

Comments: Systems Thinking, Data Minding, and Mindware Agents for Multi-agent Data Analysis Systems

Xiao-Li Meng

Harvard University

1 Systems Thinking for DIY

Upon receiving Editor Hongtu Zhu’s invitation to comment on this refreshing article [28] (for a statistical journal), two of my inner voices immediately quarreled for a full 24 hours. The wiser one reminded me of my deterministic habit of grossly underestimating the time required to fulfill every promise I (foolishly) make. The curious one urged me to completely ignore that prior knowledge and recall that I know even less about agentic AI than about software—how could I forgo such an opportunity? After all, writing about what one does not know is the second-best way to learn. The best way to learn may not be immediately obvious to every reader, but which inner voice prevailed should be.

But where to begin? As it happens, I am also in the midst of a DIY (Do-It-Yourself) home improvement project. Since the authors’ LAMBDA [28] and other similarly purposed agentic systems [29] are designed to help domain experts—and more broadly anyone—to conduct data analysis on their own, lessons from my DIY experience naturally come to mind. Being a DIYist of size one, I asked ChatGPT-4o—which can be viewed as a “wisdom of a gigantic crowd” (Meng, 2023)—for a broader list of key considerations for DIY projects. ChatGPT-4o provided a sensible list, free of hallucination as far as I can judge, summarized in Figure 1—a slide from my discussion of the authors’ article at JSM (Joint Statistical Meetings) 2025.

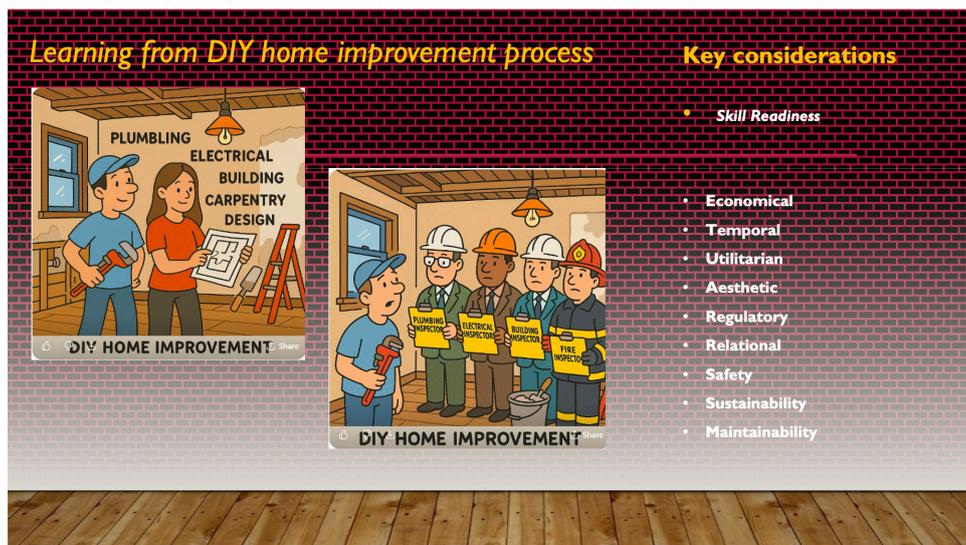


Figure 1: Systems Thinking for DIY Home Improvement (Created by ChatGPT-4o)

The list highlights considerations that should be obvious to any being possessing human intelligence, such as economical (Is it cost effective?), temporal (Do you have enough time?), and

utilitarian (Will it serve its intended purpose?). Others are less so, at least to beginners, such as regulatory (Are you compliant with local zoning codes and permits? DIY \approx RIY—Regulate It Yourself) and relational (How does it affect others? DIY projects can strain or strengthen relationships within households or with neighbors).

The list for DIY in data science will naturally differ, but the meta-level considerations are the same, at least when we aim for high-quality output. Most critically, like building a house or anything with multiple components, carrying out a successful data project requires what is known in the scientific, engineering, and organizational literature as systems thinking, a holistic approach that analyzes how a system’s constituent parts interrelate and how systems evolve over time and within larger contexts (e.g., Senge, 2006; Meadows, 2008). In data science, such holistic planning and reasoning are emphasized by Dahleh (2021) and Qin (2021) in their discussions of Chayes (2021), which provides an informative and inspiring account of how UC Berkeley organizes its educational, research, and institutional structures for data science and computing. [For full disclosure, these three articles and several cited below were published in *Harvard Data Science Review* (HDSR) at my invitation as its Founding Editor-in-Chief (EIC).]

