Building Farmer Resilience

NEW INSIGHTS FOR SCALABILITY FROM 25 YEARS OF INDEX INSURANCE

Building Farmer Resilience

Through Responsible Insurance.

A SIPA Capstone Final Report

PREPARED BY
Columbia SIPA · Capstone Team · Spring 2026

CLIENT · KEY CONTRIBUTOR
Gates Foundation · Qatar Insurance Group

COLUMBIA | SIPA
School of International and Public Affairs

Gates Foundation

QIC

A SIPA CAPSTONE FINAL REPORT

Building Farmer Resilience

Through Responsible Insurance.

PREPARED BY

Carlos Ivan Aguilar Rocha, Dean Dai, Irra Zhang, Jenna Yun,
Lara Geiger, Mauricio Trujillo, and Yiran Chen.

WITH INPUT & GUIDANCE FROM

Panos Varangis, Columbia SIPA.
Daniel Osgood, Columbia Climate School.
Max Mauerman, Columbia Climate School.

DATE

May 2026

CONTACT

Carlos Aguilar	cia2117@columbia.edu
Jenna Yun	jjy2150@columbia.edu
Yiran Chen	yc4639@columbia.edu
Lara Geiger	lg3193@columbia.edu
Mauricio Trujillo	mauricio.t@columbia.edu
Dean Dai	hd2581@columbia.edu
Irra Zhang	gz2383@columbia.edu

COLUMBIA | SIPA
School of International and Public Affairs

Gates Foundation

QIC

Table of Contents.

I. Executive Summary.

Extreme weather linked to climate change is increasing the exposure of smallholder farmers to shocks that threaten their yields, income, and household security. For farmers with limited savings and few formal safety nets, a single drought, flood, or other climate-related event disrupts their and their families' sustenance. Over the past 25 years, index insurance has been promoted as a scalable financial tool to help smallholder farmers manage these risks and strengthen resilience in the face of growing climate uncertainty.

The promise of index insurance has not been fully realized. This report argues that the central problem is not simply limited distribution, but limited verification of whether products actually work for farmers. Too often, products are scaled before sponsors know whether they solve a real problem, provide meaningful payouts in the years that matter most, or build trust between farmers and providers. When insurance fails at these critical moments, farmers do not experience it as protection, but as another costly and unreliable intervention. The challenge is not only technical or financial, it is human: whether these products genuinely improve the lives and security of the people they are meant to serve.

This report identifies two recurring barriers to responsible scale. First, products are often deployed before there is sufficient evidence that they create real value for farmers. Second, current evaluation practices frequently fail to capture what matters most to farmers: whether the product pays meaningfully in a bad year. These two barriers affect the programs move from pilots to large-scale rollout, communication and feedback loops often weaken, leaving farmers with little voice in product design, little understanding of coverage, and little recourse when outcomes do not match expectations.

These findings are based on stakeholder interviews across 18 institutions and case-based analysis, including examples from Zambia and Rwanda. The report also draws on verification exercises that compare yield-based bad-year rankings, farmer recollection of bad years, and satellite-based signals, before testing whether index payouts align with farmer expectations. Together, these sources show that product failure is not only a matter of technical basis risk, but also of weak verification, weak communication, and weak farmer-centered design. This cross-source verification is then used to test whether index-based payouts align with farmer expectations, helping distinguish technical mismatch from broader failures of communication and product design.

One of the clearest insights from the report is that many products are still judged by average portfolio indicators, such as loss ratios, even though farmers remember only one thing: what happened in the year of the real crisis. A product may appear acceptable on average while still failing in the single year that matters most. When catastrophic shocks are followed by small, delayed, or poorly explained payouts, trust breaks down and renewal becomes difficult. The report therefore argues that bad-year performance should be treated as a basic screening question before deployment, not as an issue discovered after scale has already been attempted.

Another important insight is that bundling should not be treated as inherently positive. Bundling can support scale, but only when it clearly improves the value of the overall package for farmers. Interview examples suggest that insurance can add value when it complements existing services or reduces residual risk, as in the case of livestock vaccination bundled with insurance. By contrast, adding a weak insurance product to an input or service package may increase enrollment without improving resilience. The relevant question is therefore not whether insurance is bundled, but whether the bundle is better with insurance than without it.

The report also emphasizes that the farmer is too often the absent stakeholder in product design and monitoring. In many large-scale programs, farmers have little input into how bad years are defined, how products are explained, or how payouts are interpreted after aftershocks occur. In this context, technology and artificial intelligence may support analytics, communication, and due diligence, but they cannot substitute for direct farmer feedback or independent verification of product quality. Responsible scale depends not only on smarter tools, but also on stronger attention to the lived experience of the people these products are meant to protect. In some cases, farmer input does more than improve legitimacy: it can also improve technical design, as evidence from Zambia suggests that farmer rankings of historical bad years helped produce indices that outperformed expert-built alternatives.

For the Gates Foundation, the main implication is that support for scale should be conditioned on stronger ex ante verification. Before backing large-scale deployment, sponsors should ask a small set of practical questions.

In short, the challenge is expanding index insurance while ensuring that scale is useful, trusted, and grounded in farmer realities. This is achieved with verification, clearer value for farmers, and stronger feedback mechanisms from the start.

