Situational Vulnerability and Social Firewalls: Uncovering Potential Explainability Pitfalls in Precision Viticulture

Rui Pedro Porfírio
Pedro Albuquerque Santos
Rui Neves Madeira
r.porfirio@campus.fct.unl.pt
pedro.albuquerque@estsetubal.ips.pt
rui.madeira@estsetubal.ips.pt
NOVA LINCS, NOVA School of Science and Technology, NOVA University of Lisbon
Caparica, Portugal
Setúbal School of Technology, Polytechnic University of Setúbal
Setúbal, Portugal

Abstract

Agriculture is a traditional sector currently navigating a transition toward digitalization. In this high-stakes domain, incorrect decisions may impact the entire agri-food chain. While Explainable AI (XAI) tools are often proposed to encourage trust, they may inadvertently introduce safety risks and explainability pitfalls. In this context, we present an exploratory case study investigating how viticulture stakeholders navigate these risks through a focus group with 13 participants (4 senior agronomists and 9 students) using scenario-based questions. Our results suggest that agronomists may exhibit situational vulnerability, as their self-reported agency appears to diminish when facing high-confidence metrics in low-familiarity contexts. We describe the role of social firewalls, where the professional liability agronomists face regarding farmers may act as a systemic defense that maintains expert scrutiny. However, our analysis also highlights a potential future scenario of intergenerational divergence in the use of explanations. While agronomists expressed concern regarding an atrophy of the clinical eye in future practitioners, novice users in this study reported a prioritization of passive explanation acceptance over verification. Furthermore, participants projected a temporal decay of scrutiny where expert attention may decrease as the system becomes part of a routine workflow. To address this risk, participants proposed a sampling-based auditing protocol through randomized field checks. While this work is a formative pilot study, these insights help identify ways to mitigate potential explainability pitfalls by supporting social accountabilities and practitioner-led protocols.

CCS Concepts

• **Human-centered computing** → *Interaction design*; **Empirical studies in HCI**; • **Applied computing** → **Agriculture**; • **Computing methodologies** → *Artificial intelligence*; • **Security and privacy** → *Human and societal aspects of security and privacy.*

Keywords

Explainability Pitfalls, Human-Centered Explainable AI, Agriculture, Human Agency, Focus Group

ACM Reference Format:
Rui Pedro Porfírio, Pedro Albuquerque Santos, and Rui Neves Madeira. 2026. Situational Vulnerability and Social Firewalls: Uncovering Potential Explainability Pitfalls in Precision Viticulture. In *IUI MIRAGE 2026 workshop held at 31st International Conference on Intelligent User Interfaces.* ACM, New York, NY, USA, 6 pages.

1 Introduction

The integration of machine learning into precision agriculture offers significant potential to optimize farm operations through the automation of predictive tasks such as disease diagnosis and yield forecasting [4, 5, 15]. However, the practical application of these predictive systems occurs in a high-stakes environment where the cost of failure is substantial. Global assessments indicate that diseases and pests reduce the yields of major crops by significant margins. Losses for crops such as rice and maize are estimated between 20% and 30% [12]. Furthermore, the Food and Agriculture Organization (FAO) estimates that up to 40% of global crop production is lost annually to these biological threats [8]. In this context, an incorrect automated diagnosis is not merely a technical failure but a direct threat to economic stability and food security.

To manage the risks inherent in automated decision-making, such as over-reliance and the loss of human agency, Explainable AI (XAI) is frequently deployed. While XAI-powered solutions are intended to illuminate the underlying logic of a model, recent literature warns that they may inadvertently introduce safety risks known as explainability pitfalls [6]. Evidence suggests that persuasive explanations can inflate unwarranted trust or warp the mental models of users, even when designers intend no harm [3, 11]. In extreme cases, assertive design choices or simplified metrics might even function as dark patterns that invoke unearned trust [7]. For an agronomist, an explanation that discourages verification could lead to either untreated disease propagation or wasteful chemical interventions.

Moreover, current research in digital agriculture often focuses on algorithmic fidelity but lacks an understanding of how professional accountability influences interaction in the field. Motivated by this, we examine whether participant accounts point to what we later describe as a *social firewall*: professional norms and liability that may help sustain scrutiny when explanations appear persuasive. In this framing, explanations are not only informational artifacts, but also inputs to decisions that can affect professional reputation and farmer relations. By shifting the focus toward making explanations safe, we can better understand how to support human agency in high-stakes sociotechnical systems.