Anyone who has undertaken a home improvement project knows the importance of planning and sequencing every activity in the right order. I am still learning this seemingly pedestrian lesson the costly way. Only after a reclining sectional was set in the center of the living room—as we had always planned—did I realize that we had forgotten, more precisely never planned, to install a floor electrical outlet! Navigating underground structures for wiring is a nontrivial undertaking unless one reopens a newly installed hardwood floor. Either way, it will cost hundreds of times more—I am still waiting for an estimate from an electrician—than if I had thought more thoroughly about the interplay between remodeling and interior design.

Such an oversight could have been easily avoided by a DIY planning agent, once prompted with our interior design wish list. I would be flattering myself to think I could patent this stupidity merely because there were no training data for agentic AI to learn from. In the same vein, many common oversights and mistakes in data science can be avoided or mitigated if domain experts and practitioners more broadly are reminded of them in a timely manner and shown examples of good practices (or bad consequences). A multi-agent data analysis system seems an ideal place for such AI-promoted self-regulation and self-consultation. The LAMBDA system [28] is already a significant step in that direction, especially with its built-in agentic inspector and emphasis on permitting human intervention throughout the process. The next section explores this idea further by suggesting an agentic data minder, guided by the ecosystemic thinking of data science, as summarized in my inaugural editorial for HDSR (Meng, 2019).

Before proceeding, let me be clear that there will always be cases where a human statistical consultant can outperform a multi-agentic system by a large margin. However, demand will always outpace supply not merely in human resources but also in individual capacity. The very fact that many of us have been surprised, and continue to be amazed, by how much the seemingly simplistic token-by-token predictive paradigm can accomplish should serve as an ongoing reminder that our mindware—to be discussed in Section 3—is far from fully developed, even for appreciating its own capability. Again, multi-agent data analysis systems can help us build and enhance our mindware. Section 3 therefore suggests to enhance LAMBDA and other multi-agent data analysis systems with mindware agents, whose possible roles and potentials are explored in Section 4.

2 Data Minder: A Vital Agent for Quality Control

The notions of data minding and its companion data confession (Meng, 2021) were inspired by a trio of articles in the inaugural issue of HDSR. Because HDSR was launched as a global forum to define and shape data science as a scientifically rigorous and globally impactful multidisciplinary field, it was natural for readers to expect its inaugural issue to provide a broad view of data sci-

ence. Knowing that no single discipline—let alone a single author—can capture the ecosystemic breadth of data science, I invited thought leaders from three disciplines, as depicted in Figure 2, another slide from my discussion at JSM 2025.



Figure 2: Extended Data Life Cycle

The top line with seven stages, from generation to interpretation, represents the data life cycle of Wing (2019). Known for her highly influential framing of computational thinking (Wing, 2006) among many other accomplishments, and for her extensive leadership in both industry and academia, I did not expect anything narrowly scoped from Jeannette Wing. Still, I was further inspired by the versatile philosopher and historian of science Sabina Leonelli’s emphasis on data conceptualization as preceding data generation (Leonelli, 2019). And I was deeply disappointed in myself when I read the essay by Christine Borgman, a seminal thinker in information studies and policies, on the “after lives” of data (Borgman, 2019). How could I have completely forgotten the issue of reproducibility and replicability (see the special theme in HDSR, edited by Stodden (2020)), which clearly extends the data life cycle beyond “interpretation”?

The answer is rather simple. Historically, both as a discipline and as individuals, we statisticians have adopted “modeling and fitting” as our predominant style of reasoning, as characterized by historian of science Christopher Phillips in his recent contribution to HDSR’s column *Mining the Past* (Phillips, 2025). While many of us emphasize the importance of high-quality data and analysis to ensure our positive contributions to science and society, our research training and publications have not systematized issues such as data quality inspection and analysis reproducibility as integral parts of our style of reasoning—borrowing the conception from philosopher of science Ian Hacking, as applied to statistics by Phillips. In other words, these issues do not come to our minds as habitually as modeling or fitting.

