus	Creatrix Campus is an AI-driven, cloud-based platform that combines customer relationship management with learning management systems and student information systems.

Note: Descriptions of platform features come from each vendor's website.

Table 9 — AI-powered educational platforms for management and scheduling.

Several AI powered platforms aim to help students learn languages and improve their communication skills. Some of them help educators improve their course materials. Table 6 offers a brief summary of a few such platforms.

There are few AI-based platforms that teach students coding, programming, and basic robotics. Table 7 highlights a handful of platforms.

There are several AI-powered platforms that assist students by providing personalized tutoring. Some of these platforms use gaming principles. Table 8 illustrates some platforms.

In addition to teaching and learning, AI-based solutions can assist institutes with management and resource utilization. Table 9 provides a brief summary of two such platforms.

Final Thoughts

AI's adoption by the education sector is in a nascent stage, with vast opportunities for growth as technologies and solutions mature. There are some caveats, however.

First, machine interaction should not take the place of human interaction, and we should not ignore the importance of soft skills, such as peer interaction, teamwork, and collaborative learning.

Second, there are ethical aspects of AI to be addressed, including bias and the potential for discrimination. There is evidence of human biases in data annotations that result in biased AI algorithms. The amount of data and training data sets is limited in the education sector, so reducing algorithmic biases in educational applications can be challenging.

Finally, there's a need for extensive research on the effects of the pedagogy, instructional models, and learning paths within online learning applications on the overall development of student learning.

Aspects such as empathy, experiences, use case identification, student motivation, role-model pursuit, personal interactions, and cognitive connections are all important and cannot be taught by algorithms. AI's role should be to assist educators, not replace them, while helping students learn more efficiently.

Reference

[1] "Artificial Intelligence Market in the US Education Sector 2018-2022." Research and Markets, August 2018.

Aswani Kumar Cherukuri is a Professor at Vellore Institute of Technology (VIT), India. His research interests include information security and machine learning. Dr. Ch. Aswani Kumar earned the Young Scientist Fellowship from Tamilnadu State Council for Science and Technology and was awarded the Inspiring Teacher Award from The Indian Express *(India's leading English daily newspaper). He has worked on various research projects funded by the Government of India's Department of Science and Technology, Department of Atomic Energy, and the Ministry of Human Resources Development. Dr. Ch. Aswani Kumar has published more than 150 refereed research articles in various national/international journals and conferences and is an editorial board member for several international journals. He is a Senior Member and distinguished speaker of the Association for Computing Machinery (ACM), a member of the Institute of Electrical and Electronics Engineers (IEEE), and Vice-Chair of the IEEE Taskforce on Educational Data Mining. Dr. Ch. Aswani Kumar earned a PhD in informational retrieval, data mining, and soft-computing techniques from VIT. He can be reached at cherukuri@acm.org.*

Annapurna Jonnalagadda is an Associate Professor at Vellore Institute of Technology (VIT), India. Her research interests include artificial intelligence, game theory, and social network analysis. Dr. Jonnalagadda has professional experience at both IBM and the Government of India's National Informatics Center (NiC) and more than 11 years' experience in academia. She has worked on two major research projects funded by the Government of India as co-investigator. Dr. Jonnalagadda has been published in various journals, including Physica-A, Knowledge & Information Systems, *and* Social Networks Analysis & Mining. *She earned a master's of technology and a PhD in computer science and engineering from VIT. She can be reached at jannapurna@gmail.com.*

