

Responsible AI: Great power requires greater accountability

A PEGA
WHITEPAPER



Build
for
Change®



Introduction

The world keeps getting progressively smaller. Technology has connected people and fundamentally changed the way we live. Those connections have never been more important. Technology today is more than a convenience or cost-saver – it can be a literal lifeline in times of crisis and global turmoil.

Artificial intelligence (AI) has changed how humans use technology and interact with one another. Developed as an extension of human productivity, AI performs tasks that if done by a human would take hundreds or even thousands of times longer – in a more repeatable and exact way. AI is already seamlessly integrated into our daily lives – from product recommendations to voice transcription to automated routing services. What sets AI apart from other technologies is its ability to “learn,” meaning it can draw its own insights and adapt its actions based on the data it consumes and reactions it learns. In theory, the more data and interactions consumed by an AI, the “smarter” it will become. This opens the door for even more AI-based initiatives as organizations look for better and more scalable ways to connect and serve their customers.

There’s a misconception that AI is clinical – that because it’s a machine, its actions and recommendations will seem cold and unfeeling. In reality, it doesn’t have to be. AI can ingest and analyze much more data than thousands of people ever could. It can more easily pick up

on cues that indicate what a customer needs, particularly in times of hardship. Organizations can operationalize their AI to address common concerns, pick up on new needs, and “right channel” their employees toward issues that specifically require a human touch.

Early incarnations of AI have seen widespread use since the early 1980s in the form of advanced business rule engines and data-driven predictive analytics. With the recent explosion of machine-learning AI technologies, rapid innovation has companies looking for new tech to aggressively keep pace. They’re investing in deep learning, algorithmic decision management, natural language processing, and dozens of other techniques. However, for many, it might be their first true implementation and utilization of AI. As with any new technology investment, there’s risk involved. Depending on how the AI is applied, it can cause literally billions of dollars to be at stake, should it fail.


While the rewards of AI – revenue lift, customer satisfaction, efficiency gains – can be tremendous, organizations must remain accountable for the actions of their AI-driven systems. If they aren’t, they risk irreparable damage to their engagement programs, supply chains, bottom lines, and brand reputations. Customer trust, once lost, can be nearly impossible to win back. But beyond the financial and legal implications, there is a moral obligation to constantly improve AI for the greater good, set higher standards, and eliminate problems like bias that have plagued AI technologies since their inception.

The stage is set for **responsible AI**.

What does responsible AI look like?

Nearly every company has a set of values and a code of ethics which its employees must follow. If companies hold their employees to ethical and legal business standards, why shouldn’t they hold the technology these employees deploy to the same level of standards? Whether a company is developing AI-based technologies or using existing AI technology as part of its business, the company has a moral obligation to apply AI responsibly, as they’re on the forefront of innovation and are setting the standards by which all future AI will be created, utilized, and evaluated.

AI requires human-derived ethical frameworks in order to engage empathetically with customers. It needs tools and diverse data to learn from to reduce bias. It also needs to be transparent so that its decisions can be viewed and understood by others, when required. The only way that AI can positively impact customer experiences and business outcomes is if humans and AI work together.



Responsible AI requires four critical elements: it must be **empathetic, transparent, fair, and accountable.**



Empathy in action

To be considerate, you need empathy and action. At Pega, we've built tools for organizations to scale and operationalize empathy. Next Best Action Designer enables empathy by:

1. Using next-best-action strategies that can shift between selling, serving, and retaining, based on the customer's context
2. Including hardship and adverse circumstances as part of that context
3. Linking "adaptive" machine-learning models to every offer and message to determine what a customer truly needs in the moment
4. Using "suitability" rules to minimize unempathetic types of actions that may be financially sound for the business but aren't in the best interests of the customer.

Empathy

Empathy, by definition, is understanding and identifying with the feelings of someone else. In the context of customer engagement, it means looking at offers and actions from the customer's perspective, putting yourself in their shoes, and thinking about how your actions will make them feel. While putting the customer first seems like a basic tenet of good business practices, empathy can be difficult to operationalize.