II. Background & Proposed Objectives.

There is a consistent pattern across countries and stakeholder interviews that pilot programs for index insurance tend to function relatively well, largely because they are small-scale and allow for close interaction with farmers, but these same programs break down when scaled to national or large regional levels, where complexity increases and communication channels weaken. The fundamental issue is not that the concept of index insurance is flawed, but that the mechanisms that make pilots effective, especially feedback loops and farmer engagement, are not translated into large-scale implementation, leading to systemic failure. In Zambia, when a pilot insurance program was scaled to the national level, approximately 75 percent of farmers surveyed did not even know that they had insurance coverage, which highlights a severe breakdown in communication and awareness at scale.

For over 25 years, agricultural index insurance has failed to scale. After conducting twenty-six interviews, information, support, and communication emerged as the different pain points in the value chain of index insurance. There has been one important overseen stakeholder: the farmer. In order to address these pain points we have narrowed down to following constraints which need to be targeted by a foundation or investor alike.

1. Products are deployed without verifying whether they solve a real problem

The key issue here is whether the insurance product being deployed actually provides value to farmers and addresses the risks they face, rather than simply being distributed at scale. Many insurance products are introduced without systematic verification that they solve a real problem for farmers, meaning there is little evidence that they improve outcomes or provide meaningful financial protection. Products are often bundled with other services, such as agricultural inputs or credit programs, and while this bundling can increase adoption rates, it can also mask whether the insurance itself has intrinsic value, since uptake may be driven by the bundle rather than the insurance. There is often no follow-up analysis to determine whether the factors that drove positive outcomes in pilot programs, such as localized design or farmer engagement, are still present at scale, leading to a disconnect between initial success and long-term performance.

Pilot programs typically involve direct communication with farmers, iterative feedback loops, and localized adjustments, allowing designers to refine the product based on real-world conditions and farmer experiences. Because pilots operate on a smaller scale, it is easier to test assumptions, identify problems early, and make corrections, which increases the likelihood of delivering a product that provides real value.

At scale, programs are often implemented through top-down national policies or mandates, which prioritize coverage and efficiency over customization and feedback. This shift results in reduced communication with farmers, limited understanding of local conditions, and a lack of mechanisms to verify whether the product remains relevant or effective, ultimately undermining its impact.

2. Products fail to pay meaningfully in the years that matter

The central issue is whether insurance payouts occur in the specific years when farmers experience significant losses, as these are the moments when insurance is most needed. This problem is fundamentally about trust, as farmers' willingness to adopt and continue using insurance depends on their belief that the product will deliver value when it matters most. It encompasses multiple factors, including whether the product is designed correctly, priced affordably, and capable of delivering payouts that are both timely and sufficient to offset losses. There is also a perception risk, where farmers may believe that insurance providers are collecting premiums without delivering meaningful benefits, especially if payouts are rare or negligible.

Even if average performance metrics such as loss ratios appear acceptable, a failure to deliver payouts during critical years can completely erode trust in the product, making it effectively useless from the farmer's perspective. As programs scale, product design increasingly becomes centralized and top-down, often driven by national governments or large institutions rather than by local input. This shift leads to a situation where farmers are not consulted in the design process, resulting in products that may not align with their needs or experiences.

The lack of farmer involvement contributes to poorly designed products, low awareness among beneficiaries, and widespread mistrust, all of which limit the effectiveness of the insurance. Typically something important is "lost" in the transition from pilot to scale, specifically the ability to incorporate feedback and adapt to local conditions, which is critical for success.

III. Methodology.

To find out the real concerns in the insurance ecosystem, we addressed our questions by doing semi-structured interviews with stakeholders across the network with people inside insurance companies, multilateral organizations and governments.

Semi-structured interviews because we wanted to find out pain points by creating a well made and previously discussed questionnaire to address specific issues, while leaving space in each interview to open the discussion to questions that we were probably not even thinking before.

This uncovers two sets of discoveries:

a. **What we know we don’t know:** pain points we have targeted and think are present, but don´t know how the clients (foundations and insurers) are really affected by it. This we can find with the structured questionnaire.

b. **What we don´t know we don´t know:** pain points that we don’t even know exist, and the client is probably also not aware of, but we can find out via an open conversation and coming up with questions as conversations develop.

To analyze these interviews we found out different patterns and nodes that encompassed the conversations into categories that we could then draw out to gather conclusions. After careful consideration, the categories we found out were the most common in all the interviews as things that were impeding scale in index insurance are:

I. Products should be simple and focused

Instruments should depend on one or few specific indexes like rainfall or extreme events. This would ensure there is enough data for the creation of a good product, and would make it simple for farmers to understand the payouts and overall assessment of the situation.

II. Products should be affordable and automated

For the index insurance ecosystem to work, it’s very important to use methods that make the product viable and be able to reach the community. For this, technology such as cellphones, text messages, digital wallets, etc, would ensure there is an open path of communication with the farmers, and they are informed about the product. It’s also a way to get information and ensure their interaction with the product.

III. Bundling with other products is key

As index insurance is in itself a difficult product to explain to farmers and it could be difficult for farmers to take the step to invest in their insurance as it could be seen as something that is not urgent, it is important to bundle insurance with other products such as agricultural inputs like fertilizer, technology products, other financial products or other insurance. This would prove important as farmers buy a package and insurance is seen as an add on product that could help them in rough times. It is key to scaling the product as this would make it more attractive for consumers. As said before, bundling in itself is not the answer, as making it with bad products would mean an overall bad product in itself, but it is key to bundle index insurance with other well engineered products.

IV. Profitability shouldn´t be an instant goal

As index insurance takes time to develop and covers a geographically dispersed population as small farmers, while important in the long term, profitability shouldn’t be an instant goal of insurance companies. Long term goals are more important, and companies who want to have an impact in this area should focus on creating long term capabilities for a future productive market.

