Deloitte.

Opening up to AI
Learning to trust our AI colleagues

How can we prepare for the shift that artificial intelligence (AI) is bringing, namely creating tools with autonomy?

"We don't avoid working with people because they're imperfect, we learn how to work together to complement each other."

We have used tools for the entirety of our existence as a species. Over the 200,000 years that we have been on the planet, we have improved our tools with advancing technology: from using stone tools to metal, from quill pens to the printing press, from the abacus to the computer. Across all these developments – whether swinging an axe or programming a computer – we've told the tools what to do.

But now that paradigm has begun to shift. Advancements in machine learning and artificial intelligence (AI) have created tools with more autonomy that behave more like humans – with the abilities to direct themselves, making them seem less like traditional tools and more like collaborators or partners. The arrival of Generative AI has made this particularly vivid. AI solutions are less like a hammer, and more like a co-pilot or colleague, helping us get the task complete.

Like our human counterparts, these tools are imperfect and require us to navigate their challenges. When they lack the knowledge to respond with certainty, they can make up a compelling story (an attribute called "hallucination"). They can perpetuate the bias found in the data they've been trained on. The deep learning networks that underly Generative AI are a "black box," where even the developers of the model cannot explain why the AI has come to a specific conclusion. However, shying away from AI due to these risks could prevent us from unlocking improvements in productivity, insight, and quality of life. We don't avoid working with people because they're imperfect, we learn how to work together to complement each other.

How can we learn work with an imperfect partner? Can we learn to trust AI? Working with an imperfect partner and building trust starts with three critical activities: use case and tool selection, safeguards and training.

Picking the right task and tool is as important in AI as it is in home remodeling. When cutting a piece of wood, you might consider a chop saw or jig saw, as they are likely to be well suited to the task. Similarly, it's important to focus AI efforts on use cases with high probabilities of success and manageable risks, and to select AI tools that are suitable for those use cases. You might want to use an external large language model for complex tasks, for instance, while using a smaller, local model for more simple operations.

Tools require safeguards. Table saws come with blade guards and automatic blade stops for a good reason. Just as we incorporate safety features into physical tools, we should incorporate safeguards into AI. Whether we are limiting the prompts that we submit to large language models, or review the responses we receive, it's important to focus AI applications on the tasks they were designed for. These safeguards will help us achieve solution performance, manage data privacy and security risk, as well as prevent or correct the perpetuation of bias.

Finally, training is crucial when using any new tool. You don't cut a critical part on your first use of a new saw. You review the operating manual, practice on scrap material, and build up comfort with the tool. "Measure twice, cut once" is the old adage!

The same is true of AI. When implementing AI, it's essential to thoroughly understand how it works and to test its functionality in a controlled environment before deploying it in a critical scenario. In the context of Generative AI, this can mean using prompt engineering: carefully crafting input prompts and evaluating the model's responses to ensure they're accurate, useful, and free of harmful or biased content, as well as incrementally releasing the functionality to users.

As we work with a tool over and over, we build up expertise, confidence, and ultimately, trust. While we realize the tool has limitations and even potential dangers, we know that we can use it safely within its, and our, operating limits, and we can focus on the task at hand. The same will be true with Generative AI solutions.

As we gain familiarity with these solutions, we will be able to shift our perspective of them from tools to partners and collaborators. We will likely become more comfortable letting them use their skills to help us on tasks, while being confident that they will help keep us safe.

Related reports

- [GovTech Trends 2023: A government perspective](#)
- [Tech Trends 2023](#)

Contacts

For questions regarding *Opening up to AI: Learning to trust our AI colleagues,* please contact:

Ed Van Buren
Principal
Deloitte Consulting LLP
Tel: +1 571 882 5170
Email: emvanburen@deloitte.com

Scott Buchholz
Managing Director
Deloitte Consulting LLP
Tel: +1 571 814 7110
Email: sbuchholz@deloitte.com