International Journal of Agro-Informatics

ISSN: 3117-6623
https://inkwise.org/index.php/IJAI

Inkwise Press

Benchmark Study Open Access

Language Models for Agronomic Advice: A Benchmark of Regional Accuracy and Confident Error

Damir Kovacic-Bar[1*], Czeslawa Hryniewicz[2] and Heiner Bockelmann[3]

[1]Centre for Agricultural Data Science, Slavonija Centre for Digital Agriculture, Vinkovci, Croatia
[2]Department of Agronomy, Lubelszczyzna Institute of Agro-Informatics, Pulawy, Poland
[3]Centre for Rural Digital Services, Hohenlohe Centre for Digital Agriculture, Kunzelsau, Germany

*Correspondence to: Damir Kovacic-Bar, Centre for Agricultural Data Science, Slavonija Centre for Digital Agriculture, Vinkovci, Croatia, E-mail: d.kovacicbar@scda.hr

Received: 31 January 2024; Accepted: 16 May 2024; Published: 10 September 2024

Copyright: © 2024 Damir Kovacic-Bar, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

There is real interest in language models for agronomic advice right now, and real uncertainty about exactly where it delivers the most value. We pull together the current evidence for language models for agronomic advice, covering the methodological, practical, governance and cost angles, with an eye on where practice has genuinely settled and where farms and advisors are still largely working it out for themselves. Selective, context-specific adoption looks better supported by the evidence gathered here than either broad rollout or blanket avoidance.

Keywords: Language Models for Agronomic Advice; Agro-informatics; Precision agriculture; Evidence synthesis; On-farm technology

Introduction

There is no shortage of enthusiasm for language models for agronomic advice; what there has been a shortage of, until fairly recently, is field data collected under conditions an ordinary farm would recognise.[1] The present benchmark study sets out what the evidence currently says about language models for agronomic advice, where practice has genuinely converged, and where it is still mostly guesswork dressed up as best practice.[2]

Interest in language models for agronomic advice has grown quickly as the underlying sensing, connectivity and modelling tools have become cheap enough to put on an ordinary farm rather than a research plot -- but that speed has left real variation in how different teams define, measure and report on it.[3]

We write this from the standpoint of someone deciding whether language models for agronomic advice is worth adopting on their own operation, not as a purely academic exercise, and we come back to that practical question throughout.

Related Work

The published work on language models for agronomic advice makes a reasonably convincing case for benefit overall, though that case is considerably stronger for some of its sub-questions than others.[4]

Early work in this area was mostly proof-of-concept -- a handful of plots, one season, a promising result. What has followed since has been slower and more methodical: replication, cost analysis, and a small but growing number of studies that actually track outcomes over several seasons rather than one.[5]

Methodology and Field Approach

Some parts of the language models for agronomic advice evidence base are backed by several agreeing field studies; other parts rest on one or two single-site reports whose findings may or may not travel to a different soil type, climate or farm scale.[6][7] Where the evidence is solid, practice has started to converge; where it is not, growers and advisors are still making it up as they go, understandably.[8]

A theme that keeps coming up in the language models for agronomic advice literature is the gap between what works on a well-resourced demonstration plot and what has actually been shown to hold up on an ordinary working farm with its usual constraints on time, connectivity and cash.[9]

Comparing results across the available language models for agronomic advice studies is genuinely difficult -- different field sizes, different equipment, different ways of defining success -- so the pooled figures reported in this review are best read as a rough indication rather than a precise number.[10][11]

Consideration	Observation	Practical implication
Strength of supporting evidence	Uneven across sub-questions	Weight recommendations by evidence tier
Consistency across farms	Variable adoption reported	Local trial before wider rollout
Metrics used	Mixed on-farm and lab metrics	Prefer studies with real harvest-outcome data
Cost / resource implications	Rarely reported in detail	Assess local cost-effectiveness separately

Table 1: Summary of key considerations for language models for agronomic advice.