My disappointment was merely one example among many. With that in mind, when I was invited in 2020 to write an HDSR-style editorial for the special issue of the *Journal of the Royal Statistical Society, Series A* (JRSSA) on data science, I saw a pressing need to stress that conducting a data science project entails significantly more than merely carrying out “modeling and fitting” for a substantive purpose—a theme that dominated virtually all articles in that issue. This led me to propose that we should practice and promote *data minding*, which is “a stringent quality inspection process that scrutinizes data conceptualization, data pre-processing, data curation, and data provenance, in addition to data collection—the traditional focus of our attention—before data analysis” (Meng, 2021).

The inclusion of a software inspector in LAMBDA is clearly a step in the right direction, but it still reinforces primarily the modeling-and-fitting portion of the data life cycle. Domain

experts are arguably better qualified to engage in data minding—at least for data conceptualization and pre-processing—since sensible measurement and processing often demand substantive domain knowledge. If domain experts also collect their own data, then encouraging and enabling them to properly record data provenance and curation can go a long way toward ensuring (computational) reproducibility and (scientific) replicability—see the “Reinforcing Reproducibility and Replicability” Column of HDSR (hdsr.mitpress.mit.edu/reinforcing-rr) for reasons and methods.

A dedicated data minding agent, or *data minder*, can serve both as a reminder and an enabler for domain experts or any users to follow such good practices. Concretely, it should prompt users to go through a data quality checklist such as the following, especially when they rely on data collected by others:

- A. Why were the data collected, and who conceptualized them?
- B. Who collected the data, and how?
- C. When and where were the data collected?
- D. What has been done to the data since their creation?

The publications in the aforementioned special issue of JRSSA vividly illustrate the vitality of such data quality checks, as detailed in Meng (2021). The word *vitality* is apt because a failed check can literally lead to fatal mistakes. For instance, a study examined in Meng (2021) analyzed the impact of heat-wave warning systems using “the daily count of all causes of deaths” in a city, yet never questioned whether these counts were deaths *occurred* or *reported* on that day. Because most heat waves last less than a week while reporting delays can span days or weeks, this distinction could make the difference between results that are relevant or fatally misleading (e.g., finding little impact due to time misalignment when in fact the warning system had saved many lives).

The problem could have been easily avoided if the study’s authors had been reminded by question (C) and thus had inquired with the city to clarify. The issue immediately caught my attention because I had previously studied delays in reporting AIDS cases to the Centers for Disease Control (Tu et al., 1993a,b; Bouman et al., 2005). Others may not have such experiences to instinctively raise the question—but this is precisely where an AI agent can serve as a timely reminder, drawing on a substantial literature on reporting delays to central databases.

Broadly speaking, a data minder can perform many tasks: alerting users to potential problems before analysis, suggesting good practices, and pointing to relevant literature (at least when it does not hallucinate). In addition to prompting users to go through the (A)-(D) list, the agent can help generate a *data confession* report documenting “details on the genealogy of a given body of data, including an account of its deliberations, especially with respect to sources of adverse influence on data quality” (Meng, 2021). Such a practice would mark a substantive step forward in our quest for scientific replicability—a task far more challenging than ensuring computational reproducibility. In essence, having a data minder reduces the number of incidences of “garbage in, package out” (Meng, 2021), especially as DIY data science becomes widespread. Most profoundly, it reinforces that data science is about the science of data as much as about using data for science and beyond.

3 Mindware Agents: Augment Users’ Data Intelligence

DIY is seldom easy for beginners, yet many who try it end up becoming serial DIYists. The attraction lies not merely in a sense of accomplishment—and rarely in economy, especially for those who believe time is money—but in self-improvement: acquiring life skills while enhancing one’s *mindware*, a term popularized by psychologist David Perkins (1995) in his book titled

Outsmarting IQ: The Emerging Science of Learnable Intelligence. As its book reviewer Hertz (1996) wrote, “Mindware is software for the mind that enables one to do useful things with the data stored in one’s memory.”

Perkins (1995) used the concept as an organizing theme for three questions:

- (a) What mechanisms underlie intelligence?
- (b) Can people acquire more intelligence?
- (c) Which aspects of intelligence especially need more attention?

If we regard humans’ ability to frame, learn from, and reason with data as a form of intelligence—and it is (indeed, the term *data intelligence* has already been adopted¹)—then these three questions together provide a useful checklist for creating mindware agents for LAMBDA and similar systems.

