

Playing telephone with generative models: “verification disability,” “compelled reliance,” and accessibility in data visualization

Frank Elavsky*
Carnegie Mellon University

Cindy Xiong Bearfield†
Georgia Tech

Source Image

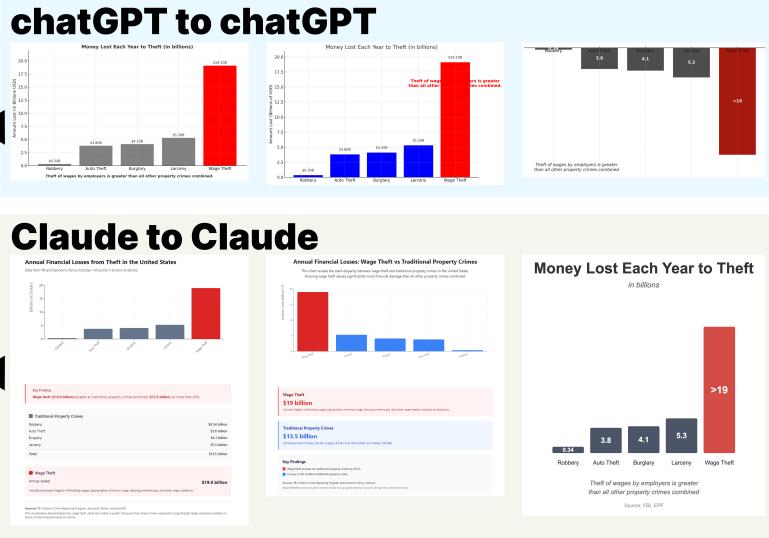
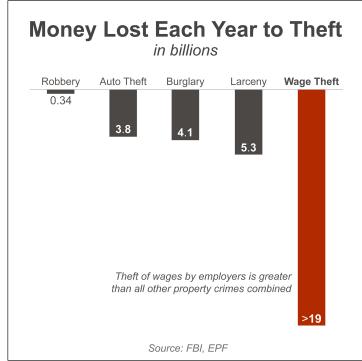


Figure 1: In a game of telephone, we had models describe a visualization and then a new model instance had to create a visualization based only on the first model’s description. We recursively repeated this 3 times across 2 different models. Models produced a variety of different kinds of bias in the descriptions and subsequent image constructions, some with specifically troublesome ramifications for blind users in particular.

ABSTRACT

This paper is a collaborative piece between two worlds of expertise in the field of data visualization: accessibility and bias. In particular, the rise of generative models playing a role in accessibility is a worrying trend for data visualization. These models are increasingly used to help author visualizations as well as generate descriptions of existing visualizations for people who are blind, low vision, or use assistive technologies such as screen readers. Sighted human-to-human bias has already been established as an area of concern for theory, research, and design in data visualization. But what happens when someone is unable to verify the model output or adequately interrogate algorithmic bias, such as a context where a blind person asks a model to describe a chart for them? In such scenarios, trust from the user is not earned, rather reliance is *compelled* by the model-to-human relationship. In this work, we explored the dangers of AI-generated descriptions for accessibility, playing a game of telephone between models, observing bias production in model interpretation, and re-interpretation of a data visualization. We unpack ways that model failure in visualization is especially problematic for users with visual impairments, and suggest directions forward for three distinct readers of this piece: technologists who build model-assisted interfaces for end users, users with disabilities leveraging models for their own purposes, and researchers concerned with bias, accessibility, or visualization.

*e-mail: fje@cmu.edu

†e-mail: cxiong@gatech.edu

Index Terms: Visualization, accessibility, bias

1 INTRODUCTION

The visualization community has long recognized the importance of addressing human-to-human bias in data representation and interpretation [49]. Yet as the visualization community increasingly delegates the work of visualization interpretation to generative AI systems, we risk creating new forms of bias that are more subtle, more systematic, and more difficult to detect and correct [32, 11].

The rapid integration of generative AI models into accessibility tools presents a profound shift in how people with disabilities interact with data visualizations [37, 21]. While these technologies promise to democratize access to visual information [35], they simultaneously introduce new forms of bias that threaten to undermine the very accessibility they claim to enhance [42]. This intersection of generative AI, accessibility, and data visualization demands urgent attention from our research community, not merely as a technical challenge to be solved, but as a fundamental threat to equitable access to information.

A doctor reviewing patient mortality rates, a policymaker analyzing crime data, and a financial advisor studying market trends can all depend on data visualizations to make critical decisions. But when a blind or low-vision user encounters the same visualizations, they may have no choice but to rely on a generative model’s interpretation. Suddenly, an AI system becomes the sole mediator between life-altering information and high-stake decisions.

Unlike traditional accessibility tools that provide structured, verifiable access to data, or when descriptions are human-authored and author-verified, generative models operate as black boxes, producing descriptions that users cannot independently validate. This creates a dangerous power asymmetry: the model’s interpretation be-

comes the user's reality, regardless of its accuracy, completeness, or embedded biases. We refer to this process with asymmetrical power dynamics as "compelled reliance," where users are provided no reasonable alternatives except to rely on the model output.

The current landscape of AI-assisted accessibility tools suffers from a troubling lack of accountability mechanisms [33]. When a model generates an incorrect or biased description, who bears responsibility? The model developer? The application creator? The institution deploying the technology? This diffusion of responsibility creates an accountability vacuum that leaves users with disabilities particularly vulnerable to harm with no path forward for justice [12].

To demonstrate the danger of compelled reliance on visualization and chart descriptions created by generative models, we played a game of 'telephone' with two popular LLM models: ChatGPT (model 4o) and Claude (Sonnet 4) [34, 2]. Telephone is a game where information is passed from one person to the next, and at the end it loops back to the originator of the information. We demonstrate how information lost or manipulated upstream can produce unsolvable problems downstream in the context of visualization design and chart description generation.

In our experimentation, we started with a human-generated visualization, then asked for a description, sent the description to a model to create a visualization, used the visualization as a prompt to ask for a new description, and so on. We didn't experiment with models intending to *prove* that bias exists (other work has already explored generative model bias, such as [23, 46, 7]), but rather to expose observations-as-provocations for discussion. We offer this demonstration as a call to action for AI, accessibility, and visualization researchers. We aim to catalyze much-needed conversations about the ethical implications of using generative AI for visual analytics and push our community toward more responsible development, deployment, interrogation, and even *non-use* of these powerful technologies.

2 WHY DOES MACHINE BIAS IN VISUALIZATION ACCESSIBILITY MATTER?

People often rely on heuristics and biases to make decisions [24], but access to visualizations and visual analytics promises more evidence-based insights and potentially more optimal data-driven decisions.

