PNAS LETTER

Not so binary or generalizable: Brain sex differences with artificial neural networks

Jeffrey W. Lockhart[a], Agustín Fuentes[b], Gina Rippon[c], and Lise Eliot[d,1]

From Victorian scientists measuring skull volumes to contemporary brain imagers using machine learning (ML), the widely publicized study by Ryali et al. (1) is merely the latest in a long history of attempts to identify the "real," or categorical brain differences between women and men. However, their actual findings are neither as novel nor definitive as stated.

As in dozens of prior studies, the "high" accuracy claimed for sorting brains into male or female categories did not generalize out-of-sample. Sex was misclassified for 18% of participants, matching the average performance over the last 12 y (2). This is neither new nor evidence of discontinuity in brain organization between sexes. To get around this problem, this study abandons out-of-sample validation, selects the measures showing the greatest within-sample differences, and then plots t-distributed stochastic neighbor embedding of only these deep neural network-derived measures. The resulting perfect separation between men and women creates the impression of clean, binary difference, which is unsupported by their classification results. With millions of parameters and just 1,500 observations, such separation is trivial and obscures the tremendous overlap in every discrete brain measure (3).

Ryali et al.'s use of the language of personalized medicine and "brain fingerprints" is also misleading. As the input features for "fingerprinting" were *designed* to discriminate between men and women, even the "individual level fingerprints" are not features unique to individuals, but "fingerprints" of their resemblance to one of two sex categories. Binarizing multiple continuous variables in this way is likelier to preclude than advance progress toward personalized or precision medicine (4).

Ryali et al. further claim to explain the algorithmic "black box," identifying *which* brain features are responsible for the purported between-sex differences. Such explanations are necessarily misrepresentations of the model: Either one needs a deep neural network with millions of parameters, high-order interactions, and nonlinear activations to model a phenomenon, or the phenomenon has a functional form and substantive interpretation human scientists can learn from, but not both (5). Moreover, the brain areas they identify (prefrontal and other heteromodal association areas) participate in so many different cognitive and emotional functions as to make these highly derived findings uninformative for practicing neuroscientists.

Inexplicably, the most clearly defined sex difference, total brain volume, was highlighted in their Introduction but never interrogated in their analysis. Similar to all nonreproductive organs, this measure varies continuously between men and women but differs substantially at the group level. Brain volume influences many features, including the ratios of white:gray matter and inter:intra-hemispheric connectivity (6). It also influences gyrification, diffusion tensor imaging measures (7), and network efficiency (8). Most notably, brain size drives ML sex classification accuracy (9), even using functional measures (10), both shown in the same data Ryali et al. (1) use. In other words, brain architecture varies with overall size, affecting both structure and functional connectivity, and strongly confounding apparent sex differences.

Given these many limitations and manipulations, it is a long stretch to conclude—as so many before have tried—that women and men think and feel using categorically different brains.

Author affiliations: [a]Department of Sociology, University of Chicago, Chicago, IL 60637; [b]Department of Anthropology, Princeton University, Princeton, NJ 08544; [c]Institute of Health and Neurodevelopment, College of Health and Life Sciences, Aston University, Birmingham B4 7ET, United Kingdom; and [d]Stanson Toshok Center for Brain Function and Repair, Chicago Medical School, Rosalind Franklin University, North Chicago, IL 60064

Author contributions: J.W.L. and L.E. analyzed data; and J.W.L., A.F., G.R., and L.E. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: lise.eliot@rosalindfranklin.edu.

Published January 2, 2025.

1. S. Ryali, Y. Zhang, C. de los Angeles, K. Supekar, V. Menon, Deep learning models reveal replicable, generalizable, and behaviorally relevant sex differences in human functional brain organization. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2310012121 (2024), 10.1073/pnas.2310012121.
2. J. Lockhart, Because the machine can discriminate: How machine learning serves and transforms biological explanations of human difference. *Big Data Soc.* **10**, 20539517231155060 (2023), 10.1177/20539517231155060.
3. D. Joel, Beyond the binary: Rethinking sex and the brain. *Neurosci. Biobehav. Rev.* **122**, 165–175 (2021), 10.1016/j.neubiorev.2020.11.018.
4. C. Sanchis-Segura, R. R. Wilcox, From means to meaning in the study of sex/gender differences and similarities. *Front. Neuroendocrinol.* **73**, 101133 (2024), 10.1016/j.yfrne.2024.101133.
5. C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **5**, 206–215 (2019), 10.1038/s42256-019-0048-x.
6. L. Eliot, A. Ahmed, H. Khan, J. Patel, Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neurosci. Biobehav. Rev.* **125**, 667–697 (2021), 10.1016/j.neubiorev.2021.02.026.
7. C. Torgerson *et al.*, Sex, gender diversity, and brain structure in early adolescence. *Hum. Brain Mapp.* **45**, e26671 (2024), 10.1002/hbm.26671.
8. C. Yan *et al.*, Sex- and brain size-related small-world structural cortical networks in young adults: A DTI tractography study. *Cereb. Cortex* **21**, 449–458 (2011), 10.1093/cercor/bhq111.
9. C. Sanchis-Segura, M. V. Ibañez-Gual, N. Aguirre, Á. J. Cruz-Gómez, C. Forn, Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction. *Sci. Rep.* **10**, 12953 (2020), 10.1038/s41598-020-69361-9.
10. C. Zhang, C. C. Dougherty, S. A. Baum, T. White, A. M. Michael, Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity. *Hum. Brain Mapp.* **39**, 1765–1776 (2018), 10.1002/hbm.23950.