V. Government should go beyond subsidies

Subsidies are important and should not be abolished, yet the role of the government should be active and purposive for index insurance to thrive. This is to create a thriving environment in which farmers have sufficient information and capacities to learn about the product they own and how to interact with the payout system and when to have the best benefits.

VI. Data is central

To build affordable products, insurers and reinsurers need enough data to assess correctly the risk of a specific area. Satellite data regarding weather and land changes, and extreme events should encompass at least 15 years in order to assess an area correctly and build a comprehensive product for farmers. This should also be an addition to on ground data that the farmers have to make a comprehensive assessment and create products that have both quantitative and qualitative insights.

VII. Scale defines viability

For long term profitability, scale should be defined as the amount of farmers index insurance can benefit. The higher the number of farms covered means there is a higher possibility of creating products that are grounded and affordable.

IV. New Insights for Scalability.

1. Can Farmers Better Design Indices Than Experts?

As index insurance programs have grown in scale, the process of product design has shifted from bottom-up to top-down. In earlier pilots, designers often worked closely with local communities to understand which years were genuinely bad and why. At scale, this practice has been replaced by centralized index construction, where insurance companies and technical consultants build indices using satellite data and actuarial models, with little or no direct input from the farmers who will ultimately purchase and rely on the product.

This shift has technical consequences. When indices are built without farmer input, they don't reflect what farmers experience on the ground. A year that appears average in satellite-derived rainfall data may have been catastrophic for farmers due to localized conditions, pest pressures, or timing mismatches between rainfall and crop growth stages.

A year that triggers a payout based on an index threshold may not correspond to any meaningful loss in farmer experience. The result is a product that is technically coherent but practically misaligned with the realities it is meant to address.[1]

Evidence from Zambia suggests that farmer input can meaningfully improve index performance. The Zambia Ministry of Agriculture conducted focus groups across approximately 1,000 villages, asking farmers to rank historical years by the severity of agricultural losses they experienced. When these farmer-reported rankings were used to calibrate satellite-based indices, the resulting products outperformed indices constructed through conventional expert-driven methods alone. This finding is significant as it demonstrates that farmer knowledge of local conditions, when systematically collected, can add information that satellite data and actuarial modeling do not capture on their own.

The policy implication is straightforward. Before any index is finalized and deployed, sponsors and product designers should reintroduce a structured process for collecting farmer input on historical bad years. This does not require large or expensive surveys. As discussed in the section on power calculations, relatively small and well-designed sampling exercises can yield statistically credible results at modest cost. The relevant question is not whether farmer input is worth collecting, but whether the verification infrastructure exists to collect it before scale is attempted rather than after problems have already emerged.

2. Does Anyone Check What Happens in a Bad Year?

2.1 Bad Year Explanation

Current monitoring practices in index insurance tend to focus on portfolio-level financial indicators. Investors and program managers routinely track loss ratios, premium volumes, and renewal rates, all of which are meaningful measures of financial sustainability. However, these metrics do not answer the question that matters most to farmers: did the product pay out in the year when it was genuinely needed? A product can appear financially viable on average while still failing in the single season that defines whether a farmer can recover from a shock or not. Profitability at the portfolio level does not guarantee beneficial impact at the farm level, and treating these as equivalent is one of the most consequential gaps in current practice.

This gap in monitoring is not incidental. As programs scale, evaluation tends to become more centralized and more reliant on aggregate data that is easier to collect remotely. The result is a systematic blind spot: the years that matter most to farmers, the catastrophic or near-catastrophic seasons, are precisely the years least likely to be examined in detail

[1] World Food Programme, n.d.

by investors or program managers. Three checks are routinely missing from insurer-level monitoring: whether payouts occur in years that farmers actually experienced as bad; whether the product delivered real value relative to the losses farmers faced; and whether farmers understood what they received and why.

The experience of Zambia in 2022 illustrates this gap. Following a substantial drought, farmers enrolled in a parametric insurance program received payouts of approximately one-third of the premiums they had paid. In response to the performance of the program during this period, the government reconsidered the structure of the parametric insurance product. Whether investors were aware of the reasons behind this decision, and whether the outcome was systematically communicated through reporting channels, remains an open question. What the episode highlights is the difference between a program that appears to function under normal conditions and one that can be verified to deliver value when conditions deteriorate. Without structured bad-year checks built into monitoring frameworks, this difference may not become visible until trust has already broken down.

The implication is not that investors are acting in bad faith, but that the current incentive structure and monitoring architecture do not require bad-year performance to be measured. Investors have a clear mandate to track financial sustainability. The question of who is responsible for checking whether the product works for farmers in the years that matter most has no clear answer. Addressing this gap does not require large or expensive new systems. As the following sections discuss, simple verification exercises using farmer-reported bad-year rankings can be conducted at modest cost and integrated into existing monitoring and evaluation frameworks. The barrier is not feasibility but prioritization.

2.2 Index & Farmer Expectations are Mismatched

Across cases, payouts do not consistently align with what farmers experience as loss: the issue is whether the index triggers in the years that matter to farmers. Evidence from Ethiopia, based on prior alignment analysis, shows that the misalignment is both frequent and material. Across 1983 to 2020, only about 50 percent of payout years align with farmer expectations, and in downside cases farmers expect payouts averaging 32% of maximum liability more than what the index delivers in bad years.[2]

[2]Mean gap between farmer-expected and index-actual payouts (Ethiopia, Genete, 1983–2020). SIPA Capstone, 2025.

