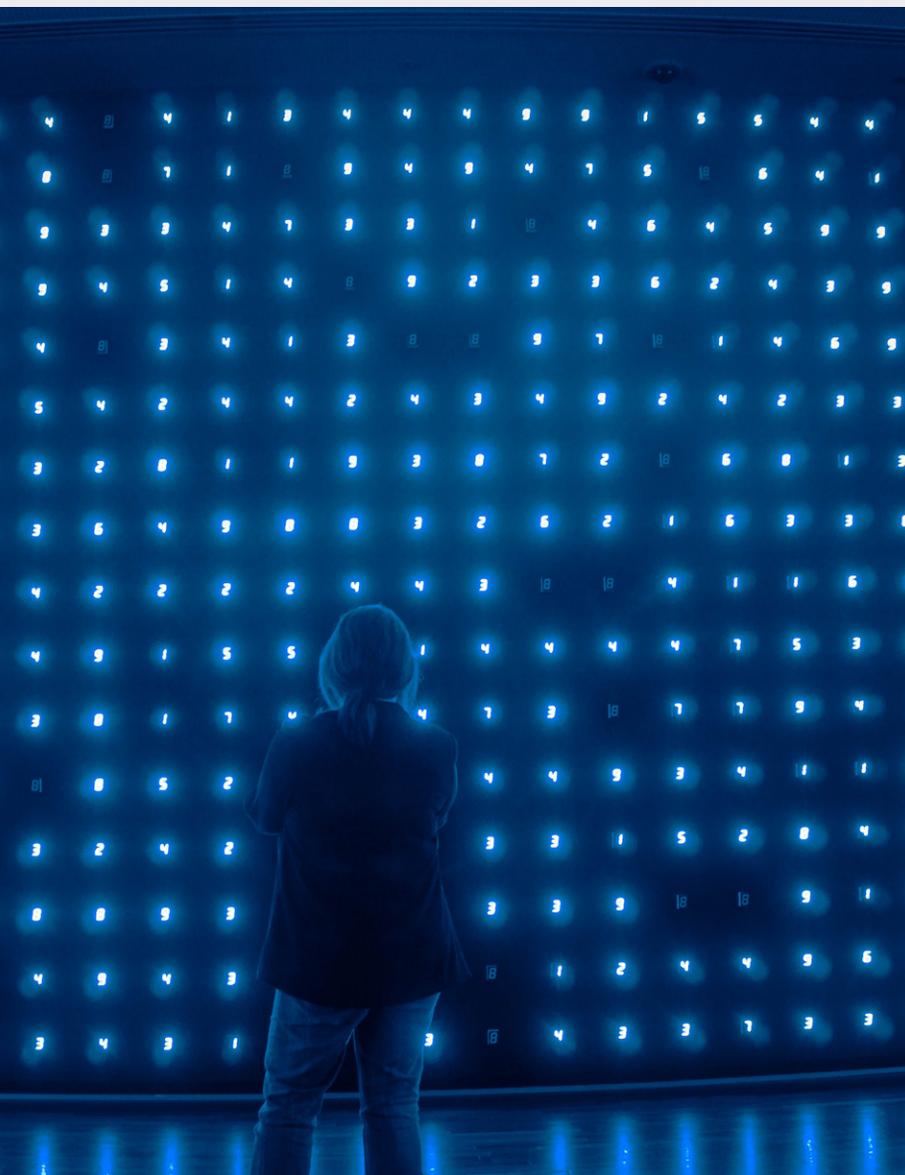


Training Data Considerations:

Producing quality training
data at scale is the biggest
obstacle to AI project success



Your Computer Vision (CV) and Natural Language Processing (NLP) systems

need to meet certain KPIs to be put into production. Your training data needs to be diverse and accurate to achieve those requirements.

Typical CV and NLP systems require enormous volumes of high-quality training data to train your model. Once a model is deployed, it needs ongoing training data for incremental improvements in confidence level and to manage model drift.

DDD // We are in the business of preparing machine learning data for corporate, government and academic AI projects.



Most data science teams produce their own training data for early project stages, where they test only one simple use case and have modest confidence level expectations.

Getting systems into production, however, involves testing many use cases, each of which requires the same or more training data. And production-level confidence demands enormous volumes of meticulously labeled data.

Human judgment is an essential element of data labeling, and preparing consistent, effective datasets requires many people, many or most of whom need specialized skills. You also need people with project management skills and with experience in managing quality and efficiency at scale.

