Artificial Intelligence in Business:
Balancing Risk and Reward

Dr. Rob F. Walker,
Vice President, Decision Management and Analytics, Pegasystems
Synopsis

AI technology is evolving faster than expected and is already surpassing human decision-making in certain instances. And sometimes, in ways we can’t explain. While many are alarmed by this, AI is producing some of the most effective and dramatic results in business today.

But there is a downside. Using uncontrolled AI for certain business functions may cause regulatory and ethical issues that could lead to liability. Optimizing AI for maximum benefit requires a new approach. This paper will consider recent advances in AI and examine how to balance safety with effectiveness through judicious control over when to use “transparent” vs. “opaque” AI.
AI: a match for human creativity?

In May 2017, Ke Jie, the world’s best player of the ancient Chinese board game Go, pictured in Figure 1a, was defeated in three straight games. This is important because he was defeated by AlphaGo, an AI computer program developed by Google DeepMind. DeepMind has since retired AlphaGo to pursue bigger AI challenges. It’s anybody’s guess what they’ll do next, as just a year earlier their Go victory wasn’t thought possible.

Winning at Go requires creativity and intuition that in 2016 were believed out of reach for today’s technology. At that time, most experts thought it would be 5-10 years before computers would beat human Go champions. While this is one of the most technologically impressive achievements for AI to date, there have been other recent advancements that the general public might find equally surprising.

Composer David Cope has been using his Experiments in Musical Intelligence (EMMY) software to fool music critics into believing they were hearing undiscovered works by the world’s great composers for the past 20 years. His most recent achievement is a “new” Vivaldi, whose works were previously thought too complex to be mimicked by software.

And in 2016, an AI learned how to counterfeit a Rembrandt painting based on an analysis of the Dutch master’s existing body of work. The results are shown in Figure 1b.

While it might not fool art critics, it would likely fool many art lovers and conveys to the less-trained eye much of the same aesthetic and emotional complexity of an original Rembrandt.

The point of these examples is that AI, at least in discrete areas, is already regularly passing the Turing Test, a key milestone in the evolution of AI. Visionary computer scientist Alan Turing, considered by many to be the father of AI, posited the test in 1950.
The Turing Test

Alan Turing (1912-1954) understood as early as the 1940s that there would be endless debate about the difference between artificial intelligence and original intelligence. He realized that asking whether a machine could think was the wrong question. The right question is: “Can machines do what we (as thinking entities) can do?” And if the answer is yes, isn’t the distinction between artificial and original intelligence essentially meaningless?

To drive the point home, he devised a version of what we today call the Turing Test. In it, a jury asks questions of a computer. The role of the computer is to make a significant proportion of the jury believe, through its answers to the questions, that it’s actually a human.

In light of Turing’s logic, what effectively is the difference between an AI being able to counterfeit a Rembrandt painting and it being a painter? Or being able to compose a “new” Vivaldi symphony and it being a composer? If an AI can pretend at this level all the time, under any circumstance, no matter how deeply it is probed, then maybe it actually is an artist, or a composer, or whatever else it has been taught to be. In any case, as Alan Turing would say, “how could we possibly prove otherwise?”

“Can machines do what we (as thinking entities) can do?”
A Turing Test for emotion

What would it take for an AI to convince a Turing Test jury that it is a human? This would certainly require much more than just being able to paint, or compose music, or win at Go. The AI would have to be able to connect with the jury members on a human level by exhibiting characteristics of human emotional intelligence, such as empathy.

Based on the examples above, perhaps it’s no surprise that AI can model and mimic the human psychological trait of empathy. Case in point: Pepper.

Pepper (shown in Figure 2) is a roughly human-shaped robot that specializes in empathy. Pepper was created by Softbank Robotics, a Japanese company, that markets Pepper as a “genuine day-to-day companion, whose number one quality is its ability to perceive emotion.” Softbank designed Pepper to communicate with people in a “natural and intuitive way,” and adapt its own behavior to the mood of its human companion. Pepper is used in Softbank mobile stores to welcome, inform and amuse their customers. Pepper has recently also been “adopted” into a number of Japanese homes.
Evidence of Pepper’s empathic abilities can be seen in Figure 3. Don’t look at the robot; look at the girl next to it. Look at her face, specifically her eyes. That is real emotion. She’s excited to have connected, on what feels like a human level, with her new friend Pepper.

