Published in Forbes France Digital:
https://www.forbes.fr/business/comment-les-entreprises-peuvent-elles-etre-a-la-pointe-de-la-creation-dune-ia-meilleure-et-plus-fiable/

A contribution by Marianne Mazaud, General Director & Cofounder at AI ON US, and Mark Ribbing, a strategic-communications executive and consultant who formerly managed the writing team at the Gates Foundation. Prior to that, he was an Obama administration appointee at the US Department of Defense.

How companies can lead in creating better, more trustworthy AI

Introduction: Consumers have real concerns about AI and many corporations do, too.

Even as artificial intelligence becomes more capable, and permeates more aspects of our daily lives, the predominant response to this development seems not to be one of celebration – but rather, one of marked uncertainty and ambivalence.

This sentiment has been captured in some comprehensive recent studies. Earlier this year, KPMG and the University of Melbourne released the results of a massive survey on global attitudes toward AI, for which researchers interviewed over 48,000 people in 47 countries.

Worldwide, a bit over half of respondents (54%) said they were wary of AI, with levels of distrust actually somewhat higher in “advanced” economies, and lower in “emergent” ones such as China, India, Nigeria, Egypt, and the United Arab Emirates.

The KPMG/Melbourne team found that overall trust in AI declined “in most countries” as the technology has become more widespread. “ This increased adoption is coupled with a trend toward people feeling more concerned about and less trusting of AI,” the report stated. “More people report feeling worried about AI and concerned about the risks, and fewer view the benefits of AI as outweighing the risks.”

Such uncertainty is by no means confined to consumers. Major corporations often herald the upside of AI in their marketing materials and public pronouncements, but a report released in July by the London-based Autonomy Institute found that in their mandatory annual disclosures of potential investor risks, companies in the S&P 500 have reported serious and growing concerns about a variety of AI-related threats to their bottom lines.

For example, just within a single year, three of every four S&P 500 firms have added or expanded upon their disclosures of risks stemming from AI. The number citing deepfakes as a business risk doubled, from 16 to 40; the number citing AI-bias hazards also doubled, from 70 to 146.

What such findings suggest is that while consumers and corporations are clearly adopting AI for an ever-broader array of needs, they are doing so while simultaneously harboring some real misgivings about what this technology represents, and where it is headed.

Trust: The necessary foundation for a free and prosperous future.

This ambivalence, among consumers and companies alike, derives from an increasingly evident truth: While the potential performance and efficiency gains associated with AI are quite compelling, the risks are at least equally so.

Those risks loom large in the concerns shared not just in these recent studies, but also in other public surveys and expert commentaries – including from within AI companies themselves.

These perceived risks include not only various forms of misinformation and algorithmic bias, but also the degradation of democratic political systems; mass job loss; the diminution of human cognitive and social capacity; the environmental toll of AI’s vast energy demands; and, most frighteningly, the subordination of human agency to a superintelligent AI whose capacities and motives are no longer under our control or even truly knowable.

This is of course not an exhaustive list of AI's risks, but it is enough to point to the scale and complexity of the challenges we face when considering the future of this technology.

In the face of all this, it's very tempting for companies to seek the seemingly expedient route of pocketing whatever near-term gains AI can offer, while leaving the addressal of risks to someone else. But while such a shortcut may seem attractive, it could soon undermine whatever hopes executives might have for AI-driven prosperity.

If the public perceives that this technology is bringing us closer to the kinds of risk scenarios outlined above, popular ambivalence and even hostility toward it – and toward those companies seen as deploying it irresponsibly – may become more pronounced. Such an erosion in consumer trust could affect both adoption levels and efficiency gains, compromising important elements of AI's overall value proposition.

Then there is the still-significant matter of attracting and retaining human talent. In the intensifying competition for the best minds, employers have a practical incentive to present themselves as proponents of sound and constructive AI practices. This may be especially true with highly skilled younger workers, who often place an especially high value on such qualities as authenticity, credibility, and overall social impact.

Shifting posture: From mere compliance to strategic creativity

If companies have such potent incentives to nudge AI toward the most trustworthy and publicly acceptable version of itself, how can they feasibly do so?

One way is by recognizing the vital importance of business's role in promoting clear and effective industry standards for the development and deployment of AI.

This is not to devalue the role of national and multilateral regulation. Such rule-setting is of course utterly necessary. But there are some very fundamental reasons why this technology demands an especially proactive business role in the delineation of optimal practices.

Among those reasons is the sheer speed of AI's evolutionary cycle. The companies developing and using AI can promulgate and hone standards far more swiftly than governmental bodies can generate meaningful regulations.

Those standards can, in turn, inform and enrich the regulations that do eventually emerge. This could reduce the hazard that governments push forth policies that are uninformed, overbroad, inconsistent, or otherwise misaligned to the problems at hand.

Another reason is that a robust and highly visible standards system would correlate with real-world incentives. Companies and research institutions seen as adhering to such a system would possess a practical means by which to signal trustworthiness to consumers, regulators, vendors, and potential employees.

This kind of validation is an especially valuable form of currency in a sector where trustworthiness is likely to become not only more highly valued (the KPMG/Melbourne study found that “the importance of organizational assurance mechanisms as a basis for trust increased in all countries”) but also more difficult to discern absent some sort of broadly accepted validation system.

Such a body of standards might use an international framework – such as ISO 42001 – as a starting point, augmenting it as needed to respond to swiftly changing technological and sectoral circumstances.

These standards would not be mere virtue-signaling mechanisms, but recognized practices for keeping the industry ahead of its own technological curve – and for fostering a competitive market that doesn’t undermine the sources of its own future vitality.

We are at a turning point. One where the most powerful technologies ever created can reinforce and exacerbate inequalities – or build a more just world. One where companies can lock themselves into minimal compliance – or seize the strategic opportunity of responsible AI.

The future is not something to endure. It is something to build. Together. Starting now.