Weekly Brief

April 01, 2025

When a Human in the Loop Is Not Enough

Two weeks ago, I wrote about the importance of senior executives staying current by practicing with AI.

As most of us intuitively understand, we also need to be careful how we implement these new technologies. We need to hold those two thoughts—experimentation and caution—simultaneously.

The centerpiece of most risk-reduction approaches is to have humans review GenAI's output. But simply asking employees to take on oversight roles—evaluating GenAI outputs for incorrect, biased, or otherwise erroneous outputs—can create a false sense of security for several reasons.

One of those reasons is automation bias, the tendency to overly trust automated systems. The more reliable a technology seems—and GenAI seems very reliable—the less critical human reviewers become, overlooking errors they previously might have caught.

Corporate incentives can also discourage oversight. Thoughtfully reviewing GenAI output takes time, cutting into the promised efficiency gains. Concerned about the negative repercussions of slowing things down, reviewers might perform only cursory reviews.

These are just two of the roadblocks to effective human oversight outlined in You Won't Get GenAI Right if You Get Human Oversight Wrong.

What's the Solution?

We often talk about the 10-20-70 rule. In implementing AI, 70% of the effort should be directed to people and processes. In other words, effective human oversight must be thoughtfully integrated into the system's design, with the riskiest outputs receiving the most attention and escalation paths being clear and simple.

Companies can establish guidelines so that reviewer qualifications match the complexity and technical nature of the GenAI systems they're overseeing. None of us would want an accountant reviewing our medical test results—and vice versa. The same holds true for GenAI reviewers. They should have the expertise to evaluate accuracy and other potential risks.

Additionally, organizations can build realistic time to review GenAI's output into their business cases. For example, a bank might occasionally want to insert an incorrect output into its system to test whether the call center worker passes it along. If companies don't account for these tests, they will overestimate the time savings from the GenAI system.

For CEOs and senior executives steering their organizations through GenAI adoption, here are a few guidelines:

- Reinforce the value GenAI can unlock so long as risks are appropriately managed.
- Set the tone that GenAI oversight matters and that your teams should feel empowered to thoroughly evaluate outputs.
- Take a risk-adjusted approach to oversight, focusing on those outputs that could most significantly affect business performance and brand strength.
- Hold users—not GenAI systems—accountable for decisions to avoid the "AI made me do it" excuse.
- Make sure your teams integrate human oversight into system design.

GenAI systems aren't perfect, but neither are humans. With proper oversight, both technology and people can realize their potential—safely, reliably, and fully—while delivering transformative impact.

Until next time,

Christoph

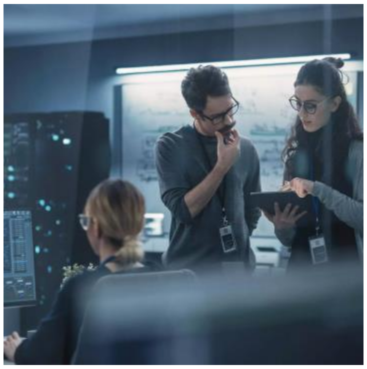
Christoph Schweizer
Chief Executive Officer

Further Insights

You Won’t Get GenAI Right If You Get Human Oversight Wrong

Generative AI presents risks, but the go-to solution—humans reviewing the output—isn’t as straightforward as executives think. Oversight needs to be designed, not delegated.

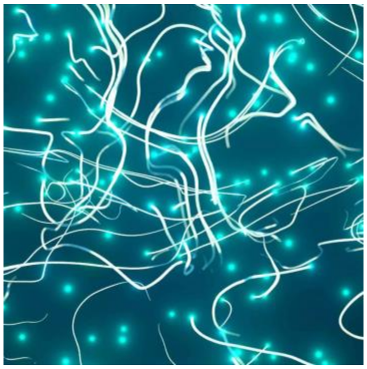
DESIGN OVERSIGHT



GenAI Will Fail. Prepare for It.

Even with comprehensive testing and evaluation, the risk of system failure with GenAI will never be zero. Organizations must respond swiftly when failures inevitably occur.

CREATE A PLAN



GenAI Can’t Scale Without Responsible AI

GenAI agents need to handle tasks responsibly, accurately, and swiftly in multiple languages, addressing potentially millions of specifications across hundreds of thousands of products.

BUILD TRUST