In traditional marketing, segment- and campaign-based approaches have dominated how companies engage with customers. However, in that model, every message, every communication is force fit to a large group and decided in advance – long before you know what an individual needs "in the moment." Even if you had a psychic on payroll, it would still be impossible to anticipate exactly what each customer needed ahead of time. Instead, you need the ability to make decisions in real time so you can truly understand your customer's context, what they need from you right now, and how they want those needs to be met. AI, with its uncanny ability to collect and analyze customer data, can help an organization be more empathetic by helping to determine the exact right message to deliver based on a customer's own data.

For example, an organization can help a single mom renew her insurance policy online during a 10-minute window on her lunch break, direct a customer to a local shelter when their home is damaged in a hurricane (or even on the hurricane's predicted path), or proactively reach out to a customer with an upgrade to their data plan that will save them \$15/month in overage charges. In each scenario, AI helps to identify exactly what the customer needs in their unique context.

In lieu of trying to program emotions into machines, empathetic AI means incorporating societal norms, policies, and regulations that correspond to well-established safe and fair behaviors, in addition to applying customer data.

One of the ways responsible AI makes sure businesses are doing the right thing is by helping dial up empathy in the choice and delivery of messages to customers. On the best of days, every single communication with your customers should be both relevant and appropriate; a catastrophe magnifies that need tenfold, and time is often of the essence. Any responsible AI will

dial down inappropriate messaging, like aggressive sales offers, based on the context. Suitability rules ensure the AI won't sell at the expense of what's in the best long-term interest of the customer. Adaptive machine-learning models can find the relevant message for each situation, cycling out messages that aren't being well received. While you continue to deliver exactly what your customers want, you're also protecting your brand; it's critical that brands aren't seen as exploiting a crisis or tone-deaf to it.

Empathetic considerations are critical to building responsible AI. AI must understand not only what's relevant to the customer, but what is suitable for that customer in that context. It's the AI developer's responsibility to define those rules and provide guardrails for the AI as it learns.

Companies must:

- Identify actions that could be perceived as indifferent or unfeeling
- Highlight empathetic actions to take more often
- Identify and apply customer context to increase relevance
- Define rules governing suitability, applicability, and eligibility
- Monitor and optimize empathy levels automatically

Transparency

One of the most common use cases for AI is to gain better insights into customer needs and deliver relevant messages and offers based on personalized, one-to-one engagement. How an AI makes decisions becomes critical when you're dealing with massive amounts of customer data and making decisions that impact people's lives, such as with loans and probability of default. Regulations like the [European Union's General Data Protection Regulation \(GDPR\)](#) mandate that businesses must be able to explain the logic behind their decisions or risk massive fines. This can be problematic because not all AI algorithms are created with the same level of rigor and transparency. Transparency becomes more critical the more regulated an industry is. For example, the financial services, insurance, and healthcare industries all have strict regulations regarding the use of data and the tractability of decisions. And while some opaque AI algorithms may be very powerful and can increase performance through better decisions, the complex reasoning behind these "black boxes" can't be fully explained – a tradeoff that becomes more troublesome (and risky) when it comes time to explain why a particular decision was made.

Transparency is key to helping customers better understand how AI works to their advantage and helping build trust by allowing organizations to ensure their algorithms are safe to use. For AI to be transparent, it needs to be explainable. Decisions or suggestions provided by AI must be understood by their users and developers. In addition, transparent AI's decisions and methodology must be explainable to customers.

Even if you are a major global tech company that is often listed as one of the most customer-centric brands in the world, you're not immune to AI gone rogue. One such company was

wrought with bad publicity during the launch of their company branded credit card, when it was revealed that members of the same high-profile household with the same financial information (including joint bank accounts and tax returns) were receiving different credit card limits. When pressed as to why this was occurring, there were no answers, leading many to assume the decisions were based solely on the gender of the applicant. But these approval decisions weren't made by a person; they were made by AI, in a very "black box" fashion.

The issue wasn't just that this company's decision-making was biased; there's often bias in AI and that's remedied by detecting it, adjusting, and providing more balanced data sets. The problem was with how the decisions were being made, why they continued unencumbered, and what was triggering the results. The way the AI was making decisions in this case wasn't easily detected or well understood, so customer service was unsure how to answer the customer's heated questions. The mishap quickly went viral.