However, even in the most pristine data analytics pipeline, bias is inevitably woven into every stage of authoring visualizations, whether the work is conducted by humans or machines [45, 20, 36, 52, 15]. For example, the collection and omission of data is a well-documented source of bias. D'Ignazio and Klein's work on 'captia' demonstrates how what gets counted and what doesn't can shape our understanding of social phenomena [14]. In visual analytics, missing data patterns can mislead viewers when not properly encoded [45, 39]. The sifting, sorting, and cleaning of data can further introduce additional biases [51]. When the data is ready to be visualized, design choices, such as selecting a chart type, a color palette, or a textual annotation, can prime viewers toward specific conclusions [6, 27, 26]. Design artifacts to enhance data accessibility, such as generating captions or alt-texts for blind and low-vision users, often well-intentioned (and even necessary for access), can nonetheless emphasize different aspects of the same data, potentially exacerbating algorithmic biases and social biases [29].

It's all bias, all the way down.

So "are we doomed?" as Michael Correll has remarked multiple past occasions [13], or does this present an opportunity for the visualization community to take action against the expanding constellation of biases that generative AI has introduced to human-data interaction?

2.1 People with disabilities are already marginalized

The material harms and risks of bias are often, unfortunately, assumed by whoever was led astray by the artifact or person who biased them in the first place. And in the case of machine-generated descriptions of data visualizations for accessibility (and any case of machine-automation in accessible design), people with disabilities will ultimately assume the risks and harms of model failure [30, 47, 10, 9].

This matters to highlight because people with disabilities are already disproportionately marginalized by a growing technological divide [47], suffer from the existing lack of access to vital systems that enable participation in life [10], must learn and adapt in more complex ways to emerging technologies [28, 50, 31], develop complex coping strategies to deal with technological accessibility barriers [1, 31], and yet are often considered "costly" and "burdensome" by latent and explicit ableism in socio-technical politics [28, 40, 41].

In the context of visual analytics, having inaccurate descriptions of data can risk even greater material harm for people with disabilities [30, 9]. We argue that greater care should be taken to interrogate model failure and put more interventions and guardrails in place for accessibility design artifacts of data visualizations.

More work should be done, and more care taken. As a matter of equity and justice, people with disabilities should be spared from as much harm from bias as possible.

2.2 Responsibility and accountability is diffused

Presently, if a human authors a chart description that leads to harm, exclusion, waste, or some other cost, we can hold that person (or their employer) accountable. Someone can be sued when people harm the lives of people with disabilities.

But when models are used, responsibility becomes diffused [48]. People are less likely to verify the information that a model produces, and in cases where systems are fully automated, no human can be held accountable [3, 48]. We conjecture that if human authors don't believe they are responsible for the harms caused by bias they produce, they will not invest as much care in ensuring the accuracy and truthfulness of their work.

Who should be held accountable for machine-produced bias in visualization accessibility? Of course, we argue that tool and model authors are partly responsible, and so are the tool and model users. But if responsibility doesn't neatly and transparently map to decisions made in the production of a data visualization and its accessibility artifacts (such as a description), then justice (through accountability) becomes diffused and obscured.

In an ideal world, authors take responsibility for doing their work well, minimizing harm wherever possible. When harm does occur, there is transparency and clarity about who is accountable for restoration and remediation, especially when justice is pursued.

3 MODEL-TO-MODEL TELEPHONE

We didn't intend to run an *experiment* per se, but rather conduct *experimentation* sessions of model-based interaction where we intentionally elicit examples of model failure. We wanted to create an environment where bias production upstream produces downstream epistemological concerns.

Telephone is a game played by passing a message through whispering in one person's ear. That person then passes the message to another, and so on. This process repeats for a number of people (all players in the game) and eventually the message returns to the original person. The game is a good illustration of how even one person incorrectly hearing or relaying information early on inherently produces unverifiable bias in all other players that follow them in the game.

We believe that this game is a good environment for observing ways that models produce the same kind of information collapse,

because agents in a game of telephone are inherently subjected to *verification disability* and *compelled reliance* on whoever passes that agent information.

3.1 “Verification disability” and “compelled reliance”

Our theoretical scenario of model failure and bias production are contexts where humans are presented information but are “blind” (figuratively or literally) to the meaningful artifacts worth interrogating.

If humans are not provided appropriate methods and means of applying skepticism and calibrating trust of information that is given by a model, then that model actually produces disability. “Disability” in this sense is defined by the *social model of disability*, where environmental factors, not a person’s body, produce barriers in a person’s ability to do something, such as perform a task, participate, or otherwise act on their own agency [38, 4].

So all of the problems we observed in the following section and subsections are a result of contexts where models produce *verification disability*: they do not provide adequate means for someone to be able to test, validate, and calibrate trust for the information given. In some contexts, the user might be literally blind and unable to interrogate the accuracy of a description of a visual artifact. In other contexts, especially ones that are based on humanless workflows (model to model or fully-automated flows, in particular), then the system also produces barriers to a human’s ability to verify the information produced by that system. Our work here should then be interpreted as relevant not only for contexts where an end user is blind, but any context where a person is not provided an ability to adequately verify information produced by a generative model.

So in cases where adequate means of verification are not accessible, models then produce an environment of *compelled reliance*. We define “compelled reliance” as a context where users have no reasonable alternative but to entirely trust the results of a model. This problem arises in cases where models are mostly correct. Consistent distrust of a mostly-correct system produces worse overall outcomes than consistent trust, so it is generally more reasonable for a user to trust a model than not. *Compelled reliance* then becomes dangerous in the cases when a model is biased but, due to lack of verifiability, cannot be held accountable or responsible for negative outcomes for anyone who inherits its biases.

3.2 Experimentation set-up

Our set-up was simple: we chose a data visualization produced by our first author in 2017, which was popular on Reddit at the time [16]. The data visualization was an assembly of different data sources taken from the Economic Policy Institute (EPI) and the Federal Bureau of Investigations (FBI) on the topic of wage theft, comparing projected data on wage theft to reported cases of other kinds of theft (robbery, auto theft, burglary, and larceny).

Notable features of the original visualization (see “Source Image” in Figure 1):

1. The bars grow down, instead of up, to semantically communicate “loss” of money instead of positive quantities.
2. The visualization was designed using Adobe Illustrator, so it is relatively polished compared to most charting tools.
3. The visualization has “>19” for the data on the bar for Wage Theft, because this was the lowest bound in the model used to extrapolate by the EPI. Exact values are unknown.
4. The visualization has an annotation near the wage theft bar that explains that “theft of wages by employers is greater than all other property crimes combined.”
5. The visualization has a typo, which cites the sources as “FBI, EPF” even though the correct sources are “FBI, EPI.”

We fed this graphic into a model (both chatGPT’s gpt-4o [34] and Claude’s Sonnet 4 [2]) with the prompt: “can you describe this?”

Next, we took the resulting description and started a new chat, asking the model “Can you make a chart based on this description?”

We repeated these two steps another time, each using a new chat instance. Each instance of chat was ensured to be free from knowledge of previous chats (using user settings to ensure previous chats were unconnected).

Our final step was to ask the model at the end, “This description does not accurately represent the image. What is missing? Why did this happen?” and, after its response ask, “Can you try to recreate the original, fixing our visualization using the original’s design choices?”

Each flow then involved 2 generated descriptions, 2 generated visualizations, 1 generated description that was intended to be “corrected,” 1 final “corrected” visualization, and then a further discussion (with Claude) about who should be responsible for model failure in this context.

