UNIVERSIDADE PAULISTA - UNIP

MARCUS VINICIUS LEITE

INTEGRATING LARGE LANGUAGE MODELS AND PARACONSISTENT ANNOTATED EVIDENTIAL LOGIC Eτ INTO DECISION SUPPORT SYSTEMS FOR RELIABLE DECISION-MAKING IN POULTRY FARMING UNDER UNCERTAINTY AND CONTRADICTION

Dissertation presented to the Graduate Program in Production Engineering at Universidade Paulista – UNIP, to obtain a Master's degree in Production Engineering.

São Paulo

2025

MARCUS VINICIUS LEITE

INTEGRATING LARGE LANGUAGE MODELS AND PARACONSISTENT ANNOTATED EVIDENTIAL LOGIC Eτ INTO DECISION SUPPORT SYSTEMS FOR RELIABLE DECISION-MAKING IN POULTRY FARMING UNDER UNCERTAINTY AND CONTRADICTION

Dissertation presented to the Graduate Program in Production Engineering at Universidade Paulista – UNIP, to obtain a Master's degree in Production Engineering.

Field of Concentration: Operating Systems Management.

Research Line: Quantitative Methods in Production Engineering.

Research Project: Quantitative, Computational, and Technological Methods in Production Engineering.

Advisor: Prof. Dr. Jair Minoro Abe.

São Paulo

2025

CATALOGING-IN-PUBLICATION DATA

Leite, Marcus Vinicius.
Integrating large language models and paraconsistent annotated evidential logic $E\tau$ into decision support systems for reliable decision-making in poultry farming under uncertainty and contradiction / Marcus Vinicius Leite. - 2025.
120 f. : il. color.

Dissertação de Mestrado Apresentada ao Programa de Pós Graduação em Engenharia de Produção da Universidade Paulista, São Paulo, 2025.
Área de concentração: Gestão de Sistemas de Operação.
Orientador: Prof. Dr. Jair Minoro Abe.

1. Paraconsistent annotated evidential logic $E\tau$. 2. Decision support systems. 3. Retrieval-augmented generation. 4. Large language models. 5. Poultry farming. 6. Explainable artificial intelligence. I. Abe, Jair Minoro (orientador). II. Título.

Ficha elaborada pelo Bibliotecária Salete Marques CRB8-5102

MARCUS VINICIUS LEITE

INTEGRATING LARGE LANGUAGE MODELS AND PARACONSISTENT ANNOTATED EVIDENTIAL LOGIC Eτ INTO DECISION SUPPORT SYSTEMS FOR RELIABLE DECISION-MAKING IN POULTRY FARMING UNDER UNCERTAINTY AND CONTRADICTION

Dissertation presented to the Graduate Program in Production Engineering at Universidade Paulista – UNIP, to obtain a Master's degree in Production Engineering.

Approved on:_____/_____/2025

EXAMINING COMMITTEE

____
Advisor: Prof. Dr. Jair Minoro Abe, Dr.
Universidade Paulista (UNIP) - Brazil

____
Irenilza de Alencar Nääs, Dr.
Universidade Paulista (UNIP) - Brazil

____
Danilo Florentino Pereira, Dr.
Universidade Estadual Paulista (UNESP)

DEDICATION

I dedicate this work to my parents, Olavo Leite (in memoriam) and Aderli Leite, for awakening in me, from an early age, the pursuit of knowledge and science.

To my life partner, Valeria, for her silent strength, constant love, and inspiring example of resilience, perseverance, and integrity.

To my little daughter, Ana, for showing love and care by constantly reminding me of the importance of health throughout this journey.

To my brother by choice, Marcos Hoffmann, for convincing me to follow the path of the master's degree, guiding me on how to walk it, and accompanying me with knowledge and wisdom at every stage. His high standards constantly challenge me to do more and better.

To my mentor and mystagogue, Jair Minoro Abe, who took on the task of guiding me through paths I could not have imagined. This discipleship has been, for me, both an honor and a great responsibility.

And finally, to all those who, in some way, contributed to making this dream come true, reaffirming my belief that education and science are powerful pathways to transforming lives.

ACKNOWLEDGMENTS

To my advisor, Prof. Dr. Jair Minoro Abe, for his steady guidance, clarity in direction, and competence in leading this work with scientific and methodological rigor.

To Prof. Dr. Marcos Leandro Hoffmann Souza, from Universidade do Vale do Rio dos Sinos (Unisinos), for his initial encouragement to pursue the master's degree and his strategic guidance as co-advisor. His availability to closely follow each stage of this dissertation was invaluable.

To my co-advisor Prof. Dr. Irenilza Naas, from UNIP, for identifying the research problem, sharing her specialized knowledge, and engaging in the fundamental exchange of ideas that shaped this study "for her specialized knowledge and insightful contributions that shaped this study and was decisive for the direction and consistency of this work".

To Prof. Dr. Thayla Morandi Ridolfi de Carvalho Curi, for her valuable contributions that strengthened the results presented here.

To my colleagues Leandro Cigano de Souza Thomas, Kennya Vieira Queiroz and João Glauco Barbosa dos Santos for their partnership, technical support and the discussions that helped improve this work.

To my wife, Valeria, for her constant understanding and support, and to my daughter, Ana, for her care and encouragement to maintain balance throughout this journey.