DDD // We employ thousands of training data annotators and project specialists around the world.





Do you have the technology you need?

Labeling large volumes of training data efficiently requires tools that are a perfect match for your project's use cases.

DDD // We supplement our workforce with a full complement of CV and NLP task and workflow management technologies.

Workers and tools don't get you very far if you don't know how to use them. Designing tasks and workflows for training against use cases is a highly specialized skillset, quite distinct from core data science skills. Managing accuracy, consistency, and bias is essential.

DDD // We bring experienced solution engineers, data analysts and project managers to your project, specialists who are accustomed to making people and technology work together to produce huge volumes of high-quality training data.



Do you have the skills you need?



If you outsource, what should you look for in a partner?

Today there are numerous organizations that provide training data services. If you decide to rely on outside help, how do you choose among these services?

Labeling and annotation is one of the last stages of training data preparation. Many organizations find they don't have enough raw data to train their system, or they discover that their datasets need cleaning and curation before they can be annotated.

Most data labeling vendors assume that you have all the data you need, and that it is in usable form without bias or errors.

DDD // *We are your single source for high-quality CV and NLP data creation, collection, cleaning, curation and annotation at scale. We assign a single team of specialists that stays with you through every stage, providing you with continuity and growing efficiency and quality.*



Can they support every stage of your training data lifecycle?



**Do they
offload the
entire data
preparation
process?**

Lots of training data vendors give you API access to their annotation tools, with the expectation that you will provide both the workforce and project management resources. Others offer you only a crowdsourced workforce, and expect you to provide their workers with tooling and project leadership.

DDD// We remove the entire burden of training data preparation. You tell us what you need, provide us with your raw data if you have it, and we return high-quality annotated training data in the volumes your system requires.

Most data labeling vendors have built their businesses around a proprietary technology platform. This strategy leaves them, and you, vulnerable to the reality of machine learning systems: no two systems are the same, and no single platform can meet the volume, quality and application-specific needs of every project. Sometimes these partners will create workarounds for unique solutions that might not work when you need to open the tap again later.

DDD// We are labeling and annotation technology agnostic. We partner with all of the leading technology providers, which lets us pick the best tool or tools for your unique use cases. If you want us to work with your in-house technology or tools you've already selected, we're happy to do that too.



**Is their tooling
appropriate
for your use
cases?**



Can they satisfy your data privacy and security requirements?

Geographically concentrated data labeling services are often challenged to meet standards for security and privacy in other geographies. And unfamiliarity with a client's locality historically leads to significant training data bias.

DDD// Our workforce is distributed across the US, Asia and Africa. We support 20+ languages and cultures. We are ISO 27001, SOC 2, GDPR and HIPAA compliant.

Most training data services feel there are too many unknowns in machine learning modeling to guarantee quality levels that meet client expectations.

DDD// We have been doing this for a long time. We guarantee our work, and won't stop until we've met it.



Do they offer a guarantee of data quality?



Changing How
the World Works.

// digitaldividedata.com

One partner, one process. Guaranteed quality.

DEDICATED CLIENT SUCCESS TEAM

We assign a team of solution engineers, data analysts, and project managers that stay with you, no changing hands.

TRANSPARENT AND METRIC-BASED PERFORMANCE GUARANTEE

Establishing SMART goals and creating an open line of communication with all members is critical to ensuring the success of your project.

WORLDWIDE HUMAN INTELLIGENCE

Our trained annotators are full-time employees located in USA, Asia, and Africa and support 20+ languages.

CUSTOM-FIT TECHNOLOGY STACK

We pick the tools that best address your project. And if there's a tool you want us to use, we'll do that too.

DATA PRIVACY AND SECURITY

Best in Class Certifications: ISO 27001, SOC 2, GDPR, and HIPAA compliant.

SOCIAL IMPACT - NON PROFIT

DDD pioneered the impact sourcing model of offering employment in data preparation services to people from underserved communities. This socially responsible approach to performing work provides these individuals with a path to economic self-sufficiency.

DDD's proven work-study methodology trains, supports and employs talented individuals around the world while providing high quality, competitively priced ML data preparation services to our clients.

TRUSTED PARTNER IN MANAGED DATA PREPARATION SERVICES FOR MACHINE LEARNING & AI





Ready to learn how
DDD can help you
achieve AI success?

SCHEDULE A CONSULT TODAY:
ddd@digitaldividedata.com