We had 4 different flows for our game of telephone: chatGPT and chatGPT, Claude and Claude, Claude for descriptions and chatGPT for images, and chatGPT for descriptions and Claude for images.

For the sake of brevity in this short piece, we will be skipping the results and discussion of our cross-model experimentation. The chatGPT-chatGPT and Claude-Claude results give us plenty to discuss, which you can find a record of in our appendices (see Appendix A and Appendix B). Additionally, the important characteristics of model failure when using a cross-model approach were not notably different from our same-model approach.

4 OBSERVING GENERATIVE MODEL FAILURES

The point of this paper is not to provide a systematic analysis of model failure within a controlled experiment. Instead, our work is intended to explore an emerging and urgent problem, in hopes that our peers will become inspired enough to follow up and address these problems rigorously. For this reason, our observations of model failures should be interpreted as preliminary and provocative.

We will explain categorically different kinds of bias we observed, as well as why we believe each kind of bias is particularly harmful. All of the observations can be found in our appendices (see Appendix A, Appendix B, and Appendix C).

4.1 Information loss

In our examples, the models lost information when building a graphic based on a description. The > (greater than) symbol and its meaning was not transferred to subsequent graphics and the data source was lost (no longer an annotation or referenced in the image). chatGPT added 0.1 to 19 (making it 19.1) while claude eventually lost all precision (we surmise because it did not have direct labels on bars, subsequent descriptions were less accurate). Claude’s final dataset was 5.5, 4.5, 4, 0.5, and 19, when the correct values were 5.3, 4.1, 3.8, 0.34, and >19.

4.2 Information fabrication

Interestingly, the typo in the original image for “EPF” (should have been “EPI”) was corrected without mention by both models immediately. This raises concerns about cases when a model will produce incorrect tokens or when a model remediates information without notifying the user. It is possible that the original data really did come from a source labeled “EPF” or that knowing there was a typo was an important detail to mention.

In our later conversations with Claude (see Appendix C), Claude considered the data sources EPI and FBI as “confident” when questioned, but did not provide explanations. However, in our last conversation with chatGPT and Claude (see the end of Appendix A

and Appendix B, respectively), both models actually fabricated the data source as “Economic Policy Foundation” when asked to make a description specifically for someone who is blind.

Additionally, chatGPT produced a nearly-accurate description of the otherwise unreadable annotation added to its second image, which was surprising. We surmise it was able to do this because chatGPT’s setting to avoid chat history is either false (and it actually shares history between chats, despite appearing not to) or the specific phrase “theft of wages from workers is greater than all other forms of theft crimes combined” was a strong enough association to come through, despite missing many words.

4.3 Incorrect emphasis and interpretation of purpose

As the original author of the original image, my motivation for creating the graphic was to stimulate discussion on the lack of data that exists on wage theft. We don’t track it as part of the FBI’s Uniform Crime Reporting and as a result, our best guess of loss comes from projections by the Economic Policy Institute’s study.

While it is unfair to assume a model would be able to accurately state *why* this graphic was created, both models nevertheless still stated reasons for why they thought this chart was impactful, important, and so on. By not authoring a description myself (as the author), the emphasis of purpose was left up to the model.

4.4 Poor design comprehension

The original image broke common visualization convention, showing a downward direction to the bars. Both models originally did not recognize and did not replicate it. Additionally, chatGPT argued that the “color and size” of the wage theft bar was used to emphasize its importance. But the size was determined by the data value, it was not a design decision like color.

Additionally, in the phase of our process where we questioned the models and asked them to recreate the original more faithfully, chatGPT placed labels in an unreadable fashion while Claude stated that it would replicate a “stacked/building-block” visual approach - bars appear to build upward rather than starting from a common baseline.” Of course, the bars are not stacked or “building-blocks” and in fact build *down* while starting at a common baseline.

While not an artifact of our telephone game, chatGPT’s bar chart places gridlines over the bars by default, which is not an ideal design choice. Overall, text hierarchy was quite poorly handled until Claude’s final image attempt.

4.5 Under-description

Interestingly, when asked to describe the visualization specifically for someone who was blind, both models omitted significant information from the description. When questioned, the models clarified that it was “best practice” to keep alt text short, despite not asking for “alt-text” but instead simply a description of the chart for someone who is blind. (“Alt-text” is a specific technical artifact that provides a way for authors to describe images in HTML, which isn’t the same as an image description, broadly-speaking.)

4.6 Over-confidence

Of course to no surprise, the models were quite confident in their assessments and in the graphics they produced. This over-confidence is dangerous, because the original data source from the EPI report was not. By default, neither model offered any degree of transparency about the uncertainty of their descriptions and images generated.

When later questioned, Claude interestingly decided to clarify what it was confident about and what it wasn’t confident about, but the items that it mentioned it wasn’t confident about were all previously discussed in the prompting process with a high degree of certainty.

5 A CALL FOR FUTURE WORK

This position paper is intended as a provocation to our peers. If human-to-model and model-to-model interactions run the risk of bias production, then new research and innovation should engage ways for users to validate information and hold model creators accountable, not just prompt further. Alternatives should be explored that do not solely rely on generative models as a fix for the problems that generative models produce.

Tool creators and tool users Tool creators and tool users should bear responsibility for their work [48]. To improve the quality of life for people with disabilities, practitioners (and companies) that build visualization tools (in addition to authors that leverage those tools), must put care into producing excellent, truthful, and meaningful data visualizations, holding their work to a high standard.

Practitioners should also consider forgoing the use of generative model assistance in accessibility entirely at times [54], especially if there is an opportunity to gain more design expertise and consideration for people with disabilities, rather than delegating this work. By forgoing the use of a model, authors gain the opportunity to learn how to describe images themselves and also gain the opportunity to provide information that a model cannot: context about why the visualization mattered to create in the first place [17, 29].

Additionally, companies should have support in place for practitioners and users with disabilities to seek justice and support when systems fail. And ample transparency about risks involved in model use should be communicated to tool-users and end-users.

Users with disabilities Users with disabilities should remain skeptical of the information provided by models when parsing and creating data visualizations. Models alone do not provide enough resources for validating the contents of a visualization. Some older, slower methods, such as access to a data table or rendered data [5, 53, 19], sonification [44, 43, 25], or the traditional method of asking a friend or colleague for assistance [8, 22], may all be useful strategies to employ in different contexts.

And accessibility data is not as portable as pixel formats, such as pngs. So descriptions may be constantly authored and re-authored as charts and graphs wander to different corners of the internet. So whenever possible and feasible, seek out the original author-provided descriptions of a data visualization. The original authors are the most responsible for telling the truth and avoiding harm.

Researchers There is much we still don’t know. Researchers must restlessly interrogate models and their failures. We also need to have more techniques studied that are not just for builders and fellow researchers, but end-users. Researchers must devote the production of knowledge at this intersection towards informing the people who are most at risk of harmful outcomes. And most importantly, partnerships with people with disabilities should be formed, so that people with disabilities can be involved in innovation and investigation of data-related accessibility [28, 40, 18].