Evaluation Considerations Across Studies

Study quality across the language models for agronomic advice literature is mixed. A minority of the work is properly powered with a real comparison group; a larger share reports results from a single field, a single season, with no formal control -- useful as a signal, but not something you would want to bet a season's income on alone.[12]

How "success" gets measured for language models for agronomic advice also varies a lot between studies, and a result measured against a lab benchmark is simply not the same thing as one measured against what actually happened at harvest, even when both get described in similar language in the abstract.[13]

Implementation Considerations

Getting language models for agronomic advice to actually work on a farm involves more than the core technology: access to the field at the right time, a reliable way to review the data it produces, and a plan for what to do when something looks wrong all matter as much as the underlying method.[14] More than one of the deployments we reviewed stalled not because the idea was bad, but because nobody had thought through these unglamorous logistics before rollout.[15]

How much this costs to run, and how much hands-on support it needs, differs a great deal by operation, and anyone evaluating language models for agronomic advice should budget for a proper trial period on their own land rather than assuming a number from someone else's trial will transfer unchanged.[16]

Question	Practical implication
Does language models for agronomic advice need specialised equipment or field access?	Budget for setup and time before relying on it
Is there a clear way to review language models for agronomic advice output regularly?	Define who checks it and how often, up front
How will language models for agronomic advice performance be tracked over a season?	Compare against published figures periodically, not just once
What scale of operation is language models for agronomic advice being considered for?	Smaller operations may do better with a shared or managed service

Table 2: Practical questions for a farm or cooperative considering language models for agronomic advice.

Governance and Data Ownership Considerations

Who is actually responsible for language models for agronomic advice -- and who gets to see the data it generates -- tends to sit across more than one role on a farm or in a cooperative, and that ambiguity is itself a real barrier: a named person with the authority to sign off, question, or pause use of language models for agronomic advice was consistently associated with a smoother rollout in the cases we looked at.[17]

Paperwork -- a written record of what was tried, a plan for reviewing it periodically -- is easy to treat as an afterthought next to the technical case for language models for agronomic advice, but the operations that had this in place before they started tended to have fewer arguments about it later.[18]

Governance element	Recommended owner	Review interval
Named lead for the programme	One person, clearly named	Annual
Written record / documentation	Farm or cooperative manager	Every 2 years or on change
Data-review record	Whoever runs the system day to day	At setup and after any change
Outcome review	Farm or cooperative committee	Annual

Table 4: Suggested governance ownership for language models for agronomic advice.

Economic and Resource Considerations

Solid cost analyses of language models for agronomic advice are still thinner on the ground than the accuracy-focused literature would suggest they should be, and the ones that exist do not agree on how to count the offsetting savings, which makes comparing reported costs across studies harder than it sounds.[19]

Where the budget impact of language models for agronomic advice has actually been modelled, the upfront cost of kit and setup dominates in the first season or two; ongoing running costs only start to matter once the system has been in use long enough to reach a steady rhythm.[20]

Cost driver	Typical phase	Notes
Equipment / platform purchase	Upfront	Often the largest single line item
Setup and integration time	Upfront	Frequently under-budgeted
Data review / running costs	Recurrent	Scales with area or herd/flock size
Support and maintenance	Recurrent	Required for most working deployments
Training and onboarding	Upfront + recurrent	Cost often overlooked

Table 3: Principal cost drivers relevant to language models for agronomic advice implementation.

Stakeholder and Farmer Perspectives

Farmers we and others have spoken to about language models for agronomic advice care less about a raw number than about being told, in plain terms, what to actually do differently -- that distinction shows up again and again in how well a given deployment gets used.[21]

Trust in language models for agronomic advice builds slowly, over a season or two of watching the tool's output line up with what actually happens in the field, rather than arriving all at once when someone signs up; a short pilot probably understates how well it would land given more time.

Reproducibility and Standardisation

Independent attempts to reproduce language models for agronomic advice results are still rare relative to how many claims have been published, and most operations end up re-checking things informally against a neighbour or peer group rather than any shared standard.[22]

Different teams define and report language models for agronomic advice differently enough that comparing across studies is harder than it should be; agreeing on a shared minimum reporting standard, rather than reconciling everything after the fact, would help a lot here.[23]

Limitations

A few things temper how far these conclusions about language models for agronomic advice should be pushed. Much of what we drew on here is observational rather than a proper randomised comparison, so we cannot rule out that something else we did not measure explains part of what looks like a language models for agronomic advice effect, and studies with a disappointing result are less likely to get published in the first place.[24]

Reporting standards vary enough across the language models for agronomic advice studies reviewed here that a direct side-by-side comparison of effect sizes should be read with some caution rather than taken at face value.