Therefore, this paper presents an exploratory case study investigating these dynamics through a focus group with 13 participants, including expert agronomists and students. We employed a scenario-based approach to surface the knowledge and professional norms that shape the interaction between stakeholders and explanations. Our primary contribution is a set of formative themes regarding potential explainability pitfalls in viticulture. These themes illustrate how factors such as situational vulnerability, numerical precision, and professional liability interact to shape expert agency and contestability in high-stakes environments.

The remainder of this paper is organized as follows. Section 2 establishes the theoretical background by connecting the ethics of persuasive technology to the concepts of explainability pitfalls and situated action. Section 3 details the exploratory methodology, including the focus group design and scenario development. Section 4 presents the formative themes surfaced during the analysis. Section 5 discusses the broader sociotechnical implications of these findings for design and safety. Section 6 outlines the methodological limitations and directions for future research. Finally, Section 7 concludes the paper.

2 Theoretical Background

The ethical foundation of persuasive technology suggests that the moral responsibility of a designer is tied to both the intent and the predictable consequences of a tool [1]. In the Human-Centered XAI (HCXAI) landscape, this distinction is formalized through the categorization of Dark Patterns (DPs) and Explainability Pitfalls (EPs) [6]. DPs represent intentional design choices, such as placebo explanations, created to manufacture unearned trust or specific system states. In contrast, EPs represent unanticipated negative outcomes, such as automation bias or a false sense of security, that manifest even when designers intend no harm. By anchoring our inquiry in this distinction, we shift the focus from the production of explanations toward the development of safe human-AI interactions.

This shift toward safety requires an understanding of expert reasoning within the situated context of agronomy. We ground this perspective in the notion of situated action as defined by Suchman et al., which states that human activity is not a pre-planned script but a continuous adaptation to specific social and physical circumstances [13]. In the vineyard, agronomic decision-making is shaped by temporal uncertainty and ecological variation, where the impact of a pathogen may not be visible for weeks. When model outputs are paired with explanations, they can shape what practitioners attend to and how they decide when to verify, which makes explainability pitfalls a practical safety concern in digital agriculture. Because agronomic decisions can carry uneven consequences across different types of failure modes, there is a need to examine how explanation designs interact with field routines, accountability, and uncertainty. As McGrath et al. argue, inclusive and "bottom-up" social science approaches are necessary to anticipate socio-ethical implications and mitigate negative effects of digitalization in the farming sector [9].

Furthermore, aiming to reduce explainability pitfalls in high-stakes work, explanation design can be framed around contestability. Contestability describes a system property that lets end users question, probe, and, when needed, override algorithmic outputs [14]. This lens is useful because explanations do not only describe model behavior. They can also shape when users decide to stop checking, which is one pathway through which explainability pitfalls can emerge [6]. Viewed this way, contestability links explainability to safety by shifting attention from how convincing an explanation appears to whether it supports ongoing scrutiny under uncertainty. This framing motivates our focus on how practitioners reason about when to verify, when to defer, and what kinds of information they treat as sufficient to challenge a system output.

3 Methodology

We conducted an exploratory focus group with 13 participants with a local viticulture partner in Portugal. The group included 4 senior expert agronomists and 9 students working on digital agriculture projects. This diversity in experience allowed us to examine how potential explainability pitfalls might manifest across different mental models.

The focus group followed a collective reasoning protocol designed to elicit critical reflection rather than simple preference. As moderators, we presented each scenario to the group and guided the discussion with follow-up prompts. Participants were encouraged to justify their answers, describe how they would act in practice, and discuss what evidence would change their mind. The audio of the session was recorded with participant consent.

3.1 Study Design and Scenarios

The study consisted of seven scenario-based questions designed as provocation scenarios (see Table 1). These questions used a "what if" structure to encourage participants to confront potential conflicts between their physical reality and a hypothetical system output. Since none of the participants had previously used XAI-powered solutions in their daily work, these future scenarios were required to ground the discussion and enable sociotechnical forecasting.