Figure 4 shows representations of the mental maps, maps, or hierarchies of emotions, Pepper uses to analyze and respond to the emotions of others. In theory, Pepper’s successors will contain mental maps that are arbitrarily deep and complex. And at some point, whether you call it pretend or not, their emotional responses will pass a Turing Test as well.
Lastly, consider Sophia seen in Figure 5. Designed to resemble Audrey Hepburn, Sophia is what her creator Dr. David Hanson calls an “evolving genius machine.” According to him, she and Dr. Hanson’s other robots have the potential to become smarter than humans and to learn creativity, empathy and compassion.

Sophia has become a media darling and given a number of high-profile interviews. In those interviews, she offers convincingly human answers to the thought-provoking questions tossed her way. The YouTube video “Sophia Awakens” shows just how human-like machines are becoming. It’s beginning to appear that we no longer need to worry about a robot passing the Turing Test, we need to worry about it pretending to fail.

In March 2016, AI researcher Stuart Russell stated that “AI methods are progressing much faster than expected, (which) makes the question of the long-term outcome more urgent,” adding that “in order to ensure that increasingly powerful AI systems remain completely under human control... there is a lot of work to do.”
Here be dragons

Building on Mr. Russel’s sentiments, some have posited that the long-term outcomes of AI and other advanced technologies are extremely dangerous. SpaceX founder Elon Musk is on record as saying that artificial intelligence is “humanity’s biggest existential threat.” And he has a kindred spirit in Sun Microsystems’ co-founder Bill Joy, who in 2000 wrote a manifesto entitled “Why the Future Doesn’t Need Us.” In it, he examined the potential threats posed by three powerful 21st century technologies—biotech, nanotech, and AI:

“The 21st-century technologies - genetics, nanotechnology, and robotics (GNR) - are so powerful that they can spawn whole new classes of accidents and abuses. Most dangerously... Knowledge alone will enable the use of them.

“Thus, we have the possibility not just of weapons of mass destruction but of knowledge-enabled mass destruction (KMD), this destructiveness hugely amplified by the power of self-replication.”

Despite such warnings, development in each of these areas has continued unabated. And the advances made thus far are truly spectacular.

In biotech, one of the things Bill Joy warns against is tinkering with DNA. He proposes government restraints on the practice. Yet, scientists from Harvard and MIT have developed a search and replace genetic editing technique called CRISPR-Cas9 that makes modifying the DNA of humans, other animals, and plants much faster and easier than ever before. According to Science magazine, “The CRISPR-Cas9 system is revolutionizing genomic engineering and equipping scientists with the ability to precisely modify the DNA of essentially any organism.”

While arguably not as advanced as biotech, a nanotech discovery received the Nobel Prize for science in 2016. A trio of European scientists developed the “world’s smallest machines.” Scientists use these nano-machines to create medical micro-robots and self-healing materials that repair themselves without human intervention. This must send a shudder down Bill Joy’s spine, as he fears that nanobots will eventually escape our control, begin to replicate rapidly and, in one example, “spread like blowing pollen and reduce the biosphere to dust in a matter of days.” This has become known among nanotechnology cognoscenti as the “gray goo problem.” A rather ignominious end to human endeavor indeed.
Regarding AI, Mr. Joy’s main concern is that we aspire to achieve a utopian future of leisure, and ultimately immortality, by gradually replacing our physical bodies and consciousness with robotic technology. So, what’s the catch? Specifically, Musk’s fear, that humanity will not survive such a close encounter with what will by then be a superior species. Joy asks, “If we are downloaded into our technology, what are the chances that we will thereafter be ourselves or even human?” In this case, the coup might be bloodless (not a Terminator-esque annihilation), but humanity as we know it might disappear nonetheless.

Of course, for AI to compete with us as a species, it would need to be a general-purpose intelligence, not an army of specialists. And while this scenario is still the stuff of science fiction, there are many who agree with Mr. Joy. In a recent survey by Pegasystems of 6000 humans in six different countries, 70% are fearful of AI and 25% believe it will take over the world and enslave humanity. While 31% also believe humans will be replaced by robots on the job, some contend that AI assistance on the job would help improve work/life balance, freeing us up to do more meaningful work and have more leisure time.