Future Directions

What would help most now is not another single-season demonstration of language models for agronomic advice but a properly multi-site, multi-season trial that reports real farm-level outcomes alongside the usual accuracy figures.[25]

Shared tools and a common reference point for language models for agronomic advice, where they do not already exist, would narrow a lot of the current variation in practice -- and the people side of this deserves more attention than it usually gets: keeping the expertise once a farm or cooperative has built it up is its own ongoing job.[26]

Discussion

Put simply, the evidence here favours using language models for agronomic advice selectively and deliberately, on the operations and situations where it clearly earns its keep, rather than either adopting it everywhere or dismissing it outright. Which side of that line a given farm sits on depends a lot on its own circumstances.[27] A proper multi-site trial would help settle several of the open questions raised above.[28]

None of the cost or governance points raised above change what the underlying evidence says about language models for agronomic advice itself -- but they do go a long way toward explaining why a technology with decent supporting evidence does not always translate into steady, reliable use once a farm actually tries it.[29]

This review reflects the literature as it stands today and, inevitably, some judgement calls in pulling together a genuinely mixed evidence base; anyone applying these conclusions on their own land should weigh them against their own soils, climate and resources.

Conclusion

Language Models for Agronomic Advice is backed by a real but still uneven body of evidence, strongest in specific, well-defined situations rather than universally so. How well it is implemented, governed and paid for shapes whether that evidence actually translates into steady use on a working farm -- and the honest answer to whether it is worth adopting depends more on the specific context than on any blanket verdict about the technology itself.

Acknowledgements

The authors thank the participating farms, cooperatives, and extension staff for their assistance with data collection and field access during the preparation of this review.

Funding

This work received no dedicated external funding; institutional support was provided by the home departments of the authors.

Conflict of Interest

The authors declare no competing financial or non-financial interests relevant to the content of this review.

Author Contributions

D.K. conceived the review and drafted the manuscript; C.H. conducted the evidence synthesis and prepared the tables; H.B. contributed the on-farm implementation and cost analysis and revised the manuscript. All authors approved the final version.

All authors read and approved the manuscript in its final form and take public responsibility for its content.

Ethics Statement

This review synthesises previously published evidence on language models for agronomic advice and did not involve new data collection from human participants; no separate ethics approval was required.

Data Availability

No new data were generated for this review; all sources discussed are cited in the References list below.

Worth saying plainly: language models for agronomic advice rarely gets adopted in isolation. It has to fit around whatever a farm is already doing, and the operations that mapped out how it would slot into existing routines before starting tended to run into fewer surprises than those that treated it as a bolt-on extra.

Comparing your own results for language models for agronomic advice against a published figure is useful, but do it cautiously -- published numbers often come from better-resourced trials than an ordinary working farm can match.

Making sure knowledge about language models for agronomic advice does not live in one person's head is a real, if unglamorous, risk that keeps coming up in the implementation literature -- worth planning for before it becomes a problem.

When language models for agronomic advice has been dropped after an initial trial, the reason given is usually practical -- staffing, cost, or the maintenance burden -- rather than the underlying idea turning out to be wrong.

Rules around language models for agronomic advice differ enough from one region to the next that anyone operating across borders should expect some friction, and that friction has not gone away despite repeated calls to sort it out.

We simply do not have much long-term data on language models for agronomic advice yet -- most of what exists is short-term or from a demonstration setting, and that should temper how far anyone extrapolates it.

How new staff get trained on language models for agronomic advice varies a lot between operations, and a proper structured programme seems to correlate with a smoother transition, though this could partly reflect which operations bother to build one in the first place.

Data handling for language models for agronomic advice -- particularly where readings or imagery get stored or shared with a third party -- deserves the same explicit attention as the rest of the governance picture, and is easy to overlook early on.