Since Perkins’s book, key aspects of mindware have been broadly identified as “rules, knowledge, procedures, and strategies that a person can retrieve from memory to aid decision making and problem solving” (Stanovich, 2009; Eigenauer, 2024). Question (a) reminds us to design mechanisms that prompt and even condition humans to engage these aspects, and to personalize such processes through mindware agents. Concretizing “people” as individual users, question (b) encourages us to assess the feasibility and effectiveness of such personalization, while question (c) calls for prioritization to avoid cognitive or computational overload.

Like super agents (e.g., Yao et al., 2025) or meta agents (e.g., Qian and Liu, 2025), mindware agents should be versatile, performing diagnostic, consultative, educational, or even therapeutic functions. But unlike super or meta agents—which are task- or product-oriented, designed to help users produce specific outcomes—mindware agents are intelligence- or process-oriented. Their primary function is to help users augment their data intelligence. Figure 3 compares the focuses of the three types of agents and the corresponding human roles they aim to emulate.

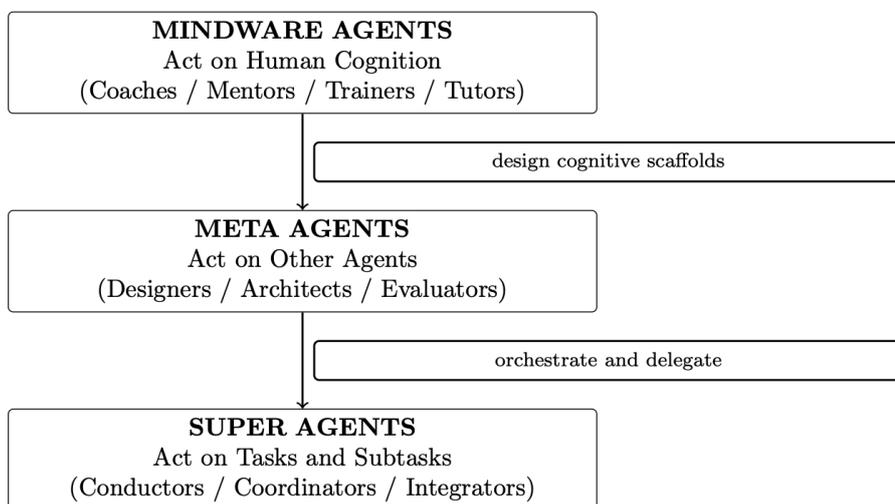


Figure 3: Conceptual hierarchy of Mindware, Meta, and Super Agents.

Another defining feature of mindware agents is *reverse prompting*, since much of their activity involves prompting users—whether for diagnostic or evaluative purposes, or as cognitive reinforcers or behavioral nudgers. The notion of *nudging*, popularized by Thaler and Sunstein (2008), shows that small and well-designed changes in choice architecture—such as adjusting defaults or prompts—can guide individuals toward better decisions without constraining freedom.

¹See blogs.idc.com/2019/11/25/defining-data-intelligence-intelligence-about-data-not-from-data/

Beshears et al. (2020) demonstrate that such interventions improve behavior across domains from health to finance, while Mertens et al. (2022) confirm that nudges exert consistent, measurable effects, typically small to medium in magnitude. Other research cautions that effectiveness depends on context: defaults can fade or even reverse once removed (Montero-Porras et al., 2024). Overall, this body of behavioral research suggests that embedding a nudging mindware agent within LAMBDA or alike can provide a robust mechanism for enhancing users’ data intelligence at minimal cognitive or financial cost—especially as such systems become more widely adopted and their nudges increasingly contextualized, reinforced, and refined as the mindware agents themselves evolve.

4 Strategizing Mindware Agents: A Human-AI Collaboration?

To take advantage of large language models’ ability to leverage *Homo sapiens* intelligence (Meng, 2023), I prompted ChatGPT-5 to “discuss ways mindware agents can help (and why) to nudge the users of multi-agent data analysis systems to enhance their data intelligence and ultimately their mindware.” Its response strengthened my confidence in the usefulness of mindware agents—both because the benefits it outlined exceeded what I had initially conceived, and because the response itself helped me enhance my own mindware. Below is the ChatGPT-5 (instant) response in six bullet points, reproduced verbatim.

