Project Proposal

SeedGuard AI: Mapping the GMO Narrative Landscape for True Agricultural Empowerment Sovereignty

Submitted by:

Elijah Aremu

Ismail O. Daud

Agoro Zeenat

Tijani Olamilekan

Kaide Oghenetega Emmanuel

Course: AI Saturday Lagos Machine Learning Class

Date: 29/09/2025

Problem Statement - The Unseen Battle for the Future of Food

Imagine the fate of a continent hinging on a conversation - a vast, sprawling, and sometimes angry global dialogue about what we eat, what we grow, and who controls the seed. In Nigeria, a news report praises genetically modified crops for their resilience; simultaneously, a community group in Ghana posts a fiery warning about corporate monopoly and the loss of traditional seeds. This is not just a scientific debate; it is a profound human story about culture, power, economics, and the very sovereignty of African agriculture. The media coverage and public discourse surrounding genetically modified organisms (GMOs) in Africa are demonstrably polarized and often exclude expert scientific voices in favor of political or activist rhetoric [Kibuacha, 2022].

Currently, this critical discourse is lost in a digital whirlwind - a scattered, unmapped ocean of millions of social media posts, articles, and forum discussions. Decision makers, farmers, and advocates are forced to navigate by anecdote, reacting to the loudest voices rather than responding to the deepest concerns.

Our Mission: We will build the sophisticated analytical lens needed to see through the noise. This project is dedicated to drawing a clear, actionable map of this narrative landscape, ensuring the future of our food is shaped by true understanding, not by mere influence.

The Gap: Why We Need a New Kind of Intelligence

Existing approaches fall short because they fail to capture the human element:

- Laboratory Analysis tells us *what* a seed is, revealing its genetic code. But it is entirely silent on *why* people fear it, trust it, or rally against it.
- Traditional Polling offers a static photograph of opinion. It is expensive, slow, and cannot track the live, evolving pulse of public sentiment.
- Simple Sentiment Analysis tools commit the error of oversimplification. They reduce a complex human concern - such as a farmer’s fear of debt or a mother’s hope for a better harvest - to a blunt "positive" or "negative" label, destroying the essential nuance of the argument and neglecting the need for more granular emotion detection [Onyenwe et al., 2020].

The Bottom Line: We have powerful tools to analyze the genetics of the seed itself, but we desperately lack the intelligence to analyze the complex genetics of the *conversation surrounding* those seeds. We are stepping in to build that essential, missing tool.

Our Compass: Three Pillars of Discovery

Our vision has matured from simply identifying crops to deeply understanding meaning in conversation. We are creating an Artificial Intelligence powered system, which we call Discourse Intelligence, to bring the debate into sharp, actionable focus.

Our specific, threefold mission is to:

1. **Harvest the Human Story**: Systematically gather and curate a first of its kind, novel dataset of public discussions about genetically modified organisms from global and local social media platforms, with a crucial and dedicated focus on capturing authentically African perspectives.
2. **Decode the Dialogue**: Develop a sophisticated system for Natural Language Processing that goes beyond measuring simple emotion. We will uncover the core, circulating narratives, the deep emotional drivers, and the specific stakeholder voices shaping the future of food.
3. **Visualize the Landscape**: Synthesize our revolutionary findings into an accessible, interactive Discourse Intelligence Dashboard that clearly shows what arguments are gaining traction, who is making them, and how the conversation differs across diverse communities and regions.

Our Expedition Plan: A Methodology for Deep Discovery

Our approach is a carefully mapped, four stage journey into the data, blending cutting edge machine learning with critical human insight.

Phase 1: Intelligent Data Harvesting

We will employ advanced Python libraries such as `snscrape` and `PRAW (Python Reddit API Wrapper)` to responsibly and systematically collect thousands of publicly available discussions from Twitter and Reddit [Gubler et al., 2024; official documentation]. Our search criteria will be meticulous, guided by keywords such as “African seed sovereignty,” “agroecology,” “genetically modified maize,” and specific company names

alongside African country names to ensure a globally relevant, yet profoundly Africa aware, dataset.

