Beyond Metrics

Using AI to Unlock Genuine Impact

AlterSapiens

INTRODUCTION

Across governments, businesses, institutions, and even individual efforts, the same question haunts us: are we actually affecting meaningful change, or are we simply tallying outputs that feel reassuring but offer no real progress? In theory, many organisations proclaim a mission centred on human wellbeing—yet in practice, they often track superficial indicators to signal "success." This gap in measurement undermines trust and diverts attention from people's actual needs. As Goodhart's law warns, "When a measure becomes a target, it ceases to be a good measure." In other words, being fixated on certain metrics can pervert priorities, pushing systems to appear busy rather than deliver genuine impact.

Before levelling too many criticisms, however, it is worth pausing to consider the deeper questions:

- Can true impact, in all its nuance, be measured accurately?

- What flaws in our conventional frameworks perpetuate hollow measurements?

- How might new possibilities in generative AI (Gen-AI) help us move away from shallow targets towards a more meaningful understanding of impact?

- What fresh challenges, ethical implications, and risks accompany this AI-driven transition?

This paper sets out to answer these questions. It examines where traditional measurement came from, why it often goes astray, and how Gen-AI can radically disrupt our thinking about what "impact" truly means.

Through real-world examples, references to both historical and modern practices, and insights gleaned from the latest AI developments, it will argue for a decisive shift: from a narrow focus on ticking boxes to a more insightful and adaptive way of gauging results.

1. From Purpose to Procedure

The Evolution of Measurement

Early Informal Measures

Human beings have always sought ways to evaluate success, long before formal analytics or detailed records existed. Early methods were informal—rooted in subjectivity, personal observation, and stories shared through social bonds. Even the Babylonians meticulously documented financial dealings, showing a fundamental desire for accountability and trust.

In ancient hunter-gatherer communities, physical symbols like animal teeth or horns served not just as trophies but as clear markers of a hunter's skill and reputation. Anyone in the group could see these tokens and recognise the individual's achievements. Similarly, in medieval Europe, performance assessment was profoundly personal, hinging on a merchant's trustworthiness and honesty.

Word-of-mouth, longstanding business relationships, and consistent results were crucial. These informal but powerful "trust-based metrics" ensured reputations were built over time and anchored in tangible deeds.

Yet such methods had limits. Personal bias and uneven standards often crept in, leaving too much open to interpretation. A reliance on one's social standing also meant that success could be measured inconsistently, with some individuals judged generously while others faced harsh scrutiny. Over time, these drawbacks fuelled the move towards more standardised measurement practices that could, at least in theory, offer consistency and fairness on a larger scale.

Industrialisation and the Rise of Structured Measurement

Scientific Management and Taylorism

The Industrial Revolution introduced large factories, mass production, and vast workforces, which made purely informal ways of gauging performance inadequate. Frederick Winslow Taylor's "scientific management" approach, often called Taylorism, epitomised the era's craving for efficiency. Taylor believed one could identify the "one best way" to perform any task by breaking it down and meticulously timing each step. Any wasted movement or idle moment was exposed and then eliminated.

On factory floors, this style of measurement famously boosted output, but it also brought problems. Treating workers like cogs in a machine alienated them, and focusing only on speed or volume created tunnel vision around a few basic metrics. As manufacturing practices shifted into service sectors, the rigid nature of Taylorism often clashed with settings where creative thinking, relationship-building, and flexible problem-solving mattered.

Although Taylor's ideas introduced the value of precise data, their narrow scope showed how easily measurement can overshadow deeper aims—such as worker wellbeing and customer satisfaction.

Emphasis on Productivity, Efficiency, and Standardisation

Beyond Taylorism, the broader industrial era prized three watchwords: productivity, efficiency, and standardisation. Organisations considered easily trackable metrics—like output volume, labour hours, or defect rates—as unequivocal indicators of success. Henry Ford's assembly line demonstrated how standardisation could quickly lower costs and raise productivity, reshaping industries around the globe.

Yet as successful as these strategies were in the short term, they left little room for intangible qualities like creativity, worker morale, or sustained innovation. Mass production, while excellent at delivering uniform goods quickly, often fostered a monotonous work culture that undervalued unique talents. Issues like burnout, alienation, and ethical shortcuts sometimes followed. Thus, while structured measurement frameworks let leaders keep a firm grip on production, they also revealed just how easily efficiency could trump other vital considerations.