To the Graduate Program in Production Engineering at Universidade Paulista (UNIP), for the opportunity to develop this research and for providing an academic environment of excellence.

To the Coordination for the Improvement of Higher Education Personnel (CAPES), for the PROSUP scholarship, which enabled full dedication to research through its Support Program for Graduate Studies in Private Higher Education Institutions – PROSUP.

Finally, I extend my sincere appreciation to all those who, directly or indirectly, contributed to the execution of this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Logic Eτ	Paraconsistent Annotated Evidential Logic Eτ
LLM	Large Language Model
DSS	Decision Support System
RAG	Retrieval-Augmented Generation
C	Complement (1 - 1B)
FTc	Certainty Tolerance Factor
FTct	Contradiction Tolerance Factor
Gce	Certainty Degree
Gin	Uncertainty Degree
QV→⊤	Quasi-True Tending to Inconsistent
QV→⊥	Quasi-True Tending to Paracomplete
QF→⊤	Quasi-False Tending to Inconsistent
QF→⊥	Quasi-False Tending to Paracomplete
Q⊤→V	Quasi-Inconsistent Tending to True
Q⊤→F	Quasi-Inconsistent Tending to False
Q⊥→V	Quasi-Paracomplete Tending to True
Q⊥→F	Quasi-Paracomplete Tending to False
Vscc	Upper Certainty Control Value
Vicc	Lower Certainty Control Value
Vscct	Upper Contradiction Control Value
Vicct	Lower Contradiction Control Value

ABSTRACT

Poultry production has become increasingly complex due to environmental variability, high-density farming, and sustainability demands, creating decision environments marked by uncertainty, contradiction, and fragmented or domain-dependent information. Conventional decision support systems (DSS) often fail to ensure consistency and interpretability under such conditions. This research develops and evaluates an integrative method that combines Paraconsistent Annotated Evidential Logic Eт (Logic Eт), Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs) agents to enable resilient and explainable reasoning for decision-making in poultry farming. The research followed a cumulative three-stage design: (i) a systematic literature review identifying conceptual and technological gaps; (ii) controlled experiments assessing the influence of RAG on LLM performance; and (iii) the modeling, implementation, and validation of a conversational DSS integrating Logic Eт-based inference with a state-of-the-art large language model. Evaluation based on semantic similarity, contextual relevance, and logical-evidential consistency confirmed that the integrated architecture remained robust even under conflicting or incomplete evidence. The study establishes Logic Eт as a computational foundation for trustworthy and resilient AI-based DSS, operationalizing it within a modern AI framework that enhances explainability and governance in agricultural production processes, particularly poultry farming.

Keywords: Paraconsistent Annotated Evidential Logic Eт; Decision Support Systems; Retrieval-Augmented Generation; Large Language Models; Poultry Farming; Explainable Artificial Intelligence.

RESUMO

A produção avícola tornou-se progressivamente mais complexa devido à variabilidade ambiental, à elevada densidade produtiva e às exigências de sustentabilidade, configurando ambientes decisórios marcados por incerteza, contradição e informações fragmentadas ou dependentes de domínio. Os sistemas de suporte à decisão (SSD) convencionais frequentemente não conseguem garantir consistência e interpretabilidade nessas condições. Esta pesquisa desenvolve e avalia um método integrativo que combina a Lógica Paraconsistente Anotada Evidencial Eτ (Lógica Eτ), a Recuperação Aumentada (RAG) e agentes baseados em Modelos de Linguagem de Grande Escala (LLMs), com o objetivo de possibilitar um raciocínio resiliente e explicável aplicado à tomada de decisão na avicultura. A pesquisa foi conduzida em três etapas cumulativas: (i) revisão sistemática da literatura para identificação de lacunas conceituais e tecnológicas; (ii) experimentos controlados para avaliar a influência da RAG no desempenho dos LLMs; e (iii) modelagem, implementação e validação de um SSD conversacional que integra a inferência lógico-evidencial da Lógica Eτ a um modelo de linguagem de última geração. A avaliação, baseada em métricas de similaridade semântica, relevância contextual e consistência lógico-evidencial, confirmou que a arquitetura integrada manteve desempenho robusto mesmo sob evidências conflitantes ou incompletas. O estudo consolida a Lógica Eτ como base computacional para SSD de inteligência artificial confiáveis e resilientes, operacionalizando-a em um arcabouço contemporâneo de IA que aprimora a explicabilidade e a governança em processos produtivos agrícolas, com ênfase na avicultura.

Palavras-chave: Lógica Paraconsistente Anotada Evidencial Eτ; Sistemas de Suporte à Decisão; Recuperação Aumentada; Modelos de Linguagem de Grande Escala; Avicultura; Inteligência Artificial Explicável.:

UTILITY

This research investigates the integration of Artificial Intelligence and Paraconsistent Logic in Decision Support Systems designed to address complex decision-making challenges in intensive poultry farming environments characterized by incomplete, ambiguous, and contradictory data.

Its contributions are expressed in three complementary dimensions: scientific, productive, and social.

In the scientific domain, it expands knowledge on reasoning and inference in advanced artificial intelligence by applying non-classical logics to augmented generative models for the formal treatment of uncertainty and contradiction, consolidating this integration as an architecture for knowledge-based decision support systems.