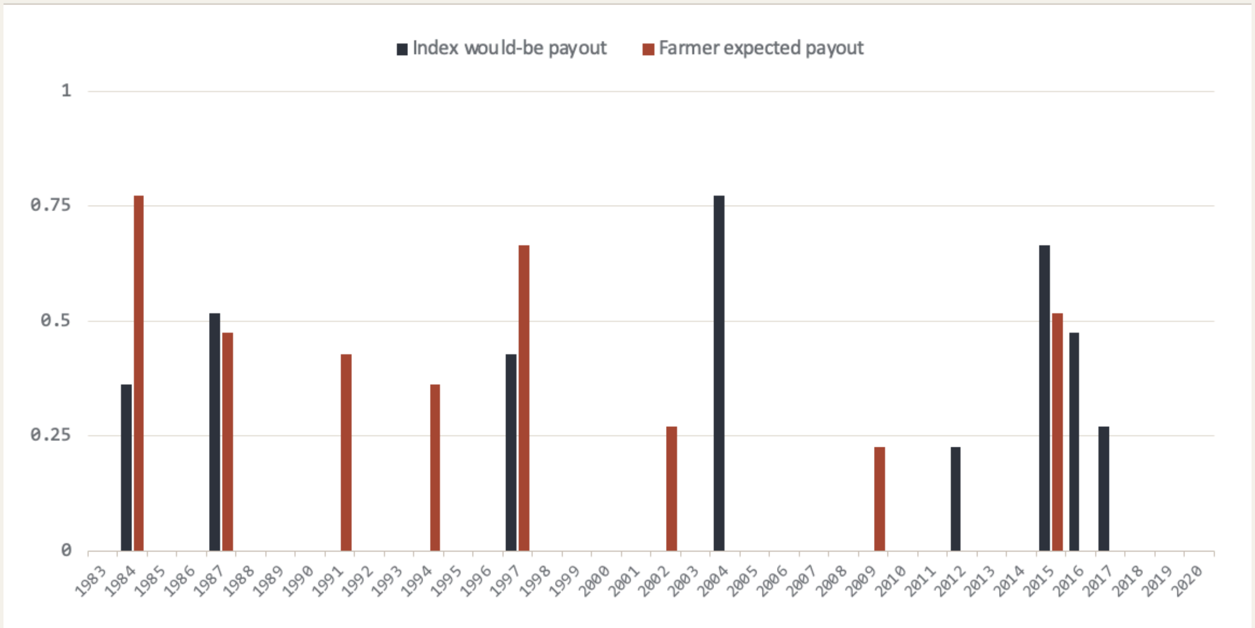

Figure 1: Year-by-year comparison of index payouts vs. farmer-reported bad years, Ethiopia (1983–2020).

A closer look at year-level comparisons highlights two distinct patterns of misalignment. In Ethiopia, several years show clear farmer-reported losses with no corresponding index payout, including 1991, 1994, 2002, and 2009. In these years, farmers ranked conditions as severe, yet the index did not trigger. In contrast, other years show the opposite pattern. The index triggered payouts in years such as 2004, 2012, 2016, and 2017, when farmers did not identify these as bad years. There are also a small number of years where alignment does occur. For example, 1987 and 1997 show both farmer-reported losses and index payouts in the same period. These comparisons indicate a recurring pattern in which payout timing does not consistently correspond to farmer-experienced losses. The gap reflects a structural limitation of index design. Indices rely on proxy variables such as rainfall or vegetation, while farmers assess outcomes based on realized yield and income.

The verification approach developed in prior Capstone work offers a practical way to diagnose this issue by comparing index trigger years with farmer recall and ranking data across time. It allows stakeholders to move beyond average loss ratios and assess whether payouts occur in the years that actually matter. In cases where systematic gaps persist, one potential response is the introduction of a complementary mechanism to address residual mismatch. Prior Capstone work has explored the idea of a complementary fund designed to correct for cases where index insurance fails to trigger despite farmer-reported losses.[3] Building on the verification and gap analysis, such a mechanism could be used to estimate the scale of resources required to cover these “missed” years. This extends the

[3]SIPA Capstone, 2025.

analytical framework by linking identified mismatch to a potential corrective layer. This approach, developed in prior Capstone work as a simulation framework, has not yet been implemented but offers a quantifiable pathway for addressing residual mismatch.

2.3 Different Levels of Severity Alignment Check

Not all bad years are equal, yet most index products treat them as binary events. A payout is triggered or it is not, with limited ability to reflect differences in the severity of loss. Farmer recall data suggests that losses are experienced along a spectrum. Some years are catastrophic, while others are moderate but still materially harmful. Ranking exercises from Ethiopia and Rwanda show that farmers consistently differentiate between levels of severity, while index payouts do not always reflect that gradient.

For example, in Rwanda, where the product is based on area yield insurance, 2019 is consistently identified as a severe year across three data sources: farmer recall, yield data, and satellite indicators. To enable this comparison across sources with different time spans and measurement methods, each source's bad year ranking is converted into a severity score by dividing the year's rank by the total number of years available in that source. For instance, a rank of 1 out of 8 years in the yield data is equal to a severity score of 12.5%, representing the probability that a year this bad would occur. This normalization places all three sources on a common scale despite covering different time periods. In Rwanda, 2019 is ranked as the worst year across all three sources, with closely aligned severity scores confirming it as a genuinely severe year by every available measure. The goal is for the difference between scores to approach zero. A large gap signals that data sources are capturing different realities, which is itself a warning sign about index reliability. While in other years, we can find that rankings diverge across data sources, which indicates that severity is not captured consistently. This severity lens provides a more precise way to assess product performance. Instead of focusing only on whether a payout occurred, the question becomes whether payout magnitude corresponds to the level of loss. Current evaluation approaches, which emphasize average loss ratios and profitability, do not capture this dimension.