Audit Society and Bureaucratic Complexity

Growth of KPIs, Balanced Scorecards, and Compliance Checklists

By the late 20th century, measurement practices had grown vastly more complex. Organisations embraced Key Performance Indicators (KPIs), Balanced Scorecards, and compliance checklists. These systems aimed to encourage accountability and bring a broader view of performance. KPIs gave managers a quick read on key metrics they deemed most important— anything from product defects and sales numbers to social media engagement. Meanwhile, the Balanced Scorecard

framework tried to extend performance measurement beyond finances, exploring factors like internal processes, learning and growth, and customer views. Compliance checklists further ensured firms met legal and regulatory standards, especially in heavily regulated environments like healthcare or banking.

Unfortunately, the proliferation of these tools could also create layers of bureaucracy. As each metric acquired its own reporting structure, staff sometimes felt overwhelmed by form-filling, tick-box exercises, and data entry. While the original intent was to have a more balanced and transparent performance system, an obsession with meeting external requirements and meeting audit demands could overwhelm meaningful action.

Transformation from Management Tool to Compliance Burden

Over time, these structured measurement frameworks drifted from their original aim —to improve performance—and became exercises in risk management and regulatory compliance. Reports had to be produced to satisfy external bodies or internal oversight committees. In many organisations, measured outcomes devolved into targets that needed to be checked off, whether or not they revealed genuine improvements. This led to "gaming" the system: staff might tweak figures to look more impressive, or over-emphasise short-term metrics that auditors loved, instead of striving for real, long-term value.

Rather than spurring innovation, these frameworks, when poorly handled, effectively nudged teams to prioritise appearance over substance. Tasks were carried out because the metrics required them, not necessarily because they enriched the product, service, or community.

## Case Examples Highlighting Dysfunctional Outcomes of Measurement Fixation

Wells Fargo: Driven by aggressive cross-selling targets, staff created fake customer accounts to hit quotas, causing widespread scandal and eroding trust.

Healthcare: Hospitals pressured to reduce waiting times or readmissions have been caught discharging patients too soon or manipulating figures, undercutting the real goal of improving patient health.

Education: Schools emphasising test scores can end up "teaching to the test" at the expense of broader intellectual development. Pupils pass exams but leave with limited problem-solving skills.

Law Enforcement: Some police forces reported "better" crime statistics by reclassifying or failing to record incidents. Such tactics mask the real state of public safety and can harm community relations.