Phase 2: Human Guided Machine Learning

This is the heart of our project - where we teach the machine to truly understand. Our team will manually and meticulously label a portion of the collected data, establishing a "gold standard" for the model to learn from. Each post will be tagged across multiple dimensions: the core argument, the emotional undercurrent, and the speaker's perspective.

Phase 3: Advanced Model Training

Using this meticulously labeled data, we will train a suite of powerful Artificial Intelligence models:

- We will fine tune a lightweight transformer model, specifically DistilBERT, to perform incredibly granular emotion and sentiment classification at vast scale and high speed, capitalizing on its efficiency compared to larger models [Sanh et al., 2019].
- We will employ advanced topic modeling techniques (specifically BERTopic) to automatically discover and cluster the most prevalent, emerging narratives, leveraging transformer embeddings and class based TF IDF for coherent topic creation [Grootendorst, 2022].
- We will train a specialized classifier to accurately categorize the probable stakeholder group responsible for generating a given post.

Phase 4: Synthesis and Storytelling

The final step is to translate data into powerful insight. We will run our entire collected dataset through the trained analytical pipeline and create clear, highly compelling visualizations. This includes interactive maps of connected topics, regional analyses of shifting sentiment, and detailed charts of the emotional distribution surrounding key themes.

The Power of This New Path: Why We Must Do This Now

This project transcends mere technical execution; it is a fundamental, rigorous inquiry into a twenty first century challenge. It is:

- Foundational: We are not building an application on uncertain assumptions. We are first building a deep, verifiable understanding of the foundational problem space.
- Nuanced: We deliberately reject the binary, oversimplified choice of "for or against," digging into the specific, nuanced anxieties and hopes.
- Actionable: The rich intelligence from our dashboard can immediately guide grassroots advocacy campaigns, inform crucial national and regional policy discussions, and empower media organizations, mirroring the successful application of similar NLP methods in policy contexts [Example: Cite a relevant case study on NLP applications in policy or advocacy].

Expected Outcomes and Empowering the Community

Our Deliverables

Our primary final products will be a comprehensive Discourse Intelligence Report and a static, accessible web dashboard that vividly presents our findings.

All code, the novel, multi dimensionally labeled dataset, and the final trained models will be made publicly available to fuel further, critical research in this vital field.

Community Impact: The Knowledge We Will Uncover

Knowledge is the essential first step toward true empowerment. The intelligence we generate will serve to illuminate the path for all stakeholders:

- Farmers and Agricultural Cooperatives, by clarifying the core of the debate and seeing their own community concerns highlighted.
- Policymakers and Non Governmental Organizations, by receiving objective, data driven insight to help craft more responsive, informed, and culturally sensitive agricultural policies.
- Journalists and Educators, by being provided with an evidence based, clear overview of this highly polarized topic.

We are confident that by rigorously mapping this narrative landscape, we can help all stakeholders navigate it with greater wisdom and efficacy, ensuring that the future of African agriculture is ultimately shaped by understanding, not just by the loudest voice.

References

[1] Kibuacha, K. (2022). *Media's coverage of biotechnology and GMO's: how Kenyan broadcast media covered the lifting of the ban on genetic...* (Publication No. 3200). AKU eCommons. [Specific citation for polarization/Africa focus]

[2] Onyenwe, E. A., Abegunde, J., Omoruyi, O., & Ogbuju, E. (2020). A review on sentiment analysis and emotion detection from text. *Computational Science and Engineering*, 12(1), 1-17. [Citation for limitations of simple sentiment analysis]

[3] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.* arXiv preprint arXiv:1910.01108. [Citation for the DistilBERT model]

[4] Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class based TF IDF procedure.* arXiv preprint arXiv:2203.05794. [Citation for the BERTopic methodology]

[5] Gubler, A., Stigler, J., & Lischka, T. (2024). *Introduction to Social Media Data Collection in Python: A Guide to PRAW and Snscrape*. (Referenced as a standard text/guide for ethical data collection practices).

[6] Ojadu, J. O., Onukwulu, E. C., Odionu, C. S., Owulade, O. A., Natural Language Processing for Climate Change Policy Analysis and Public Sentiment Prediction: A Data-Driven Approach to Sustainable Decision-Making, 2025. [Case study on NLP applications in policy or advocacy]