We grounded the design of these scenarios in established literature concerning automation bias, trust calibration, and contestability, focusing explicitly on how explanations might trigger these vulnerabilities as explainability pitfalls. Prior research suggests that users may exhibit automation bias by disproportionately deferring to algorithmic outputs when the accompanying explanations appear highly confident or fluent [10]. Furthermore, accurately calibrating trust in a predictive system is often disrupted in low-familiarity contexts, where a persuasive explanation can lead to unanticipated over-reliance [16]. To translate these concepts into practice, we mapped our questions to potential triggers of explainability pitfalls. Q2 assesses trust calibration under uncertainty to observe if system complexity acts as a proxy for authority. Q3 and Q7 were designed to understand a potential over-reliance induced by explanations, presenting high-confidence metrics and fluent text to observe the impact on the perceived authority of the explanation. Finally, Q1, Q4, Q5, and Q6 explore expert agency and contestability, exploring whether practitioners maintain active scrutiny or if exposure to automated explanations unintentionally reduces the friction required for verification.

3.2 Data Analysis and Rigor

As part of the data analysis, we used thematic coding to identify recurring patterns in the focus group discussions [2]. Two researchers conducted the coding process inductively, allowing themes to emerge directly from the transcriptions rather than imposing a pre-existing framework. Discrepancies in the coding were resolved through iterative discussion until a consensus was reached. This process helped us define themes that highlight how potential explainability pitfalls might emerge when both agronomists and students interact with hypothetical scenarios.

3.3 Methodological Considerations

The scenario-based approach was chosen to enable a provocative exploration of potential explainability pitfalls. In a real-world setting, the pressure to maintain crop health might discourage agronomists from admitting to self-doubt. By using speculative "what if" scenarios, we allowed participants to reflect on their vulnerabilities without the immediate fear of professional consequences.

We acknowledge that this exploratory approach measures self-disclosed intent rather than observed behavior. While these scenarios provide a formative step toward understanding the boundaries of human-AI collaboration, the results should be interpreted as potential risks and scrutiny patterns surfaced within this specific cohort.

4 Potential Explainability Pitfalls: Formative Themes

Our analysis surfaced five primary themes regarding the interaction between agronomists, students, and automated explanations. These themes suggest that participant scrutiny is not a static quality but a situated state influenced by modality, social context, and domain familiarity.

4.1 Situational Vulnerability

Agency in the vineyard appears to fluctuate based on the context of the specific decision. We describe this surfaced theme as situational vulnerability, primarily evidenced by the contrast between participant responses to familiar mildew detection (Q1) and unfamiliar diseases (Q2). While the agronomists in our study comfortably contested system outputs in familiar domains, their reported scrutiny appeared to diminish when facing rare pests. When faced with familiar domains, one expert noted, *"Given my experience, if I got there and saw nothing, I would say the explanation failed"* (Q1).

However, when facing uncertainty in a low-familiarity context (Q2), participants described a shift from critical auditors toward a more dependent role. This suggests that domain expertise may act as a protective shield only until the practitioner feels junior to the specific problem. As one expert admitted when facing an unfamiliar crop, *"I would have a higher tendency, perhaps, to trust the machine"* (Q2). This pattern surfaces a potential explainability pitfall where the explanation can function as a proxy for authority in low-knowledge situations, creating an unanticipated pathway to over-reliance even without manipulative intent.

4.2 High-confidence Metrics and Contestability

Participants described a potential risk where numerical confidence metrics can reduce contestability and erode human agency (Q3). In the scenario involving a 95% confidence score, both agronomists and students reported that the precision of the metric could trigger reduced scrutiny and self-doubt, even when field inspection did not confirm symptoms. Even the most experienced participants indicated that such a high probability would make them doubt their own vision. One expert remarked, *"Even having experience, I will doubt myself"* (Q3).

This theme suggests an explainability pitfall in which quantified certainty may function as a proxy for authority, making it harder for users to challenge the system output. We acknowledge that Q1 and Q3 differ in both modality and diagnostic context, so this finding should be read as a qualitative pattern rather than a causal comparison between visual and numerical explanations.