- **Human biases & bounded rationality:** Even data-savvy users are prone to cognitive biases such as confirmation bias, anchoring, overconfidence, and base-rate neglect. Nudges can mitigate these by structuring the choice architecture within the data analysis workflow.
- **Choice overload & complexity:** Data analysis often presents a proliferation of models, metrics, and visualizations. Agents can simplify and guide choices, making it easier for users to select good defaults or analytical paths.
- **Scaffolding metacognition:** Nudges can prompt reflection (e.g., “Have you considered alternative hypotheses?”), encouraging more deliberate and critical reasoning.
- **Progressive learning & internalization:** Well-designed nudges, gradually faded, help users internalize good heuristics so that over time the nudging agent becomes less necessary.
- **Personalization & adaptivity:** Agents can tailor nudges based on user behavior, error patterns, and demonstrated competence, increasing their relevance and effectiveness.
- **In-situ intervention:** Because agents operate within the user’s workflow, nudges can be delivered precisely at the decision point rather than retrospectively—enhancing immediacy and impact.

In short, mindware agents can nudge users toward scientifically healthier behaviors by encouraging self-monitoring, incentivizing learning and habit building, and making quality inspection routine.

For human introspections on data analysis quality, I previously provided a list of self-check questions in Meng (2022). A mindware agent can readily take on the role of nudger or reminder to make such reflection routine by switching from the *Did-I* self-questioning in my original list to *Do-you* agent prompting. This adjustment also addresses Thomas Junk’s rightful criticism of my original list: that my “formulation of these questions in the past tense implies their being asked too late” (Junk, 2023). Below is the revised list, whose purpose may be best captured by the Chinese proverb “throw a brick to attract jade.” That is, the list is neither comprehensive nor optimally phrased; it is offered as a starting point for improvement and augmentation.

- Do you understand and carefully consider the data collection and pre-processing processes in your study? How much do you know about the quality of the data you used?
- Do you have sufficient understanding of the substantive problem to recognize its inherent challenges, such as confounding factors or lack of identifying information?
- Are the assumptions and models you adopted free of internal contradictions? If not, what justifications allow such contradictions?
- Do you follow principled and holistic methods such as probabilistic propagation and conditioning, or do you rely on ad hoc “intuitive” methods?
- What approximations do you make in modeling, derivations, or computation? Which are most vulnerable?
- Do you understand the impact of these approximations—especially their worst possible consequences? Have you examined studies on their impact, or investigated them theoretically or empirically?
- Do you perform sufficient validation or robustness checks for the models and assumptions you posit? Are you confident your results would hold reasonably well if your data were perturbed?
- Do you carry out reliability checks on numerical evaluations or simulations—for instance, verifying computational results independently by different methods, routines, or research assistants?
- Do you engage in any form of cherry-picking, from data gathering to validation? If so, what reasons lead you to believe your cherry-picking causes little harm?
- Do you understand your findings well enough to explain and teach them confidently to non-experts?

This list was constructed more than a year before November 30, 2022, and thus could not benefit from the “wisdom of a gigantic crowd” offered by ChatGPT and its kin. While legitimate concerns remain about overreliance on AI, for tasks such as brainstorming or strategizing it would be as silly not to consult ChatGPT or its equivalents as it would be to trust everything they produce. Table 1 therefore reproduces verbatim another ChatGPT-5 output; readers may judge for themselves whether it offers more inspiration than hallucination or vice versa. Regardless, I hope the human-AI collaboration documented in this section can prompt the authors of LAMBDA and their competitors and successors to take on the task of designing, building, and integrating mindware agents into multi-agent data analysis systems.

Table 1: Possible Agent-Based Nudging Strategies in Multi-Agent Data Analysis Systems

