Data is contingent, constructed, value-laden

Contingent → Datasets are contingent on the social conditions of creation

Constructed → Data is not objective; 'Ground truth' isn't truth

Value-laden → Datasets are shaped by patterns of inclusion and exclusion

---

Our data collection and data use practices should reflect this

Timnit, et al. (2018). Datasheets for datasets
Holland et al. (2018). The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards
Bender and Friedman (2018). Data Statements for NLP: Toward Mitigating System Bias and Enabling Better Science

H/t Emily Denton, Google

CHRISTOPHER NEWPORT UNIVERSITY
DEPARTMENT OF PHYSICS,
COMPUTER SCIENCE & ENGINEERING