San Murugesan (BE [Hons], MTech, PhD; FACS) is a Senior Consultant with Cutter Consortium's Data Analytics & Digital Technologies practice and a member of Arthur D. Little's AMP open consulting network. He is also Director of BRITE Professional Services and former Editor-in-Chief of IT Professional. *Dr. Murugesan has four decades of experience in both industry and academia, and his expertise and interests include artificial intelligence, quantum computing, the Internet of Everything, cloud computing, green computing, and IT applications. He offers certificate training programs on key emerging topics and keynotes. Dr. Murugesan is coeditor of a few books, including* Encyclopedia of Cloud Computing *and* Harnessing Green IT: Principles and Practices. *He is a fellow of the Australian Computer Society, Golden Core Member of IEEE Computer Society, and Life Senior Member of IEEE. Dr. Murugesan has held various senior positions at Southern Cross University, Australia; Western Sydney University, Australia; the Indian Space Research Organization, Bangalore, India; and also served as Senior Research Fellow of the US National Research Council at the NASA Ames Research Center. He can be reached at experts@cutter.com.*

A LITTLE THIS, A LITTLE THAT

How Will AI Transform Everyday Life?

by Jayashree Arunkumar

Technology is not the solution to all our problems, but artificial intelligence (AI) does have the potential to transform the way we conduct our everyday lives. Which begs the question: just how powerful are today's AI solutions, and which areas are showing the most promise? The answer lies in five interesting AI trends and how that technology is being slotted into real-world use. It also lies in several AI projects aimed at helping the environment and how similar efforts might be used to improve healthcare.

5 AI Trends: 2021 and Beyond

1. Composite AI

Composite AI involves combining multiple AI techniques to achieve a better outcome. For example, graph-accelerated machine learning (ML) helps optimize AI models and speed up AI processes.[1]

There are three steps to building graph-accelerated ML models:

1. Aggregate data from a variety of sources using a tool such as Apache Spark (open source).
2. Build out the graph and view possible data relationships using a graph database tool such as Neo4j.
3. Send the completed graph to the ML pipeline.

Leading organizations like Amazon, Google, and Facebook are using graph-accelerated ML to develop recommendation systems, combinatorial optimization, and computer imaging, respectively.[2] The context that graphs can add to AI applications means we're likely to see an increase in roles like graph AI engineer, ontology engineer, and AI ethicist in the near future.

Here are a few use cases for data science graphs:

- **Query-based knowledge graphs**
  - AstraZeneca uses knowledge graphs to build its understanding of disease.[3]
  - NASA uses knowledge graphs to extract knowledge from its Lessons Learned database.[4]
- **Query-based feature engineering**
  - Hetionet is an open source heterogeneous information network of biomedical knowledge. It uses query-based feature engineering to help scientists predict whether a compound will have an effect on a disease, helping them explore new uses for existing drugs.[5]
- **Graph embeddings**
  - Amazon's DGL-KE helps make knowledge graph embeddings easier to use.[6]
- **Graph neural networks**
  - Twitter uses Fabula AI to detect social network manipulation.[7]
  - Alibaba leverages AliGraph for e-commerce recommendations.[8]
  - Uber Eats uses GraphSAGE for recommending dishes and restaurants to users.[9]
  - Google Maps leverages DeepMind to improve Google Maps services recommendations.[10]

2. Generative AI

Generative AI refers to the technique of generating new sample data (picture, voice, or text) from a training data set. The model looks at a sample training data set to evaluate how it was created and uses the probability distribution to generate completely new samples. For example, given a set of images of people's faces, a generative model can create photos of imaginary people by inferring from the probability distribution of the training data set.

There are two neural networks involved in this process. One is called the "generator"; the other is the "discriminator." During training, the generator creates the fake data, and the discriminator classifies it as fake. The

generator iteratively improves the quality of the fake data until the discriminator identifies the fake content as real.