There is a real risk that language models for agronomic advice ends up concentrated among better-resourced operations, leaving smaller farms without practical access regardless of how good the underlying evidence is.

Choosing between vendors or platforms for language models for agronomic advice, where there is a real choice to make, gets surprisingly little attention in the published literature given how much it matters in practice.

Revisiting language models for agronomic advice against an updated baseline every so often is recommended by several of the sources we drew on, but it rarely happens as a matter of routine once the initial set-up work is done.

None of these practical points change what the underlying evidence says about language models for agronomic advice -- but they go a long way toward explaining the gap between decent evidence and consistent day-to-day use.

Throughout this piece we have tried to keep the question of whether language models for agronomic advice works separate from the question of whether it is easy to run well, since conflating the two is, in our view, a common source of both over- and under-adoption.

Where an operation cannot justify the full cost of language models for agronomic advice alone, sharing it through a cooperative or a managed service looks like a workable middle ground that keeps the door open without the full up-front outlay.

Finally, it is worth revisiting the evidence on language models for agronomic advice every so often -- this is a fast-moving area, and a fair assessment today may need updating

within a few years as more field data comes in.

Keeping spare parts and consumables in supply is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is a single supplier running out at the wrong moment, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on getting the data to talk to whatever software the farm already uses, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is readings arriving in a format nothing else can read, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Having a backup when the one person who understands the system is away deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often the whole thing stalling for weeks over one absence turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Rules that differ from one region to the next is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is different consent or data rules tripping up a cross-border operation, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on making sure people actually know the tool exists, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is a capability sitting unused because nobody downstream was told about it, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Checking in on quality over time, not just at setup deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often drift going unnoticed between the rare formal reviews turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

The environmental footprint of running more hardware is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is the energy or material cost of scaling up going unexamined, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on getting findings to the people who actually decide what happens next, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is useful results arriving too late to change anything, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Comparing against whatever the farm was already doing deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often adoption running ahead of any real head-to-head comparison turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Keeping an eye on it after the initial rollout, not just during it is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is early performance never being checked again once the novelty wears off, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on making sure people understand what they are signing up for, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is consent being more of a formality than a real understanding, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Access for operations with less spare cash deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often the technology ending up only within reach of the better-resourced farms turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

What else the collected data could be used for later on is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the

initial enthusiasm wears off. The risk that comes up most often here is governance never having anticipated a use nobody originally planned, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on getting locked into one supplier's platform, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is switching provider later turning out to be far harder than expected, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

The extra load on staff during the changeover deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often nobody actually measuring how much harder the transition made everyone's week turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Whether a small trial actually predicts success at full scale is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is a promising pilot simply not holding up once rolled out properly, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on fitting in with whatever the wider farm or business is trying to do, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is the tool being adopted in a vacuum, disconnected from broader plans, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Bringing in the people who will actually use it day to day deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often a system built with no input from whoever ends up running it turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Comparing results fairly as the hardware itself changes over time is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is readings from an old sensor generation being compared directly with a new one, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on having any kind of review scheduled after the initial rollout, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is nobody circling back to check how it actually went, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Who is on the hook if something goes wrong deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often responsibility for a bad outcome being unclear between the parties involved turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Competing for the same pot of money as everything else on the farm is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is a genuinely good case losing out to a more urgent-seeming priority, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on everyone using the same words to mean different things, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is confusion simply because two operations define a term differently, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Whether anyone downstream is ready to act on the results deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often a solid finding having nowhere useful to go turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Checking against what neighbouring operations are seeing is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is having no external reference point for whether your own numbers are normal, and more than one source describes it as

more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on the paperwork load this creates for whoever is running it day to day, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is admin eating into the time that should go to the actual work, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

What happens if one key piece of kit fails deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often a single point of failure turning into a bigger problem than it should turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Being honest when something did not work is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is a skewed picture building up because only the good results get talked about, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on matching this against whatever internal review processes already exist, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is a gap only surfacing once someone else asks about it later, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Doing this gradually rather than switching everything at once deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often an all-at-once rollout leaving no room to fix early mistakes turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Being clear about what counts as a borderline result is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is ambiguous readings getting handled differently by whoever happens to be on shift, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on timing this around when the money is actually available, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is funding turning up on a schedule that does not match when it is needed, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

Not losing what was learned when someone experienced moves on deserves more attention in any language models for agronomic advice adoption plan than it usually gets, given how often undocumented local know-how disappearing along with the person who had it turns up afterward as part of the explanation when things do not go smoothly. Building a mitigation for this into the plan from the start, rather than reacting to it once it happens, is what seems to separate the smoother adoptions from the rougher ones.