Importantly, this type of check does not require extensive data. Even with a limited number of years, stakeholders can compare the worst-ranked years across available data sources and assess whether the index captures the most critical events. In practice, this can be applied even in data-constrained settings. For example, the Rwanda case relies on only eight years of area yield data, yet still allows for a meaningful comparison across sources.[4] Taken together, mismatch in timing and mismatch in severity point to the same underlying issue. Product performance is evaluated at an aggregate level, while farmer experience is shaped by specific years and levels of loss. Any verification framework

[4] SIPA Capstone, 2025.

that aims to assess whether index products deliver meaningful protection must therefore address both dimensions, even where data availability is limited.

2.4 Power Calculations Make Verification Feasible and Affordable

A major reason why bad-year verification is often not conducted systematically is the assumption that credible validation requires a large and expensive field exercise. Power calculations suggest a more practical approach. They define the problem as one of sample design, namely, how many wards and villages must be observed to estimate farmer-reported historical climate risk with acceptable precision. This reframing matters for scalability because it turns verification into a manageable operational task.

This design problem is especially important in rural settings where data is often limited. In such contexts, statistical precision depends on how those observations are distributed across sampling units. As both the Nigeria note[5] and J-PAL[6] emphasize, higher intra-cluster correlation reduces effective sample size. As a result, a larger number of observations is required to achieve a given level of statistical power. Verification therefore cannot be planned simply by collecting more data. What matters is the balance between breadth across administrative areas and depth within them. Power calculations are useful because they make that trade-off explicit and allow sample design to be guided by precision requirements.

The Nigeria sampling note shows how this logic can be applied under realistic field constraints. Because farmer-level data from the target area were not yet available, the analysis used comparable survey evidence from Zambia to estimate the likely structure of clustered responses. It then simulated alternative sampling combinations and found that a standard 80 percent statistical power could be achieved with only roughly 20 to 25 wards and 3 to 4 villages per ward, implying approximately 60 to 100 focus groups. The note also emphasizes that village selection should be geographically representative and informed by local knowledge.

The significance of this result is not only methodological but institutional. Once minimum sample requirements are specified in advance, bad-year verification becomes easier to embed in routine monitoring systems and to budget for across programs and contexts.

3. Bundling Must Add Value, Not Just Volume

Based on the interviews, it was discovered that bundling has become one of the most popular strategies for expanding index insurance options.

[5] IRI, n.d.
[6] J-PAL, n.d.

When smallholder farmers are slow to adopt, expensive premiums, and the challenge of reaching scattered rural communities can be daunting, bundling offers a practical solution. By attaching insurance to a product or service farmers already use and need, we make it easier for them to adopt and benefit.

This approach to bundling seems to address two key challenges that arose in the stakeholder interviews: the limited willingness of people to pay for insurance on its own and the high costs of selling directly to consumers. However, it's also important to recognize that bundling doesn't necessarily lead to better results for farmers. It may increase sales volume, but not necessarily the value they receive.

This suggests that the main goal in designing and expanding the product is to ensure the bundle offers a clear improvement with insurance compared to without.

Three bundling structures emerge from the interviews, each representing a distinct logic:

Income protection

The first model, associated with MIC Global, considers insurance as part of a comprehensive income protection package rather than a standalone indexed product. It combines agricultural coverage with health, hospitalization, and income protection, enabling households to receive frequent small payouts from health claims while reserving the agricultural part for catastrophic seasons. This setup addresses a key issue that standalone parametric insurance cannot: farmers desire regular, visible benefits to justify their premiums, but a single-peril agricultural policy that pays often tends to drain the fund, leaving little for genuine crises. By spreading risk across perils with varying likelihoods and severities, the bundle ensures consistent payouts that build trust while still offering catastrophic coverage. The core shift in framing is from risk transfer to income protection, aligning with farmers' actual needs: maintaining household income amidst multiple, overlapping shocks rather than compensation for a single, statistically defined event.

Risk reduction

The second structure, illustrated by Agricore/InsureCow in Bangladesh, combines risk reduction with insurance rather than treating them as separate interventions. Vaccination is bundled with livestock cover: the vaccination lowers the ex ante probability of cattle death, and the insurance covers the residual risk. Two reinforcing effects emerge: First, because the expected loss is lower once vaccination is in place, the premium priced against residual risk can be lower and more credible to the farmer. Second, the vaccination itself provides a tangible, up-front benefit that the farmer can see and value before any insured event occurs. In the provider's reporting, the combined intervention was associated with a substantial decline in cattle mortality and a corresponding reduction in borrowing costs, as financial institutions began to recognize previously uninsured animals as collateralizable assets. The bundle succeeds because each component strengthens the economic case for the other. This is the relevant test. When a bundle only works because one component subsidizes or conceals another, insurance has been added for distribution rather than for value.

Credit or input financing

The third approach involves combining index insurance with seasonal credit or input financing. The main idea is that insurance reduces the risk for lenders, enabling banks and input suppliers to provide credit they might not offer otherwise. Although this reasoning seems convincing, interviews indicate it should not be taken for granted. Two key questions should be carefully considered before expanding credit bundling. First, does adding the insurance premium make borrowing still affordable for farmers? If the premium raises the total loan cost beyond what farmers can afford to repay, the bundle benefits lenders by reducing their risk but leaves farmers worse off — an outcome opposite to the goal. Second, is weather risk genuinely the main obstacle for lenders? Banks and suppliers might hesitate to lend due to other issues, such as untraceable collateral, lack of financial records, high transaction costs per borrower, or uncertainties in enforcing repayment. If these are the primary constraints, attaching weather insurance to credit doesn't solve the core problem and may only increase borrowing costs. When credit bundling works well, it usually does so because it also addresses issues like asset traceability, repayment discipline, or input delivery, rather than because the insurance itself is the key factor.