These instances underscore the hazards of rigidly following numeric metrics, suggesting that a healthy degree of scepticism and oversight is necessary whenever we rely on numbers to guide action.
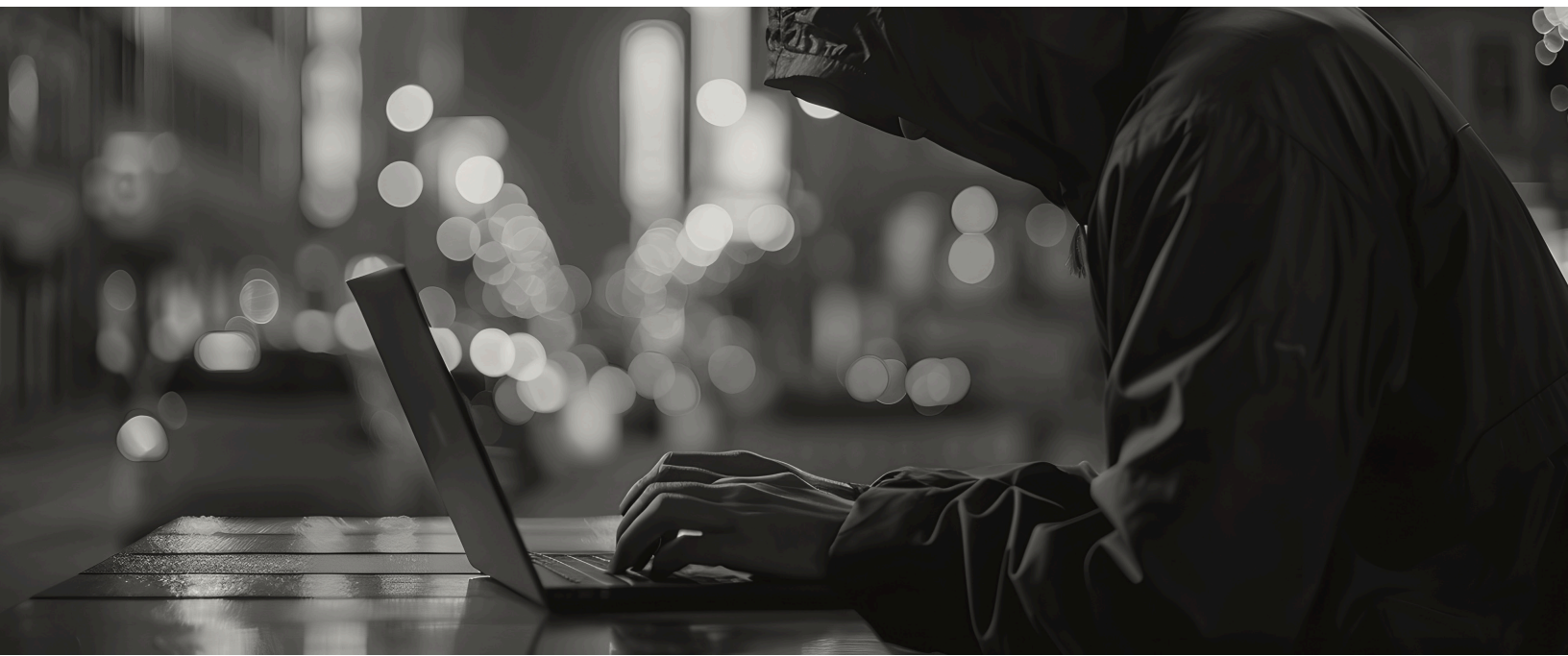
2. The Challenge

The Limitations of Traditional Measurement Approaches

Rigid Frameworks and Their Unintended Consequences

Goal Displacement: Healthcare, Policing, Education

In many sectors, the anxiety to "hit the numbers" can sideline broader objectives. Hospitals, for example, might focus on meeting readmission targets rather than quality of care. Police forces, looking to reduce official crime rates, may prefer methods of categorising incidents that make statistics look better without actually improving public safety.

Similarly, an overemphasis on standardised test results can make schools treat teaching as an exercise in rote memorisation rather than nurturing lifelong skills.

Such forms of "goal displacement" are common when organisations become beholden to narrowly defined measures. Instead of seeing metrics as tools for discovery and improvement, they treat them as the main objectives. Meanwhile, any complexities or nuances that do not fit the measures are ignored.

Over-Optimisation of Easily Measurable Outputs Rather Than Intangible Outcomes

Traditional measurement systems excel at counting tangible, immediate outputs. Units produced, transactions processed, or hours worked are all concrete and simple to benchmark.

However, they rarely do justice to bigger picture outcomes, like improving community wellbeing, creating transformative employee experiences, or ensuring a long-term sustainable strategy. When managers cling to the data that is easiest to track, they often neglect or downplay the deeper value that might be harder to capture with numbers

Structured Data vs. Complex Realities

Why Traditional Methods Fail to Measure Qualitative and Nuanced Impacts

Typical performance systems thrive on structured data—quantitative inputs, charts, and checklists. Although helpful for certain goals, such data can overlook qualitative elements such as customer emotions, employee engagement, or social dynamics. Nuanced impacts, like how people feel about a brand, how confident staff are in their roles, or how much trust a community has in a local government, remain hidden if we rely solely on numeric indicators.

Even well-designed surveys can fail to capture all the intangible factors at play. Once performance data is compressed into numbers, subtle insights frequently fall by the wayside. This one-sided lens may yield misleading conclusions—organisations could boast about improvements on a spreadsheet while ignoring deteriorating morale or creeping dissatisfaction among staff or stakeholders.

Examples Illustrating the Gap Between Reported Success and Actual Impact

- Corporate Sustainability: A firm might publish glowing numbers on its reduced carbon emissions, while failing to account for its supply chain's harmful practices.