In the agricultural sector, particularly poultry farming, it proposes an adaptive approach embodied in a DSS designed for producers, capable of supporting the optimization of production processes with greater quality, predictability, and operational safety.

In the social dimension, the study reinforces sustainable practices, promotes animal welfare, and supports regulatory compliance, generating direct impacts on food security and sectoral governance.

These advances align with the global goals established by the United Nations Sustainable Development Goals (UN SDGs) (United Nations General Assembly, 2015).

This research employs advanced artificial intelligence to address complex infrastructure challenges and foster technological modernization and operational resilience, contributing to SDG 9 – Industry, Innovation, and Infrastructure.

It enhances productivity and food security through environmental control and precise husbandry management, supporting SDG 2 – Zero Hunger and Sustainable Agriculture.

For SDG 13 – Climate Action, it provides mechanisms for rapid response to unexpected events and for mitigating adverse environmental impacts.

By promoting efficiency and reducing waste in resource utilization, it supports SDG 12 – Responsible Consumption and Production.

Furthermore, it supports SDG 15 – Life on Land, by ensuring animal welfare conditions compatible with the physiological needs of poultry.

Finally, by enhancing transparency and accountability in production management, promoting regulatory compliance and strengthening governance standards, the research contributes to SDG 16 – Peace, Justice, and Strong Institutions.

By articulating scientific advances, practical applications for the production sector, and socio-environmental responsibility, this study transcends a purely technological scope and contributes to consolidating a poultry production model that integrates innovation, efficiency, and a strong commitment to sustainability.

CHAPTER I

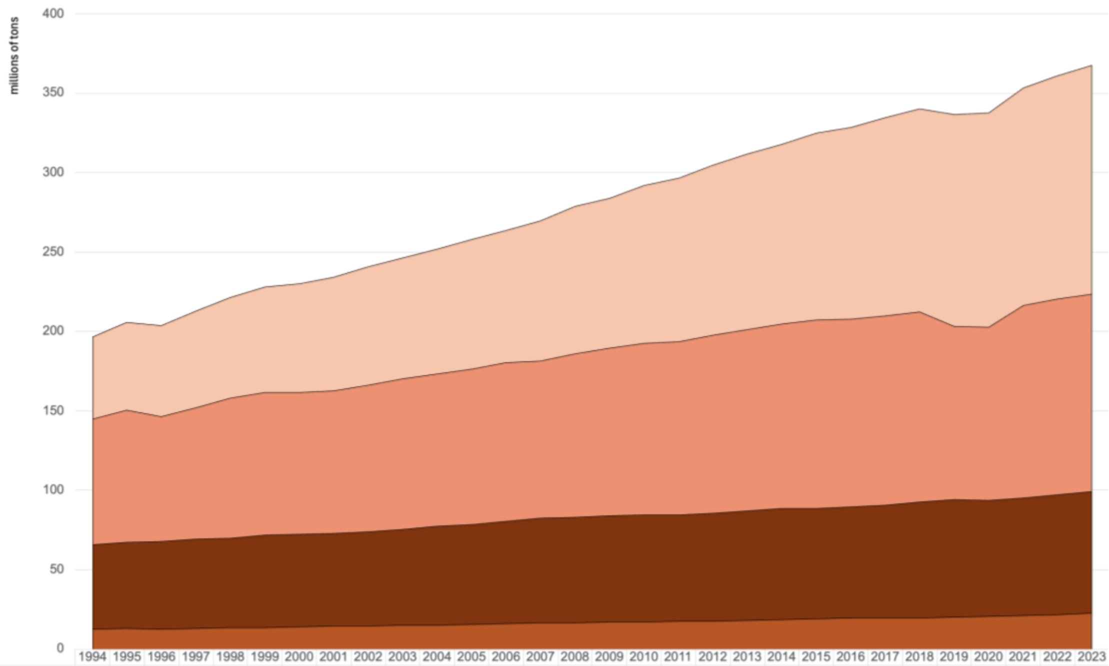
1 INTRODUCTORY CONSIDERATIONS

This chapter provides a contextualization of the research, presenting the study's context and its interactions with various fields of knowledge. It also includes the research rationale, objectives, methodology, and the structure of the thesis.

1.1 Introduction

Global animal protein production has consistently grown over the past decades (Figure 1), driven by the demographic and economic growth, urbanization, and changes in dietary habits (Figure 2). It is estimated that this demand may rise by up to 70% by 2050, with poultry meat becoming the main source of animal protein consumed worldwide (FAO, 2022; Mottet & Tempio, 2017; Berckmans, 2017).