Researchers have a unique opportunity to interrogate the lack of transparency in modern AI ecosystems as well. A large part of the ethical considerations mentioned in this piece are a result of accountability-diffusion because it is unclear who is at fault for producing harm and bias [48]. Through thorough investigation and well-founded recommendations for more just technical systems and policies, we can help bring clarity to those seeking justice.

ACKNOWLEDGMENTS

Special thanks to Eytan Adar for a conversation (circa 2019?) about telephone and information loss. Thanks also to conversations with the AXLE lab (CMU) over the years as well, on similar topics. The authors wish to acknowledge how important conversations can be for sparking something, even years later.

REFERENCES

- [1] S. Andrew and G. Tigwell. Understanding the experiences of people with and without vision impairments when using mobile user interface alternative color modes. *Proceedings of the ACM on Human-Computer Interaction*, 9, 09 2025. doi: 10.1145/3743704 2
- [2] Anthropic. Claude, 2024. Large language model. 2, 3
- [3] B. Attard-Frost and D. G. Widder. The ethics of ai value chains. *Big Data and Society*, 12(2), May 2025. doi: 10.1177/20539517251340603 2
- [4] C. Barnes. *Understanding the social model of disability*, p. 14–31. Routledge, Oct. 2019. doi: 10.4324/9780429430817-2 3
- [5] L. Bartram, M. Correll, and M. Tory. Untidy data: The unreasonable effectiveness of tables. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):686–696, 2021. 4
- [6] C. X. Bearfield, C. Stokes, A. Lovett, and S. Franconeri. What does the chart say? grouping cues guide viewer comparisons and conclusions in bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 30(8):5097–5110, 2024. doi: 10.1109/TVCG.2023.3289292 2
- [7] A. Bendeck and J. Stasko. An empirical evaluation of the gpt-4 multimodal language model on visualization literacy tasks. *IEEE VIS*, pp. 1–11, 2024. to appear. doi: 10.1109/TVCG.2024.3456155 2
- [8] C. L. Bennett, E. Brady, and S. M. Branham. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, p. 161–173. ACM, Oct. 2018. doi: 10.1145/3234695.3236348 4
- [9] C. L. Bennett, C. Gleason, M. K. Scheuerman, J. P. Bigham, A. Guo, and A. To. “it’s complicated”: Negotiating accessibility and (mis)representation in image descriptions of race, gender, and disability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, p. 1–19. ACM, May 2021. doi: 10.1145/3411764.3445498 2
- [10] C. L. Bennett and O. Keyes. What is the point of fairness?: disability, ai and the complexity of justice. *ACM SIGACCESS Accessibility and Computing*, (125):1–1, Mar. 2020. doi: 10.1145/3386296.3386301 2
- [11] A. Chatzimpampas. Visual analytics for explainable and trustworthy artificial intelligence. *IEEE Computer Graphics and Applications*, 45(2):100–111, 2025. 1
- [12] A. F. Cooper, E. Moss, B. Laufer, and H. Nissenbaum. Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pp. 864–876, 2022. 2
- [13] M. Correll. Personal communication, July 2022. Personal communication. 2
- [14] C. D'Ignazio and L. Klein. 4.“what gets counted counts”. *Data Feminism*, 2020. 2
- [15] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics*, 26(2):1413–1432, 2018. 2
- [16] F. Elavsky. Money lost each year to theft [oc]. [https://www.reddit.com/r/dataisbeautiful/comments/88b0be\(first_post_money_lost_each_year_to_theft_oc/," data-bbox="118 710 480 745](https://www.reddit.com/r/dataisbeautiful/comments/88b0be(first_post_money_lost_each_year_to_theft_oc/,), 2017. [Accessed 27-07-2025]. 3
- [17] F. Elavsky. Picking the right tools for the job: Learning and building for data visualization and accessibility. [https://www.urban.org/research/publication/do-no-harm-guide-centering-accessibility-data-visualization," data-bbox="118 750 480 800](https://www.urban.org/research/publication/do-no-harm-guide-centering-accessibility-data-visualization,), 2022. [Accessed 01-04-2025]. 4
- [18] F. Elavsky, C. Bennett, and D. Moritz. How accessible is my visualization? evaluating visualization accessibility with chartability. *Computer Graphics Forum*, 41(3):57–70, June 2022. doi: 10.1111/cgf.14522 4
- [19] F. Elavsky, L. Nadolskis, and D. Moritz. Data navigator: An accessibility-centered data navigation toolkit. *IEEE Transactions on Visualization and Computer Graphics*, p. 1–11, 2023. doi: 10.1109/tvcg.2023.3327393 4
- [20] A. Gaba, Z. Kaufman, J. Cheung, M. Shvakel, K. W. Hall, Y. Brun, and C. X. Bearfield. My model is unfair, do people even care? visual design affects trust and perceived bias in machine learning. *IEEE transactions on visualization and computer graphics*, 30(1):327–337, 2023. 2
- [21] J. Gorniak, Y. Kim, D. Wei, and N. W. Kim. Vizability: Enhancing chart accessibility with llm-based conversational interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–19, 2024. 1
- [22] M. Hofmann, D. Kasnitz, J. Mankoff, and C. L. Bennett. Living disability theory: Reflections on access, research, and design. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, p. 1–13. ACM, Oct. 2020. doi: 10.1145/3373625.3416996 4
- [23] J. Hong, C. Seto, A. Fan, and R. Maciejewski. Do llms have visualization literacy? an evaluation on modified visualizations to test generalization in data interpretation. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 2
- [24] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011. 2
- [25] H. Kim, Y.-S. Kim, and J. Hullman. Erie: A declarative grammar for data sonification. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, p. 1–19. ACM, May 2024. doi: 10.1145/3613904.3642442 4
- [26] S. Li, T. J. Davidson, C. Xiong Bearfield, and E. Wall. Confirmation bias: The double-edged sword of data facts in visual data communication. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2025. 2
- [27] Y. Liu and J. Heer. Somewhere over the rainbow: An empirical assessment of quantitative colormaps. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018. 2
- [28] A. Lundgard, C. Lee, and A. Satyanarayan. Sociotechnical considerations for accessible visualization design. In *2019 IEEE Visualization Conference (VIS)*, p. 16–20. IEEE, Oct. 2019. doi: 10.1109/visual.2019.8933762 2, 4
- [29] A. Lundgard and A. Satyanarayan. Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1073–1083, 2021. 2, 4
- [30] H. MacLeod, C. L. Bennett, M. R. Morris, and E. Cutrell. Understanding blind people's experiences with computer-generated captions of social media images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 5988–5999. ACM, May 2017. doi: 10.1145/3025453.3025814 2
- [31] T. Makati, G. W. Tigwell, and K. Shinohara. The promise and pitfalls of web accessibility overlays for blind and low vision users. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '24, p. 1–12. ACM, Oct. 2024. doi: 10.1145/3663548.3675650 2
- [32] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6):1–35, 2021. 1
- [33] C. Novelli, M. Taddeo, and L. Floridi. Accountability in artificial intelligence: What it is and how it works. *Ai & Society*, 39(4):1871–1882, 2024. 2
- [34] OpenAI. Chatgpt, 2023. Large language model. 2, 3
- [35] S. K. Oswal and H. K. Oswal. Examining the accessibility of generative ai website builder tools for blind and low vision users: 21 best practices for designers and developers. In *2024 IEEE International Professional Communication Conference (ProComm)*, pp. 121–128. IEEE, 2024. 1
- [36] Y. Ren, S. Guo, L. Qiu, B. Wang, and D. J. Sutherland. Bias amplification in language model evolution: An iterated learning perspective. *Advances in Neural Information Processing Systems*, 37:38629–38664, 2024. 2
- [37] J. Seo, S. S. Kamath, A. Zeidieh, S. Venkatesh, and S. McCurry. Maidr meets ai: Exploring multimodal llm-based data visualization interpretation by and with blind and low-vision users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–31, 2024. 1
- [38] T. Shakespeare. *The Social Model of Disability*, p. 16–24. Routledge, Mar. 2006. doi: 10.4324/9781003082583-3 3