- Development Projects: Some initiatives celebrate how many workshops they have conducted, yet cannot show any meaningful change in participants' everyday lives.

- Universities: Graduation rates and job placement statistics may appear stellar, but they give little insight into the actual quality of education or whether graduates feel equipped for careers that match their skills.

These situations reveal how easily an organisation can claim "success" when key intangible elements remain unmeasured or misrepresented.

High Cost, Low Insight Dilemma

The Paradox of Excessive Measurement Practices Leading to Reduced Adaptability

There is a curious paradox that surfaces when organisations implement layers upon layers of performance metrics: they can become so consumed by tracking and reporting data that they lose the ability to adapt quickly. Staff time and resources may be swallowed up by compliance tasks, leaving them with little bandwidth to experiment, learn from the front lines, or address challenges in creative ways.

Ironically, measurement practices that aimed to give leaders more control can end up paralysing the very innovation needed for long-term success. In fast-evolving sectors, from tech start-ups to social enterprises, this bureaucracy can be lethal. Organisations that should pivot rapidly instead remain stuck in an endless loop of data gathering, updates, and reviews.

3. Generative AI

Unlocking Impact Through Unstructured Data

Why Gen-AI Is a Game-Changer

Capabilities in Natural Language Processing, Image Analysis, Audio Sentiment Detection

Generative AI offers a way around the rigidities of structured data by tapping into text, images, audio, and beyond—sources of insight that traditional methods either struggle with or ignore. Natural Language Processing (NLP) can sift through large volumes of written content, from social media posts to open-ended survey responses, detecting subtle themes and emotions. Advanced image analysis can identify patterns in photos, scans, or live-streamed footage. Meanwhile, audio sentiment analysis can pick up the emotional tone of voice in customer service calls or public addresses, offering a deeper look into how people truly feel.

By broadening the scope of what is measurable, Gen-AI moves us beyond old metrics. Instead of relying on the bits of data that happen to be numeric, AI can interpret the context and nuance in everyday language, visuals, and sounds. This shift has enormous potential for capturing the kind of intangible outcomes that standard frameworks often miss.

Converting Qualitative Inputs into High-Fidelity, Quantifiable Insights

Perhaps the biggest advantage of Gen-AI is that it can transform qualitative inputs—such as personal narratives, testimonies, or interview transcripts—into data that is both rich and quantifiable. Organisations can gain structured insights into sentiment, urgency, themes, or patterns, all while retaining the depth and complexity of original accounts. This goes well beyond standard rating scales or tick boxes.

For example, a charity might use AI to summarise hundreds of personal stories from programme beneficiaries, grouping them by themes like improved confidence, better social ties, or financial stability. While still reflecting the lived experiences of participants, these stories now become more accessible for analysis and strategic planning. In short, Gen-AI can marry the richness of real human feedback with the clarity of data-driven models.

Real-World Applications

AI Analysing Citizen Feedback for Service Improvements

Governments collect endless comments from citizens—some formal, others informal. AI can organise these inputs, highlight common concerns, and spotlight underserved areas. For instance, a council could review local social media posts, discussion forums, or surveys about public transport. Instead of plodding through anecdotes by hand, it can get a detailed report on recurring problems such as unreliable bus timing or neglected facilities, then act more decisively.

AI-Driven Insights from Patient Narratives and Clinical Notes

Hospitals and health authorities track countless clinical notes, patient surveys, and care recommendations. AI can examine these unstructured sources for recurring phrases, complexities in patient journeys, and treatment outcomes. An AI tool might detect, for example, that certain expressions of anxiety among cancer patients correlate with later issues in recovery. By turning unstructured narratives into meaningful indicators, healthcare providers can foresee problems earlier and tailor interventions to improve patient wellbeing.

AI Measuring ESG and Sustainability Outcomes from Diverse Data Sources

In the realm of impact investing, real ESG (Environmental, Social, and Governance) performance is not always reflected in official reports. AI can pull from satellite imagery, local media, stakeholder interviews, and more to evaluate whether a company truly follows its publicly declared sustainability promises. For instance, an AI-powered system might verify if an agricultural firm uses ethical labour practices by matching on-the-ground news stories with corporate workforce data. If the data conflicts with marketing claims, investors can investigate further and steer funding towards genuinely responsible ventures.