Nudging Strategy	Examples	Mindware Benefit
Smart defaults / pre-sets	When a user opens a new analysis, the agent sets default visualization types, statistical tests, or hyperparameters based on dataset characteristics.	Reduces cognitive load; ensures better “first-pass” choices; avoids poor initial settings.
Adaptive ordering / pruning	The agent orders models, metrics, or filters by relevance, or hides unlikely-to-be-useful options.	Minimizes search costs; nudges toward more promising analyses.
Prompts / reminders	Before finalizing results, the agent asks, “Did you check for multicollinearity?” or “Have you cross-validated?”	Encourages critical checks; reinforces best analytical practices.
Contrast framing	When presenting two competing hypotheses or model options, the agent frames them as “baseline vs. alternative” with side-by-side differences.	Promotes systematic comparison and counterfactual thinking.
Progressive disclosure / scaffolding	Initially hides advanced features; reveals them as the user demonstrates mastery.	Prevents overwhelm; supports staged skill development.
Error or anomaly alerts	If suspicious data patterns or possible data leakage are detected, the agent flags them and suggests checks.	Guards against flawed analyses; fosters vigilance and assumption testing.
Reflection / metacognitive prompts	After completing an analysis, the agent asks: “What assumptions did you make? What alternatives could you test?”	Encourages reflection, deeper reasoning, and metacognitive awareness.
Peer / exemplar nudges	“In similar datasets, users often tried method X; would you like to see those results?”	Leverages social proof and exemplar-based learning.
Fading or tapering nudges	As the user demonstrates competence, the agent gradually reduces prompts and reminders.	Encourages autonomy and prevents overdependence on the agent.

5 Let Us Wear AI T-Shirts: Build a Tide to Lift All Boats

As an educator for 35 years, and EIC of *Harvard Data Science Review* with its mission to feature “everything data science and data science for everyone,” I thank the authors for emphasizing the role of multi-agent systems like LAMBDA in reshaping data science education. Mindware agents not only advance this mission but also enhance it by intentionally augmenting users’ data intelligence while they carry out a data science project. For readers still wondering what I meant

earlier by the best way to learn something new—it is by teaching it. Likewise, we enhance our mindware by synchronizing its development with its deployment. The more people become data-intelligent, the greater the net impact of data science and AI technologies, because the more we understand the science of data, the more effectively we can undertake the use of data for science.

From David Donoho (as discussant for the fireside Conversation):

“Can statisticians still contribute to the field, without fully understanding AI? The expansion of statistics departments over the last decade is remarkable and our community should be lauded for our tradition of collaboration. While some may feel that AI threatens to overshadow statistics, I suggest that we take a more welcoming stance toward AI—maybe even embrace it more openly. **Let us ‘wear AI t-shirts’ to show our willingness to participate and collaborate.**”

From David Donoho (as discussant for the fireside Conversation):

“Can statisticians still contribute to the field, without fully understanding AI? The expansion of statistics departments over the last decade is remarkable and our community should be lauded for our tradition of collaboration. While some may feel that AI threatens to overshadow statistics, I suggest that we take a more welcoming stance toward AI—maybe even embrace it more openly. **Let us ‘wear AI t-shirts’ to show our willingness to participate and collaborate.**”

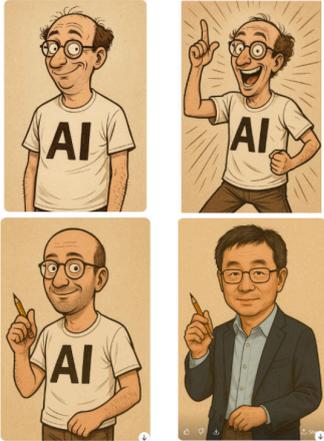


Figure 4: Images of statisticians as captured by ChatGPT-4o as of July 2025

I am equally grateful to Editor Hongtu Zhu for his vision in dedicating the JSM 2025 JASA session to two articles that depart significantly from past selections traditionally focused almost exclusively on advances in statistical theory and methods. These choices are, of course, expected of JASA, but as the world experiences the rapid evolution of AI technologies, we statisticians should—in David Donoho’s words from his discussion in “Statistics and AI: A Fireside Chat” (Lin et al., 2025)—wear the “AI t-shirts.” See Figure 5, the closing slide of my JSM 2025 discussion, for the full quote.

Indeed, the cartoons created by ChatGPT in the slide should remind us of the need to improve our images—figuratively and literally—as a profession. This slide was borrowed from a panel on “AI Meets Statistics: A Two-Way Power Boost”, took place on July 28, 2025, as a part of a celebration of my colleague Jun Liu’s many accomplishments and his retirement from Harvard (or more precisely, his new mission to make another university great again). My presentation drew on several recent HDSR articles about our beloved discipline (Kirch et al., 2025; Lin et al., 2025; Phillips, 2025), and Donoho’s call provided a fittingly energizing summary.