Here are some generative AI use cases:

- **Dental restoration.** Generative AI can be used to design dental crowns, improving fit and reducing the number of times the patient must visit the dentist during the process.
- **Life support.** A NASA contractor used this technique to optimize the design of astronauts' life support backpacks. The design engine allowed the engineering team to explore multiple options that fit within the project's defined constraints to generate the most effective design.[11]
- **Improve space medicine.** NASA scientists developed a way to use AI-synthesized biosensor data to simulate potential health conditions that could impact astronauts.[12]
- **Automated videos.** Reuters worked with AI startup Synthesia to create a fully automated presenter-led sports news summary system.[13]

The downside of generative AI is its potential use by hackers and other bad actors. For example, the CEO of a UK energy company transferred money to a supplier based on a phone call in which he believed he was talking to his boss at the firm's parent company.[14]

To prevent this type of crime, organizations will need tools such as Microsoft Video Authenticator, which can help users detect manipulated photos and videos.[15]

3. Edge AI

Smart devices are expected to generate as much as 175 zettabytes of data by 2025.[16] This volume will create challenges for cloud storage, data transmission, and data processing.

Edge computing eases this burden and lowers latency by bringing storage and processing closer to the location where it's needed. The data generated still needs to be analyzed, of course, so AI capabilities are being developed and hosted on local edge servers.

5G networking will be needed to support these connections. Although still in its nascent stage, 5G will mature considerably and become mainstream in the next five years or so. The combination of edge computing, edge cloud, 5G, AI, and open source will drive advanced solutions just now being considered.

The downside of generative AI is its potential use by hackers and other bad actors.

Edge AI has four domains:

1. Edge caching
2. Edge training
3. Edge inference
4. Edge offloading

Table 1 describes the most common edge use cases and the edge domains associated with them.

The biggest challenge in developing edge AI solutions is achieving high performance, given the constraints of today's Internet of Things devices. The next challenge is scaling complex applications. For example, hundreds of cameras and sensor nodes can be installed in a city setting, but scaling the edge solution in such a setup is complex. To address these challenges, researchers have proposed a combination of specialization design methods (on-device training, software design, hardware design, automation) and co-design methods (software/

	Edge Domains			
Edge Use Cases	**Edge Caching**	**Edge Training**	**Edge Inference**	**Edge Offloading**
Connected sensors in health monitoring	Applicable	n/a	n/a	Applicable
Connected cameras in crime prevention	n/a	Applicable	Applicable	Applicable
Connected cars in smart traffic	n/a	Applicable	Applicable	Applicable

Table 1 — Common edge use cases and the edge domains associated with them.

hardware co-design, software/compiler co-design, hardware/compiler co-design).

Technical guidelines around privacy, security, and environmental considerations for future developments in edge AI are also in the exploratory stage.

4. Artificial General Intelligence

Currently, AI solutions are fairly narrow. For example, the AI designed for self-driving cars can't be used to drive a truck or motorcycle. To get to the next level, AI must be able to educate itself and learn like humans, a characteristic usually referred to as artificial general intelligence (AGI). (AI that can surpass human intelligence is referred to as artificial superintelligence.) Table 2 shows the three stages of AI development.

In its simplest form, AGI consists of thousands of ML models working together to solve a complex problem. AGI is all about building a human-like system. Indeed, the AI solutions we see today are not really AI; they don't have flexible, general-purpose intelligence.

However, there are some interesting developments in this area. Given the pace of development in the last decade, futurist Ray Kurzweil predicts we're about 10 years away from AGI. Kurzweil predicts AGI will likely pass a valid Turing test by 2029 and surpass human intelligence by 2045.[17]

The most prominent AGI systems right now are OpenNARS, OpenCog, and AERA:

- **OpenNARS (Open Non-Axiomatic Reasoning System)** is an open source, general-purpose AI system that focuses on building a thinking machine.
- **OpenCog** is an open source project focused on building AGI capabilities that are equivalent to, or better than, human capabilities. It is novel architecture for AGI, based on a hypergraph knowledge store called AtomSpace.
- **AERA (auto-catalytic endogenous reflective architecture)** demonstrates numerous operational features necessary to achieve AGI using domain-independent learning, cumulative incremental learning, transfer learning, time-sensitive resource management, and long-term scalability.

On the corporate side, Microsoft is working with OpenAI to develop human-like solutions, and Google DeepMind is working to advance AGI.