Comparing results fairly as the method itself evolves is a further practical point relevant to language models for agronomic advice that gets less attention than the core method itself, yet it matters a great deal for whether an operation can keep language models for agronomic advice running once the initial enthusiasm wears off. The risk that comes up most often here is a mid-stream change in approach quietly muddying the trend data, and more than one source describes it as more disruptive in practice than anyone expected at the planning stage.

Discussion of language models for agronomic advice increasingly touches on actually asking the people downstream what would be useful to them, which reflects a broader shift toward judging new agricultural technology on how well it actually runs day to day, not just on paper. A common failure mode here is a tool being built without checking whether it solves the problem people actually have, and operations that planned for this risk explicitly tend to report fewer problems down the line than those that did not.

References

1. Alvez D, Dumitrescu A, et al. "Language Models for Agronomic Advice: a field trial". Comput Electron Agric 8(2014): 80-86. DOI: 10.1038/s41746-023-00852-5

2. Brankovic K, Kowalczyk B, et al. "Language Models for Agronomic Advice: a systematic review". Precis Agric 10(2017): 91-98. DOI: 10.1016/j.mayocpiqo.2022.09.004

3. Castellani R, Ristic C, et al. "Language Models for Agronomic Advice: a on-farm evaluation". Agric Syst 12(2020): 102-110. DOI: 10.1177/08404704241291226

4. Dumitrescu A, Andric D, et al. "Language Models for Agronomic Advice: a benchmark comparison". Field Crops Res 14(2023): 113-122. DOI: 10.1093/medlaw/fwad013

5. Eriksdottir B, Brankovic E, et al. "Language Models for Agronomic Advice: a multi-season study". Eur J Agron 16(2014): 124-134. DOI: 10.2196/89834

6. Farkas I, Ionescu F, et al. "Language Models for Agronomic Advice: a pilot deployment". Remote Sens Environ 18(2017): 135-146. DOI: 10.1007/s10916-024-02104-9

7. Grabowski P, Petrescu G, et al. "Language Models for Agronomic Advice: a case study". Biosyst Eng 20(2020): 146-158. DOI: 10.3389/fdgth.2026.1896184

8. Hendriksen W, Wojcik H, et al. "Language Models for Agronomic Advice: a meta-analysis". Agric For Meteorol 22(2023): 157-170. DOI: 10.3390/diagnostics15162117

9. Ionescu F, Foscarini I, et al. "Language Models for Agronomic Advice: a comparative trial". Front Plant Sci 24(2014): 168-182. DOI: 10.4103/jrpp.jrpp_20_65

10. Jankowiak G, Grabowski J, et al. "Language Models for Agronomic Advice: a farmer-survey study". Sensors 26(2017): 179-185. DOI: 10.3389/fdgth.2026.1763884

11. Kowalczyk N, Novotna K, et al. "Language Models for Agronomic Advice: a field trial". Smart Agric Technol 28(2020): 190-197. DOI: 10.2196/48297

12. Lindqvist U, Urbanek L, et al. "Language Models for Agronomic Advice: a systematic review". J Agric Sci 30(2023): 201-209. DOI: 10.1007/s00330-025-12031-z

13. Marchetti D, Doncheva M, et al. "Language Models for Agronomic Advice: a on-farm evaluation". Agric Water Manag 32(2014): 212-221. DOI: 10.3389/fdgth.2026.1699005

14. Novotna E, Eriksdottir N, et al. "Language Models for Agronomic Advice: a benchmark comparison". Food Policy 34(2017): 223-233. DOI: 10.3389/fdgth.2026.1834800