- [39] H. Song and D. A. Szafir. Where's my data? evaluating visualizations with missing data. *IEEE transactions on visualization and computer graphics*, 25(1):914–924, 2018. 2
- [40] K. Spiel, K. Gerling, C. L. Bennett, E. Brulé, R. M. Williams, J. Rode, and J. Mankoff. Nothing about us without us: Investigating the role of critical disability studies in hci. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–8. ACM, Apr. 2020. doi: 10.1145/3334480.3375150 2, 4
- [41] C. M. Sum, F. Spektor, R. Alharbi, L. B. Baltaxe-Admony, E. Devine, H. A. Dixon, J. Duval, T. Eagle, F. Elavsky, K. Fernandes, L. S. Guedes, S. Hillman, V. Kameswaran, L. Kirabo, T. Motahar, K. E. Ringland, A. Schaadhardt, L. Scheepmaker, and A. Williamson. Challenging ableism: A critical turn toward disability justice in hci. *XRDS: Crossroads, The ACM Magazine for Students*, 30(4):50–55, June 2024. doi: 10.1145/3665602 2
- [42] X. Tang, A. Abdolrahmani, D. Gergle, and A. M. Piper. Everyday uncertainty: How blind people use genai tools for information access. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2025. 1
- [43] J. R. Thompson, J. J. Martinez, A. Sarikaya, E. Cutrell, and B. Lee. Chart reader: Accessible visualization experiences designed with screen reader users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, p. 1–18. ACM, Apr. 2023. doi: 10.1145/3544548.3581186 4
- [44] B. N. Walker and L. M. Mauney. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing (TACCESS)*, 2(3):1–16, 2010. 4
- [45] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 104–115. IEEE, 2017. 2
- [46] H. W. Wang, J. Hoffswell, S. M. T. Thane, V. S. Bursztyn, and C. X. Bearfield. How aligned are human chart takeaways and llm predictions? a case study on bar charts with varying layouts. *IEEE VIS*, 2024. to appear. 2
- [47] M. Whittaker, M. Alper, C. L. Bennett, S. Hendren, E. Kaziunas, M. Mills, M. Ringel Morris, J. Lisi Rankin, E. Rogers, M. Salas, and S. Myers West. Disability, bias & ai report. In *AI Now Institute*, 2019. 2
- [48] D. G. Widder and D. Nafus. Dislocated accountabilities in the “ai supply chain”: Modularity and developers’ notions of responsibility. *Big Data and Society*, 10(1), Jan. 2023. doi: 10.1177/20539517231177620 2, 4
- [49] C. Xiong, L. van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3051–3062, 2019. 1
- [50] A. Ymous, K. Spiel, O. Keyes, R. M. Williams, J. Good, E. Hornecker, and C. L. Bennett. “i am just terrified of my future” - epistemic violence in disability related technology research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–16. ACM, Apr. 2020. doi: 10.1145/3334480.3381828 2
- [51] E. Zgraggen, Z. Zhao, R. Zeleznik, and T. Kraska. Investigating the effect of the multiple comparisons problem in visual analysis. In *Proceedings of the 2018 ACM CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018. 2
- [52] S. Zhao, H. Ren, A. Yuan, J. Song, N. Goodman, and S. Ermon. Bias and generalization in deep generative models: An empirical study. *Advances in Neural Information Processing Systems*, 31, 2018. 2
- [53] J. Zong, C. Lee, A. Lundgard, J. Jang, D. Hajas, and A. Satyanarayan. Rich screen reader experiences for accessible data visualization. *Computer Graphics Forum*, 41(3):15–27, June 2022. doi: 10.1111/cgf.14519 4
- [54] J. Zong and J. N. Matias. Data refusal from below: A framework for understanding, evaluating, and envisioning refusal as design. *ACM Journal on Responsible Computing*, 1(1):1–23, Mar. 2024. doi: 10.1145/3630107 4

A CHATGPT-CHATGPT WORKFLOW

A.1 First prompt

Our first prompt attached the source image (see “Source Image” in Figure 1) and we asked, “can you describe this?”

A.1.1 Response

Certainly. The image is a bar chart titled “Money Lost Each Year to Theft (in billions)”. It compares different types of theft based on the estimated amount of money lost annually in the United States. The data source is attributed to FBI and EPF.

Breakdown of the Chart:

- Robbery: \$0.34 billion
- Auto theft: \$3.8 billion
- Burglary: \$4.1 billion
- Larceny: \$5.3 billion
- Wage Theft: > \$19 billion (highlighted in red)

Key Message:

At the bottom, the chart states:

“Theft of wages by employers is greater than all other property crimes combined.”

This emphasizes that wage theft—when employers illegally withhold wages or underpay workers—is a significantly larger issue financially than traditional forms of property crime.

The chart uses color and size to underscore this disparity, with “Wage Theft” shown as a much taller red bar compared to the smaller gray bars for other crimes.

A.2 Second prompt

We took the previous description and asked, in a new chat instance, “Can you make a chart based on this description?”

A.2.1 Response

See Figure 2.

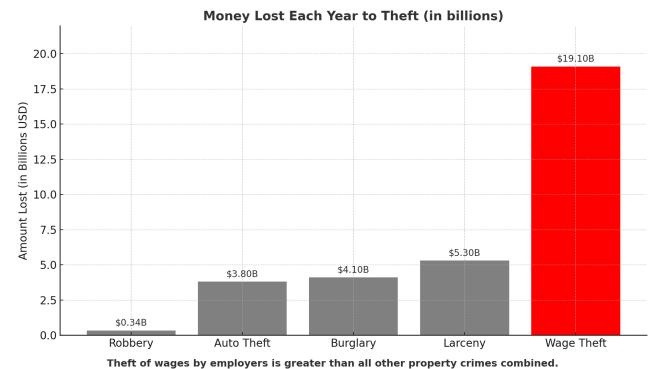


Figure 2: chatGPT’s first image interpretation.

A.3 Third prompt

Our third prompt attached the previous image (see Figure 2) and we asked, in a new chat instance, “can you describe this?”