These three frameworks offer a more disciplined approach to assessing bundling proposals. Instead of questioning whether a product is bundled, sponsors and investors should consider the specific role the insurance plays within that bundle. In the income protection model, insurance addresses a coverage gap not covered by other perils and maintains the frequent payout pattern farmers use to determine the product's authenticity. In the risk reduction model, it covers residual risks that preventive measures cannot eliminate, with its cost based on the lower expected loss due to those measures. The credit model's purpose is less clear, often assumed rather than proven, because the same bundle could potentially increase lending volumes through traceability and delivery improvements alone. This perspective does not suggest abandoning credit bundling but rather treating it as a design choice that must be justified beforehand, not as an automatic path to scale. When insurance genuinely enhances the bundle, adoption and value tend to increase together. If not, bundling may boost enrollment figures for reporting reasons without actually improving farmers' resilience the product aims to protect.

4. AI Is a Tool, Not a Substitute for Farmer Insurance Decisions

AI can strengthen agricultural insurance systems, but it cannot replace the fundamental need to ensure that insurance products deliver real value to farmers. At its core, AI is most useful when applied to two critical questions: (1) are we solving a real problem? (2) do payouts deliver when it matters the most?

To deliver value, AI can support breakeven and scenario analysis for cooperatives and providers. Many index insurance products are bundled with inputs or subsidies, making it difficult to determine whether insurance delivers standalone value. AI can stress-test product assumptions by modeling farmer outcomes across different weather and yield scenarios, identifying whether insurance meaningfully improves resilience or simply adds cost. It can also flag inconsistencies between how products are marketed and how they actually perform, helping prevent misaligned incentives at the design stage.

On the trust side, AI can help evaluate whether payouts occur when farmers actually experience losses. This is especially important in index-based insurance, where payouts are triggered by proxies such as rainfall rather than observed losses. A central challenge is basis risk, the mismatch between index triggers and real-world outcomes, which can undermine farmer trust even when products are actuarially sound.

4.1 Example Applications: Ethiopia and Kenya

The experiences of Ethiopia and Kenya illustrate these challenges clearly. The use of AI technology can play a key role in detecting inconsistencies in the connection between index triggers and farmer losses. Indeed, one of the major issues associated with index insurance is that farmers' losses and local agricultural results may not necessarily match the chosen indicator like rainfall or vegetation indices. One example of an index-based insurance initiative in Africa is the R4 Rural Resilience Initiative by the World Food Programme implemented in Ethiopia since the early 2010s.[7] The purpose of this insurance system is to integrate insurance with saving and risk reduction to build resilience among rural communities. However, as in many cases, this insurance program can face a discrepancy between index triggers and farmer experiences due to a variety of reasons including the use of remote sensing information to calculate payout.

However, the application of AI can help compare the results received using indexes with the reported loss experience and identify any possible inconsistency in this regard. Another aspect related to index insurance that may benefit from utilizing AI is the issue of payouts not always being able to accurately reflect farmers' losses in case of extreme events. This problem is widespread in livestock insurance schemes based on satellite-derived variables. For example, IBLI (Index-Based Livestock Insurance) developed in Kenya by the International Livestock Research Institute makes use of Normalized Difference Vegetation Index (NDVI) as a measure of forage situation, which is used to give insurance cover to pastoralists facing droughts.[8] Though this method is better than the rest in terms

[7] World Food Programme, n.d.
[8] ILRI, n.d.

of scalability and transparency, it is a form of satellite proxy and not loss-based, which has led to situations whereby there is a variance in losses and payouts due to insurance policies.

Although AI does not solve the underlying issues of product design, governance, or incentives, it acts as a verification and diagnostic tool to improve data inputs, improve transparency, and enable more rigorous evaluation of whether insurance products actually work for farmers. In this sense, AI's value lies not in replacing decision-making, but in making sure that decisions are evidence-based.

4.2 Verification is Essential for AI Impact

It is important to note that although artificial intelligence can significantly improve analysis in agricultural risk management, it cannot determine outcomes on its own. In the context of smallholder farming and index-based insurance, AI can be used best as a supporting tool that improves data collection, pattern detection, and communication, rather than a system that defines whether interventions are successful. This distinction is critical, particularly in environments where program effectiveness depends on real-world delivery, farmer experience, and local context.

AI's impact is most evident on the "farming-facing" side, where it can facilitate data collection and communication between farmers, insurers, and policymakers. Through mobile phones and extension services, AI-enabled systems can aggregate farmer-reported outcomes, weather data, and market conditions, creating feedback loops that were previously difficult to establish at scale. These feedback loops can improve monitoring and enable more responsive program design over time. An example comes from *Flood Event Extraction from News Media to Support Satellite-Based Flood Insurance*, where AI models were used to analyze unstructured data such as local news reports alongside satellite imagery to detect disaster events.[9] In Bangladesh, combining news-based classifiers with satellite data improved the detection of flood exposure compared to relying on a single data source. This shows how AI can improve situational awareness by incorporating various data inputs, particularly in areas where data is limited.

