CONVENING BRIEF

Digital Tools and the Future of International Development

This brief provides a summary of the convening's key discussion points and offers four forward-looking recommendations for advancing the effective, ethical, and inclusive application of machine learning in international development programming.

SEPTEMBER 2019

About the Convening

In May 2019, the United States Agency for International Development (USAID)'s Global Development Lab with Duke University's DevLab@Duke and NORC at the University of Chicago convened nearly 90 academic experts, digital practitioners, development partners, and USAID colleagues in Washington, D.C. to discuss the **latest research on machine learning (ML) and its implications for international development**. The convening featured high-level panel discussions and 10-minute "lightning talks" that unpacked the potential of **three types of ML applications: machine vision, natural language processing, and network analysis**. Discussions centered around specific applications of ML in development contexts, the role of the private sector, and the path toward building a global ML workforce, all while underscoring critical ethical considerations.

WHAT IS MACHINE LEARNING?

Machine learning (ML) uses computers to identify patterns in existing data to make predictions.

ML models can be especially effective at finding complex, nonlinear relationships, and for making sense of unstructured visual, audio, and text data.

For shorthand, you could think of ML as **"data-driven predictions."**

Source: Amy Paul, Craig Jolley, and Aubra Anthony, Reflecting the Past, Shaping the Future: Making AI Work for International Development (Washington, D.C.: U.S. Agency for International Development, 2018).

Photo credits: Jack Devine/Research Technical Assistance Center

BY THE NUMBERS

05.17.19
CONVENING DATE

86
PARTICIPANTS

25
EXPERT PANELISTS

5
PRECONFERENCE PAPERS

9
LIGHTNING TALKS

4
RECOMMENDATIONS FOR ACTION

RESOURCES

- CONVENING AGENDA
- SPEAKER BIOS
- LIGHTNING TALK PRESENTATIONS
- FRAMING PAPERS

Machine Learning Has Promise for International Development

Machine vision (MV), natural language processing (NLP), and network analysis are promising for global development applications. Each of these ML subfields is already being used in ways that can support and accelerate development outcomes, but they also present challenges both for those who create them and those who benefit from their use.

MACHINE VISION

MV is the use of devices to receive and interpret an image in order to obtain information and/or control machines or processes. MV systems acquire data from images through remote sensors (usually cameras) and analyze and interpret that data by identifying patterns. This feature allows for rapid filtering of large image databases and can help identify signals that are less obvious.

Machine vision can help:

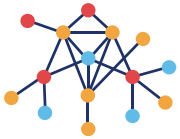
- Monitor and study crops, and track and identify diseases.
- Map changing landscapes, including from urbanization and natural disasters, as well as climate impacts on natural resources and man-made infrastructure.
- Allow both people and computers without a common language to share information.

NATURAL LANGUAGE PROCESSING

NLP draws on computer science, linguistics, statistics, and probability to help machines understand, process, and produce human language. NLP can enable computer systems to interpret speech and text as people naturally speak and type it—ultimately providing usable data and insights from unstructured text and/or audio.

NLP can help:

- Translate written language (think "Google Translate").
- Answer questions and respond to verbal commands (think Amazon's Alexa, and Apple's Siri).[1]
- Support mobile banking and payment systems in rural and developing regions with low literacy or high linguistic density (for example, a country or region where very different languages are natively spoken in a small area).[2]



NETWORK ANALYSIS

Network analysis examines the structure of relationships and communication between entities or people. Social media and other information and communication technologies have changed how people communicate, and all these activities leave "digital traces." ML for network analysis seeks to develop maps and create algorithms of activities to study information flow and social influence.

Network analysis can help:

- Map information flows about particular topics, such as childhood vaccinations or breastfeeding, to inform public health campaigns.
- Understand human responses to natural disasters and other emergencies, both for real-time decisionmaking during crisis events and for after-action reports and research.
- Recognize and counter misinformation and identify targeted disinformation campaigns and their sources.[3]

Key Considerations When Using Machine Learning for International Development

The goals of ML for international development should include ownership by low- and middle-income country (LMIC) institutions and individuals.

Societies do not become wealthier when they receive technologies; they become wealthier when they own and produce these technologies.[4] ML has already helped to improve the functioning of many development sectors, such as agriculture, education, and health. However, the technology is expensive, which makes initial investments in ML difficult for countries and organizations with limited resources. To maximize economic development, developing countries need to be producers of ML, not just consumers, which will require strengthening local systems to have the capacity and resources needed to develop their own ML solutions.

Education and training programs should focus on equipping LMIC individuals with relevant technical skills.