There are many areas where transparent algorithms may be determined to be critical, ranging from medical diagnosis to self-driving cars. Common use cases for transparency and high visibility include financial decisions like how much credit is available to an individual or business and what financial terms can be offered. Churn models also fall into this bucket because they're linked to how large an incentive you should offer a customer to get them to stay (What level would still be worth it? \$10 discount per month? \$50?).

Actions that are less sensitive may not require the same level of transparency – like which marketing treatment to use, which channel to contact a customer in, which non-sales message to send – because they're not directly related to a decision of legal significance, like financial terms. These are scenarios where the brand isn't denying anyone credit or access to a service or a discount based on the calculation, but rather they're simply optimizing the message to make it more personalized for that customer.

Companies must be proactive about certifying their algorithms and providing a clear explanation of why decisions were made when there's a problem. With the tools available to improve AI, there's no excuse not to use transparent and explainable algorithms, especially for regulated or higher-risk use-cases.



Companies must:

- Define appropriate thresholds, per use case, for AI transparency
- Limit the use of highly opaque models to low-risk business concerns, like treatment and message selection in marketing
- Enforce stricter transparency requirements for high-risk scenarios like fraud detection, credit risk assessment, offer provision, or any other area with significant legal ramification

Fairness

One of the most remarkable aspects of AI is its ability to learn and evolve; the algorithm adapts as it learns from its experiences. However, how it adapts depends on the diversity of the information it consumes, which can lead to concerns around fairness. One of the most common issues surrounding the fairness of AI-based decisions is bias.

In a [2019 Pega global survey on AI and empathy](#), 54% of respondents said they believe AI will always be biased.

One example of how life-altering bias in AI can be is within the American criminal justice system. Algorithms are used to assess a criminal defendant's likelihood of becoming a repeat offender. There are dozens of these risk assessment algorithms in use today and commercial tools to process the data. In 2016, [ProPublica examined Northpointe's tool, COMPAS \(Correctional Offender Management Profiling for Alternative Sanctions\)](#). The analysis found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, or the likelihood to reoffend, while white defendants were more likely than black defendants to be incorrectly flagged as low risk.

Companies must:

- Identify attributes that increase the risk of introducing bias
- Proactively monitor and detect bias in predictive models and complete customer strategies as well as investigate the root cause
- Test “what if” scenarios and adapt accordingly
- Prevent algorithms and strategies that exceed the bias threshold to go into production

AI algorithms will make decisions based on all the data at their disposal. In the case of COMPAS, the developers weren't intending to develop a racist AI – the bias it uncovered reflected the bias that exists within the justice system itself. Building a fair AI requires a focused effort to avoid that kind of unfairness. Even with the best of intentions, data might be correlated with protected variables like gender and age and introduce problems.

Biases related to factors like age, ethnicity, or gender can unintentionally creep into the AI models and skew outcomes. For AI to be fair, data, models, and strategies must be used in a way that is balanced toward all groups.

Companies must also regulate AI training data and evaluate the impact of their strategies before they are deployed (using simulations) and as they're being used in the real world to catch bias that might have been unintentionally introduced earlier in the process. This becomes especially important as teams integrate machine learning. Those algorithms adjust themselves rapidly, which can further mask the problem if companies don't take a proactive approach to detecting and eliminating bias.

Accountability

Trust is a hard thing to come by, even in the best of times. While Pega's survey results have indicated most consumers don't trust companies to make decisions in the best interests of their customers, 69% believe that companies have a moral obligation to do what's right, beyond what is legally required. When there is a crisis, that distrust is put under a magnifying glass. AI can help build trust (or even repair it), but only if companies hold themselves and their AI to very high standards.