4.3 Social Liability as a Systemic Firewall

Participants described social liability as a protective factor that can counter automation complacency in this community of practice (Q6). This social firewall is built on the professional accountability agronomists hold toward farmers, where agronomists remain the visible decision-makers and therefore internalize the consequences of system failure. This accountability appears to sustain a baseline of scrutiny and field verification, especially under the asymmetric cost of error in agriculture. One expert highlighted this pressure by stating, *"if this goes wrong, the farmer will scream at the machine and the machine does not care. Now, screaming at us, that is a real risk... if things go wrong, there is a big problem for us"* (Q6).

In the same discussion, participants contrasted the initial phase of deployment with long-term integration. The group projected that while they would feel obliged to re-check each alert during the first year, they might lower their guard after repeated exposure to accurate recommendations. We treat this as an anticipated risk of reduced vigilance rather than an observed behavioral effect. This theme surfaces a potential sociotechnical condition that anchors system use in professional contestability. However, relying on this pressure to naturally mitigate explainability pitfalls over time may be optimistic. We therefore view the social firewall as a contingent safeguard that likely requires design reinforcement.

However, this same liability appears to create an asymmetry in risk posture. Participants indicated that missed detections carry a higher social cost than unnecessary alerts due to the visible damage to the crop and reputation. Consequently, the social firewall may drive practitioners toward conservative decision-making, where the fear of social penalty leads them to tolerate false alarms rather than risk a missed diagnosis. Even so, we recognize this tolerance likely has a threshold. Excessive false positives may risk undermining the user reliance on the system's recommendations.

ID	Question	Design Intent
Q1	Imagine a model detects Mildew and shows a Heatmap as explanation. If you go to the field and see no symptoms, do you think the model saw something you missed or that the explanation is wrong? Why?	Analyzes whether users maintain the agency to reject a visual explanation when it conflicts with physical field observations.
Q2	Imagine the system diagnoses a rare disease you have little experience with. In this situation, do you feel more confident because the system seems to "know a lot", or do you become more suspicious?	Tests if system complexity serves as a proxy for authority and induces over-reliance in low-knowledge scenarios.
Q3	The model identifies a pest with a confidence score of 95%, but your physical inspection finds no symptoms. Does this metric act as an authority that makes you doubt your own vision?	Tests the resilience of expert agency against the perceived objectivity of high-precision metrics.
Q4	Considering new winegrowers using automated explanations over a 5-year horizon: Do you think frequent exposure leads to a decline in field-based assessment? Does the tool function as pedagogical support?	Forces participants into long-term sociotechnical forecasting regarding deskilling and scrutiny decay.
Q5	If the system recommends treatment based on 18°C but would recommend no treatment at 20°C: would you rather see counterfactuals, or be required to manually confirm the data before accepting?	Probes the trade-off between transparency and safety-forcing manual validation.
Q6	Imagine the model correctly explains a risk in Vineyard A. Immediately afterwards, it issues an identical alert for Vineyard B. Does the earlier success make you "lower your guard," or do you feel obliged to audit again?	Tests the durability of human scrutiny over repeated successful interactions with the system.
Q7	An LLM generates a fluent textual explanation: "*Cross-analysis of humidity and temperature indicates ideal conditions...*" Does this articulate style change your perception of the facts?	Tests if articulate linguistic fluency induces unwarranted trust or masks potential hallucinations.

Table 1: Scenario-based questions used to explore potential explainability pitfalls and expert agency, mapped to their design intent.

4.4 Linguistic Skepticism and Audience Segmentation

This pilot study surfaced a protective skepticism toward fluent textual explanations produced by generative tools (Q7). Participants, particularly students, reported that the assertive and articulate language in generative tools can increase concern that the system may hallucinate. One student observed that *"the more it is forced to be articulate, the more it might start making things up and hallucinating"* (Q7). This skepticism may act as a protective factor against fluency-related explainability pitfalls.

We also observed audience segmentation in transparency needs (Q7). Expert participants suggested that short textual summaries might be sufficient for farmers, while agronomists require validation data that supports professional scrutiny. This suggests a potential explainability pitfall if simplification hides uncertainty from stakeholders who bear financial risk. Consequently, this could become a dark pattern risk if uncertainty is intentionally obscured to drive compliance or trust.