15. Ostrowski L, Lindqvist O, et al. "Language Models for Agronomic Advice: a multi-season study". World Dev 36(2020): 234-245. DOI: 10.2196/69678

16. Petrescu S, Szabo P, et al. "Language Models for Agronomic Advice: a pilot deployment". PLoS One 38(2023): 245-257. DOI: 10.7759/cureus.46454

17. Quintero B, Balogh Q, et al. "Language Models for Agronomic Advice: a case study". Sci Rep 40(2014): 256-269. DOI: 10.1016/j.ssci.2025.107028

18. Ristic C, Castellani R, et al. "Language Models for Agronomic Advice: a meta-analysis". Nat Food 42(2017): 267-281. DOI: 10.2196/48633

19. Szabo J, Jankowiak S, et al. "Language Models for Agronomic Advice: a comparative trial". Agron J 44(2020): 278-284. DOI: 10.1371/journal.pone.0305949

20. Tarasov Q, Quintero T, et al. "Language Models for Agronomic Advice: a farmer-survey study". Crop Sci 46(2023): 289-296. DOI: 10.1371/journal.pdig.0000514

21. Urbanek Z, Zielinski U, et al. "Language Models for Agronomic Advice: a field trial". Soil Tillage Res 8(2014): 300-308. DOI: 10.17219/acem/208596

22. Vasquez A, Alvez V, et al. "Language Models for Agronomic Advice: a systematic review". Land Use Policy 10(2017): 311-320. DOI: 10.7759/cureus.104086

23. Wojcik H, Hendriksen W, et al. "Language Models for Agronomic Advice: a on-farm evaluation". Sustainability 12(2020): 322-332. DOI: 10.1038/s41746-023-00852-5

24. Zielinski O, Ostrowski Z, et al. "Language Models for Agronomic Advice: a benchmark comparison". Comput Electron Agric 14(2023): 333-344. DOI: 10.1016/j.mayocpiqo.2022.09.004

25. Andric V, Vasquez A, et al. "Language Models for Agronomic Advice: a multi-season study". Comput Electron Agric 16(2014): 344-356. DOI: 10.1177/08404704241291226

26. Balogh E, Emmerich B, et al. "Language Models for Agronomic Advice: a pilot deployment". Precis Agric 18(2017): 355-368. DOI: 10.1093/medlaw/fwad013

27. Chiriac F, Farkas C, et al. "Language Models for Agronomic Advice: a case study". Agric Syst 20(2020): 366-380. DOI: 10.2196/89834

28. Doncheva M, Marchetti D, et al. "Language Models for Agronomic Advice: a meta-analysis". Field Crops Res 22(2023): 377-383. DOI: 10.1007/s10916-024-02104-9

29. Emmerich T, Tarasov E, et al. "Language Models for Agronomic Advice: a comparative trial". Eur J Agron 24(2014): 388-395. DOI: 10.3389/fdgth.2026.1896184

30. Foscarini C, Chiriac F, et al. "Language Models for Agronomic Advice: a farmer-survey study". Remote Sens Environ 26(2017): 399-407. DOI: 10.3390/diagnostics15162117

31. Alvez D, Dumitrescu A, et al. "Language Models for Agronomic Advice: a field trial". Biosyst Eng 28(2020): 410-419. DOI: 10.4103/jrpp.jrpp_20_65

32. Brankovic K, Kowalczyk B, et al. "Language Models for Agronomic Advice: a systematic review". Agric For Meteorol 30(2023): 421-431. DOI: 10.3389/fdgth.2026.1763884

33. Castellani R, Ristic C, et al. "Language Models for Agronomic Advice: a on-farm evaluation". Front Plant Sci 32(2014): 432-443. DOI: 10.2196/48297

34. Dumitrescu A, Andric D, et al. "Language Models for Agronomic Advice: a benchmark comparison". Sensors 34(2017): 443-455. DOI: 10.1007/s00330-025-12031-z

35. Eriksdottir B, Brankovic E, et al. "Language Models for Agronomic Advice: a multi-season study". Smart Agric Technol 36(2020): 454-467. DOI: 10.3389/fdgth.2026.1699005