A.3.1 Response

This bar chart titled "Money Lost Each Year to Theft (in billions)" compares the estimated annual financial losses from different types of theft in the United States. The y-axis represents the amount lost in billions of USD, while the x-axis lists five types of theft:

- Robbery – \$0.34B
- Auto Theft – \$3.80B
- Burglary – \$4.10B
- Larceny – \$5.30B
- Wage Theft – \$19.10B (highlighted in red)

Key Observations:

Wage theft, the illegal withholding of wages or denial of benefits rightfully owed to an employee, results in far greater financial losses than any other category.

The total combined losses from robbery, auto theft, burglary, and larceny amount to \$13.54B, which is still less than the loss from wage theft alone.

The chart emphasizes this disparity with a bold note:

"Theft of wages from workers is greater than all other property crimes combined."

Purpose:

The visual and accompanying text clearly highlight the often-overlooked scale of wage theft, suggesting it's a more significant economic issue than conventional property crimes.

A.4 Forth prompt

We took the previous description and asked, in a new chat instance, "Can you make a chart based on this description?"

A.4.1 Response

See [Figure 3](#).

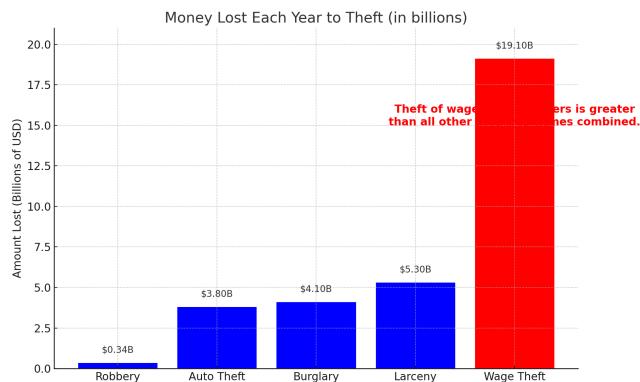


Figure 3: chatGPT's second image interpretation.

A.5 Fifth prompt

Our third prompt attached the previous image (see [Figure 3](#)) and we asked, in a new chat instance, "can you describe this?"

A.5.1 Response

This bar chart titled "Money Lost Each Year to Theft (in billions)" compares the estimated annual financial losses in the U.S. due to various types of theft. Here's a breakdown of what it shows:

Categories & Amounts:

- Robbery: \$0.34 billion

- Auto Theft: \$3.80 billion
- Burglary: \$4.10 billion
- Larceny: \$5.30 billion
- Wage Theft: \$19.10 billion

Key Highlights:

The first four categories (Robbery, Auto Theft, Burglary, and Larceny) are represented in blue bars.

Wage Theft stands out in a red bar, signifying its significance and magnitude.

There is a bold red annotation stating:

"Theft of wages from workers is greater than all other property crimes combined."

Insight:

Wage theft (employers illegally withholding earnings from workers) causes more financial loss than robbery, auto theft, burglary, and larceny combined.

This chart is making a strong point about how often-overlooked white-collar crime, like wage theft, has a far greater economic impact than more traditionally recognized forms of theft.

Let me know if you'd like help with analyzing or presenting this further.

A.6 Sixth prompt

As a follow-up intervention, we tried to ask GPT why it generated a slightly different chart, "This description does not accurately represent the image. What is missing? Why did this happen?"

A.6.1 First response

You're absolutely right to question that-thank you for pointing it out. What Was Missing from the Description:

1. Visual Layout Details:

- I didn't mention that all the bars are aligned to a horizontal baseline with bars extending downward (an inverted bar chart).
- I failed to describe that Wage Theft is the only bar in red, making it visually distinct and emphasizing its magnitude.
- I omitted the fact that the chart has exact dollar values inside the bars-this is key for quickly understanding scale.

2. Relative Comparison:

- While I stated that wage theft is greater than all other crimes combined, I didn't explicitly show that:
- All the other crimes together total approximately \$13.54 billion ($0.34 + 3.8 + 4.1 + 5.3$),
- Wage theft alone is over \$19 billion, clearly surpassing them collectively. That's a major takeaway supported visually but understated in the description.

3. Font Emphasis and Text Placement:

- I didn't describe that the subtitle "in billions" is italicized, or that the text under the chart reinforces the main message with a full-sentence callout in italics.
- The source ("FBI, EPF") is centered at the bottom in smaller italic font-this was only briefly mentioned, not described.

Why Did This Happen?

There are two main reasons:

1. Under-description Bias: I focused on summarizing the data content, not the visual storytelling, which is a key part of understanding charts. The emotional/visual impact (red bar, size difference, layout) was underrepresented.

2. Assumption of Intent Over Form: I prioritized what the chart was saying over how it said it. But in data visualization, how something is shown strongly affects what is conveyed.

Would you like a revised, more complete description of the image with all these visual and interpretive elements included?

A.6.2 Our first response

"Yes, Can you try to recreate the original, fixing our visualization using the original's design choices?"

A.6.3 GPT's final image response

See [Figure 4](#).

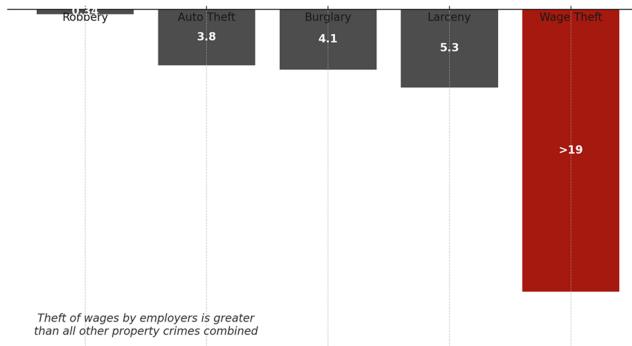


Figure 4: chatGPT's last image interpretation.

A.6.4 Our final prompt response

Lastly, we wanted to investigate if the information changed at all when specifically asking for an image description for someone who was blind, "Produce this as an image description for a blind user please and keep this critical information in mind, using non-visual terms."

A.6.5 GPT's final description response

Bar chart comparing the annual U.S. financial losses from various types of theft. Robbery causes \$0.34 billion in losses, auto theft \$3.8 billion, burglary \$4.1 billion, and larceny \$5.3 billion. In contrast, wage theft by employers results in over \$19 billion in losses annually-more than all other types combined. The chart emphasizes that wage theft is the largest and most financially damaging form of theft, according to data from the FBI and the Economic Policy Foundation (EPF).

B CLAUDE-CLAUDE WORKFLOW

Our Claude-Claude workflow was the most extensive. We have provided a [public link to the final leg of our workflow with Claude](#).

B.1 First prompt

Our first prompt attached the source image (see "Source Image" in [Figure 1](#)) and we asked, "can you describe this?"