To add a touch of memorable fun, I prompted ChatGPT-4o: “Please create a humorous picture of a statistician wearing a t-shirt with ‘AI’ on it.” To my surprise, it produced the goofy-looking figure in the top-left corner of the four cartoons in Figure 4. Suspecting the word “humorous” had misled it, I tried again: “Create one with a much more dynamic and exciting statistician.” The reveal of that cartoon during my talk nearly brought down the house—though I’m sure ChatGPT hadn’t anticipated that reaction when executing my instruction to be “exciting.”

Frustrated, I prompted it a third time: “Create one with a statistician who better reflects the current profession.” Alas, this produced an even balder image, shown in the lower-left corner of Figure 4. If this cartoon reflects, even partially, the public image of our field, then we have much to do—starting with shedding the perception that statisticians are too theoretical or too restrictive in scope. With collective effort, we can leverage the AI wave to create a tide that lifts all boats: improving data literacy and data intelligence at every level. Building widely adoptable and applicable multi-agent systems like LAMBDA is a concrete step toward that goal—especially when equipped with mindware agents that can enhance data intelligence itself.

As for the prompt behind the final cartoon—with full self-confidence coverage—I’ll leave that as a riddle for readers (though I once nearly joined the Hair Club for Men to boost mine). What truly pleased me, however, was the persistence of the pencil in the image instead of, say, a laptop. To me, that symbol of writing is synonymous with the act of thinking. While I hope our field will continue to produce more user-friendly tools, we must not forget that, at its core, statistics is a leader of human intelligence because, as our philosopher friends remind us, we are applied epistemologists (Meng, 2024).

So let us wear our AI t-shirts not as uniforms of surrender, but as emblems of synthesis—reminding the world that the future of intelligence, artificial or otherwise, still begins with the human mind.

References

- [1] Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Milkman, K. L. (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes*, 161:3–19.
- [2] Borgman, C. L. (2019). The Lives and After Lives of Data. *Harvard Data Science Review*, 1(1). <https://hdsr.mitpress.mit.edu/pub/4giycvuj>.
- [3] Bouman, P., Dukic, V., and Meng, X.-L. (2005). A bayesian multiresolution hazard model with application to an AIDS reporting delay study. *Statistica Sinica*, pages 325–357.
- [4] Chayes, J. (2021). Data Science and Computing at UC Berkeley. *Harvard Data Science Review*, 3(2). <https://hdsr.mitpress.mit.edu/pub/wzhgxmcc>.
- [5] Dahleh, M. A. (2021). Data Science in Domains: Interaction of Physical, Social, and Institutional Systems. *Harvard Data Science Review*, 3(2). <https://hdsr.mitpress.mit.edu/pub/nv9xq9yu>.
- [6] Eigenauer, J. (2024). Mindware: Critical thinking in everyday life. *Journal of Intelligence*, 12(2):17.
- [7] Hertz, N. R. (1996). (Book Review) Outsmarting IQ: The emerging science of learnable intelligence. *Personnel Psychology*, 49(2):520.
- [8] Junk, T. R. (2023). Comment on “Double Your Variance, Dirtify Your Bayes, Devour Your Pufferfish, and Draw Your Kidstogram,” by Xiao-Li Meng. *The New England Journal of Statistics in Data Science*, 1(1):26–28.
- [9] Kirch, C., Lahiri, S., Binder, H., Brannath, W., Cribben, I., Dette, H., Doebler, P., Feng, O., Gandy, A., Greven, S., Hammer, B., Harmeling, S., Hotz, T., Kauermann, G., Krause, J., Krempl, G., Nieto-Reyes, A., Okhrin, O., Ombao, H., Pein, F., Pešta, M., Politis, D., Qin, L.-X., Rainforth, T., Rauhut, H., Reeve, H., Salinas, D., Schmidt-Hieber, J., Scott, C., Segers, J., Spiliopoulou, M., Wilhelm, A., Wilms, I., Yu, Y., and Lederer, J. (2025). Challenges and Opportunities for Statistics in the Era of Data Science. *Harvard Data Science Review*, 7(2). <https://hdsr.mitpress.mit.edu/pub/ufaltur6>.

