How To Build Responsible AI Step 3: Resilience

How To Build Responsible AI Step 3: Resilience

Artificial intelligence is now an integral component of the processes and systems that drive our organizations. As AI practitioners, we must be intentional about developing, deploying and managing responsible AI — minimizing risk and removing bias while working toward our objectives.

I recently defined a framework of six essential elements of responsible AI that any organization can use as a guide:

1. Accountability
2. Impartiality
3. Resilience
4. Transparency
5. Security
6. Governance

In a series of articles, I am exploring each of these elements in detail. In this article, I'll focus on the third theme listed above: resilience. My philosophy for resilience in AI is: Small thinking is a disaster waiting to happen. If you think small, you will only prepare your organization for threats that are immediately apparent. Think big to be prepared to weather any storm. What unforeseen dangers — physical, intangible, financial — might you encounter? How can you fortify your AI to withstand them?

What Is Resilience?

Straightaway, to be resilient, both the AI and the practitioner must:

1. Adapt to situations and recover quickly

Resilience is comparable to grit. AI needs to have some grit to be resilient when facing changing circumstances. To use a sports analogy, in a fast-paced game like basketball a player has to adapt to situations and recover quickly. One way they can do this during a game is to shift to zone defense, covering a certain area instead focusing on one-on-one defense. Their resilience depends on how well they can adapt to each new situation, how well they can recover from obstacles and how they can sync with other players as the game progresses.

2. Engineer conditions that explore the full solution space and feasible scenarios available

AI offers an endless array of potential scenarios and solutions. Resilient AI requires you to create conditions that will play out different possibilities and what your responses would be for each. As a basketball coach, this would mean thinking through many possible actions from your players: How could they run trick plays? How could they inbound or rebound plays differently? Where could they be more adaptable?

3. Account for the effects of local constraints, and overcome these conditions to mitigate premature stopping (overfitting a model to data)

Resilient AI balances your short- and long-term objectives. How do you take local constraints and their consequences into consideration? How does the model housing the data limit you in what you are trying to achieve? How can you overcome those conditions to reach your ultimate goal? In a basketball game, this might mean benching your star player for a whole quarter when he's in foul trouble. This move will hurt in the short term, but it will pay off in the fourth quarter when he shoots a game-winning three-pointer.

Systems Will Fail

Every system will fail. This is not an indictment of the software developers, data scientists or machine learning engineers you work with. It's just a fact that these are complex systems, and changes and challenges are inevitable. Planning for systems to fail is a far better approach than hoping that they won't. Anyone who tells you otherwise either doesn't have real experience building systems or is trying to sell you something that will overpromise and underdeliver.

Adversarial AI Looms

As AI becomes more ubiquitous, so does the danger of "adversarial AI," misleading inputs to an AI system designed to deceive models or trigger misclassification. Developing

How To Build Responsible AI Step 3: Resilience

resilience is a way to account for and combat adversarial AI. Are you aware of the technical debt your organization carries? Have you also built a technical margin to hedge against threats like adversarial AI? What defense mechanisms do you have in place to protect yourself?

Three Key Features Of Resilience

To measure, benchmark and roadmap your journey toward more resilient AI, consider the following three key features.

1. Understanding Risks

Operational risk management — or understanding risks, as I call it — involves weighing the likelihood and severity of potential threats. When I was in the Marine Corps, I was responsible for calculating the operational risk involved for activities including taking units to the rifle range or even combat deployments to locations in Iraq and Afghanistan, asking: What's the likelihood that event will happen? What would be the severity if it did? While it's likely for someone to sprain an ankle on a march to the rifle range, the severity is minimal. The likelihood of an accidental discharge, on the other hand, while low, carries high severity and makes preventative risk measures necessary to mitigate both likelihood and severity.

In your AI, how well have you identified the likelihood and severity of risks? How have you actively mitigated the unacceptable risks?

2. Response Planning

Failure Mode and Effects Analysis (FMEA) helps identify possible points of failure and stop quality problems before they arise. Once you understand the risks, how do you develop a system to review the process, potential failure modes, potential effects, severity, occurrence rankings and detection rankings, then capture the risk priority number (RPN)? FMEA software and practices abound in mechanical, civil and systems engineering — understanding and applying FMEA across fields of AI is upon us. In other words, once an inevitable bad thing happens, what will be your response?

3. Effects Tuning

Forecast value added (FVA) is a metric used to evaluate the performance of activities and participants in the forecasting process to see if they add value or not. I like to add reinforcement learning to FVA to create a metric I call "effects tuning," which allows you to incorporate constant feedback to monitor changing conditions. Which touchpoints or people are adding value to your forecast? Which are detracting from it? You see effects tuning in action in dynamic pricing models. If you search for airline tickets now and two months from now, the prices will fluctuate due to seasonality, demand, fuel prices and other factors. Companies are using effects tuning to price plane tickets much higher just before Christmas than in late February.

Risk and resilience planning is nothing new — we're simply modernizing it to fit the AI world. At the end of the day, the main questions you need to ask to build a resilient AI system are: Do you understand the risk? Do you have a response plan for the risk? How are you tuning your plan over time to adjust for effects, anticipated or not? Be wise.

Aaron Burciaga, CAP, ACE

Chair of the Analytics Certification Board. Roles have included VP, CTO, and Practice Director of Fortune 500s and co-founder of data and technology startups.

Aaron is a Forbes contributor, frequently invited keynote and speaker, and Certified Analytics Professional (CAP). He is an appointed member of the U.S. Department of Commerce's National Technology Information Service (NTIS) advisory board. Aaron received his M.S. Operations Research from the Naval Postgraduate School and his B.S. from the US Naval Academy.