INFUSING 'STATISTICAL, DATA SCIENCE AND AI LITERACY' FOR THE GENERAL LEARNER POPULATION: CONCEPTUAL, INSTRUCTIONAL, AND SYSTEMIC CHALLENGES

Iddo Gal
University of Haifa, Israel, iddo@research.haifa.ac.il

Focus Topics: AI and Data Science Competencies, AI and Data Science Curricula and Implementation in School

Introduction

How can we prepare *all* school graduates (i.e., the general learner population, not just high-ability students) to function effectively in a world increasingly infused with messages and results of processes involving statistics (S), data science (DS), machine learning (ML), and systems involving artificial intelligence (AI)? This broad question underscored the AIDEA_2025 symposium in which this talk was given, and has motivated the current talk, in which I have problematized and explored several related conceptual, instructional, and systemic challenges.

Recent developments and expansion of applications in statistics, data science, machine learning and AI (hereby: S/DS&ML/AI) and in related areas have been causing profound rethinking regarding needed changes at the school and tertiary levels (Wilkerson & Polman, 2020; Engel et al., 2022; Casal-Otero et al., 2023; Fielding et al., 2025). Related directions for change involve developing new curricular content and pathways that aim to expand and deepen the knowledge of relatively advanced learners in mathematics and STEM areas, in order to prepare them for formal academic studies and future jobs in S/DS&ML/AI. Such pathways involve, for example, adding courses or elements that develop advanced coding skills and/or the ability to analyze big data, either in school-based statistics and mathematics education or other contexts (e.g., computer science).

The directions sketched above pose many educational challenges on their own, hence have been the focus of many talks at the AIDEA_2025 symposium. However, this talk has argued that such directions, while important to pursue, should be seen as insufficient, for two reasons: (a) current directions are focused on preparing students to cope with 'producer tasks' rather than with 'consumer tasks' (see below), thus not preparing all students for coping with key tasks they can expect to encounter as adults, and (b) current directions serve only a *minority* of school learners, those with relatively advanced mathematical or computer skills and related interests (the upper 20%-30% of students), and do not serve the needs of the majority of the school-age population.

These issues are explained below in more detail, and some of the associated educational and systemic challenges are then sketched in broad strokes. This talk was founded in part on prior work related to the development of statistical literacy, probability literacy, and adult numeracies (Gal et al., 2020) and informed by recent discussions of the connections between mathematics and statistics education with citizenship roles (Geiger et al, 2023). Accordingly, the talk aimed to contribute to extant thinking regarding both conceptual and systemic issues which are seldom discussed in a coordinated manner in extant literature.

Preparing learners to cope with consumer tasks vs. producer tasks

Typically, mathematics and statistics education at the school level focus on teaching learners to address producer tasks, i.e., tasks where they have to produce or 'do' mathematics or produce statistics in order to solve or address various types of demands of instructional situations. Yet, in many contexts of life, people do *not* have to 'do' mathematics or statistics, in the sense of computing, estimating, collecting or analyzing data, etc. Rather, in key important life contexts people have to serve as (critical) *consumers of information* and interpreters of messages with embedded mathematical or statistical elements. The talk sketched three principal contexts: (1) *workplace* (e.g., in agriculture, manufacturing, service industries), (2) *service consumption* (when acting as customers of, e.g., financial, health-related, travel and hospitality services, interpreting information on websites, in letters and leaflets to customers, etc.), and (3) *news and information consumption* (e.g., when interacting with websites, digital news services, or social networks).

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

Overall, I argue that consumer tasks encountered in such key life contexts do *not* require adults to analyze any data (which is what schools teach). Rather, they (a) involve *text-based message*s, with (b) a broad range of statistical constructs and ideas that *go above and beyond* material in standard statistics instruction at the school level (e.g., related to dynamic, multivariate and multi-source statistics, projections and forecasts), and (c) require *critical interpretation* of text-based messages and arguments, and understanding of related ethical issues. For details and illustrations, see works regarding Civic Statistics (Engel et al., 2022; Gal, 2022; Gal et al., 2022), numeracy demands on citizens (Gal et al., 2020; Gal, 2024a), ideas related to statistical models and modeling that all citizens should know (Gal, 2024b), and AI literacy more generally (e.g., Casal-Otero et al, 2023).

Which learners are we targeting? The majority of students are in risk of being neglected

This talk promotes the need to develop educational solutions that are suitable not just for school learners with relatively advanced mathematical or computer skills and related interests (the upper 20%-30% of students), but for the *majority* of the school population (the other 70%-80%), who may lack such skills, *yet still need to cope with a broad range of consumer tasks involving statistics, data science, and some aspects of ML&AI*. The figures regarding the size of the minority and majority groups are derived from results of the most recent PISA assessment (OECD, 2023). I argue that it is crucial to examine two interrelated parts of the PISA reports (which are also echoed in results from recent TIMSS and PIAAC surveys): (a) results regarding p*roficiency distributions* across reported performance levels in two key and relevant domains: *mathematics* (or numeracy in PIAAC) and *reading* (or literacy in PIAAC), and (b) the associated *level descriptors*, which explain and illustrate what students at each proficiency level can do or cope with in mathematics and reading.