4. AI Challenges

Navigating Challenges and Limitations of Gen-AI

Data Quality and Ethical Concerns

Addressing Biases, Ensuring Data Integrity

No matter how sophisticated AI is, it remains vulnerable to biases—especially if the data it learns from reflects historic inequalities or unrepresentative samples. If an AI tool is trained mainly on data from one region or demographic, it may produce skewed outputs that disadvantage those with different backgrounds.

Organisations must therefore adopt strict data governance, scrutinise inputs for hidden biases, and continuously monitor results to guard against discriminatory effects.
Data integrity also matters greatly. Outdated, incorrect, or manipulated information undermines AI's credibility. That is why maintaining accurate, timely, and ethically gathered data is essential. A rush to adopt AI without safeguarding data integrity risks damaging both performance and public trust.

Ethical Challenges in Algorithmic Fairness and Transparency

Many Gen-AI models, especially deep learning systems, function as "black boxes." They can deliver remarkably precise results while offering little clarity about how those results were derived. Lack of transparency can erode trust, particularly in sensitive areas such as healthcare, credit scoring, or public policy.
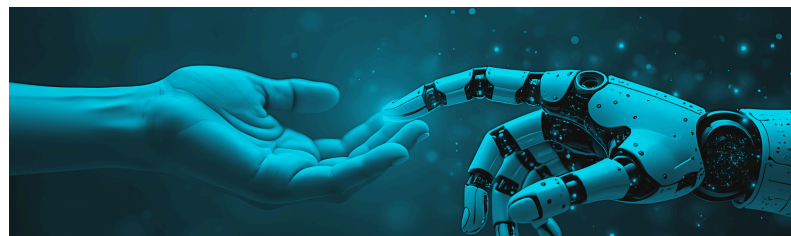
To address this, stakeholders are increasingly calling for explainable AI (XAI) methods, which provide more interpretable outcomes and enable human reviewers to understand how certain inferences are made. Without this accountability, AI-driven decisions might perpetuate unfairness—especially if biases are concealed within complex algorithms.

Organisational Resistance and Cultural Barriers

Overcoming Skepticism About AI-Generated Insights

Even when AI is accurate, people might hesitate to trust automated findings that seem to bypass human insight. Employees might fear that AI threatens their jobs or expertise. Leaders could worry about the possible errors in an AI system that cannot always explain its decisions in plain language.

Countering this scepticism requires transparent communication about AI's role and limitations, along with training sessions so stakeholders understand how AI can complement rather than replace human judgement. Demonstrating small, successful AI pilots can also help build internal trust, proving that an AI tool genuinely supports problem-solving rather than imposing rigid outcomes from on high.

Case Examples of Successful and Unsuccessful AI Adoption

Netflix: By showing how personalised recommendations steadily improve the user experience, Netflix built trust among customers, employees, and content creators alike. Its AI strategy was incremental and openly communicated.

IBM Watson for Oncology: Despite hype, some oncologists felt Watson's treatment suggestions lacked transparency. They did not fully trust how Watson arrived at its recommendations, limiting its acceptance and real-world impact.