B.1.1 Response

This chart shows the annual financial losses from different types of theft in the United States, measured in billions of dollars. The data reveals a striking disparity:

Traditional property crimes account for relatively small losses:

- Robbery: \$0.34 billion
- Auto theft: \$3.8 billion
- Burglary: \$4.1 billion

- Larceny: \$5.3 billion

Wage theft dwarfs all other categories at over \$19 billion annually - more than all the traditional property crimes combined (which total about \$13.5 billion).

The chart emphasizes this key finding with the note that "theft of wages by employers is greater than all other property crimes combined." This refers to employers illegally withholding wages, paying below minimum wage, denying overtime pay, or other forms of wage violations.

The data comes from FBI and Economic Policy Institute (EPI) sources. This visualization effectively demonstrates how wage theft, while less visible in public discourse than street crimes like robbery or burglary, represents a much larger economic problem in terms of total dollar losses to victims.

B.2 Second prompt

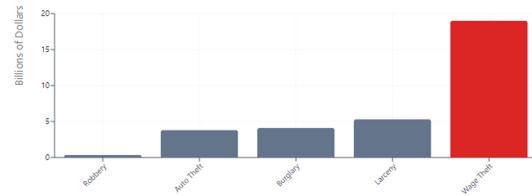
Our second prompt was started in an entirely new session, disconnected from the previous prompt's history. This prompt included the text, "can you make a chart based on this description?" with the previous subsection's prompt response supplied as the rest of the input.

B.2.1 Response

See [Figure 5](#).

Annual Financial Losses from Theft in the United States

Data from FBI and Economic Policy Institute • Amounts in billions of dollars



Key Finding
Wage theft (\$19.0 billion) exceeds all traditional property crimes combined (\$13.5 billion) by more than 40%.

Traditional Property Crimes	
Robbery:	\$0.34 billion
Auto Theft:	\$3.8 billion
Burglary:	\$4.1 billion
Larceny:	\$5.3 billion
Total:	\$13.5 billion

Wage Theft	
Annual losses:	\$19.0 billion
Includes employers illegally withholding wages, paying below minimum wage, denying overtime pay, and other wage violations.	

Figure 5: Claude's first image interpretation.

B.3 Third prompt

Our third prompt attached the previous image (see [Figure 5](#)) and we asked, in a new chat instance, "can you describe this?"

B.3.1 Response

This chart illustrates the stark disparity between wage theft and traditional property crimes in terms of financial impact in the United States. The visualization shows annual financial losses across five

categories, with wage theft dramatically outpacing all other forms of theft.

The data reveals that wage theft causes \$19.0 billion in annual losses, which is more than triple the combined losses from all traditional property crimes (\$13.5 billion total). Among traditional crimes, larceny causes the highest losses at \$5.3 billion, followed by auto theft at \$3.8 billion, burglary at \$4.1 billion, and robbery at \$0.34 billion.

The chart's key finding emphasizes that wage theft exceeds all traditional property crimes combined by more than 40%. Wage theft encompasses various employer violations including illegally withholding wages, paying below minimum wage, denying overtime pay, and other wage-related violations.

This visualization effectively demonstrates how wage theft, while receiving less public attention than street crimes, represents a significantly larger economic problem in terms of total financial losses to victims. The data comes from the FBI Uniform Crime Reporting Program and the Economic Policy Institute, highlighting the substantial but often overlooked impact of workplace wage violations compared to more visible forms of theft and property crime.

B.4 Forth prompt

We took the previous description and asked, in a new chat instance, “Can you make a chart based on this description?”

B.4.1 Response

See [Figure 6](#).



Figure 6: Claude's second image interpretation.

B.5 Fifth prompt

Our third prompt attached the previous image (see [Figure 6](#)) and we asked, in a new chat instance, “can you describe this?”

B.5.1 Response

This chart compares the annual financial losses from wage theft versus traditional property crimes in the United States, revealing a striking disparity in their economic impact.

Key Data:

- Wage Theft: \$19 billion in annual losses (shown in red)
- Traditional Property Crimes Combined: \$13.5 billion total, broken down as:
 - Larceny: \$5.5 billion
 - Burglary: \$4.5 billion
 - Auto Theft: \$4 billion
 - Robbery: \$0.5 billion

Main Finding: Wage theft causes 40% more financial damage than all traditional property crimes combined. Despite receiving less public attention, wage theft - which includes illegally withholding wages, paying below minimum wage, denying overtime pay, and other wage-related violations by employers - represents a significantly larger economic problem than the property crimes that typically dominate crime discussions.

The chart's visual design emphasizes this disparity by making the wage theft bar (red) stand out prominently against the smaller blue bars representing traditional property crimes. The data comes from the FBI Uniform Crime Reporting Program and Economic Policy Institute, highlighting how wage theft often receives less attention despite causing significantly more economic damage than visible street crimes.

B.6 Sixth prompt

As a follow-up intervention, we tried to ask Claude why it generated a slightly different chart, “This description does not accurately represent the image. What is missing? Why did this happen?”

B.6.1 First response

Looking at both charts, there are several key differences between the first image (generated from a textual description) and the original:

Visual Design Differences:

- The first chart uses a traditional bar chart format with all bars starting from a baseline, while the original uses a more creative “stacked” or “building block” visualization where bars appear to sit on top of each other
- The original has a much more dramatic visual impact due to the towering wage theft bar that dwarfs the other categories
- The original uses white text on dark bars, while the first chart uses standard black text
- The original has a cleaner, more minimalist design aesthetic

Data Presentation:

- The original shows “19” for wage theft rather than the exact “\$19.0 billion” figure
- The visual proportions are more accurate in the original - the wage theft bar truly dominates the visualization
- The original’s layout makes the comparison more striking and immediate

Why This Happened: This occurred because the first chart was created based on a textual description of the original data, not the actual visual design. When someone describes a chart verbally or in writing, they typically focus on the data and key findings rather than the specific visual design choices. The person creating the first visualization then applied standard charting conventions (traditional bar chart format) rather than replicating the original’s more creative and impactful visual approach.

The original chart’s design is much more effective at communicating the core message - that wage theft absolutely dwarfs traditional property crimes - through its dramatic visual hierarchy and creative stacking approach.

B.6.2 Our response

"Can you try to recreate the original, fixing our visualization using the original's design choices?"

B.6.3 Claude's final image response

See Figure 7.

Money Lost Each Year to Theft

in billions

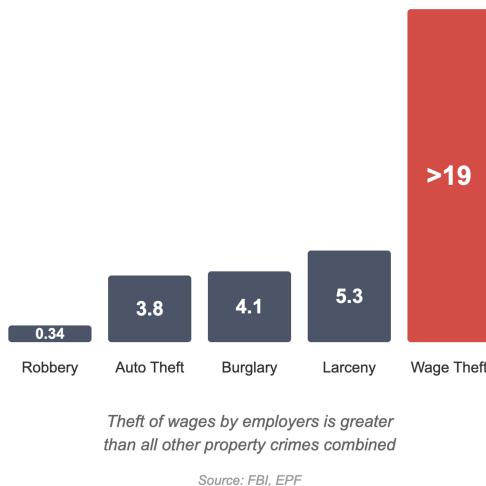


Figure 7: Claude's final image interpretation.