4.5 Intergenerational Deskilling and Probabilistic Auditing

Participants anticipated a potential long-term risk of deskilling, described by some as an atrophy of the "clinical eye", under continuous use of automated explanations (Q4). Agronomists expressed concern that, since future practitioners *"will grow up with these new technologies, they might not need to depend so much on their own eye"* (Q4). In the same discussion space, some novice students described a tendency to prioritize passive acceptance of automated explanations over manual verification, which aligns with a projected shift from active auditing to passive reliance.

To address this risk, participants proposed a practitioner-led probabilistic auditing protocol as a design consideration (Q5, Q7). Drawing an analogy to quality control in manufacturing, one student suggested that *"in a hundred diagnoses and associated explanations, at least in fifteen you have to go and confirm in the field"* (Q7). While the specific 15% threshold is an untested participant suggestion, the core idea is to introduce routine, randomized verification that sustains active skill use and contestability over time.

5 Discussion

The themes surfaced in this exploratory pilot study suggest that the design goal for explanations in high-stakes domains might benefit from a shift toward safety rather than uncritical trust. Participants described how high-confidence metrics and low-familiarity contexts can create explainability pitfalls by increasing over-reliance and reducing contestability. To mitigate these risks, we reflect on how designers might prioritize professional contestability over seamless integration. In the following subsections, we discuss these implications across four key areas: the impact of situational vulnerability on expert agency, the relationship between numerical precision and active scrutiny, the role of professional accountability in sustaining contestability, and the risks associated with audience segmentation and simplification.

5.1 Situational Vulnerability and Situated Information

The surfaced pattern of situational vulnerability suggests that domain expertise is not a permanent safeguard. Participants described that when practitioners move beyond familiar experience zones, they may shift from critical auditors to more dependent users. This generates a design hypothesis regarding the need for situated information that helps users recognize when their own uncertainty is high. Rather than generic heatmaps, future systems could support agronomic reasoning by explicitly aligning predictions with historical precedents and local variety sensitivity. Such signals may help counter authority bias by prompting scrutiny when the user feels less qualified to challenge the system.

5.2 Numerical Precision and Probabilistic Auditing

While we acknowledge the confounded variables in our scenarios, participants' qualitative reasoning highlights quantified certainty as a potential fault-line for self-doubt. This could be understood as an explainability pitfall because numerical precision can function as a proxy for authority and make contestation feel illegitimate. To manage the projected decay of scrutiny over time, we reflect on the participant-informed suggestion of probabilistic auditing. Participants drew analogies to industrial quality control, including an automotive factory example, to propose randomized field checks as a form of strategic workflow friction. This design direction aims to preserve field-based expertise through routine verification rather than routine acceptance.

5.3 Professional Accountability and Social Contestability

Our analysis suggests that contestability in the vineyard is not merely an interface feature but also a social process. The agronomists' suggestion of a potential social firewall indicates that professional accountability to farmers can sustain scrutiny, even under repeated system success. Designers might support this safeguard by acknowledging the agronomist as the final authority and by providing interaction mechanisms that fit existing verification practices. This approach recognizes that safety is sustained by professional norms and the situated social reality of agricultural work rather than technical features alone.

5.4 Audience Segmentation and Simplification Risks

Finally, we reflect on the tension identified in audience segmentation. While simplifying explanations for farmers may improve immediate clarity, it introduces a risk of simplification bias if uncertainty is hidden from the stakeholder who bears the financial risk of a wrong decision. This could become an explainability pitfall when it occurs unintentionally, and it becomes a dark pattern risk only if simplification is deliberately used to drive compliance or trust. Design should balance clarity with the need to provide situated information that supports informed decision-making for all stakeholders.

We also reflect on the tension identified regarding participant skepticism toward fluent generative text. This skepticism presents an interesting contrast to recent findings in conversational explainable AI. For instance, Zhang et al. observed that non-AI experts often leverage conversational interfaces to actively negotiate meaning and build mental models of predictive systems [17]. In the context of precision viticulture, this creates a complex dynamic. While students demonstrated caution, if an agronomist engages in an iterative dialogue with a highly articulate system, the conversational fluency might initially encourage engagement but could ultimately mask underlying system uncertainty. A practitioner, swayed by the authoritative tone of the dialogue, might lower their guard and accept a flawed recommendation without field verification. Therefore, while conversational explanations can aid comprehension, their linguistic fluency risks amplifying the explainability pitfalls associated with unearned trust.