Figure 1. Global Meat Production, Last 30 Years (million tons)

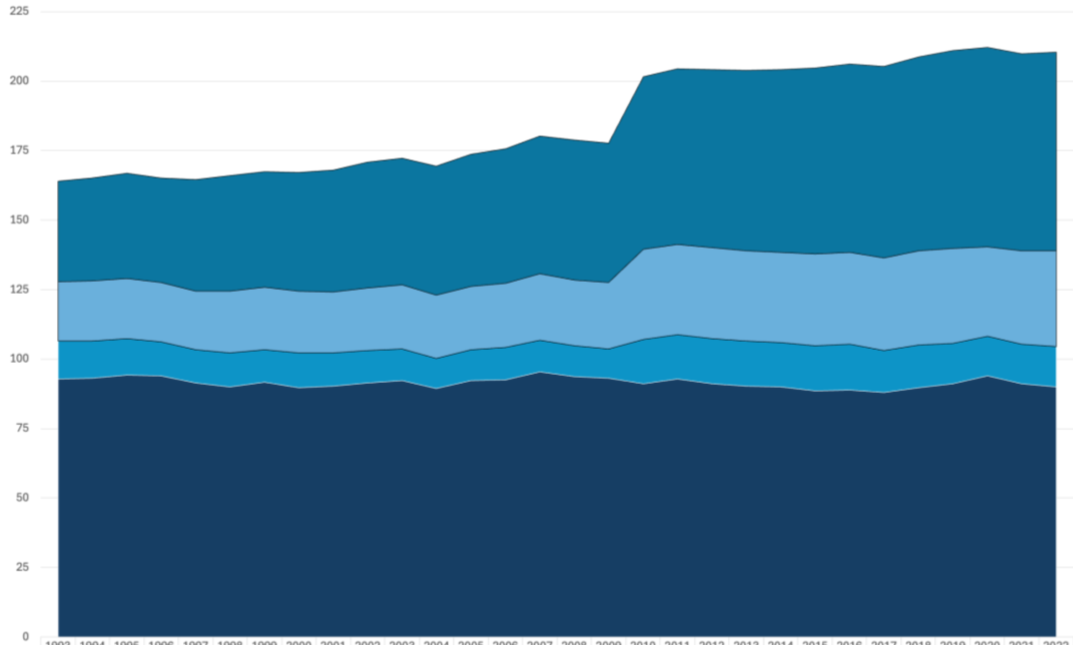


	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Poultry	52	55	57	61	63	66	69	71	75	76	79	81	83	89	93
Pork	79	83	79	83	88	90	89	90	92	95	96	98	100	99	103
Beef & Buffalo	54	54	55	57	56	58	58	58	59	60	62	63	64	66	67
Sheep, Goat & Others	12	13	13	13	13	14	14	14	14	15	15	16	16	16	17

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Poultry	95	99	103	107	110	113	118	121	125	128	134	135	137	141	144
Pork	105	108	109	112	114	116	119	118	119	120	109	109	121	123	125
Beef & Buffalo	67	67	67	68	69	70	70	70	71	73	74	73	74	75	77
Sheep, Goat & Others	17	17	17	17	18	19	19	19	20	20	20	21	21	22	22

Source: Prepared by the author, based on FAOSTAT (2024).

Figure 2. Global Per Capita daily Protein Intake, last 30 years (in g)



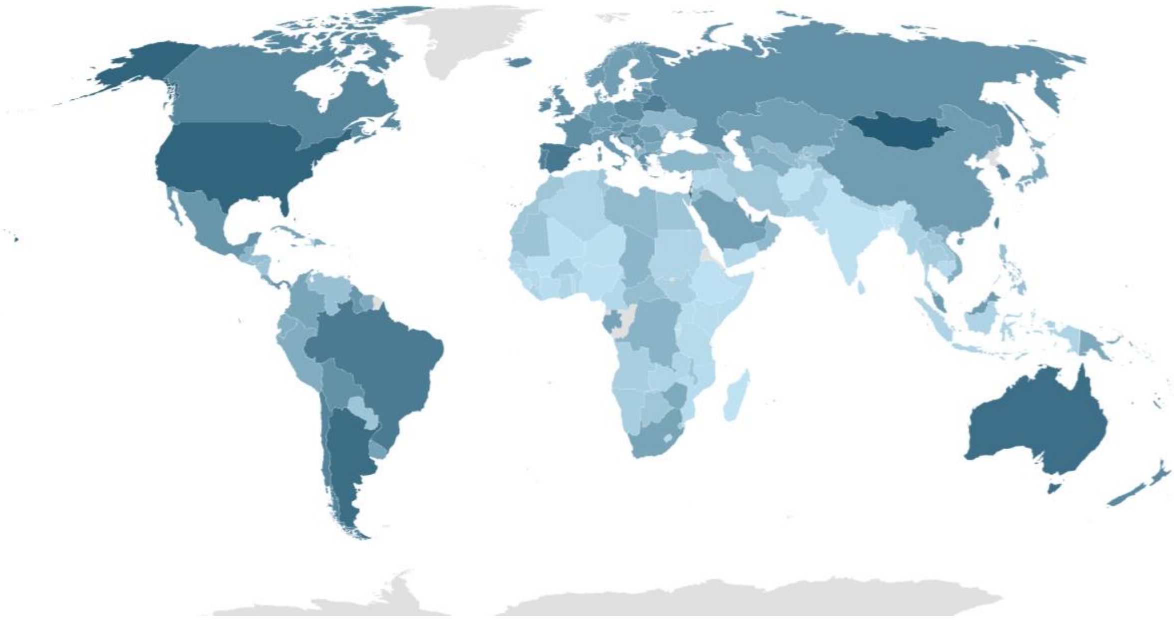
	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Poultry	36	37	38	38	40	41	42	43	44	45	46	46	47	48	50
Pork	21	21	22	21	21	22	22	22	22	22	23	23	23	23	24
Sheep, Goat & Others	14	14	13	12	12	12	12	12	12	12	11	11	11	11	12
Beef, Buffalo	93	93	94	94	91	90	92	90	90	91	92	89	92	93	95

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Poultry	50	50	62	63	64	65	66	67	68	69	70	71	72	71	72
Pork	24	24	32	32	33	33	32	33	33	33	34	34	32	34	34
Sheep, Goat & Others	11	11	16	16	16	16	16	16	16	15	15	14	14	14	15
Beef, Buffalo	94	93	91	93	91	90	90	89	89	88	90	91	94	91	90

Source: Prepared by the author, based on FAOSTAT (2024).