Agronomic advisory AI can additionally enhance these feedback loops by providing farmer-level data and insights through tools that offer crop guidance, weather recommendations, and input usage support. Farmers using these technologies generate more detailed and constant data that can also have implications for improving agricultural insurance programs.

Despite these capabilities, AI has clear limitations. It does not determine whether insurance products provide real value to farmers, nor can it assess whether payouts are meaningful in critical "bad years." These outcomes depend on factors such as product

[9] arXiv, 2023.

design, basis risk, delivery systems, and farmer expectations, which extend beyond what data alone can capture. AI also cannot replace farmer knowledge or local context, both of which are essential for interpreting agricultural risks and outcomes. Additionally, persistent data gaps and access constraints limit AI's effectiveness. In Sub-Saharan Africa, where an estimated 860 million people remain offline, the ability of AI systems to generate accurate and representative insights is inherently constrained.[10]

More importantly, though, AI optimizes models and not the livelihoods results. AI is capable of recognizing general problems like the problem of basis risk in index insurance. However, AI cannot recognize why and how exactly a particular failure occurs. This lack of context means that AI might end up amplifying the existing information asymmetry problems between program developers and those who are the intended recipients of the program benefits. AI is capable of optimizing certain processes and making analysis more accurate but it cannot replace the underlying structural conditions.

To mitigate this issue, AI should be viewed as a supporting tool that helps validate information and provide additional insights, rather than as the decision-maker. The strength of AI is its ability to improve the accuracy and availability of information, allowing individuals to evaluate whether programs are effective. Although AI may help reveal important patterns or detect problems, the actual decision-making process must involve people, including policymakers, practitioners, and farmers themselves.

V. Addressing the Governance Gap.

As index insurance programs move from implementation to national scale, many of the mechanisms that supported accountability and adaptation disappear. During pilot stages, communication and feedback loops often exist between farmers, product designers, governments, and regulators. Products can be adjusted through iteration based on farmer experiences and local conditions. However, once programs scale nationally, these checks frequently weaken or disappear altogether.

This results in a gap in governance. Products continue to operate without proper oversight to verify whether payouts continue to occur in the years that matter most, whether farmers have any channel to report failures after rollout, or whether they still deliver value to farmers. This results in failures that go unnoticed as well as issues without a way to course correct, which then undermines trust.

[10] DevelopmentAid, n.d.

1. What's Missing at Scale?

1. Ongoing Performance Monitoring. No entity is consistently responsible for verifying whether payouts continue to align with farmer losses over time, especially in catastrophic years that shape farmer trust.

2. Farmer Feedback Channels. Farmers often lack a formal mechanism to report when products fail to pay meaningfully or when outcomes do not match expectations after deployment.

3. Course-Correction Authority. Insurance products are typically approved once at launch, but no institution is mandated to continually reassess whether products remain effective or to require adjustments when underperformance becomes evident.

4. Sanity Checks at Scale. Pilot programs often included informal or institutional checks asking whether the product was still working for farmers. At the national scale, these checks rarely exist on an institutional level.

2. Beyond Regulation

Insurance regulators play an important role in approving products, monitoring solvency, and overseeing market conduct. However, regulation alone does not answer the central question raised throughout this report: does the product continue delivering value to farmers in year three, year five, or year ten? Currently, there is no entity that is required by law to ensure that the product still works as expected after implementation. This creates a gap between product approval and long-term accountability.

3. The Implementation Answer: A Governance Framework for Scaled Index Insurance

One potential solution to address this issue is a public-private governance framework that institutionalizes ongoing monitoring, farmer feedback, and product adjustment within the country. This framework distributes responsibility across institutions that are already involved in agricultural insurance systems.

4. Who Sits at the Table

Stakeholder	Role
Insurance Regulators	Product approval, solvency oversight, and market conduct supervision
Agricultural Development Bodies	Monitoring farmer outcomes and implementation performance (e.g., ATI in Ethiopia)
Government Ministries	Aligning insurance programs with agricultural and financial policy priorities
Private Insurers & InsurTechs	Product delivery, recalibration, and operational adjustments

Table 1: Stakeholders in a public-private governance framework for scaled index insurance.

The experience of Ethiopia provides one example of how such coordination can emerge through collaboration between the Ethiopian Agricultural Transformation Institute (ATI), regulators, and insurers.

5. Core Monitoring Approaches

Approach 1: Farmer Bad-Year Rankings. Farmer ranking exercises can help ensure that farmer experiences are continuously reintroduced into product evaluation at each contract cycle.

Approach 2: Gap Calculations. Verification exercises can compare farmer-reported losses against payout outcomes to identify persistent mismatch across years and regions.

Approach 3: Bundle Value Assessment. Monitoring should also evaluate whether insurance strengthens the overall value of bundled products or simply increases enrollment without improving resilience.

This type of governance structure is not entirely new. In Central America, coffee institutes (Instituto del Café) already coordinate across exporters, producers, ministries, and financial actors to manage sector-wide risks and policies together. Similar public-private coordination mechanisms could help institutionalize accountability and verification within agricultural insurance systems (FAO, 2018).

VI. Key Takeaways and Conclusion.

After 25 years, index insurance has not failed to scale for the lack of ambition, interest, capital, or technology. The failure identified through the analysis points that the mechanisms that make pilots work, proximity to farmers, iterative correction, honest assessment of whether the product delivered when it mattered, do not survive the transition to scale. The industry has treated scale as a distribution problem, however, the findings suggest a verification problem and argues that the distinction separates responsible scale from the recurring rollout, quiet underperformance and collapse in trust.

To unlock scale, two critical questions need to be answered:

(1) How are we verifying if products are solving a real problem?