“AI assistance on the job would help improve work/life balance, freeing us up to do more meaningful work”
Mr. Joy makes some hard-to-refute points. For the first time in history, with AI at least, humans may be confronted with another intelligence. And peering inside the “mind” of that intelligence in its nascent stages gives a better sense of where some of the concern is coming from.

Enter Bob Ross, the afro’d purveyor of painting pedagogy. He was the creator and host of “The Joy of Painting,” an instructional television program that aired from 1983 until his death in 1994. With his relaxed style and amusing phraseology (“Maybe in our world there lives a happy little tree over there”), he gained a cult following that has since followed him to YouTube and Netflix.

In 2016, Google ran a Bob Ross instructional painting video though their DeepMind tool called DeepDream, which enables neural networks to view and try to “process” imagery. Figure 6 shows some of the strange associations the AI is making to try to guess at what the objects in the video are. In the YouTube video, entitled “Deeply Artificial Trees,” images morph from association to association, and strange sounds can be heard as the AI tries to guess at the sounds Bob Ross is making. Its guesses are based on images and sounds Google has “trained” the AI with.

In Google’s words, this video was designed to highlight “the unreasonable effectiveness and strange inner workings of deep learning systems.” Researchers believe this AI’s system of guessing is roughly equivalent to the first steps in human visual perception.

The AI that watched this video is not the if-then logic of a computer that beat humans at chess 20 years ago. It’s more akin to an alien mind trying to understand the world and learning as it goes. This is the intelligence that beat us at Go.

Figure 6: Alexander Reben, Deeply Artificial Trees, 2017
An artificial COO?

The neural network that went to work on Bob Ross is probably not something most CEOs would want responsible for managing anything particularly critical or sensitive in their business. There is clearly tremendous power in the thinking/learning/creating potential of modern AI, but how can you trust it? How do you keep it from giving in to its Bob Ross-like sensibilities when making decisions about, say, a customer? More human control is required.

To show the form this control could take, consider two concepts: Opaque AI and transparent (or trusted) AI. Opaque AI includes neural networks, deep learning, genetic algorithms, ensemble models, etc. What these methods have in common is that the “logic” behind their predictions and decisions can’t be easily expressed. Describing how AlphaGo does its “magic” in terms of if-then-else rules, is a bit like trying to define a circle using straight lines. A very simple “explanation” of a circle would be a square, a hexagon is slightly better, an octagon better still. But it takes an infinite number of lines to perfectly describe a circle using only lines. The same applies when the inner-workings of an opaque AI are described using transparent if-then-else rules. The sheer number of rules involved (and their inter-dependencies) make the description incomprehensible to any human (not to mention impossible to code). This doesn’t mean that decisions made by an opaque AI are ineffective, often quite the contrary. It just means that because its reasoning is not understood, its decisions may potentially be a liability.

Transparent AI, on the other hand, can explain how it came to a decision in a way that is understandable to a human. The models and patterns it uses can be expressed in transparent if-then-else or decision tree formats. In fact, if-then-else rules or decision trees can be used to explain opaque AI, in which case, the complexity of the explanation becomes a measure of opacity. A transparent explanation of an opaque model represents something akin to a pen portrait trying to represent a high definition photo; it is and can only be an approximation.

Now imagine you have an AI running your business, or at least making some of its critical decisions. And this AI has some sort of a switch, let’s call it a T-Switch, for Trusted or Transparent. With this T-Switch, you can change the methods your AI uses to make decisions from opaque to transparent.
In some areas, take marketing for instance, there may be a CMO (Chief Marketing Officer) who is perfectly happy to let an opaque AI select the location for her billboards, or the TV shows and web sites that will run her ads. As a result, she may be getting tremendous return on her marketing investment. And perhaps that’s good enough for her. On the other hand, she may fear that because she doesn’t understand how the AI makes its decisions, it may have unintended consequences for the brand. For example, the AI may decide it doesn’t want to sell widgets to Asian women over the age of 45 for some reason. And because the AI is opaque, the CMO may never know that it’s behaving in this way.

