The 95 percent is a measurement problem

At the second GenAI Enterprise Day, at IBM in Zürich, five practitioners showed what separates AI that pays off from AI that only demos well: a baseline, a measured workflow and evaluation that keeps running.



PUBLISHED

November 26, 2025

READ ONLINE

samuylov.ai · The 95 percent is a measurement problem

A project that never measured its starting point cannot show a return. By November 2025, that was the more useful reading of the 95% headline.

The headline came from MIT NANDA's report *The GenAI Divide*. Measurement was the theme of the second GenAI Enterprise Day, *GenAI Reality Check: Business Wins, Fails & the Metrics in Between*, at the IBM Auditorium in Zürich on 26 November 2025, part of GenAI Enterprise Hub. Denis opened the evening with Lars Mallien of IBM Switzerland and moderated the closing panel.

Five stories

- **Measured at scale.** Markus Danhel, IBM, on IBM as its own first client, with productivity tracked across HR, IT support, finance, procurement and supply chain.
- **Scaled in stages.** Katarina Granath, SIX, on three years from a first bot to an enterprise platform, with governance and AI Integration Leads inside the business.
- **Measurement first.** Alexander Taboriskiy, Mentiora AI, on defining, measuring and improving a system against business requirements.

- **Responsible by design.** Anne-Marie Buzatu, ICT4Peace Foundation, on risk assessment, safeguards and staff capability.
- **Stable when it counts.** Manuel Flurin Hendry, Silicon Stories, on keeping real-time inference steady during a live performance.



SIX's adoption journey, 2023–2025.

What measurement changed

The clearest example came from Mentiora. On a business evaluation set, the pass rate rose from 31% to 91% with the same underlying models. What changed was a written definition of what a right answer looks like, and a calibrated model judge checking every answer against it.



The gap between a demonstration and dependable use.

All five returned for the closing panel, under three rules set at the start of the evening: no sales pitches, ask questions, share experiences.



The closing panel at the IBM Auditorium.

What it means for your project

Before you change a workflow, measure it: volume, handling time, error rate and review effort. Write down what a right answer looks like, and check against it before and after deployment. Then a pilot either shows its value or shows exactly where it falls short.

Our document-classification system builds that check in: every document is classified twice, independently, and an AI judge settles any disagreement.

Evaluating an assistant or agent? Start with the baseline. We set it up with your team before the build and keep it running after. See workshops & training, or Start a conversation.



Denis Samuylov
– Founder & CEO

Founder-led, team-delivered – Denis leads the engagement and stays on it after the pitch.

contact@samuylov.ai +41 79 136 38 26 Book an intro call →