5. AI and Coding

AI has some interesting roles to play in coding. The first is in finding and fixing human errors in code to help products get to market faster and with fewer problems. This approach frees up engineers to work on contextualization, customization, and problems involving deep logic requiring human intervention. Here are some examples:

1. **SketchAdapt** is a framework developed by MIT that combines pattern matching and symbolic search techniques to generate high-level program structure and low-level detailed coding. An MIT study found the framework performs better than Microsoft DeepCoder.[18]
2. **Code processing tools like Eclipse and Visual Studio** have built-in language models that can help engineers by proposing pluggable code snippets. The problem is that these tools don't currently guard against hackers injecting variables into the snippets that can harm the application in production. IBM and MIT co-developed a tool that can spot the weak points in the code generated by code processing tools and ensure robustness against adversarial attacks.[19]
3. **Intel worked with MIT and the Georgia Institute of Technology** to develop an automated engine that can improve engineering productivity by

AI Stage	Description
Artificial narrow intelligence	Today's AI systems/applications — domain-specific, narrow AI (e.g., image recognition, speech recognition)
Artificial general intelligence	Human-like AI that combines multiple models to solve a complex problem
Artificial superintelligence	AI that is superior to the human brain; machines acquire enormous knowledge and skills, surpassing human intellectual capacity

Table 2 — The three stages of AI development.

learning what the code is trying to accomplish and providing recommendations for optimal ways to get to the goal.[20]

The second role for AI is in no-code platforms that allow citizen developers to create apps on their own, without IT. Google, Apple, Microsoft, and Amazon all rolled out no-code AI solutions between 2017 and 2020, and 12 coding/AI-related venture capital–backed companies received funding in 2020 (amounts ranged from US $2 million to $750 million).[21]

How AI Helps the Environment

In 2015, the United Nations set forth a series of Sustainability Development Goals (SDGs) intended to "achieve a better and more sustainable future for all."[22] Digital technology will be critical to meeting those goals, and AI has the potential to dramatically speed up their delivery. AI can accelerate the pace of delivering SDG goals at scale. Here are two examples:

1. **Google.** AI applications in the area of wind and solar power from Google DeepMind is a step change toward renewable energy goals. In 2019, Google launched an accelerator program that supports tech startups that support SDG goals.[23]

2. **Microsoft.** Glacier melting is directly related to climate change, perhaps today's most pressing global concern. Microsoft's AI for Good Research Lab is working with several organizations to understand the extent of glacier melting in the Himalayas and how to minimize its impact.[24] Microsoft has also outlined plans to further align its technology initiatives to SDGs.[25]

These efforts are excellent first steps, but there's a great deal of work still to be done to harness the power of AI in addressing environmental issues. For example, McKinsey analyzed 160 AI social-impact use cases and identified 10 domains where adding AI could have a significant impact.[26] Unfortunately, the framework includes only one use case in life below water and two use cases across affordable and clean energy and clean water and sanitation. That leaves multiple SDGs still needing to be looked at from an AI-application perspective.

Here are some potential environmental use cases identified in a 2020 workshop hosted by Stanford University's Institute for Human-Centered AI (HAI):[27]

- Use satellite tracking and digital data streams to identify ships involved in human trafficking.
- Create food-choice recommendations to enable consumers to help protect the environment.
- Determine optimal water allocation based on environmental constraints.
- Optimize agriculture returns through early detection of crop disease and other issues.

Unprecedented collaboration between academia, corporations, and governments is needed to accelerate the use of AI to solve our environmental challenges.

How AI Could Aid Healthcare

There's an enormous potential for AI in patient care. One example is intelligent equipment logistics, something that would have been extremely useful during the pandemic when hospital beds, oxygen tanks, and respirators were in short supply. Many of these resources could have been moved from place to place as cases spiked if better and more integrated logistics systems were available. Human resources like physicians and nurses could have been tracked and "allocated" nationwide rather than every healthcare system having to grapple with staffing shortages on their own. Vaccines could also be expedited more efficiently using AI systems. Indeed, during the spring 2021 COVID outbreak in India, several startups began piloting conversational AI solutions aimed at improving logistics.[28]

Unprecedented collaboration between academia, corporations, and governments is needed to accelerate the use of AI to solve our environmental challenges.

