Measuring AI Inputs: Challenges and Opportunities

Prasanna (Sonny) Tambe, tambe@wharton.upenn.edu

DRAFT. Comments welcome.

1. The current state of play with regards to AI measurement, including links to your own work if applicable.

Why are some firms better at producing and using AI?

Differences in tangible and *intangible* factors may generate different patterns of AI investment and use among US firms. Some of our work has begun to explore how firms' accumulation of these underlying, foundational assets drives competitive differences in a firm's abilities to "produce AI capital" (Rock et al 2024). For instance, although it is clear that some firms, such as Google, have been particularly successful when using AI, these firms may have some type of "secret sauce", or alternatively and more likely, may simply have developed the right assets to support development of downstream tools related to algorithmic decision making. Here, I focus on the key measurement issues around three of the most critical categories of AI investment firms may be making to drive AI-enabled productivity: i) infrastructure (data, software, and computing), ii) workforce training, and iii) high-skill talent.

1. Infrastructure: Data, software, and computing. Access to key factors needed for effective AI use -- data, software, and processors -- is unevenly distributed. Certainly the volume and variety of data matters. Then, increasingly, with large AI projects, hardware (chips) itself has become a notable supply chain constraint. Much software at the AI frontier is available through open source platforms, but firms differ in their ability to adapt and deploy these frameworks in their production context, partly due to differences in access to talent, because of the costs imposed by legacy systems, uncertainty around regulatory environments in some industries, or other factors.

2. Workforce training. A particularly important category of AI investment is the workforce training and reskilling investments that are likely to be needed for AI. Much of the attention on AI and labor has been focused on the potential for labor displacement. This stream of work forecasts occupations most vulnerable to job displacement based on the task content of that work. Then, given the occupational distribution of employment in the US and other countries, the approach enables analysts to estimate the potential impact of AI technologies on the demand for different occupations.

But, in the short-run, the larger workforce changes employers will need to make may be the reskilling of existing occupations to enable human-AI collaboration. As AI diffuses into a larger number of jobs, workers in occupations ranging from sales to accounting to customer service will need new skills to be effective. The idea that human-AI collaboration will be the dominant mode of AI use in the medium-term has been underscored by the entry of large language models into the public conscience. Call centers of the future will still need customer service agents, but they will look different than they do today.

3. Talent. Finally, the talent pool for AI development is a critical input to driving AI progress. This includes not only highly paid software developers, upon whom the media likes to focus, but also the development of a robust semiconductor workforce that can provide support for the investments in physical capital related to the 2022 CHIPs Act. Universities like Ohio State and Purdue have initiated training programs specific to this kind of work, but whether the US can successfully supply well-trained workers to these reshored semiconductor initiatives in a way that allows them to replicate foreign manufacturing capabilities is an open question.

A proper accounting of all of these factors -- data, software, and computing, as well as talent differences -- can improve our understanding of why firms differ in their prospective AI capabilities.

2. The conceptual and empirical gaps and opportunities.

Infrastructure: data, software, and computing. Of the various factors separating firms trying to deploy AI technologies, few are measurable and fewer still are measured. How should we account for the value of firms' **databases**? The recency of its data? Its variety? How quickly does data depreciate and how does this vary across industries? Data on consumer preferences may depreciate quickly, but the value of databases of images of human x-rays might retain their value longer.

Another complication is that AI **software**, even frontier software in this domain, is increasingly made available through open source channels. State of the art LLM capabilities can be accessed through Meta's LLaMa project, for example, and it is free to use even for commercial purposes. So, software expenditures in the AI space provide a less useful signal of AI capabilities than figures on ERP spending would for supply chain capabilities. When firms contribute to, download, or use AI software obtained through open source channels, they do produce some digital trails, such as Github forks/clones, but it is not easy to convert these indicators of digital interest to measures of meaningful software use. Any exercise that attempts to equate these measures to real outcomes is likely to suffer from severe measurement problems.