Transparency about how an AI does its magic isn’t always important, but in highly regulated areas, such as credit risk, or circumstances in which data privacy is a concern, it may be critical or even mandatory. For instance, the EU General Data Protection Regulation (GDPR) is slated to take effect in May 2018. This new regulation mandates that companies be able to explain exactly how they reach certain algorithmic-based decisions about their customers. In this case, companies with a T-Switch capability will have an advantage, enabling them to better and more easily comply with the legislation.

But the question of when to use opaque AI and when to use transparent AI isn’t just about satisfying regulators. It’s a discussion that veers quickly into ethics and morality.

“...companies with a T-Switch capability will have an advantage, enabling them to better and more easily comply with the legislation.”
Taking the moral high-ground

Without a T-Switch, opaque AI is extremely difficult to control. For example, Microsoft and the University of Cambridge published some research in 2013\(^1\) that shows that even transparent algorithms are able to predict, with an impressive degree of accuracy, personal things about you using a very small number of your Facebook Likes. And by personal we mean: gender, political affiliation, sexual orientation, religion, favorite intoxicants, and perhaps a few things you didn’t even know about yourself.

The bars in these graphs in Figure 7 show the AI’s accuracy in predicting each of the traits listed. While not 100% accurate, it’s easily superior to human judgment.

To summarize their findings, if the algorithm sees 10 of your Likes, it will know you better – in terms of these traits – than your co-workers know you. With 70 Likes, it knows you better than your friends do. With just 150 Likes, it knows you better than your parents. And with 300, better than your partner. Already, with an algorithm that has been around for a long time (such as the one used in this study), the level of predictive power is superior to human. Yet because it’s a transparent algorithm, we’re able to understand the associations it makes and can reverse engineer how it came to its conclusions. So even without any spooky “Bob Ross”-type associating, algorithms are more powerful than human minds when it comes to predictions about human behavior.

---

1. Source: Private traits and attributes are predictable from digital records of human behavior, Michal Kosinski, David Stillwell, and Thore Graepel
The larger point, however, is that today’s AI cannot be controlled with data. Some might think that limiting the input data, for instance, to only Facebook Likes, will prevent it from “going rogue.” Unfortunately, that viewpoint is naïve. For example, taking “gender” out of the data is no guarantee that an AI won’t internally conceptualize gender and use it as a factor in its reasoning. Remember the example of an AI deciding not to sell to Asian women over the age of 45? This may happen without the AI being fed data on race or age.

Consider an opaque algorithm that has been given much more data than just Facebook Likes - such as a bank using an opaque AI for its lending processes. The bank feeds the AI what it believes are neutral and limited data about the applicants. Nevertheless, before you can say “adjustable rate mortgage,” the AI becomes racist or misogynist. With no way to monitor or control its AI, the bank has no idea this is happening until it’s slapped with a class action lawsuit, or sees a headline in the news.

Finally, imagine feeding even more data - Big Data - into an opaque AI. The AI may develop a value system of its own that humans cannot understand. As with human intuition, it cannot be reverse engineered or explained. In the lending scenario, AI will make significant decisions about people’s financial lives without any explanation trail. The GDPR would not approve, and neither should your compliance officer.

When applying AI to your business, having a T-Switch will enable you to make a conscious decision about where to allow opaque AI and where to insist on transparency.

“...having a T-Switch will enable you to make a conscious decision about where to allow opaque AI and where to insist on transparency.”
Exploring the human/artificial intelligence continuum

Most businesses today rely on people for much of their decision-making. And it’s undeniable that human brainpower alone has built some incredibly successful, profitable, and powerful businesses. But given enough data, AI can make a huge contribution. And in the realm of behavior prediction, which is one of the keys to good decision-making, human judgment isn’t even in the ballpark.

To illustrate the positive power of opaque AI, look at the antenna in Figure 8.

NASA needed an antenna that could receive commands and send data to Earth from the Space Technology 5 satellites. These satellites help scientists study magnetic fields in Earth’s magnetosphere. The antenna had to be very small, light, and consume little power, while also being strong and robust. AI was used because it can develop designs much faster than a human could under the same circumstances. It is also able to invent designs that no human designer would ever think of.
But for all its power, opaque AI also has a downside. In 2016, Microsoft conducted an experiment by releasing a Twitter chatbot, called Tay, based on an opaque deep-learning algorithm. In less than 24 hours, Twitter trolls completely corrupted Tay, essentially turning it into a racist, misogynist Hitler devotee.

