Surveying Surveys: Surveys' Role in Evaluating AI's Labor Market Impact

Anonymous Author(s)
Affiliation
Address
email

Abstract

A majority of Americans believe that AI will reduce the number of jobs in their industry [1], with concerns about AI use rising by 40% between 2021 and 2023 [2]. Additionally, 37% of US CFOs reported that AI tools have automated tasks previously performed by workers [3]. We only know this information because of surveys. To deepen our understanding of AI's effects on humans, we should use more in-depth, more frequent, and broader surveys that evaluate the human experience of AI's impacts, especially in the workplace. By integrating survey data with technical performance evaluations, we can better understand AI's real-world impact on workers and forecast future disruptions. Alongside technical evaluations, survey evaluations can inform policy responses aimed at mitigating the negative economic consequences of AI adoption.

1 Introduction

Current evaluations focus on technical performance at the model or task level, while political discourse around model safety emphasizes pre-deployment concerns. This approach often neglects critical human factors, such as adoption rates, integration processes, and organizational personnel decisions. Although there has been progress in evaluating short-term labor market impacts—such as comparing AI-generated output to human work or assessing potential reductions in labor costs—these evaluations fall short in assessing realized impacts [4]. This challenge arises because such impacts are driven more by human and organizational decisions than by AI models themselves.

2 The Need for Surveys

This is why we need additional tools to evaluate AI's realized impact, and surveys are one such tool. We should pay more attention to existing surveys to understand the current state of AI implementation and societal sentiments. Furthermore, investing in new, regularly conducted, internationally representative surveys could provide valuable insights that current evaluations cannot capture. Expanding the use of surveys as an evaluation method would enhance assessments in three critical areas:

1. Measuring models' direct and indirect impacts on workers and organizations;
2. Informing data-driven policymaking in the public and private sectors; and
3. Guiding model development to create fairer, more equitable outcomes in the labor market.

Historically, we failed to mitigate the negative economic impacts of technology-driven job displacement. While this inaction was detrimental for many, the overall economic impact was limited because previous displacement tended to be slow-moving and geographically concentrated. In contrast, AI-driven displacement is likely to be faster, more widespread, and affect a broader range of workers [5].

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

Given these differences, it is crucial for surveys to evaluate AI's human impact by linking technical capabilities with real-world outcomes. This includes gathering post-integration data on AI's effects on work, workers, and organizations, as well as assessing AI's collective impact across industries. Surveys also offer a relatively low-cost, low-effort method to gather insights from a representative set of workers and organizations.

3 Survey Evaluations as a Policy Tool

The Michigan Survey tracks consumer sentiments and is used by economic policymakers to assess current conditions and forecast future trends [6]. Similarly, we need a mechanism that enables policymakers to make informed decisions regarding regulations, social safety nets, and other policy areas affected by AI-driven changes in the labor market. By combining insights from technical evaluations with survey data, researchers and policymakers can more accurately identify and forecast labor market disruptions. Survey insights can be used to:

- **Trigger policy responses** by continuously surveying workers, particularly those in AI-exposed professions, as seen in recent surveys conducted in Denmark [7];
- **Forecast the need for responses**, drawing on data such as the June 2024 Fed/Duke University CFO Survey, which reveals employers' future plans to automate tasks [3]; and
- **Assess the efficacy of policies** by embedding surveys into policy performance measurements to understand if the policy's intended human outcomes are achieved.

Policy responses should specifically address challenges related to job displacement, such as providing financial safety nets and reskilling opportunities. It is crucial for policymakers, researchers, and technologists to integrate surveys into current evaluation frameworks to ensure timely and effective responses to labor market disruptions.

4 Advancing Surveys

In addition to these limitations, there are overarching gaps in current survey data that need to be addressed. To fill these gaps, we propose developing new surveys or expanding existing ones to provide:

4.1 Regularly collected time-series data

Given the rapid pace of AI adoption and the evolving nature of AI capabilities, it is essential to track changes over time. For example, adding questions about AI-driven automation and job displacement to the US Census Bureau's Current Population Survey (CPS) would create a monthly record of employment changes tied to demographic data.

4.2 Forward-looking expectations

It is important to understand how workers and organizations adjust their behavior in response to AI integration. Currently, we lack data on how workers plan to respond to AI-related economic shifts (e.g., reskilling, financial planning). Polling efforts like the EU's Eurobarometer could incorporate questions on this topic to help shape future policy responses.

5 Limitations

The proposed solution is based on findings from a limited number of existing surveys (Appendix A) that do not yet exemplify human impact. Moreover, this research makes the assumption that survey data is accurate. Survey data—and self-reported data in general—are imperfect for measuring many aspects of the human impact of AI. Administrative data would be ideal, but we recognize that it can be more difficult and slower to attain and is possibly costlier to collect.