The skill set required to develop ML technologies is still globally scarce. The number of people in developing countries who have the technical knowledge and skills to create and deploy ML-based tools needs to expand. While universities are an obvious choice for such development, they often lack the appropriate resources, faculty, curriculum, computing capacity, and training data. Several programs are working to address these limitations, such as the "Western and Central Africa Higher Education Centers of Excellence Project" that is strengthening the capacities of participating African universities to deliver high-quality training and conduct applied research in areas such as data mining, with applications for ML.

To invest in local talent development, it is also essential to think beyond universities, and break away from traditional channels of knowledge transfer. For example, policymakers should be open to, and research should examine the promise of, training programs that are nimble and can exist outside of standard academic programs. Online education programs, both within and outside of university programs, will play a key role in satisfying the need for accessible and adaptive learning opportunities.

Initiatives to train local talent must also focus on talent retention. Universities and start-ups must compete for talent in a global labor market that includes large technology companies in the Global North. As LMIC universities increase capacity, many students may be drawn to jobs in the United States and Europe, and it is important to establish financially competitive and creatively rewarding local opportunities.

WHAT IS "TRAINING DATA"?

Training data is data used to develop an ML model. A learning algorithm will find patterns and relationships in training data and use them to define rules for new predictions.

Source: Amy Paul, Craig Jolley, and Aubra Anthony, Reflecting the Past, Shaping the Future: Making AI Work for International Development (Washington, D.C.: U.S. Agency for International Development, 2018).

High quality, context-specific data is needed to advance accurate and efficient ML tools for use in development.

All data is context specific. For example, a diagnostic tool programmed for use in Uganda might suggest that a child with a fever has malaria, given the prevalence of the disease in that region; however, the same tool programmed for the United States, might diagnose a different illness for similar symptoms. This reality underscores that local context and locally representative training data are crucial for ML to be useful for development.[5] Collecting and coding this training data is a significant opportunity for workers, firms, and governments in the developing world. Labeling training data, however, should not be the sole opportunity for developing country actors.

To maximize ML's possible utility for development, data collection and application should focus on developing countries' needs and resources. Local partners and staff should be engaged from the beginning of ML design and development to encourage ownership and ensure training data is carefully selected, context specific, and knowledgeably applied by those who understand the setting. The many different necessary inputs required for selecting, collecting, and applying data for the creation of ML tools suggest opportunities for partnerships between technical experts and context-specific experts.

NLP at present is best suited for digital text, which creates a barrier to leveraging NLP in communities whose language is predominantly spoken rather than written. There are over 7,000 languages in sub-Saharan Africa alone, and more than half of them have neither a standardized written format nor documented, teachable grammar—both prerequisites for most NLP technologies. Development actors must address this barrier to foster inclusive access to these digital tools.

Privacy and ethical use is paramount.

Concerns about privacy and the ethical use of information occur across all domains but are particularly prominent regarding networked data, where traditional notions of anonymity might not be enough to protect identities. The privacy and safety of the people who generate this data (the network members) should be of foremost concern. Even anonymized data in network graphs can reveal unintended patterns and associations, making it extremely important for those using these data to carefully consider how the data are created, interpreted, and communicated to decisionmakers and the public.

As with other research methods, ML can also propagate bias, errors, and misuse. For example, facial recognition, a form of MV, can be used for security systems and as an investigation tool. The same technology, however, can be used to profile populations based on race or gender, and enforce authoritarianism. And any bias in an ML system's data, algorithm, or application will proliferate as quickly the system operates.

When expanding the use and reach of ML, it is important to consider the ethics of the data that are collected, the models that are used, and the outcome of their use. The conversation about ML ethics should go beyond privacy concerns to include context, power, and accountability.[6]

- **Context:** Repurposing an algorithmic solution made for one context to be used in a different context does not always work.
- **Power:** Donor governments and organizations will always have asymmetrical power relationships with the developing world, and the cost of this unbalanced relationship can affect informed consent, fairness, and privacy. These hierarchies will be reinforced and heightened if ML models for use in developing country contexts are created and owned by those from more developed countries.
- **Accountability:** While many organizations, such as Google and IBM, have their own ML ethical guidelines, currently no universal, enforceable, ethical standard for generating and applying ML-backed tools exists.