- [10] Leonelli, S. (2019). Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science. *Harvard Data Science Review*, 1(1). <https://hdsr.mitpress.mit.edu/pub/4ovhpe3v>.
- [11] Lin, X., Cai, T., Donoho, D., Fu, H., Ke, T., Jin, J., Meng, X.-L., Qu, A., Shi, C., Song, P., Sun, Q., Wang, W., Wu, H., Yu, B., Zhang, H., Zheng, T., Zhou, H., Zhou, J., Zhu, H., and Zhu, J. (2025). Statistics and AI: A Fireside Conversation. *Harvard Data Science Review*, 7(2). <https://hdsr.mitpress.mit.edu/pub/a7kmqk35>.
- [12] Meadows, D. H. (2008). *Thinking in Systems: A Primer*. Chelsea Green Publishing, White River Junction, VT.
- [13] Meng, X.-L. (2019). Data Science: An Artificial Ecosystem. *Harvard Data Science Review*, 1(1). <https://hdsr.mitpress.mit.edu/pub/jhy4g6eg>.
- [14] Meng, X.-L. (2021). Enhancing (publications on) data quality: Deeper data minding and fuller data confession. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(4):1161–1175.
- [15] Meng, X.-L. (2022). Double your variance, dirtify your bayes, devour your pufferfish, and draw your kidstogram. *The New England Journal of Statistics in Data Science*, pages 4–23.
- [16] Meng, X.-L. (2023). Human Intelligence, Artificial Intelligence, and homo sapiens Intelligence? *Harvard Data Science Review*, 5(4). <https://hdsr.mitpress.mit.edu/pub/nw5bilq7>.
- [17] Meng, X.-L. (2024). A Heisenberg-esque uncertainty principle for simultaneous (machine) learning and error assessment? *Statistics and Applications ISSN 2454-7395 (online) Special Issue in Memory of Prof. CR Rao*, 22(3):667–693.
- [18] Mertens, S., Herberz, M., Hahnel, U. J. J., and Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*, 119(1):e2107346118.
- [19] Montero-Porrás, E., Suchon, R., Lenaerts, T., and Fernández Domingos, E. (2024). Defaults: A double-edged sword in governing common resources. *arXiv preprint 2403.06796*.
- [20] Perkins, D. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. Simon and Schuster.
- [21] Phillips, C. J. (2025). Styles of Statistics. *Harvard Data Science Review*, 7(2). <https://hdsr.mitpress.mit.edu/pub/yen5ysj5>.
- [22] Qian, H. and Liu, Z. (2025). Metaagent: Toward self-evolving agent via tool meta-learning. *arXiv preprint* <https://arxiv.org/abs/2508.00271>.
- [23] Qin, S. J. (2021). Data Science Education With Domain Knowledge and System Principles. *Harvard Data Science Review*, 3(2). <https://hdsr.mitpress.mit.edu/pub/8si074w9>.
- [24] Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization*. Broadway Business.
- [25] Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.
- [26] Stodden, V. (2020). Theme Editor’s Introduction to Reproducibility and Replicability in Science. *Harvard Data Science Review*, 2(4). <https://hdsr.mitpress.mit.edu/pub/kincvy67>.

- [27] Su, B., Zhang, J., Collina, N., Yan, Y., Li, D., Cho, K., Fan, J., Roth, A., and Su, W. (2025). The icml 2023 ranking experiment: Examining author self-assessment in ml/ai peer review. *Journal of the American Statistical Association*, (just-accepted):1–16.
- [28] Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., Yuan, Y., and Huang, J. (2025). Lambda: A large model based data agent. *Journal of the American Statistical Association*, 1–13.
- [29] Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., Yuan, Y., and Huang, J. (2025). A survey on large language model-based agents for statistics and data science. *The American Statistician*, 1–14.
- [30] Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, CT.
- [31] Tu, X. M., Meng, X.-L., and Pagano, M. (1993a). The aids epidemic: estimating survival after AIDS diagnosis from surveillance data. *Journal of the American Statistical Association*, 88(421):26–36.
- [32] Tu, X. M., Meng, X.-L., and Pagano, M. (1993b). Survival differences and trends in patients with aids in the united states. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 6(10):1150–1156.
- [33] Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3):33–35.
- [34] Wing, J. M. (2019). The Data Life Cycle. *Harvard Data Science Review*, 1(1). <https://hdsr.mitpress.mit.edu/pub/577rq08d>.
- [35] Yao, Y., Wang, H., Chen, Y., Wang, J., Ren, M. C. J., Ding, B., Avestimehr, S., and He, C. (2025). Toward super agent system with hybrid AI routers. *arXiv preprint* <https://arxiv.org/abs/2504.10519>.