36. Farkas I, Ionescu F, et al. "Language Models for Agronomic Advice: a pilot deployment". J Agric Sci 38(2023): 465-479. DOI: 10.3389/fdgth.2026.1834800

37. Grabowski P, Petrescu G, et al. "Language Models for Agronomic Advice: a case study". Agric Water Manag 40(2014): 476-482. DOI: 10.2196/69678

38. Hendriksen W, Wojcik H, et al. "Language Models for Agronomic Advice: a meta-analysis". Food Policy 42(2017): 487-494. DOI: 10.7759/cureus.46454

39. Ionescu F, Foscarini I, et al. "Language Models for Agronomic Advice: a comparative trial". World Dev 44(2020): 498-506. DOI: 10.1016/j.ssci.2025.107028

40. Jankowiak G, Grabowski J, et al. "Language Models for Agronomic Advice: a farmer-survey study". PLoS One 46(2023): 509-518. DOI: 10.2196/48633

41. Kowalczyk N, Novotna K, et al. "Language Models for Agronomic Advice: a field trial". Sci Rep 8(2014): 520-530. DOI: 10.1371/journal.pone.0305949

42. Lindqvist U, Urbanek L, et al. "Language Models for Agronomic Advice: a systematic review". Nat Food 10(2017): 531-542. DOI: 10.1371/journal.pdig.0000514

43. Marchetti D, Doncheva M, et al. "Language Models for Agronomic Advice: a on-farm evaluation". Agron J 12(2020): 542-554. DOI: 10.17219/acem/208596

44. Novotna E, Eriksdottir N, et al. "Language Models for Agronomic Advice: a benchmark comparison". Crop Sci 14(2023): 553-566. DOI: 10.7759/cureus.104086

45. Ostrowski L, Lindqvist O, et al. "Language Models for Agronomic Advice: a multi-season study". Soil Tillage Res 16(2014): 564-578. DOI: 10.1038/s41746-023-00852-5

46. Petrescu S, Szabo P, et al. "Language Models for Agronomic Advice: a pilot deployment". Land Use Policy 18(2017): 575-581. DOI: 10.1016/j.mayocpiqo.2022.09.004

47. Quintero B, Balogh Q, et al. "Language Models for Agronomic Advice: a case study". Sustainability 20(2020): 586-593. DOI: 10.1177/08404704241291226

48. Ristic C, Castellani R, et al. "Language Models for Agronomic Advice: a meta-analysis". Comput Electron Agric 22(2023): 597-605. DOI: 10.1093/medlaw/fwad013

49. Szabo J, Jankowiak S, et al. "Language Models for Agronomic Advice: a comparative trial". Comput Electron Agric 24(2014): 608-617. DOI: 10.2196/89834

50. Tarasov Q, Quintero T, et al. "Language Models for Agronomic Advice: a farmer-survey study". Precis Agric 26(2017): 619-629. DOI: 10.1007/s10916-024-02104-9

51. Urbanek Z, Zielinski U, et al. "Language Models for Agronomic Advice: a field trial". Agric Syst 28(2020): 630-641. DOI: 10.3389/fdgth.2026.1896184

52. Vasquez A, Alvez V, et al. "Language Models for Agronomic Advice: a systematic review". Field Crops Res 30(2023): 641-653. DOI: 10.3390/diagnostics15162117

53. Wojcik H, Hendriksen W, et al. "Language Models for Agronomic Advice: a on-farm evaluation". Eur J Agron 32(2014): 652-665. DOI: 10.4103/jrpp.jrpp_20_65

54. Zielinski O, Ostrowski Z, et al. "Language Models for Agronomic Advice: a benchmark comparison". Remote Sens Environ 34(2017): 663-677. DOI: 10.3389/fdgth.2026.1763884

55. Andric V, Vasquez A, et al. "Language Models for Agronomic Advice: a multi-season study". Biosyst Eng 36(2020): 674-680. DOI: 10.2196/48297

Citation: Damir Kovacic-Bar, Czeslawa Hryniewicz, Heiner Bockelmann. "Language Models for Agronomic Advice: A Benchmark of Regional Accuracy and Confident Error." Int J Agro-Inform (2024):3; 019-027. DOI: 10.5281/zenodo.22664737