Home care is another example of an area that's especially important as populations age. Systems that combine sensors (monitoring blood pressure, blood sugar levels, etc.) with intelligent analytics could be used to create better outcomes for home-bound patients, not to mention the boon they'd be to home health workers and family caregivers. Likewise, today's

telemedicine solutions are fairly simplistic and could be developed into interactive solutions for virtual care using AI.

The pharmaceutical industry can also benefit from AI, as demonstrated by AI startup PostEra, which organized COVID Moonshot.[29] This crowdsourced initiative invited submitted drug designs, used its ML tools to determine which ones should be tested, and arrived at a potential antiviral in 48 hours, a task that would have taken weeks using traditional methods.

If there are any doubts about the potential for AI to permanently change our lives, we have only to look at the most recent technology developments.

Recent AI Developments

If there are any doubts about the potential for AI to permanently change our lives, we have only to look at the most recent technology developments. To close out this article, here are just a few:

1. **Facebook AI** recently made public a data set consisting of 45,186 videos of 3,011 humans having conversations. The tool is expected to help researchers better understand racial bias in technology and explore solutions to ensure fairness in AI.[30]

2. **Facebook** also released a recommendation system that uses 12 trillion parameters to speed AI-model training time by 40x.[31]

3. **Waymo** (formerly Google's self-driving car program) has released what it calls the largest interactive data set yet released for research into behavior prediction and motion forecasting for autonomous driving.[32]

4. **GPT-3 and Eleuther**, both open source language algorithms, are capable of writing coherent articles in English when given a text prompt.[33]

Disclaimer: The ideas expressed in this article are based on the author's industry experience. Wipro does not subscribe to the substance, veracity, or truthfulness of said opinion.

References

[1]Hodler, Amy E. "AI and Graph Technology: 4 Ways Graphs Add Context." Neo4j, 29 July 2019.

[2]Ivanov, Sergei. "Top Applications of Graph Neural Networks 2021." *Criteo R&D Blog*, 14 January 2021.

[3]Bendtsen, Claus, and Slavé Petrovski. "How Data and AI Are Helping Unlock the Secrets of Disease." AstraZeneca, 1 November 2019.

[4]Meza, David. "How NASA Finds Critical Data Through a Knowledge Graph." Neo4j, 17 May 2017.

[5]Hetionet.

[6]Zheng, Da. "Amazon's Open-Source Tools Make Embedding Knowledge Graphs Much More Efficient." Amazon Science, 6 August 2020.

[7]Fabula AI is a London, UK-based fake news detection company owned by Twitter; see Wikipedia's "Fabula AI."

[8]Zhu, Rong, et al. "AliGraph: A Comprehensive Graph Neural Network Platform." Cornell University, 23 February 2019.

[9]GraphSAGE on GitHub, accessed May 2021.

[10]DeepMind.

[11]"Jacobs Takes Product Design to New Heights with Generative Design in Aerospace." Case study, PTC, accessed May 2021.

[12]Mackintosh, Graham. "AI Applications for Astronaut Health." NASA, 7 October 2020.

[13]Chandler, Simon. "Reuters Uses AI to Prototype First Ever Automated Video Reports." *Forbes*, 7 February 2020.

[14]Damiani, Jesse. "A Voice Deepfake Was Used to Scam a CEO Out of $243,000." *Forbes*, 3 September 2019.

[15]Khan, Faisal. "'Video Authenticator' Is Microsoft's Answer to Deepfake Detection." Technicity, 11 September 2020.

[16]Coughlin, Tom. "175 Zettabytes by 2025." *Forbes*, 27 November 2018.

