Hey! What's New? 2026-36

Agentic Data Quality – The Beginning of a Beautiful Friendship

Dr. John Talburt writes in *The Data Administration Letter*, that "Agentic AI is becoming a game changer. It extends Generative AI from simple prompt-and-response interaction to the ability to deploy agents that can autonomously observe, analyze, evaluate, make decisions and coordinate with other agents. Perhaps most importantly, these agents can access and use external tools and data. Agents can even create their own helper agents!"

The impact of applying agentic AI to data quality, i.e., agentic data quality (ADQ), will be enormous, he says. "ADQ has the potential to automate the detection of data anomalies, autonomously remediate certain data errors and minimize the number of human-in-the-loop interactions, dramatically increasing data quality management efficiency and effectiveness. ADQ pipelines have the potential to dramatically reduce processing costs while improving data quality that leads to better decision making."

Talburt notes that "we are now entering an age where we can replace many rule-based, batch-oriented validation routines with real-time agents that can detect data anomalies like missing value, invalid values for the item domain, misfielded item and out-of-range values. Not only that, but these agents can decide which specialized agent should try to correct the problem. If agents are unable, or not allowed, to make the correction, they can escalate the problem to a person to resolve the issue. Through the escalation process, agents can learn new solutions for their specialization. Agents can also log and report all their decisions and actions for later analysis and audits."

Properly deployed and governed, he adds, "agents can be like having an army of additional data stewards in your organization. Agents can increase the level of automation in both data anomaly detection and correction, lower the volume of manual decisions and interventions thus speeding up processing. Having agents learning new ways to remedy problems reduces the reduces the time and effort needed to release new code as required with rule-based systems."

While LM performance is constantly improving, he points out that there are drawbacks to scaling agentic AI applications. One "is model inference time, and second is error cascading. One problem of agentic scaling is that LM inferencing time is on the order of seconds instead of milliseconds for traditional rule-based systems. This makes scaling to process tens of millions of records a challenge."

The key strategies here, Talburt advises, "are to run on local models, not remote API calls. Whenever possible, employ graphic processing units (GPUs) and large memory spaces. And don't abandon rules! Combining deterministic rules and inferencing where appropriate can create efficient hybrid solutions. Another strategy is to employ distributed parallel processing where the worker nodes employ small language models (SLMs) or distilled LLMs. Many applications such as master data management (MDM) and semantic search use the power of LMs through embedding models where text records are transformed into high-dimensional vectors for faster processing."

There are many architectural issues for ADQ such as modification, maintenance, reusability, interoperability, vulnerability and security, but these will have to wait, Talburt says. “However, the parting thought for this column is governance. ADQ and agentic AI in general have elevated the need for governance to a new level. Because agents can make decisions and take actions, they must be subject to the same governance policies as employees. New AI frameworks continue to make it easier for non-technical employees to build and deploy agents. Rogue agents can cause tremendous damage to organizational security, regulatory compliance, integrity and reputation.”

Therefore, he stresses, “human oversight is still important for critical decisions. Strict, least-privileged access to data by agents must be enforced to prevent unauthorized data exposure and agents must be limited to well-defined tasks. It is also important to maintain an inventory of agents and document their actions, ensuring decisions are understood. There should be clear human ownership, responsibility and accountability for each agent.”

We can never be sure of what the future holds, but Talburt is “hopeful that ADQ will turn out to be one of our best friends.”

For considerably more, check out Data Speaks for Itself: Agentic Data Quality – The Beginning of a Beautiful Friendship - Dataversity.