Microsoft launched Tay as an experiment in conversational understanding. According to Microsoft, the more you chat with Tay, the smarter it will become. The goal was to teach it to engage people through “casual and playful conversation.” But only two hours into the experiment, Tay began insulting humankind (Figure 9a). Just three hours later, Tay is ready to do away with humanity altogether (Figure 9b).

Microsoft ended its experiment after less than 20 hours, promising to make some “adjustments” to Tay. It hasn’t been re-launched in over a year.

Did Microsoft send Tay out into the online world unprepared for trolls? Were they truly surprised by the affect the trolling had? Not all of Tay’s responses can be explained by the garbage in/garbage out axiom, and just as we may never know what makes a person turn down a dark path, Microsoft may never know what went wrong with Tay.

These examples clearly show the power and risk of opaque AI. As businesspeople, it’s critical to understand the potential rewards and inherent risks when applying this technology. In the case of Tay, what if Microsoft had launched it as a customer service chatbot? Disaster.
Optimizing AI through control & collaboration

Just a bit of human supervision can greatly reduce the risks of implementing AI for business. Supervision doesn’t mean second-guessing, it just means providing a guideline for what is acceptable and what is not in a given scenario.

One way to achieve this control is through a Customer Decision Hub – a centralized decision-making engine – that includes a transparency setting (such as a T-Switch). The setting would allow companies to decide which level of AI transparency to use for which business purposes.

For example, Figure 10a shows model transparency sliders. The slider is used to set the desired level of model transparency in a particular area of the business.

Predictive models, generated by AI or more traditional means, are rated from 1-5, with 1 being very opaque, and 5 being completely transparent.

Using this Decision Hub, AI techniques can be combined to create the desired level of transparency or opacity. By adjusting this setting, users can actively block the deployment of certain algorithms.

Figure 10a: Pega Customer Decision Hub with model transparency sliders
Each of the tiles in Figure 10b represents a predictive model. In the lower right-hand corner of each tile, the model's rating is shown. A Neural Network and Random Forest are rated a 1 (opaque), whereas the Scorecard is rated a 5 (transparent).
AI Algorithm Quality Assurance (QA) Process

Companies need more than just a T-Switch to reduce AI risks. They also need a more robust QA and approval process for model development and execution – one that is augmented with AI-specific best practices.

First, in the Explanation phase, the selected model explains how its logic works. Transparent AI provides an exact explanation. Opaque AI, by definition, gives only an approximate explanation at best, using the aforementioned pen portrait.

Next, in the Testing phase, bias tests show if the model has developed any undesirable tendencies. These include what might appear to be discriminatory practices or ethical violations, any sort of biases a regulator might see as a red flag in the model’s predictions, classifications, or actions. A notable shortcoming, however, is that bias tests are only possible on available data. For example, you’ll only know if an AI is discriminating against a particular racial group if you have data on the race of at least a sample of your customers.

In the Approvals phase, managers provide ethical sign-off for opaque AI based on the explanations and bias test results. This is to confirm that the model’s results are in line with corporate policies, ethical correctness, brand considerations, regulatory requirements, and so forth.

Lastly, in the Production phase, the benefits of AI/human collaboration can be tested. In environments, such as call centers, retail branches, or even vehicles, the AI’s decisions are delivered via an interface used by a human operator. In this way, the AI acts as a “wingman,” helping the operator achieve better outcomes – from providing better recommendations to customers, to closing a loan, or parallel-parking a car. And in this final phase, the added value of AI collaboration can be measured. As always, monitoring for real-world effects is critical.
Exactly how can this benefit my business?

In Figure 12, Forrester ranks 13 AI technologies they believe add the most value to human decision-making. The vertical axis represents the value-add to business, and the horizontal axis represents the maturity of the technology.