Looking broadly at these information sources, and acknowledging variation between and within participating countries, I have argued that *only students at PISA levels 4 and above* (out of six main proficiency levels in mathematics and reading) may be able to cope effectively with advanced materials related to S/DS&AI *and* with tasks requiring *critical* reading and evaluation of complex data-based textual claims. Overall, PISA results show that only 20-30% of students (and sometimes even below that, depending on the country) perform at Level 4 and above in mathematics and reading.

S/DS&ML/AI literacy: Needed conceptual developments

The talk has argued that given the realities sketched by the PISA results, and based on a combined societal and economic rationale, school systems should direct much more attention to a new goal that will enable school graduates below Level 4 in mathematics in reading (again, 70%-80% of the entire school population!) to develop by the end of high-school an enhanced literacy, termed here *S/DS&ML/AI literacy*. Building on prior work regarding statistical and probability literacy (e.g., Gal, 2002, 2005, 2024b) and other sources, S/DS&ML/AI literacy is tentatively defined as follows:

> **S/DS&ML/AI literacy:** *The ability to comprehend and critically interpret messages and actions from systems that process or involve statistical, data science, and selected ML & AI elements*.

This definition aims to highlight the need for all citizens to comprehend the statistical and mathematical *products* (see Geiger & Gal, 2022) *of S/DS&ML/AI systems* ─ but not necessarily the details of their internal working processes and analytic mechanisms, which mostly remain as a black box. The term “systems” should be viewed broadly, and encompass both products of traditional producers such as official statistics agencies, news media, and so forth, as well as more recent and rapidly emerging systems based on DS/ML/AI (e.g., recommender systems, algorithmic systems and those using Large Language Models) that release to their consumers *data-based* messages and decisions. (See Fielding et al, 2025 regarding data-based aspects of some of these systems)

The development of S/DS&ML/AI literacy is necessitated from a societal and economic rationale, yet poses a serious challenge from a *conceptual* perspective, given the many conceptual elements it involves or subsumes (Note: some are not explicit in the definition itself, but emerge only in examining the underlying sub-constructs). This definition draws on ideas from all the disciplines involved, which have somewhat separate intellectual or scholarly communities, and from others such as conceptions of data literacy in communication science and media studies. In traditional statistics, which is where I have started this intellectual journey given that everything is based on 'data', key big ideas refer to data, variability, distribution, randomness and predictability, sample and population, generalization and inference, and so forth. While such constructs are central in their own right, they provide only a partial preparation at best for the wider and more complex set of ideas that underlie the subfields underlying S/DS&ML/AI. I argue that we need to develop, test, and implement further 'conceptual continua' that cut across all the subfields involved, such as (but not limited to):

Procedure <> Formula <> Model <> Algorithm <> Heuristic <> Large language model

In addition, keeping in line with the logic underlying consumer tasks and roles, the development of S/DS&ML/AI literacy requires the advancement of learners' capacities for *criticality* such as those that underlie statistical and probability literacy (Gal, 2002, 2005, 2022; Weiland, 2017; Bailey & McCulloch, 2023), but going beyond them to accommodate the proliferation of data-based messages across news media, social networks, workplace systems, and related channels. This in turn requires the development of additional relevant conceptual continua, such as (but not limited to):

Features <> Limitations <> Bias (Statistical or ethical) <> Misinformation <> Fake news

Further, conceptual targets that can guide instruction related to S/DS&ML/AI literacy should go beyond specification of cognitive knowledge and describe *dispositions* that have to be developed. Examples are the *beliefs, attitudes, and critical stance*, and appreciation of *ethical* issues, required for critical and balanced interpretation of messages, actions, decisions, or recommendations by systems based on ML&AI that utilize data (e.g., ML&AI are increasingly used in various industries (e.g., health, finance) to assess probabilities and risk levels). Prior conceptual frameworks in this regard that relate to statistical and probability literacy (Gal, 2002, 2005; Weiland, 2017) and related habits of mind (Bailey & McCulloch, 2023), or to AI literacy (e.g., Casal-Otero et al., 2023), have to be extended to cover all elements subsumed in S/DS&ML/AI literacy.

S/DS&ML/AI literacy: Systemic and instructional challenges

The development of S/DS&ML/AI literacy poses formidable challenge for educational systems (at the national and school levels) and for teachers. S/DS&ML/AI literacy is a complex educational target because it involves or requires *integrated* and *coordinated* attention from scholars and practitioners alike to many underlying "big ideas" (some of which, but not all, were sketched above). Learning and understanding of all the related constructs should be nurtured and augmented across the school years, yet the preferred ordering and positioning of key big ideas across the school grades, and the related learning trajectories, are quite unclear or contested.