On the **computing** front, the growing role of cloud computing for AI projects complicates our ability to measure and track firms' access to computation. In comparison with expenses on internal servers owned by the firm that can be tracked, cloud expenses are more elusive to gather and it is not even clear how any figures that are obtained on cloud computing expenditures can be converted to meaningful quantities. Depreciation matters here a great deal as well. Early figures on AI-specific chips suggest particularly high levels of utilization and a need for frequent re-architecting to take advantage of software innovations, so depreciation for this hardware may be high.

Talent. For other domains of national importance, such as the R&D inputs needed for scientific progress, scholars have paid close attention to the role of high skill individuals such as scientists in driving innovation, productivity, and growth outcomes. These economic activities are often tracked through databases on patenting activity or federal funding that are informative about the career activities of scientists. Just as with basic science, certain individuals and organizations appear to have played an outsized role in pushing the AI frontier forward, but there may be no ready equivalent for the types of databases that have been so useful for understanding key players in the scientific ecosystem.

This is because software developers, including star engineers, rarely leave reliable patent trails. Moreover, funding to AI development initiatives is often provided internally by corporations, rather than through public institutions like universities, making it harder to track what money is being allocated to research and development for AI projects. As a result, we are in a very different data environment when it comes to understanding the relationship between the supply of skilled talent and AI progress -- including the role of AI scientists, skilled software developers, semiconductor engineers, cybersecurity experts, and other tech workers who will eventually be critical for the AI ecosystem.

Workforce training. Finally, reskilling and job redesign are significant investments for employers and are critical for firms to be able to realize productivity growth from AI technologies. A key measurement gap with respect to these investments relates to the extent of workforce-level, sub-occupational change occurring in firms. The question here is: in a model of production where AI works with humans to generate productivity improvements, how should workers be reskilled to realize these gains? For instance, will workers be taught to discriminate when LLM output is likely to be trustworthy? Or, will effective salespeople need to be trained in prompt engineering?

As employers learn the answers to these questions and begin to make these workforce adjustments, there will remain large gaps in our ability to measure these changes because statistical agencies focus data collection at the occupational level. Short-run opportunities for AI-augmentation, on the other hand, are more likely to require changes to how workers are trained (Tambe 2024).

For instance, software "co-pilot" tools are expected to significantly change the way in which developers will spend their time, whether or not they change the number of developers that are needed to do the work. This margin of change is currently difficult to track, despite the fact that upskilling and retraining costs have always comprised a significant component -- often the largest component -- of the financial cost of IT transformation. Further, because AI is viewed by most as a general purpose technology and therefore relevant to the work content of most jobs, the costs of workforce adjustment that employers will incur may be even larger than the costs of IT-enabled work reengineering that prior waves of business technology have required.

Measurement gaps related to workforce needs were already large prior to 2022, but have likely only been amplified by rapid growth in the corporate use of generative AI tools and large language models. The use of generative AI tools requires workers to 1) appropriately structure prompts to get the types of output they want and 2) to critically assess whether LLM output is usable. There are still many unknowns about how workers will need to be retrained to work effectively with LLMs, but it is clear that changes will be needed.

Why are these measurement gaps, related to databases, workers, and other factors necessary for building AI systems, important?

Understanding how these assets are distributed among producers, as well as how they are accumulated by firms and how quickly they depreciate, can be useful for explaining why AI has a much larger impact on some sectors than others, and on some firms than others. It is also important for understanding how any productive gains from these investments can be expected to last as these assets depreciate and require ongoing investment (Rock et al 2024). Another implication of this gap in measurement is a limitation in our ability to forecast whether AI technologies are likely to be democratizing or whether they are instead likely to drive inequalities. Particularly given the interest in sources of competitive advantage for big tech or “superstar” firms in general, it would be useful to evaluate whether firm size matters for accessing these critical resources.

3. Short, medium, and long-term next steps.

How do we address these measurement gaps?

Workforce training and talent. The easiest place to start may be with workforce and talent measurement gaps.

Fortunately, the challenge of tracking how workers are being retrained to effectively use AI technologies, or measuring the supply of skilled talent, can already be partially met by corporate data sources. Data sets, already widely used, from providers like LinkedIn, Revelio, and Lightcast provide a window into this type of technical change (Horton & Tambe 2015).