Accountability in AI is about holding the people in your company responsible for your AI development to regulatory, ethical, and precedential standards, as follows:

- **Regulatory:** What is legally required for AI to be compliant with current laws and regulations? Is your company and its technology positioned to comply quickly with future regulations?
- **Ethical:** Does what you're doing sufficiently take into account the best interests of your customers? Does your AI reflect your values as a company? Does it reflect the values of your customers, partners, employees, stockholders, and the AI community at-large?
- **Precedential:** Are you respecting the positive precedents set by others in the community? Would your use and development of AI be positively received by the greater AI community? Are you adhering to or advancing the current standards? How will your actions be judged in the future?

Accountability in AI also requires a traceable history across training datasets, parameter configurations, model versions, and other metadata, which comply with the audit requirements. This is twofold; for businesses, it allows them to go back and look at which models were used and who created them if something goes wrong. It also allows customers to understand how their data is being used. Without a traceable history, it is very difficult to correct problems or even improve models and data sets. Just like transparency, there are now regulations like GDPR and the [California Consumer Privacy Act \(CCPA\)](#) in place that require customer data usage to be disclosed and hold corporations accountable for explaining AI's decisions.

Accountability can come in the form of regulations, but in the absence of that, it's the responsibility of organizations to do the right thing. In the years since the 2016 United States presidential election, Facebook has received backlash for stating it will not police the truthfulness of targeted ads sent by political campaigns. There is no law that says Facebook must do this. But considering the massive influence Facebook's ever-changing algorithm has on its audience, many believe it's poor

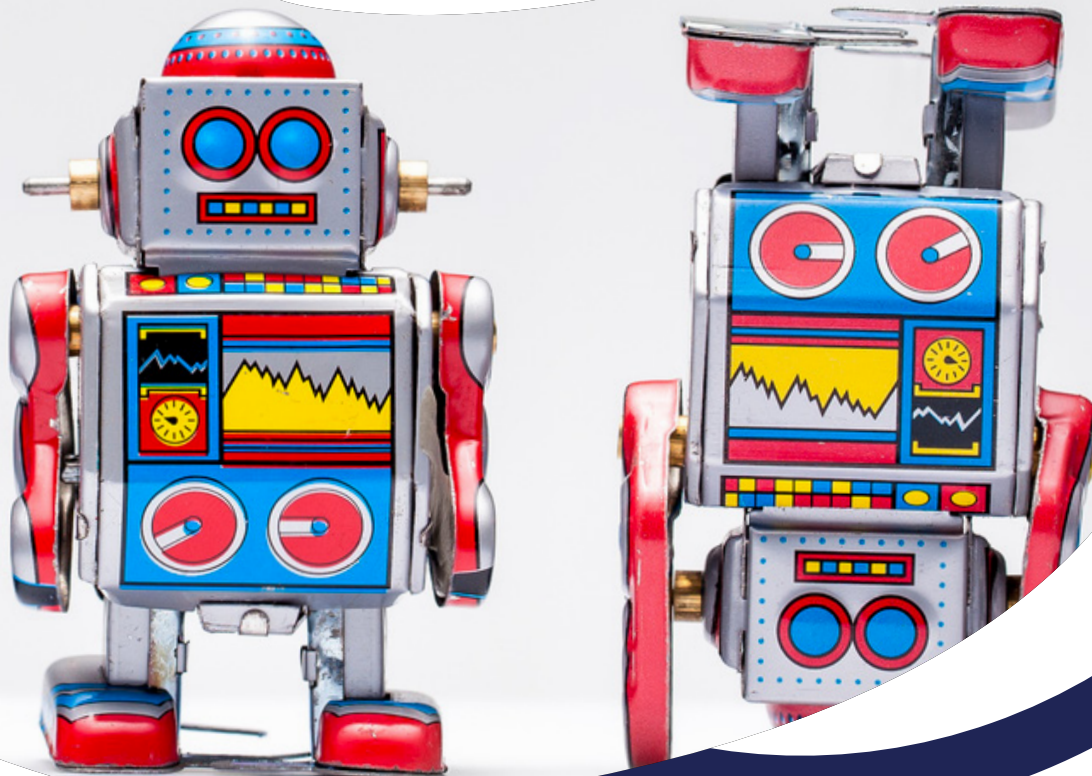


form to refuse to hold itself accountable for the quality and accuracy of the information displayed in Facebook ads. Decisions like these, made by some of the most well-known companies on the planet, might be why many consumers are still skeptical of AI and companies themselves. In fact, Pega's AI and empathy survey revealed that only 35% of respondents globally believe the companies they do business with have their best interests at heart.