To enable the development of S/DS&ML/AI literacy by all students, not just elite students, beyond deciding on curricular trajectories we need to examine and transform many practical systemic issues. Examples are the need to develop the content knowledge and pedagogical knowledge of teachers (many of whom lack familiarity even with basic statistics), develop attitudes of current and future pre-service teachers towards addressing a seemingly new literacy that cuts across regular disciplinary boundaries and consider professional development schemes in this regard, address funding levels, and revise assessment frameworks. Each of these requires sustained attention from policy makers and school leaders, since each could function as either a barrier or facilitator that may affect the promotion of S/DS&ML/AI literacy in a cohesive and dynamic manner.

To illustrate the systemic complexity, I note that statistics instruction, which is only one aspect of S/DS&ML/AI, is often done within mathematics education at the school level, yet often (still) focuses on procedural aspects and shuns attention to critical interpretation of text-based statistical messages, even though these are very prevalent in the life of workers, citizens, and service recipients. The search is still on for effective ways to promote statistical and data literacy across the school years (Engel et al., 2022; Gal, 2022; Schreiter et al., 2024). Adding solid extensions related to DS&ML/AI, in an integrated way, pushes the limits for schools and teachers alike, but is essential in order to prepare all school graduates to engage the vast and rapidly growing array of messages, services, products, and systems with embedded S/DS&ML/AI, where adults act as critical/smart *consumers of information* and do not analyze or manipulate any data.

The many challenges sketched above involve numerous issues that have not been explicated or discussed in the literature in a systematic and comprehensive manner. I hope that this talk will inform a constructive interdisciplinary dialogue and future research directions in this regard.

References

Bailey, N. G., & McCulloch, A. W. (2023). Describing critical statistical literacy habits of mind. *The Journal of Mathematical Behavior*, *70*, 101063. https://doi.org/10.1016/j.jmathb.2023.101063

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: a systematic literature review. *International Journal of STEM Education*, *10*(1), 29. https://doi.org/10.1186/s40594-023-00418-7

Engel, J., Nicholson, J., Louie, J. (2022). Preparing for a data-rich world: Civic Statistics across the curriculum. In J. Ridgway (Ed.), *Statistics for Empowerment and Social Engagement: Teaching Civic Statistics to Develop Informed Citizens.* Springer, https://doi.org/10.1007/978-3-031-20748-8_18

Fielding, J., Makar, K. & Ben-Zvi, D. (2025). Developing students' reasoning with data and data-ing. *ZDM–Mathematics Education*, 57, 1–18. https://doi.org/10.1007/s11858-025-01671-6

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I. (2005). Towards 'probability literacy' for all citizens. In G. Jones (ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 43-71). London: Kluwer Academic Publishers.

Gal, I. (2022). Critical understanding of Civic Statistics: engaging with important contexts, texts, and opinion questions. In J. Ridgway (Ed.), *Statistics for Empowerment and Social Engagement: Teaching Civic Statistics to Develop Informed Citizens*. https://link.springer.com/book/10.1007/978-3-031-20748-8

Gal, I. (2024a). Adult education in mathematics and numeracy: A scoping review of recent research. *ZDM–Mathematics Education*, *56*(2), 293-305. https://doi.org/10.1007/s11858-024-01549-z

Gal, I. (2024b). What do citizens need to know about real world statistical models, and the teaching of data modeling. In S. Podworny, D. Frischemeier, M. Dvir and D. Ben-Zvi (Eds.), *Reasoning with Data Models and Modeling in the Big Data Era* (pp. 91-99). Minerva Foundation and Paderborn University. https://doi.org/10.17619/UNIPB/1-1815

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: a study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics*, *111*, 5-28. https://doi.org/10.1007/s10649-022-10151-7

Gal, I., Grotlüschen, A., Tout, D., & Kaiser, G. (2020). Numeracy, adult education, and vulnerable adults: a critical view of a neglected field. *ZDM–Mathematics Education*, *52*(3), 377-394. https://doi.org/10.1007/s11858-020-01155-9

Gal, I., Nicholson, J., & Ridgway, J. (2022). A conceptual framework for Civic Statistics and its educational applications. In J. Ridgway (Ed.), *Statistics for Empowerment and Social Engagement: Teaching Civic Statistics to Develop Informed Citizens.* https://doi.org/10.1007/978-3-031-20748-8_13

Geiger, V., Gal, I., & Graven, M. (2023). The connections between citizenship education and mathematics education. *ZDM–Mathematics Education*, *55*(5), 923-940. https://doi.org/10.1007/s11858-023-01521-3

OECD (2023), *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing, Paris. https://doi.org/10.1787/53f23881-en.

Schreiter, S., Friedrich, A., Fuhr, H., Malone, S., Brünken, R., Kuhn, J., & Vogel, M. (2024). Teaching for statistical and data literacy in K-12 STEM education: a systematic review on teacher variables, teacher education, and impacts on classroom practice. *ZDM–Mathematics Education*, *56*(1), 31-45. https://doi.org/10.1007/s11858-023-01531-1

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, *96*(1), 33-47. https://doi.org/10.1007/s10649-017-9764-5

Wilkerson, M. H. & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, *29*(1), 1-10. https://doi.org/10.1080/10508406.2019.1705664