



The Training Data Project

FOR IMMEDIATE RELEASE

April 3, 2024

NEW AI NONPROFIT SEES A DIFFERENT FUTURE FOR AI TRAINING DATA

Without good data, even the best machine-learning algorithms cannot do well. Training data is the key to AI, and trust is the key to good training data. The Training Data Project (TDP) is a new Nonprofit with a mission to promote trustworthy training data for AI initiatives across the public and private sectors. By advancing open-source tools, developing guidance and best practices, and establishing standards and interoperability for data labeling, the TDP aims to help create an AI-led future where Human-Machine Teaming (HMT), powered by high-quality training data, can serve the common good.

The TDP recognizes the explosive growth of AI in society today and moving into the future. Total Federal spending on AI has increased more than 2.5 times since 2017. In a recent Government Accountability Office (GAO) report, 20 agencies reported over 1,200 AI use cases- specific challenges or opportunities that AI may solve. With a \$1.8B request for AI in FY25, the US Defense Department is constantly exploring applications to support everyone from analysts to warfighters. The potential use of AI capabilities continues to grow.

While generative AI has accelerated the demand and importance of AI, it has also highlighted the importance of training data. Globally, more than half of organizations have increased their investment in generative AI programs over the past year. However, the rise of this transformational technology and of other AI capabilities presents a massive data challenge. Harnessing the benefits and mitigating the dangers of Generative AI and other AI capabilities begins with focusing on training data as AI fuel.

The data used to train AI must have integrity to ensure that people- the Humans in Human-machine-teaming- benefit from this technology as much as possible. To improve trust, high-quality training data is not just a goal, but a necessity. A recent GAO report- Artificial Intelligence in Natural Hazard Modeling- emphasized the difficulties of adoption and trust in AI by first responders in the Forest Service and at the Federal Emergency Management Agency. The roadblock- limited access to the data sources used to train the AI. Through open-source tools, industry standards, accessibility, and transparency, the TDP aims to enhance training data quality across the public and private sectors, and to expand access to trustworthy AI for all.

"We're in the midst of an AI revolution, one where machines that can learn, think, and even create are no longer just science fiction. It's an exciting time, but it's also a moment that calls for responsibility, especially when it comes to transparency and data foundations. Bad data is often worse than no data." - Dave Cook, Founder TDP

ABOUT THE NONPROFIT:

The Training Data Project was founded in 2023 by Dave Cook, who has over 25 years of experience working with AI/ML, advanced analytics, data engineering, and geospatial intelligence. Driven by a mission to build a more inclusive and diverse approach to AI initiatives, the TDP sees high-quality training data as both an opportunity and imperative for all.

Media Contact

press@trainingdatapoint.org

703-989-8409

www.trainingdatapoint.org