References

[1] Ray, J. (2024, August 27). *Americans express real concerns about artificial intelligence.* Gallup. https://news.gallup.com/poll/648953/americans-express-real-concerns-artificial-intelligence.aspx.

[2] Faverio, M., & Tyson, A. (2023, November 21). *What the data says about Americans' views of artificial intelligence*. Pew Research Center. https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/.

[3] Duke University, Federal Reserve Bank of Richmond, & Federal Reserve Bank of Atlanta. (2024, June 20). *U.S. companies ramp up automation and AI as inflation persists*. The CFO Survey.

[4] Solaiman, I., Talat, Z., Agnew, W., Ahmad, L., Baker, D., Blodgett, S. L., Chen, C., Daumé III, H., Dodge, J., Duan, I., Evans, E., Friedrich, F., Ghosh, A., Gohar, U., Hooker, S., Jernite, Y., Kalluri, R., Lusoli, A., Leidinger, A., ... Struppek, L. (2024). *Evaluating the Social Impact of Generative AI Systems in Systems and Society*. Preprint under review. https://arxiv.org/abs/2306.05949.

[5] Kochhar, R. (2023). Which U.S. workers are more exposed to AI on their jobs? Pew Research Center.

[6] Survey Research Center. (n.d.). Surveys of Consumers. University of Michigan. https://data.sca.isr.umich.edu/fetchdoc.php?docid=24774.

[7] Humlum, A., & Vestergaard, E. (2024). The adoption of ChatGPT. University of Chicago, Booth School of Business, & University of Copenhagen, Department of Economics. https://doi.org/10.1257/rct.12527-1.0.

A Appendix / supplemental material

Some surveys and polling are beginning to fill this gap. Yet, there are areas in which these surveys could go deeper to provide us with more insightful data around AI's impacts. Examples of surveys that have come out over the past few months that do fill some gaps include:

Table 1: Appendix A: Surveys related to AI and work (June-August 2024)

Survey	Population	Organization	Key Takeaways	Follow-up Questions
GenAISurvey –2024	225 global C-Suite and senior business leaders	KPMG	71% of business executives say they are using GenAI to leverage data in decision making, 52% say it is shaping competitive positioning, and 47% say it is opening new revenue opportunities.	What types of decisions are these leaders using AI for?
Americans Express Real Concerns About Artificial Intelligence	5,835 US adults	Bentley University & Gallup	75% of Americans believe AI will reduce the number of jobs over the next 10 years. 77% trust businesses "not much" or "not at all" to use AI responsibly. 85% expressed concern about AI making hiring recommendations. 57% believe that business transparency around AI use would reduce their concerns.	What were workers' personal experiences with AI or their expectations for their own jobs?
YouGov Survey: AI and Jobs	1,098 US adults	YouGov	48% of respondents think AI will decrease job opportunities in their industry. 1 in 3 are concerned about AI-induced job reduction or loss. 56% believe that the government should regulate AI in the workplace.	Which types of workers are already affected by income loss?
AI Survey: Four Themes Emerging	200 global companies	Bain	Poor performance and output quality were the top reasons that GenAI did not meet companies' expectations.	Did AI's performance change its implementation within organizations?
AI at Work 2024: Friend and Foe	13,102 global workers	BCG	42% of workers fear AI-related job loss (up 6 percentage points from last year); managers and leadership are more confident about GenAI.	Are these workers changing their behavior in response to this predicted job market shift?
U.S. Companies Ramp Up Automation and AI as Inflation Persists	2,200 US CFOs across businesses	Duke University, Federal Reserve	37% of CFOs said that AI tools automated tasks previously done by workers, with 54% planning to do so in the next 12 months.	What was the actual impact of this automation on workers?
Harvard Undergraduate Survey on Generative AI	326 US undergraduate students	Harvard Undergraduate Association	50% of students are concerned that AI will negatively impact their job prospects. Students are worried about economic inequality and extinction risk.	Have these concerns altered students' planning for the future?
Most workers think AI will affect their jobs. They disagree on how.	35,000 global private-sector workers	ADP	85% of workers believe AI will impact their job in the next two to three years. Workers who think AI will help them have more confidence in their skills and are more likely to say they have the skills necessary to advance their career.	Does the administrative data match concerns?

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

Justification: The abstract and introduction are a basic overview of the claims made in the paper.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations section points out which assumptions underly the claim made in the provocation.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.

- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.

- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.

- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: I have read and abided by the code.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: The paper does not discuss negative societal impacts of surveying because their current negative societal impacts are limited.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification:

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification:

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA] .

    Justification:

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA] .

    Justification:

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA] .

    Justification:

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.