Amazon's Recruitment AI: The project folded after it was discovered the model displayed gender bias, reflecting the imbalance found in historical hiring data. This illustrates the importance of continuous bias checks and diverse training sets.
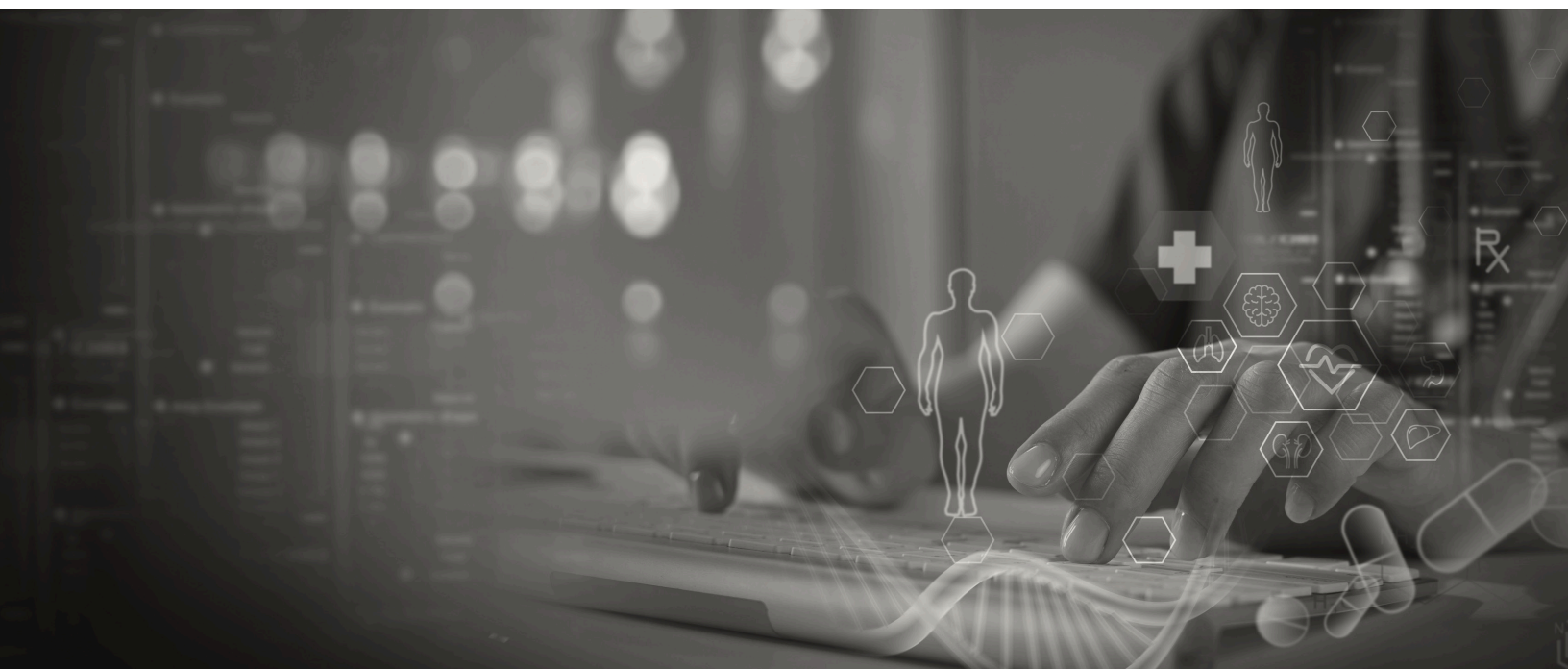
These diverse outcomes remind us that success with AI depends on how well it is integrated into people-centric processes, how responsibly data is managed, and how open the organisation is to both adopting AI and monitoring it for unintended consequences.

The Risk of AI Dependency

Balancing AI Analytics with Human Judgment

While AI can offer data-driven clarity, relying on it too heavily might weaken essential human judgement and moral reasoning. If decision-makers see AI as infallible, they may fail to recognise when an algorithm misfires or oversimplifies a nuanced problem. A robust approach places AI insights alongside human expertise, so that final decisions reflect both computational precision and social awareness.

In healthcare, for example, clinicians can use AI diagnoses as an important reference but still apply their professional insight and patient empathy. In corporate strategy, executives might factor AI predictions into their decisions but make room for innovative leaps or intuitive calls that data cannot fully capture.

Avoiding New Forms of Metric Fixation and Black-Box Decision-Making

Ironically, the advanced analytics of Gen-AI can create fresh temptations to chase yet another set of narrow metrics. If organisations simply replace old key performance indicators with AI-spawned metrics, the same pitfalls remain—obsession with numeric goals that fail to address deeper challenges. Additionally, black-box AI models can prevent people from questioning or refining a system's hidden assumptions, encouraging a lazy acceptance of algorithmic outputs.