However, academic measurement exercises based on these data sources tend to be ad-hoc, with no standard approach to measurement, making it difficult to compare measures across papers or time periods. For example, which skills and tools are “AI”? One study on the effects of AI technology might include “regression”, while another one may not. The lack of a clear taxonomy mapping technologies to skills has been a long-standing obstacle for all types of measurement in this area.

A useful objective would be to develop a standardized process (and taxonomy) through which to measure how these skills spread within the US workforce. This would allow for comparable and reproducible measures to be used across studies, industries, and periods.

Infrastructure. Measurement of AI inputs that are not embodied may pose a larger challenge. The data, software, and computing “stack” required to implement AI technologies has already changed rapidly over the last few years and this trend will continue and perhaps even accelerate. New tools, like “AutoML”, automate the machine learning process and reduce the importance of some factors relative to others. Macro-factors that affect supply of these infrastructure resources, such as geopolitical conflict or the introduction of new export controls, may also become important.

The first (short-term) steps for measuring AI infrastructural inputs are to identify the tangible and intangible assets of interest. If there is agreement that AI productivity reflects an accumulation of a common set of factor inputs, then developing consensus on what these critical inputs are in successful AI initiatives is a good starting point.

A longer term goal is to develop a framework for accounting for these assets. This itself is a considerable challenge, as the measurement and accounting of these different assets is not easy. Challenges with accounting for intangible assets, of course, have a long history. Nonetheless, this is a worthy project, and one could perhaps use proxy measures for different inputs -- like Oracle DBA developers to measure investment into Oracle databases -- even though there are likely to be a lot of measurement problems with this approach.

A better, and longer-term, goal is to improve reporting requirements for firms (e.g. software investment), as well as methods to value other inputs that allow for consistent measurement of a firm's digital assets.

Note that it is particularly important that any measurement attempts, whether they be related to the workforce or the infrastructural components, be reproducible across time periods because the capabilities of the tools themselves present a moving target. LLMs are significantly more capable than they were even a few months ago, so any prognostication related to organizational needs has to first build an accurate forecast of the evolution of these technologies.

Another source of uncertainty is regulation. We are entering a period of sustained uncertainty in which how AI technologies are used will need to be adjusted to meet changing regulatory requirements, so the diffusion path of these technologies, and how these technologies will be used in production, is uncertain. For instance, despite its quick start, current concerns about copyright violation may significantly slow LLM use in some industries.

We also need to acquire a much better understanding of how workers will use AI tools and what AI and human collaboration will look like, which in turn could be informative about how workers need to be reskilled for this form of collaboration.

For these reasons, establishing a benchmark now, when the adoption of most AI tools is still limited, and development of a repository of data sources, taxonomies, and approaches to measure this type of workforce change, can drive a much better understanding of how the US workforce is changing to meet the need for human-AI collaboration.

4. Possibility for your suggestions to inform the measurement of other critical and emerging technologies.

AI is one of many technologies likely to drive social change in the coming decades. It is probably the most important. However, there are other emerging technologies that will have a significant economic impact like virtual reality systems for work environments (such as that represented by Apple's Vision Pro and the ecosystem it will produce). VR systems have already significantly impacted some domains (e.g. robotic surgery systems such as DaVinci).

The approaches outlined above are not specific to AI. Any attempt to measure the impact of emerging technologies on workforce change can benefit from a well-supported infrastructure for measurement. Moreover, new emerging technologies are very likely to be closely interwoven with existing systems, including those that house data and that drive AI prediction. For instance, VR systems will integrate AI into their core functions. A proper accounting of AI assets is a first step that can serve as a foundation for better measurement for all technologies that build upon it.

References

Horton, J. J., & Tambe, P. (2015). Labor economists get their microscope: Big data and labor market analysis. *Big data*, *3*(3), 130-137.

Rock, D., P. Tambe, and Z. Wang. Deep capital for deep learning. Working Paper.

Tambe, P., Hitt, L., Rock, D., & Brynjolfsson, E. (2020). *Digital capital and superstar firms* (No. w28285). National Bureau of Economic Research.

Tambe, P. (2024). Algorithmic Bilinguals. Working paper.