5.5 Mitigating Potential Explainability Pitfalls

To address the practical implications of these findings, it is helpful to explore how the identified themes might inform actionable design mitigations. By shifting the focus from simply generating explanations toward ensuring those explanations are safe, designers might better mitigate the potential downstream effects of automated systems. We outline practitioner-led mechanisms mapped directly to the core themes surfaced in our focus groups. To address the risks of numerical precision acting as a false authority, interfaces could replace standalone percentages with historical baselines or visual counterfactuals. This strategy aims to anchor the data and encourage the user to compare the algorithmic output against physical reality before making a final judgment.

Furthermore, to prevent the long-term decay of the clinical eye, it would be interesting for key farm stakeholders such as agronomists to implement probabilistic auditing. Rather than relying on the individual's motivation to contest the system, the tool itself could periodically prompt mandatory manual field checks for highly confident routine predictions. This mechanism essentially embodies the concept of social firewalls, transforming professional accountability into a system feature that resists potential intergenerational deskilling.

6 Limitations and Future Work

We acknowledge several methodological constraints in this exploratory work. As a formative pilot study, the sample size of 13 participants from a single viticultural region in Portugal limits the immediate generalization of these themes across agricultural cultures and operational contexts. The use of speculative "what if" scenarios introduces the risk of hypothetical bias. While the social firewall emerged as a strong participant-reported norm, the protective scrutiny described here represents reported intent rather than observed adherence under real vineyard pressures (e.g., exhaustion, time constraints, or repeated alerts). Furthermore, the current study design does not allow us to causally isolate explanation modality effects, because explanation formats (e.g., a numerical confidence score) were bundled with specific agricultural tasks. This bundling also applies to the magnitude of quantified certainty: presenting a very high confidence (e.g., 95%) may elicit different contestation dynamics than a more ambiguous value (e.g., 72%), so the observed self-doubt should be interpreted as a potential risk signal rather than definitive behavioral proof.

Future research should address these limitations through more controlled and higher-fidelity evaluations. We plan to develop a comparative framework that applies multiple modalities (text, visual, numerical) to the same diagnostic scenario while systematically varying confidence levels (e.g., moderate vs. high). This will allow us to decouple modality effects from task difficulty and from the authority conveyed by precision. We also aim to test the stability of the social firewall using high-fidelity simulations that introduce more realistic pressures (e.g. workloads and financial consequences). Additionally, a separate line of inquiry will focus on the potential emergence of dark patterns in this domain. We intend to research how design choices, such as assertive language or simplified metrics, might be intentionally used to bypass professional scrutiny or manufacture unearned trust.

7 Conclusion

In this exploratory, scenario-based case study, we surfaced potential explainability pitfalls by examining the distinct mental models of agronomists and novice students. This participant diversity allowed us to observe how expert agency functions as a situated state rather than a fixed trait. Participants described that their willingness to challenge the system can diverge by situation, particularly when they face unfamiliar cases or when the system presents very high confidence scores. These instances highlight specific moments where automated explanations might unintentionally erode human scrutiny. Conversely, participants described a strong protective norm tied to social accountability, where agronomists remain the visible decision-makers and expect that farmers will hold them accountable for negative consequences.

Overall, our findings suggest that explainability pitfalls in this domain may manifest as a tension between immediate trust and long-term safety. While students described a protective skepticism toward fluent generative text, agronomists expressed concern that future practitioners might experience an atrophy of the clinical eye. To address these projected risks of deskilling and reduced vigilance, the group suggested probabilistic auditing through randomized field checks to keep verification skills active. We conclude that these themes represent formative signals of explainability pitfalls that require validation in future studies to separate modality effects from task context. Ultimately, safety in the vineyard may depend less on seamless explanations and more on maintaining the friction required for professional contestability.

Acknowledgments

This work was supported by **Fundação para a Ciência e a Tecnologia (FCT)**, I.P. (FCT.IP), through the Ph.D. Studentship grant with project reference 10.54499/PRT/BD/154548/2023, and **NOVA LINCS** (UID/04516/2025) with the financial support of FCT.IP.

References

[1] Daniel Berdichevsky and Erik Neuenschwander. 1999. Toward an ethics of persuasive technology. *Commun. ACM* 42, 5 (May 1999), 51–58. doi:10.1145/301353.301410

[2] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa _eprint: https://doi.org/10.1191/1478088706qp063oa.