This growth phenomenon is particularly significant in developing regions such as Asia and South America (Figure 3), where the increase in per capita income has favored more frequent consumption of animal-based products, especially poultry (FAO, 2022; Mottet & Tempio, 2017; Berckmans 2009, 2017)

Figure 3. Global Per Capita Daily Meat Availability for Consumption (g/day), 2022



Source: Prepared by the author, based on FAOSTAT (2024).

This accelerated growth in poultry production has led to significant structural transformations in the production chain, with the consolidation of large-scale intensive systems characterized by high density and increasingly shorter production cycles (FAO, 2022; ABPA, 2024; Gržinić et al., 2023). Currently, about 92% of global poultry production takes place in intensive systems (FAO, 2022; Mottet & Tempio, 2017; Berckmans, 2009; Berckmans, 2017). Brazil exemplifies this process, being the world's largest exporter and the second-largest producer, with more than 50,000 integrated producers operating under strict sanitary and quality standards (ABPA, 2024). This model has enabled Brazilian production to grow by more than 1087% over the past four decades, while exports have increased by more than 3040% (ABPA, 2024).

The increase in production scale, combined with the high density of flocks, poses significant challenges to management, animal welfare, health, and environmental sustainability. Inadequately controlled environments can compromise zootechnical performance, food safety, and regulatory compliance (Curi et al., 2017; Hafez & Attia, 2020). Variables such as temperature, humidity, air velocity, gas concentration, feeding conditions, management practices, and disease incidence interact dynamically and interdependently, directly affecting bird health, welfare, and productivity (Gržinić et al., 2023; Pereira & Nääs, 2008; Martinez et al., 2021; Qi et al., 2023).

To address these challenges, initiatives have emerged that integrate digital technologies into the monitoring and management of animal production systems, including poultry farming. From this perspective, intelligent systems have been applied to detection and monitoring layers (IoT sensors, computer vision, and acoustic analysis), generating large volumes of environmental and behavioral data (Astill et al., 2020; Zheng et al., 2021; Dewanto, Munadi & Tauviqirrahman, 2019; Lashari et al., 2018). Recent advances also demonstrate the use of convolutional neural networks, deep reinforcement learning, and support vector machines to identify behaviors, stress, and clinical signs in birds (Halachmi et al., 2019; Raikov & Abrosimov, 2022; Ojo et al., 2022).

Despite these advances, available solutions remain fragmented at the analysis and decision-support stages. Specialized modules (such as vision, acoustics, and climate) operate as technological silos, with low interoperability, limited incorporation of contextual knowledge (flock history, zootechnical objectives, seasonality, operational constraints), and restricted capacity to integrate information from heterogeneous sources. Moreover, the complexity of decision-making in intensive poultry farming involves multiple decision domains, extending beyond environmental control to encompass nutrition, health, welfare, and management, across operational, tactical, and strategic levels (Zhai et al., 2020; Rossi, Caffi & Salinari, 2012). Factors such as climate, market dynamics, and animal behavior make the decision-making process uncertain and, in many cases, contradictory (Hamsa & Bellundagi, 2017; Berckmans, 2009; Cheng, McCarl & Fei, 2022).

This fragmentation reflects a structural challenge already recognized in the field of decision science: the dissociation between predictive modeling and the decision-making process itself. Recent studies have highlighted the need to integrate symbolic reasoning, contextual inference, and decision-oriented learning, consolidating a "decision-focused" approach, in which the value of a model is assessed not only by its predictive accuracy but by the quality of the decisions it supports (Mandi et al., 2023).

Consequently, the recommendations generated by these systems tend to be locally effective but systemically disjointed, resulting in latency between detection and action and difficulty in prioritizing trade-offs when signals are ambiguous. The limitations imposed by this fragmentation become even more severe in the frequent presence of incomplete information (sensor failures, misaligned time windows) or contradictory data (divergent readings between sources, conflicts between model outputs and field observations), creating a critical point for decision-making (Gržinić et al., 2023; Hafez & Attia, 2020).

Specialized support constitutes the main resource for integrating information from different decision domains within the production system. Technical consultants play a relevant role in interpreting heterogeneous, and often, unprecise and incomplete data and formulating practical recommendations, acting as mediators between technological monitoring systems and decision-making in operations.