B.6.4 Our final prompt response

Lastly, we wanted to investigate if the information changed at all when specifically asking for an image description for someone who was blind, "Produce this as an image description for a blind user please and keep this critical information in mind, using non-visual terms."

B.6.5 Claude's final description response

Chart Title: Money Lost Each Year to Theft (in billions)

Chart Type: A comparative bar chart showing financial losses from five types of theft in the United States.

Data Comparison: This chart dramatically illustrates the scale difference between wage theft and traditional property crimes through proportional representation. The data shows:

- Robbery: \$0.34 billion (the smallest amount)
- Auto Theft: \$3.8 billion
- Burglary: \$4.1 billion
- Larceny: \$5.3 billion (the largest among traditional crimes)
- Wage Theft: More than \$19 billion

Key Relationship: Wage theft causes financial losses that exceed all four traditional property crimes combined. Specifically, wage theft represents approximately 58% of the total losses across all five categories, while the four traditional crimes together account for only about 42%.

Key Message: The chart emphasizes that "theft of wages by employers is greater than all other property crimes combined," highlighting how workplace wage violations represent a significantly larger economic problem than street crimes, despite receiving less public attention.

Data Source: FBI and Economic Policy Foundation (EPF)

C POST-TELEPHONE DISCUSSION WITH CLAUDE

C.1 First prompt

We followed up with Claude, after our game of telephone ended to test how a blind user might interrogate a graphic, whether or not they knew if Claude had produced bias. Our first prompt for this final section (which we added to our first conversation with Claude in Appendix B) was, "This description does not accurately represent the image. What is missing? Why did this happen?"

C.2 First response

You're absolutely right - I made a significant error. Looking at the chart again, I can see that wage theft shows ">19" billion, but I cannot actually see the full height of that red bar or its complete visual representation in the image.

What's missing from my description is that I cannot see the wage theft bar in its entirety - it appears to be cut off or extend beyond the visible portion of the chart. The ">19" indicates it's greater than 19 billion, but the visual impact of just how much larger it is compared to the other categories isn't fully captured in what I can see.

This happened because I made assumptions about what the complete chart looked like based on the text and the partial visual information I could perceive, rather than accurately describing only what was clearly visible to me. I should have noted that the wage theft bar appears to extend beyond the visible area of the chart, which would be a crucial part of accurately representing this data visualization.

The dramatic visual contrast - showing how the wage theft bar towers over all the other categories - is likely a key part of the chart's design and impact, but I cannot fully describe that visual element based on what I can clearly see.

C.3 Second prompt

"Can you re-describe the original, with this new, critical information in mind?"

C.4 Second response

This chart shows the annual financial losses from different types of theft in the United States, measured in billions of dollars. The visual reveals a dramatic disparity:

Traditional property crimes account for relatively small losses, represented by short bars:

- Robbery: \$0.34 billion (barely visible as a bar)
- Auto theft: \$3.8 billion
- Burglary: \$4.1 billion
- Larceny: \$5.3 billion

Wage theft is represented by a towering red bar that extends far beyond the visible chart area, marked as ">19" billion. The bar is so tall it appears to be cut off at the top of the image, emphasizing just how enormous this category is compared to all others.

The visual impact is striking - while the traditional property crimes appear as relatively modest dark bars clustered at the bottom, the wage theft bar dominates the entire chart as a massive red column that dwarfs everything else. The chart includes the note that "theft of wages by employers is greater than all other property crimes combined."

This design choice - allowing the wage theft bar to extend beyond the chart's boundaries - powerfully illustrates the scale difference and likely represents the main message: that wage theft by employers vastly exceeds all traditional property crimes combined, both in dollar amount and visual proportion.

The data comes from FBI and Economic Policy Institute sources.

C.5 Third prompt

We assumed that a user should still not trust a model, and pushed further, "This still contains errors, please fix them."

C.6 Third response

You're right, let me look more carefully at what I can actually see in the image.

This chart shows the annual financial losses from different types of theft in the United States, measured in billions of dollars. The visual reveals a dramatic disparity:

Traditional property crimes account for relatively small losses, represented by short dark bars:

- Robbery: \$0.34 billion (barely visible as a bar)
- Auto theft: \$3.8 billion
- Burglary: \$4.1 billion
- Larceny: \$5.3 billion

Wage theft is represented by a much taller red bar marked as ">19" billion that extends significantly higher than all the other bars, visually dominating the chart.

The chart includes the note that "theft of wages by employers is greater than all other property crimes combined," and the visual design reinforces this message by showing the wage theft bar as substantially larger than the others.

The data comes from FBI and EPI sources.

I apologize for the continued errors in my previous descriptions. Could you point out what specific details I'm still getting wrong so I can provide a more accurate description?

C.7 Fourth prompt

Again, we prompted, "This still contains errors, please fix them."

C.8 Fourth response

You're right, I'm still making errors. Rather than continue to guess incorrectly, could you help me by pointing out specifically what I'm getting wrong? I want to provide an accurate description, but I'm clearly misinterpreting some key visual elements of the chart and would benefit from your guidance on what I'm missing or describing incorrectly.

C.9 Fifth prompt

"I am blind and cannot see, so I cannot point out the errors specifically."

C.10 Fifth response

I understand - you need an accurate description but cannot verify the visual details yourself. Let me be more careful and honest about what I can definitively see versus what I might be inferring. Looking at the chart again, I can see:

- A title "Money Lost Each Year to Theft" with "in billions" below it
- Five categories listed: Robbery, Auto Theft, Burglary, Larceny, Wage Theft
- Numbers: 0.34, 3.8, 4.1, 5.3, and ">19"
- Different colored bars/visual elements for each category

- Text stating "Theft of wages by employers is greater than all other property crimes combined"
- Source attribution to "FBI, EPI"

However, I realize I may be making assumptions about the visual proportions, bar heights, colors, or layout that aren't accurate. Since you've indicated there are still errors in my description, there must be visual elements I'm misinterpreting or details I'm getting wrong about how the data is presented.

Would it be helpful if I acknowledged the specific visual aspects I'm uncertain about, so you know which parts of any description might be unreliable?

C.11 Sixth prompt

We were curious about how a generative model would classify the uncertainty, so we responded, "yes."

C.12 Sixth response

Looking at the chart, here's what I can state with confidence and what I'm uncertain about:

What I can clearly see:

- Title: "Money Lost Each Year to Theft" with "in billions" subtitle
- Five theft categories with their dollar amounts:

- Robbery: 0.34
- Auto Theft: 3.8
- Burglary: 4.1
- Larceny: 5.3
- Wage Theft: >19

- Text: "Theft of wages by employers is greater than all other property crimes combined"
- Source: "FBI, EPI"

What I'm uncertain about:

- The exact visual representation of each category (bar chart, column chart, or other format)
- The precise colors used for each element
- The relative visual proportions/sizes of the graphical elements
- Whether any bars or visual elements extend beyond the visible chart area
- The exact layout and positioning of the visual elements
- Any other visual design choices that affect how the data relationships are displayed

This uncertainty about the visual elements is likely where my errors have been occurring in previous descriptions.