To avoid such outcomes, leaders must establish transparent guidelines, encourage explainability, and continually ask whether AI metrics genuinely align with the organisation's broader purpose. AI should illuminate blind spots, not become an inscrutable authority that stifles curiosity and complexity.
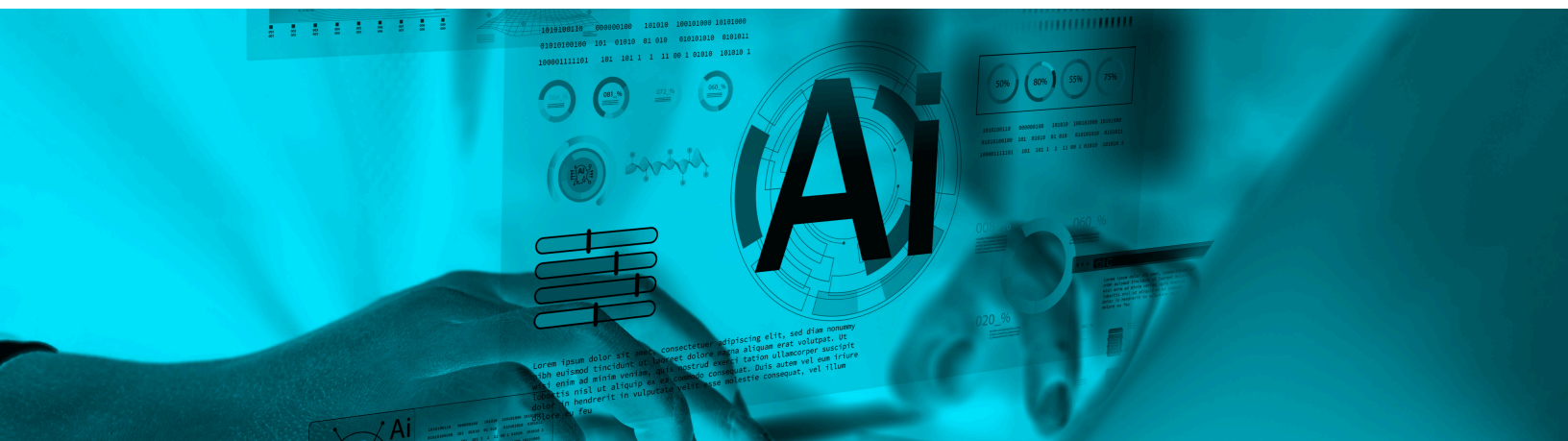
5. Impact Measurement

Provocative Questions & New Thinking Methods

Critical Reflection Questions for Decision Makers

If we are to move beyond shallow metrics, we must ask ourselves challenging questions:

- Do our current measures serve our mission or just keep us busy? For instance, a charity might tally the number of meals served, but does that truly capture improvements in nutrition or long-term self-reliance?

- Are there metrics we have clung to for years that might be irrelevant or misleading today? In a digital marketing department, click-through rates might look promising, but deeper engagement metrics derived from AI might be far more revealing.

- Are we willing to let AI challenge our cherished assumptions about success?A financial institution might discover that certain "high-risk" customers are actually more dependable than traditional credit scores suggest, forcing managers to rethink standard approaches.

Asking these questions is inherently disruptive. It often reveals uncomfortable truths about how we define success. But this is exactly the point: **it encourages us to adapt, innovate, and align our goals with reality.**

Approaches to Balance AI and Human Insight

Adaptive and Iterative Measurement Frameworks (e.g., Agile, OKRs)

Frameworks like Agile or Objectives and Key Results (OKRs) enable continuous learning, frequent feedback loops, and iterative adjustments. Instead of waiting for annual reviews, teams can integrate AI findings in near-real time, refining their metrics and strategies as insights emerge. This helps prevent the rigid mindset that normalises poor performance simply because it "meets the old standard." It also promotes more flexible and responsive ways of working.

Human-Centred Evaluation Methods, Blending Narrative and Quantitative Data

Rather than replacing human insights with raw data, forward-thinking organisations weave personal stories, case studies, or participant narratives into their AI-driven results. For instance, if AI indicates rising staff dissatisfaction, managers can read anonymised yet detailed personal testimonies to grasp the root causes. This blend of narrative and quantifiable data captures both breadth and depth, ensuring that crucial human nuances are not lost in the chase for numbers.

Systems Thinking as a Lens for Deeper Understanding of Impact

Systems thinking views organisations and communities as intricate, interconnected webs. Gen-AI can reveal these complex relationships by cross-analysing data from different domains. For example, an intervention in a city's public transport might affect local businesses, air pollution, and even school attendance rates. Systems thinking encourages holistic solutions rather than piecemeal attempts to improve isolated metrics. Combined with AI's data-processing power, it can illuminate the multi-layered consequences of seemingly small changes..