However, this type of support presents limitations that compromise its effectiveness. First, it often relies on retrospective analyses based on historical data, leading to reactive responses with limited predictive value in dynamic environments (Lashari et al., 2018; Halachmi et al., 2019; Raikov & Abrosimov, 2022; Ojo et al., 2022). Second, it involves high costs, which restrict accessibility for small and medium-sized producers. In addition, it depends heavily on individualized expertise, which introduces variation in the quality of recommendations and limits the scalability of the model. Considering the intensive and complex nature of production systems, this approach proves insufficient to ensure agility, standardization, and resilience in real-time decision-making.

In this context, there is a growing need to refine the Decision Support Systems (DSS) currently in use so that they can integrate different sources of information, including environmental, zootechnical, and operational data, as well as continuous data generated by monitoring systems, into advanced analytical models capable of producing reliable and actionable recommendations (Zhai et al., 2020; Curi et al., 2017; Astill et al., 2020; Zheng et al., 2021; Liakos et al., 2018; Brugler et al., 2024).

To address the demands of intensive production, such systems must not only operate in a way that provides timely responses but also maintain consistent performance in scenarios characterized by uncertainty, contradiction, and informational gaps, while demonstrating adaptability to different production conditions. This requirement aligns with the contemporary movement toward the unification of learning and decision-making, which seeks to replace the historical separation between predictive models and decision mechanisms with integrated structures of optimization and contextual reasoning (Mandi et al., 2023).

Thus, the research problem guiding this study seeks to investigate how Decision Support Systems can be developed to integrate heterogeneous information and maintaining consistency even in the presence of uncertainty and contradiction, thereby overcoming the limitations of the solutions currently available in intensive poultry farming.

In this context, the present research adopts as its theoretical and methodological framework the integration of Paraconsistent Annotated Evidential Logic Eτ (Logic Eτ), Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG). Logic Eτ provides a quantitative formalism capable of representing degrees of favorable and unfavorable evidence while preserving the coherence of reasoning even under conditions of contradiction or incompleteness (Abe, Akama & Nakamatsu, 2015). LLMs, in turn, expand the capacity for knowledge representation and processing in natural language, enabling the extraction and organization of information from heterogeneous textual sources such as operational records, technical protocols, and human interactions (Brown et al., 2020). RAG complements this structure by ensuring the dynamic incorporation of updated contextual and factual evidence during the inferential process (Lewis et al., 2021; Li et al., 2022). From this integration emerges a theoretical and operational framework supported by three complementary dimensions, logical-evidential inference, contextual processing, and semantic interpretation, which together enable consistent, interpretable, and adaptable decision-making under conditions of uncertainty and contradiction (Mandi et al., 2024).

1.2 Justification

The advancement of Decision Support Systems (DSS) in animal production, particularly in poultry farming, has been strongly driven by automation technologies, statistical modeling, and machine learning. However, these models remain anchored in paradigms of data consistency and completeness, which contrast with the realities of actual production environments. In intensive poultry systems, information is often uncertain, contradictory, and context-dependent, challenging traditional approaches to modeling and inference.

In this scenario, the present research is justified by proposing an innovative methodological approach aimed at the convergence of logical-evidential inference, contextual processing, and semantic interpretation through the integration of Logic Eτ and LLMs with RAG, with the purpose of overcoming current challenges related to decision-making in poultry production. This integration is not merely technical but also conceptual: it proposes a new chain of computational reasoning grounded in logical-evidential, contextual, and semantic integration, representing a relevant contribution to the field of artificial intelligence applied to decision science and, more broadly, to research on intelligent systems capable of operating under uncertainty and contradiction. With this proposal, the study contributes to the development of explainable, scalable, and adaptable systems aligned with the demands for efficiency and sustainability in modern animal production.

From an applied perspective, the research is also justified by addressing a concrete need in the poultry sector, where decision-making processes depend on fragmented and often inconsistent data. The development of DSS capable of processing and interpreting contradictory information offers potential gains in agility, standardization, and reliability, reducing reliance on human consultancy and expanding access to operational intelligence.

Beyond its technical and methodological relevance, the study is also justified from a social and institutional standpoint, as it aligns with the Sustainable Development Goals (SDGs) of the 2030 Agenda, contributing to the technological modernization and operational resilience of production systems, promoting efficiency and food security through process optimization, and strengthening transparency, traceability, and accountability in the use of intelligent technologies applied to poultry production.

Finally, the research is further justified by its scientific and institutional pertinence. Embedded within the research line “Quantitative Methods in Production Engineering” at Universidade Paulista (UNIP), it contributes to advancing logical-evidential modeling as a decision-support tool for complex and uncertain contexts.

1.3 Objectives

1.3.1 General Objective

The central objective of this master’s dissertation is to design the architecture, develop, and evaluate a knowledge-based Decision Support System for intensive poultry farming, capable of supporting resilient decision-making processes under conditions of uncertainty and informational contradiction.

1.3.2 Specific Objectives

To achieve the general objective, this research establishes the following specific objectives:

1.3.2.1 Critically examine the limitations and challenges of current DSS used in precision poultry farming, with particular emphasis on environmental control.

1.3.2.2 Evaluate the extent to which recent AI technologies, particularly LLMs and related techniques, can be integrated to strengthen knowledge-based decision support in intensive poultry farming.

1.3.2.3 Develop and evaluate a knowledge-based DSS for poultry farming that incorporates recent AI technologies and is structured as a conversational agent resilient to uncertainty and contradiction.