[17]See Wikipedia's "Predictions Made by Ray Kurzweil."

[18]Martineau, Kim. "Toward Artificial Intelligence That Learns to Write Code." *MIT News*, 14 June 2019.

[19]Srikant, Shashank, et al. "Generating Adversarial Computer Programs Using Optimized Obfuscations." OpenReview.net, 28 September 2020.

[20]"Intel, MIT, and Georgia Tech Deliver Improved Machine-Programming Code Similarity System." Intel Newsroom, 29 July 2020.

[21]"No-Code AI Startups Are Raising Funding & Getting Acquired. Why App Developers & Big Tech Cos Think Plug-and-Play AI Coding Tools Could Be a Game Changer." Research Brief, CB Insights, 19 April 2021.

[22]"Take Action for the Sustainable Development Goals." United Nations, accessed May 2021.

[23]"Sustainable Development Goals." Google for Startups, accessed May 2021.

[24]Fleming, Seán. "Tracking the Effects of Glacial Melting at the Top of the World." *On the Issues*, Microsoft, 12 January 2021.

[25]Art, Jean-Yves, et al. "Microsoft and the United Nations Sustainable Development Goals." Microsoft, September 2020.

[26]Chui, Michael, et al. "Notes from the AI Frontier: Applying AI for Social Good." McKinsey Global Institute, December 2018.

[27]Maher, K. "Environmental Intelligence: Applications of AI to Climate Change, Sustainability, and Environmental Health." Stanford University, Human-Centered Artificial Intelligence (HAI), 16 July 2020.

[28]Bhavani, Divya Kayla. "Conversational IT Becomes the Unlikely Hero During the COVID-19 Pandemic." *The Hindu*, 11 May 2021.

[29]"COVID Moonshot." PostEra, accessed May 2021.

[30]"Shedding Light on Fairness in AI with a New Data Set." Facebook AI, 8 April 2021.

[31]Alford, Anthony. "Facebook Announces ZionEx Platform for Training AI Models with 12 Trillion Parameters." InfoQ, 4 May 2021.

[32]Hawkins, Andrew J. "Waymo Is Disclosing More Autonomous Vehicle Data for Research Purposes." The Verge, 10 March 2021.

[33]Knight, Will. "This AI Can Generate Convincing Text — and Anyone Can Use It." *Wired*, 29 March 2021.

Jayashree Arunkumar leads software engineering transformation teams for a portfolio of US clients. She has delivered multiple global transformation programs involving products and platforms across banking, media, and healthcare clients. Ms. Arunkumar has vast experience in developing and delivering business applications via Agile software development and builds high-performance software engineering teams for value-based delivery. She is passionate about giving back to the organization and the community. Ms. Arunkumar nurtures engineering teams globally through a high-performance software engineering culture and mentors engineering leads toward active participation in their community and capability building. Ms. Arunkumar also leads a solution development program for breast cancer awareness, which has been designated AWS Get Started service credits for implementation. She can be reached at jayashree.arunkumar@wipro.com.

CUTTER Business Technology Journal

About Cutter Consortium

Cutter Consortium is a unique, global business technology advisory firm dedicated to helping organizations leverage emerging technologies and the latest business management thinking to achieve competitive advantage and mission success. Through its research, training, executive education, and consulting, Cutter Consortium enables digital transformation.

Cutter Consortium helps clients address the spectrum of challenges technology change brings — from disruption of business models and the sustainable innovation, change management, and leadership a new order demands, to the creation, implementation, and optimization of software and systems that power newly holistic enterprise and business unit strategies.

Cutter Consortium pushes the thinking in the field by fostering debate and collaboration among its global community of thought leaders. Coupled with its famously objective "no ties to vendors" policy, Cutter Consortium's *Access to the Experts* approach delivers cutting-edge, objective information and innovative solutions to its clients worldwide.

For more information, visit www.cutter.com or call us at +1 781 648 8700.