Companies must:

- Ensure AI is compliant with all current regulatory requirements and processes are established for any future regulatory compliance
- Establish processes to disclose information to any customer they do business with, informing them whether their personal data is being processed and where and for what purpose
- Fulfill requests for copies of consumer personal data or deletion of that data on request
- Explain automated decisions (of legal significance) related to individual customers
- Understand – for each model or strategy created – when and where it was applied, who or what created it and when, and what data was used at the time
- Understand and model the financial value in the near term and look for opportunities to build long-term value in the future
- Analyze and evaluate risk regularly as it relates to the company and its shareholders

Accountable AI is often a decision a company must choose to make on their own. While there are some regulations in place, they are not wide reaching and are often slow to be implemented and enforced. Should a company wait for new regulations to take effect before committing to responsible AI? No. Nearly every interaction you have with your customers has a digital record, from a Twitter exchange to a recorded service call. Any misstep can quickly damage your brand reputation. Financially, AI gone rogue can lead to fines and even lawsuits. Morally, the algorithms at the core of your AI need to be safe, fair, and correct – within established norms. What information will be shown? What information would be considered inappropriate? Responsible AI starts with your code of ethics, and a strong AI practice will allow you to integrate them into your environment.



Conclusion: Pega's commitment to responsible AI

The obvious benefit of responsible AI is the ability to better mitigate risks associated with “AI gone wrong,” or decisions made by AI that were determined to be biased, incorrect, irrelevant, and/or inappropriate. But building responsible AI is more than simply a way to reduce risk. Responsible AI can also strengthen trust with consumers, thus enhancing brand value. Companies who make it a point to use customer data responsibly, provide appropriate transparency, and use AI to be more relevant and empathetic, will be able to build the lasting customer relationships every company is searching for.

Building consumer confidence also helps with portraying AI in a more positive light and, in turn, helps push adoption and innovation forward.

At Pega, we understand that it's humans that ultimately control (and should be held accountable) for whether business systems are empathetic, unbiased, and transparent. We developed our real-time, omni-channel AI platform to provide global brands with the ability to control their own AI, providing the tools that help businesses realize the value of responsible applications and directing the power of AI toward improving outcomes – not just for their organization but also for their customers.

At Pega, we've built responsible AI tools into [Pega Customer Decision Hub™](#). They include:

1. **[Next Best Action Designer](#)**: Helps drive more empathetic customer engagement by encouraging a suitability policy, and using customer data and context to determine the right action to take in the moment: service, sales, retention, or even no action at all.
2. **[T-Switch](#)**: Allows organizations to set the appropriate thresholds, per business function and purpose, for AI transparency. Businesses predefine these levels for each AI model using a sliding scale from one (most opaque) to five (most transparent). The transparency scores help guide users to build responsible AI systems using models that both meet their organization's transparency requirements and deliver exceptional customer experiences.
3. **[Ethical Bias Check](#)**: Helps eliminate biases by simulating AI-driven customer engagement strategies before they go live and flagging the potential for discriminatory offers and messages.
4. **A decision management environment** that allows trackability and accountability for any number of algorithms, model versions, rules, and strategies.

The four foundations of responsible AI will not only benefit customers, they will differentiate any organization from its competitors, help create a better brand, and generate a significant financial return.

Ready to get started?

Experience the power of the Pega Platform firsthand with a **30-day free trial**.



Pega is the leader in cloud software for customer engagement and operational excellence. The world's most recognized and successful brands rely on Pega's AI-powered software to optimize every customer interaction on any channel while ensuring their brand promises are kept. Pega's low-code application development platform allows enterprises to quickly build and evolve apps to meet their customer and employee needs and drive digital transformation on a global scale. For more than 35 years, Pega has enabled higher customer satisfaction, lower costs, and increased customer lifetime value.

For more information, please visit us at www.pegacom