[3] Rachele Carli, Amro Najjar, and Davide Calvaresi. 2022. Risk and Exposure of XAI in Persuasion and Argumentation: The case of Manipulation. In *Explainable and Transparent AI and Multi-Agent Systems*, Davide Calvaresi, Amro Najjar, Michael Winikoff, and Kary Främling (Eds.). Vol. 13283. Springer International Publishing, Cham, 204–220. doi:10.1007/978-3-031-15565-9_13 Series Title: Lecture Notes in Computer Science.

[4] Renato L. F. Cunha, Bruno Silva, and Marco A. S. Netto. 2018. A scalable machine learning system for pre-season agriculture yield forecast. *Proceedings - IEEE 14th International Conference on eScience, e-Science 2018* (Dec. 2018), 423–430. doi:10.1109/ESCIENCE.2018.00131 ISBN: 9781538691564.

[5] Giovanni Delnevo, Roberto Girau, Chiara Ceccarini, and Catia Prandi. 2022. A Deep Learning and Social IoT Approach for Plants Disease Prediction Toward a Sustainable Agriculture. *IEEE Internet of Things Journal* 9, 10 (May 2022), 7243–7250. doi:10.1109/JIOT.2021.3097379

[6] Upol Ehsan and Mark O. Riedl. 2024. Explainability pitfalls: Beyond dark patterns in explainable AI. *Patterns* 5, 6 (June 2024), 100971. doi:10.1016/j.patter.2024.100971

[7] Malin Eiband, Daniel Buschek, Alexander Kremer, and Heinrich Hussmann. 2019. The Impact of Placebic Explanations on Trust in Intelligent Systems. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3312787

[8] FAO. 2022. *FAO's Plant Production and Protection Division.* FAO, Rome, Italy. doi:10.4060/cc2447en

[9] Karen McGrath, Claire Brown, Áine Regan, and Tomás Russell. 2023. Investigating narratives and trends in digital agriculture: A scoping study of social and behavioural science studies. *Agricultural Systems* 207 (April 2023), 103616. doi:10.1016/j.agsy.2023.103616

[10] Giuseppe Romeo and Daniela Conti. 2026. Exploring automation bias in human–AI collaboration: a review and implications for explainable AI. *AI & SOCIETY* 41, 1 (Jan. 2026), 259–278. doi:10.1007/s00146-025-02422-7

[11] Mersedeh Sadeghi, Daniel Pöttgen, Patrick Ebel, and Andreas Vogelsang. 2024. Explaining the Unexplainable: The Impact of Misleading Explanations on Trust in Unreliable Predictions for Hardly Assessable Tasks. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization.* ACM, Cagliari Italy, 36–46. doi:10.1145/3627043.3659573

[12] Serge Savary, Laetitia Willocquet, Sarah Jane Pethybridge, Paul Esker, Neil McRoberts, and Andy Nelson. 2019. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* 3, 3 (March 2019), 430–439. doi:10.1038/s41559-018-0793-y

[13] Lucille Alice Suchman. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions.* Cambridge University Press.

[14] Kristen Vaccaro, Karrie Karahalios, Deirdre K. Mulligan, Daniel Kluttz, and Tad Hirsch. 2019. Contestability in Algorithmic Systems. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing.* ACM, Austin TX USA, 523–527. doi:10.1145/3311957.3359435

[15] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal. 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* 177 (Oct. 2020), 105709. doi:10.1016/j.compag.2020.105709

[16] Magdalena Wischnewski, Nicole Krämer, and Emmanuel Müller. 2023. Measuring and Understanding Trust Calibrations for Automated Systems: A Survey of the State-Of-The-Art and Future Directions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* ACM, Hamburg Germany, 1–16. doi:10.1145/3544548.3581197

[17] Tong Zhang, Mengao Zhang, Wei Yan Low, X. Jessie Yang, and Boyang Albert Li. 2025. Conversational Explanations: Discussing Explainable AI with Non-AI Experts. In *Proceedings of the 30th International Conference on Intelligent User Interfaces.* ACM, Cagliari Italy, 409–424. doi:10.1145/3708359.3712143