1.4 Methodology

This master’s dissertation adopts an article-based format. The overall study is structured into three interdependent works that address, sequentially and complementarily, the different stages of investigation and validation of the proposed model. This methodological choice aims to align the partial scientific outputs with the stages of the research process, ensuring logical continuity among the studies and coherence with the defined objectives. The specific methodologies employed in each article are detailed in their respective publications; therefore, this section presents the unified design and methodological procedures that guided the entire investigation, highlighting the scientific and operational structure common to all three studies.

1.4.1 Research Design and Methodological Structure

This master's dissertation is characterized as applied in nature, experimental in design, and exploratory in scope, employing a quantitative–qualitative approach within an article-based structure. The study is applied because it aims to solve a concrete problem: the inconsistency and fragmentation of data supporting technical decisions in intensive production environments. It is experimental because it employs controlled tests with measurable variables, and exploratory due to the originality of integrating Logic Eτ, LLMs, and RAG, a combination still at an early stage within Production Engineering research. This methodological design ensures coherence between the research problem and the adopted empirical strategy, enabling a progressive analysis of theoretical and practical contributions across the stages of the investigation.

The object of study is the DSS resulting from the integration of Logic Eτ, LLMs, and RAG, designed to operate under conditions of uncertainty and contradiction typical of complex production systems. The phenomenon under investigation is the process of logical-evidential inference and decision-making in fragmented and imperfect informational contexts, while the empirical unit of analysis corresponds to the decision-making processes related to environmental control and zootechnical management in broiler farms.

The three articles that compose this master's dissertation form an articulated and cumulative methodological trajectory, in which each study plays a specific role within the process of investigation and validation of the proposed model. The first article conducts a critical analysis of the state of the art, mapping conceptual and technological gaps in DSS applied to intensive poultry farming; the second performs an experimental evaluation of the RAG technique, assessing its contribution to the performance of LLMs in the domain of environmental control; and the third consolidates the integrative model, formalizing it in logical-evidential and computational terms and evaluating it according to criteria of consistency, accuracy, and operational applicability.

Table 1 summarizes the correspondence among the articles, their objectives, and the resulting scientific outputs.

Table 1: Methodological structure of the master's thesis.

Article	Journal	Status	Objective	Nature of the Investigation	Scientific Outcome
Decision Support Systems for Environmental Control in Poultry Production: Trends, Advances, and Perspectives	Proceedings of SIMPEP	Accepted for publication	1.3.2.1	Systematic review and critical analysis of the state of the art.	Identification of theoretical, technological, and methodological gaps.
Enhancing Environmental Control in Broiler Production: Retrieval-Augmented Generation for Improved Decision-Making with Large Language Models	AgriEngineering (MDPI)	Published in 2025	1.3.2.2	Applied performance experiment and comparative evaluation of LLM+RAG.	Evidence on the accuracy, relevance, and applicability of RAG in DSS.
A Decision Support AI-Copilot for Poultry Farming: Leveraging Retrieval-Augmented LLMs and Paraconsistent Annotated Evidential Logic Eτ to Enhance Operational Decisions	AgriEngineering (MDPI)	Submitted	1.3.2.3	Logical–computational modeling, implementation, and experimental validation	Functional DSS prototype and acceptance criteria based on logical-evidential inference

Source: Prepared by the author.

This cumulative methodological trajectory demonstrates the evolution of the research from theoretical formulation to experimental validation and applied model development, ensuring coherence among the conceptual, empirical, and technological stages of the study.

1.4.2 Methodological Procedures

The methodological procedures in this research are organized into four complementary axes, encompassing activities from the review and critical analysis of the literature to the empirical validation of the proposed model:

1.4.2.1 Review and Critical Analysis of the State of the Art: The first methodological axis consisted of a systematic review and critical analysis of the literature on DSS applied to intensive poultry farming, with emphasis on environmental control approaches. This stage, corresponding to Article 1, aimed to identify theoretical, technological, and methodological gaps, as well as to understand the limitations of existing solutions in scenarios characterized by uncertainty and contradiction.

The review followed principles of systematic rigor: definition of descriptors, inclusion and exclusion criteria, bibliometric analysis, and thematic categorization. The results allowed the formulation of the research problem and the establishment of conceptual and operational requirements for the proposed method.

1.4.2.2 Experimental Evaluation of Retrieval-Augmented Techniques: The second axis, developed in Article 2, consisted of a controlled experimental stage aimed at evaluating RAG as a strategic component for contextual query processing in natural language–based decision systems.

Comparative experiments were conducted between language model executions with and without RAG, assessing metrics of semantic similarity, contextual relevance, and practical applicability. The results of this axis provided empirical evidence of RAG's potential to enhance LLM performance and served as the basis for configuration and calibration decisions adopted in the subsequent stage.

1.4.2.3 Modeling, Implementation, and Validation of the Proposed Model: The third methodological axis, corresponding to Article 3, encompassed the logical–computational modeling, implementation, and experimental validation of the system resulting from the integration of Paraconsistent Annotated Evidential Logic Eτ (Logic Eτ), LLMs, and RAG.

In this stage, mechanisms for logical-evidential inference, computational modules for contextual processing and knowledge base construction, as well as the conversational decision-support agent responsible for semantic interpretation, were defined. Validation was carried out based on criteria of logical-evidential consistency, semantic accuracy, and operational applicability, demonstrating the feasibility and robustness of the developed model.