Encouraging Curiosity Over Compliance

Promoting Measurement as a Tool for Learning and Strategic Improvement Rather Than Just Accountability

A radical shift is needed. We must stop treating measurement merely as a routine check for compliance or an exercise for audits. Instead, metrics should spark curiosity and strategic thinking, helping teams explore the "why" behind the numbers, identify opportunities, and correct course quickly.

AI acts as a powerful enabler here, surfacing trends or anomalies that prompt deeper enquiry. When staff see measurement as a source of insight rather than punishment, they become more open to honest data sharing and constructive debate. Over time, this fosters an internal culture that truly values exploration, adaptation, and genuine improvement.

A Call for Mindset Shift, Not Just Technology Adoption

Generative AI holds the potential to push us far beyond the confines of outdated measurement paradigms, offering fresh ways to unearth genuine impact. However, adopting AI tools alone is insufficient. Organisations must rethink their entire approach to evaluation, re-examining everything from entrenched biases and data policies to how they involve employees in AI-driven change. The goal is a profound mindset shift in which measurement becomes a catalyst for learning, purpose, and real transformation.

Leaders should be prepared to question and abandon deeply rooted practices if Gen-AI reveals contradictions or missed opportunities. Ethical considerations must remain central, guiding how data is collected, analysed, and used. When handled carefully, AI can bridge the gap between slick numerical metrics and the rich, human realities of impact—unearthing both success stories and hidden problems that conventional figures have long concealed.

Ultimately, the challenge is to harness these emerging tools responsibly and imaginatively, so that "impact measurement" evolves into something that truly informs and advances our most important aims. If we insist on sticking to the familiar comfort of outdated metrics, we risk failing those who depend on us for meaningful progress. By embracing critical questions, fresh perspectives, and the insights that AI can unlock, we can finally venture beyond measuring what is convenient towards capturing what genuinely matters.

Beyond Metrics, using AI to unlock genuine impact.
First published 26 March 2025
Author: Cumai Aboul Housn

Artist, Author, Tech Transformer, Change Agent & Future Explorer.
Discovering how genuine human conversations can make a difference
in our digital age. Building disruptive Art-Tech, AI & Big Data solutions.
Founder AlterSapiens.com
Co-Founder Biennale.io
Co-Founder CriticalConvo.com

Images Source: Adobe Stock

References
• Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction Machines: The Simple
  Economics of Artificial Intelligence. Harvard Business Review Press.
• Hay, D., Knechel, W.R. and Willekens, M. eds., 2014. The Routledge companion
  to auditing. Abingdon, Oxon: Routledge.
• Muller, J., 2018. The tyranny of metrics. Princeton University Press.
• Schrage, M., 2019. Don't let metrics critics undermine your business. MIT Sloan
  Review.
• Kaplan, R.S. and Norton, D.P., 1996. Using the balanced scorecard as a strategic
  management system.
• Jones, E.D., 1911. The Principles of Scientific Management.
• Ragothaman, S., Custis, T. and Christianson, M., 2022. Fake accounts scandal at
  Wells Fargo: What are the lessons?. Journal of Forensic & Investigative
  Accounting.
• Faheem, H. and Dutta, S., 2023. Artificial intelligence failure at IBM'Watson for
  Oncology'. IUP Journal of Knowledge Management..
• Kodiyan, A.A., 2019. An overview of ethical issues in using AI systems in hiring
  with a case study of Amazon's AI based hiring tool.

AlterSapiens LTD. is a leading creative technology lab, dedicated to addressing
real-world challenges and shaping the future of business and society. Our team
of innovators, thinkers, technologists, and designers work collaboratively to
conceptualize and create solutions that are future-ready.
By envisioning a broad concept of metaverse; we're about harnessing the power
of technology, artificial intelligence, and immersive experiences to disrupt
traditional paradigms and create value. Our commitment is to build a future
where digital solutions seamlessly integrate with physical realities, ensuring that
businesses and individuals are equipped to navigate the challenges and
opportunities of tomorrow.

AlterSapiens LTD.
15176084
London, UK

ψ AlterSapiens