Recommendations for Action

USAID and other development actors, including other donors, foundations, partner governments, and the private sector, should continue to support investment in creating and using ML for international development, but must simultaneously work to mitigate the negative impacts that ML can have on society—especially vulnerable or traditionally marginalized populations. Development actors should:

Continue to explore how ML can benefit communities around the world

- Investigate new uses of ML and approach existing problems from new perspectives.
- Use ML to improve development programs and integrate ML tools into the programmatic toolbox.
- Pilot-test how ML could be used in program monitoring, evaluation, and learning (MEL). As with all MEL, ML consideration should happen as early in the program cycle as possible.[7]

Invest in generating quality training data

- Partner with researchers and developers who are expanding the base of quality, labeled training data that represents a wide variety of contexts.
- Focus on generating the information needed to create tools to benefit LMICs and where data is especially scarce.
- Support projects to map the languages of developing regions and support collaborations with technology firms that will create NLP for local languages.

Promote equitable, fair, and sustainable progress through development and use of ML technology.

- Integrate ethical considerations into the main conversation about ML and development.
- Ensure that capacity building is done in a way that promotes attention to the unique concerns about ethics, power balances, and accountability that might arise or be exaggerated in LMIC contexts.
- Work across sectors to establish and adhere to high ethical standards for ML development and use.

Promote ownership of and capacity for ML in developing countries

- Train local talent in ML fields and create incentives to encourage talent retention.
- Modernize computer science training programs (including updating high-performance computing infrastructure) at developing country universities and support non-university and non-accredited paths to technology careers.
- Support policies that make ML technology and data accessible to nongovernmental organizations and other institutions in lower-resource settings.
- Educate policymakers about ML so that they can make informed decisions about critical aspects of ML technology.

Acknowledgments

This brief provides key insights and recommendations from the May 17, 2019 USAID convening titled "Digital Tools and the Future of International Development." It draws on information and ideas featured in the five pre-conference framing papers, the expert panels and lightning talks, and a summary paper produced by the DevLab@Duke.

This document was produced by Population Reference Bureau (PRB) under the Research Technical Assistance Center (RTAC). RTAC is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the terms of contract no. 7200AA18C00057. The contents of this document are the sole responsibility of RTAC and NORC at the University of Chicago, and do not necessarily reflect the views or recommendations of USAID or the United States Government. PRB would like to recognize and thank the following USAID colleagues for their careful review of and helpful contributions to this document: Aubra Anthony, Shachee Doshi, Ticora Jones, Megan Miller, Amy Paul, Peter Richards, and Brent Wells.

References

1 Nicholas Monaco and Samuel Woolley, "Natural Language Processing and Global Development: A Future-Focused Primer," framing paper for Convening, University of Texas at Austin, May 17, 2019.

2 Monaco and Woolley, "Natural Language Processing and Global Development."

3 Kate Starbird, "Analyzing 'Social' Data to Inform Crisis and Humanitarian Response in a Networked Age," framing paper for Convening, University of Washington, May 17, 2019.

4 Kentaro Toyama, Machine Vision Lightning Talks: Algorithmic Audits, University of Michigan School of Information, May 17, 2019.

5 Erik Wibbels, "Key Takeaways, Global Development Lab's Machine Learning Conference," conference report, DevLab@Duke, June 2019.

6 Mark Latonero, "Ethical Considerations for AI Applications in International Development," framing paper for Convening, The Carr Center for Human Rights, Harvard Kennedy School, May 17, 2019.

7 Wibbels, "Key Takeaways, Global Development Lab's Machine Learning Conference."

Framing Papers

The series of papers produced in conjunction with the convening include:

Mark Latonero, "Ethical Considerations for AI Applications in International Development," framing paper for Convening, The Carr Center for Human Rights, Harvard Kennedy School, May 17, 2019.

Nicholas Monaco and Samuel Woolley, "Natural Language Processing and Global Development: A Future-Focused Primer," framing paper for Convening, University of Texas at Austin, May 17, 2019.

Mehdi Oulmakki, "Creating a Conducive Environment for AI and ML in the Developing World," framing paper for Convening, African Leadership University, Rwanda, May 17, 2019.

Kate Starbird, "Analyzing 'Social' Data to Inform Crisis and Humanitarian Response in a Networked Age," framing paper for Convening, University of Washington, May 17, 2019.

Raju Vatsavai, "Machine Vision," framing paper for Convening, North Carolina State University, May 17, 2019.

Erik Wibbels, "Key Takeaways, Global Development Lab's Machine Learning Conference," conference report, DevLab@Duke, June 2019.

SUGGESTED CITATION

Rachel Yavinsky and Carolyn Rodehau. 2019. Digital Tools and the Future of International Development. Convening Brief. Research Technical Assistance Center: Washington, D.C.

USAID FROM THE AMERICAN PEOPLE

PRB

NORC at the UNIVERSITY of CHICAGO

DEVLAB @ Duke