When considering vendors to help build out your AI capability, look for those that have invested in the highest ranked AI building blocks, unified them into one platform, and can provide proven results and scalability.2

---

AI can deliver tremendous business value, but it must first be operationalized.\textsuperscript{2} Real-time customer engagement is one area in which AI has been successfully operationalized and is delivering unprecedented business benefits. A real-time environment in which the models receive fast feedback on decisions and lots of data provides the perfect conditions for machine learning. In fact, since the early 2000s, the true leaps forward in AI have come as a result of the sheer availability of data - Big Data.

Many companies can now test and perfect AI models in real-time with the data from millions of customer interactions per day. And every day the models get smarter, faster, and more sophisticated. Today's big computing power also helps make the increases in complexity more feasible. Super fast hardware combined with a continuous flood of real-time data from multiple channels enables AI to drive unprecedented business results.

Here is an example of the kinds of results Pega has achieved for its customers using AI:

- 10 to 30% reductions in customer attrition
- 30 to 50% increases in Net Promoter Score
- 2.5 million additional yearly net adds (new customers)
- 500 to 800% increase in upgrade rates

Setting aside any dystopian visions, AI underpins some of the smartest software available to business today. And with the new features and controls discussed – such as the T-Switch and Pen Portraits – the outcomes will not only remain unprecedented, but safe and ethical to achieve.

\textsuperscript{1} Pega has proven use cases of operationalizing different forms of AI in the areas of sales, service and marketing. Beginning in the late 90s with decision management and moving into more sophisticated models with machine learning in the early 2000s.
Further Pega AI reading/viewing

**Pegasystems**
PegaWorld 2017 Keynote: Rob Walker, Pega AI in Customer Engagement: Balancing Risk and Reward –Rob Walker gives his Pegaworld 2017 keynote address, which provides the source material for this paper.

**Pegasystems on CustomerThink**
AI in CX articles: Vince Jeffs
http://customerthink.com/author/vjeffs/

Other AI collateral

**Forrester Research**
TechRadar™: Artificial Intelligence Technologies, Q1 2017
In this report, Forrester analyzes the 13 most important AI technologies that will “Augment Your Enterprise Applications, Amplify Your Intelligence, And Unburden Your Employees.”

**Microsoft @Tay andYou**
“Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism [Updated]”
https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/
Read Techcrunch.com’s take on the Microsoft Tay A.I. experiment, what may have gone wrong and what Microsoft had to say about it.

**Bill Joy, cofounder and Chief Scientist of Sun Microsystems**
“Why the Future Doesn’t Need Us” –Wired.com, April 4, 2000
Read the full text of Bill Joy’s manifesto entreatting humanity to restrain R&D in AI, nano and biotech.

**Hanson Robotics**
Watch “Sophia Awakens” on YouTube:
https://www.youtube.com/watch?v=LguXfHKsa0c
For more information about Hanson Robotics and Sophia:
http://www.hansonrobotics.com/

**SoftBank Robotics**
Learn more about Pepper the robot here:
https://www.ald.softbankrobotics.com/en/cool-robots/pepper

**David Cope and EMMY**
Listen to one piece of David and EMMY’s latest composition “Zodiac,” created in the style of Vivaldi:
https://www.youtube.com/watch?v=2kuY3BrmTfQ

**The Next Rembrandt**
Learn more about The Next Rembrandt project and how this new portrait by the Dutch master was created:
https://www.ing.com/Newsroom/All-news/Rembrandt-goes-digital-.htm

**Deeply Artificial Trees**
Watch DeepDream trying to process Bob Ross here:
https://www.youtube.com/watch?v=mEhqFyvcpu8

**Bob Ross**
Here’s Bob Ross’ channel on YouTube:
https://www.youtube.com/user/BobRossInc
Watch Bob and judge for yourself how distorted DeepDream’s view of reality is.
ABOUT PEGASYSTEMS

We are Pegasystems, the leader in software for customer engagement and operational excellence. Our adaptive, cloud-architected software – built on the unified Pega® Platform – empowers people to rapidly deploy, and easily extend and change applications to meet strategic business needs. Over our 30-year history, we've delivered award-winning capabilities in CRM and BPM, powered by advanced artificial intelligence and robotic automation, to help the world's leading brands achieve breakthrough results.

For more information, please visit us at WWW.PEGA.COM

© 2017 Pegasystems, Inc. All rights reserved. All trademarks are the property of their respective owners.