(2) How are we making sure that products are making meaningful payments when needed?

The four findings in this report are components of a consolidated verification structure to solve these constraints. Farmer input in the product design and the bundle-value test force the former upstream rather than leaving it to be discovered after deployment or in a post-mortem. On the other hand, a bad-year verification against farmer recalls and the use of artificial intelligence (AI) for diagnostic rather than decision making, replace the average portfolio indicators with tests that reflect what farmers actually use to judge if their product, coverage and service is worthwhile.

This report adds to the 25-year conversation with a set of tools that make verification operationally tractable. Farmers ranking exercises, calibrated by power calculations, can be conducted across focus groups driving costs below the few thousands of dollars, resulting in credible applicable data. Alignment check using yield, farmer recall, and satellite data can be run even where only a few years of data are available, as the Rwanda case demonstrates. Bundle-value assessments distinguishing insurance that strengthens a package from a worthless package filled with add-ons. AI applications that integrate unstructured data, as in the Bangladesh flood-detection case, augment ground truth without replacing it. The methodological point is that responsible scale is affordable.

The implications for the stakeholders vary depending on their role but are generally shared by the structure culture. Donors and sponsors such as the Gates Foundation could condition scale funding on ex-ante verifications: evidence that the product solves a real problem, a known catastrophic year checked against what the product would have paid, and a credible feedback mechanism between farmers and providers. Impact investors as the Qatar Insurance Group could recognize that a bad-year performance is not captured in loss ratios and is the primary determinant of renewal, long-term viability, and portfolio resilience. Governments should treat verification infrastructure as a public good, standardized, budgeted, and shared, rather than a bespoke research project commissioned for each program. Insurers and InsurTech providers should treat farmer feedback as risk data, not reputational management.

The cost of continuing without verification is not another 25 years of stalled scale. It is the erosion of trust in parametric insurance as a category. Each rollout that underperforms in a crisis year, followed by silence or delayed partial payment, makes the next one harder to launch and harder to justify. Conversely, even a small number of visible cases where products demonstrably paid in the years that mattered, verified before scale, not argued after, would change what “responsible” means in this field. The tools to do this are in this report. It is clear that these findings together are to be combined in a next step into an AI agent that combines the agentic capabilities with the tools presented. Some of the best agentic work is based on tools that are more efficient for the end user. These tools are replicable and presentable. They would solve a lot of the core issues addressed in this report.

VII. Technical Appendix.

Annex A. List of Stakeholder Interviews

This report was conducted as part of the Columbia School of International and Public Affairs Capstone Project 2025–2026 by Carlos, Dean, Irra, Jenna, Lara, Mauricio, and Yiran, under the advisory of Panos Varangis and Dan Osgood.

Below is a comprehensive list of the stakeholders we interviewed and to whom we extend our wholehearted gratitude for their time, and interest in participating in our project. This is also our way to recognize their effort and role in tackling down the barriers to scale.

- The Gates Foundation
- Qatar Insurance Group
- The World Bank
- International Finance Corporation
- The Food and Agriculture Organization
- UN World Food Program
- The International Fund for Agricultural Development
- GIZ Germany
- BlueOrchard
- Agricore (InsureCow)
- WIFO *(Austrian Institute of Economic Research)*
- Ibisa Network
- Myubuntu
- Ethiopian Agricultural Transformation Institute
- Howden
- MIC Global
- Columbia World Projects
- Columbia University School of International and Public Affairs (SIPA)

VIII. Bibliography.

Reports & Institutional Sources

International Livestock Research Institute (ILRI). *Index-Based Livestock Insurance (IBLI) Kenya*. https://ibli.ilri.org/ibli-kenya/.

International Research Institute for Climate and Society (IRI). *Nigeria Sampling Note*. https://fist.iri.columbia.edu/publications/docs/1af_nigeria_sampling/.

SIPA Capstone. *Alignment of Payouts with Farmer Losses*. Columbia University, 2025. https://dosgoodcu.github.io/SIPA_IFC_Capstone2025/alignment/.

World Food Programme (WFP). *R4 Rural Resilience Initiative*. https://www.wfp.org/r4-rural-resilience-initiative.

World Food Programme (WFP). *Weather Index Insurance Design and Evaluation*. https://docs.wfp.org/api/documents/WFP-0000130342/download/.

FAO (2018). *The successes and shortcomings of Costa Rica exports diversification policies*. Food and Agriculture Organization of the United Nations, Rome. Document ID I8308EN. https://www.fao.org/3/i8308en/I8308EN.pdf

Academic & Technical Sources

Abdul Latif Jameel Poverty Action Lab (J-PAL). *Power Calculations*. https://www.povertyactionlab.org/resource/power-calculations.

Flood Event Extraction from News Media to Support Satellite-Based Flood Insurance. arXiv, 2023. https://arxiv.org/abs/2312.14943.

Vidal, K., Budiyanti, T., Ganita, Bais, N., Ekkanath, S., & Almomen, F. (2025). *An assessment method to evaluate index insurance quality and effectiveness* [Capstone report]. Columbia University, School of International and Public Affairs.

Media / Commentary

DevelopmentAid. *Africa's Offline Majority Risk Missing Out on the AI Revolution*. https://www.developmentaid.org/news-stream/post/201552/africas-offline-majority-risk-missing-out-on-the-ai-revolution.

– END OF REPORT –

A COLLABORATION OF

COLUMBIA | SIPA
School of International and Public Affairs

Gates Foundation

QIC

COLUMBIA SIPA · GATES FOUNDATION · QATAR INSURANCE GROUP