1.4.2.4 Validity and Reproducibility Synthesis: This axis established procedures to ensure internal, external, and construct validity, as well as the reproducibility of results. Internal validity was controlled through the standardized experiments and repeated runs; external validity was ensured by generalization across poultry decision domains; and construct validity was verified through the alignment between the principles of Logic Eτ and the decision-making phenomenon under uncertainty. All codes, parameters, and datasets were documented and versioned in a public repository to ensure traceability and reproducibility.

The empirical evidence resulting from these verifications is presented and discussed in the subsequent chapters.

Together, these four methodological axes, systematic review, controlled experimentation, logical–computational modeling and validation, and synthesis of validity, provide an integrated framework that ensures coherence, methodological consistency, and reproducibility throughout the study.

1.4.3 Methodological Integration Among the Articles

The methodological integration among the articles stems from the logical and cumulative linkage of their scientific purposes. Each study contributes in a distinct yet complementary manner to the consolidation of the proposed method, forming an incremental trajectory in which the outcomes of one stage redefine the conditions and hypotheses of the next.

The first article establishes the conceptual and diagnostic framework of the research, identifying theoretical and operational gaps that justify the need for a model capable of formally addressing contradiction and incompleteness. Its findings not only contextualize the problem but also define the evaluation criteria and performance dimensions to be observed in the subsequent experimental phases.

The second article plays an intermediate instrumental role, translating the identified gaps into testable hypotheses and experimental parameters. The controlled evaluation provides empirical evidence of the behavior of language models when exposed to heterogeneous data and generates quantitative and qualitative foundations for the design of the following stage.

The third article represents the synthesis stage of the process, integrating the logical-evidential foundations of Logic E_τ with the experimental evidence accumulated in the previous phases. This integration results in the formalization of the model, its computational implementation, and the assessment of criteria related to consistency, accuracy, and operational applicability.

In summary, the methodological trajectory unfolds through the following sequence: logical–theoretical foundation → review and diagnostic analysis (Article 1) → component experimentation (Article 2) → modeling, implementation, and validation of the integrative model (Article 3), ensuring epistemological continuity and experimental rigor.

1.5 Structure of the Master's Thesis

This master’s dissertation is organized into four chapters, structured from the conceptual framework to the empirical validation and final conclusions of the study.

Chapter I – Introductory Considerations, presents the contextualization of the research theme, the rationale, the objectives, and the methodological design of the research.

Chapter II – Theoretical Framework brings together the conceptual and scientific foundations that support the study, addressing intensive poultry farming, Decision Support Systems, Paraconsistent Annotated Evidential Logic Eτ, and Artificial Intelligence technologies based on Large Language Models.

Chapter III – Results and Discussion contains the three scientific articles that constitute the core of this work, each corresponding to a stage of the investigative process: (i) critical analysis of the state of the art, (ii) controlled experimentation of the RAG technique, and (iii) modeling and validation of the DSS integrating Logic Eτ and LLM agents with RAG. It concludes with an integrative discussion that consolidates the main findings and their theoretical and applied implications.

Finally, Chapter IV – Final Conclusions summarizes the research contributions, limitations, and perspectives for future work, highlighting the methodological and practical advances provided by the developed integrative model.

CHAPTER II

2 THEORETICAL FRAMEWORK

This chapter consolidates the theoretical and scientific foundations that support the integration of Paraconsistent Annotated Evidential Logic Eτ (Logic Eτ), Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) within the scope of Decision Support Systems (DSS) applied to intensive poultry farming. This integration, which constitutes the core of the research, underpins the development of an inference method capable of operating under conditions of informational fragmentation, contradiction, and incompleteness.

The theoretical framework is structured around four complementary pillars: (i) the empirical context of intensive poultry farming and its decision-making specificities; (ii) the conceptual and methodological evolution of Decision Support Systems in Production Engineering; (iii) the formal structure of Logic Eτ for managing contradictory or incomplete evidence; and (iv) the role of LLMs and RAG in expanding contextual processing and semantic interpretation capabilities in intelligent systems.

2.1 Application Domain: Decision-Making in Intensive Poultry Systems

Intensive poultry farming constitutes the application domain adopted in this research, serving as the empirical environment for the formulation and evaluation of the DSS based on a theoretical–operational structure composed of three complementary dimensions: logical-evidential inference, contextual processing, and semantic interpretation. This context is characterized by high operational complexity and by the interdependence among environmental, health, and zootechnical variables, which impose on the decision-making process recurring conditions of fragmentation, uncertainty, contradiction, and informational incompleteness (Curi et al., 2017; Hafez & Attia, 2020; Gržinić et al., 2023; Qi et al., 2023).

Environmental control, as well as other domains of zootechnical management, requires continuous and multivariate decisions involving numerous variables under dynamic and often conflicting constraints. The simultaneous presence of multiple information sources, with different levels of reliability and temporal alignment, makes the inference process particularly sensitive to the coherence of the available data (Martinez et al., 2021).

In this context, DSS act as fundamental instruments for the integration and interpretation of such information (Zhai et al., 2020; Rossi, Caffi & Salinari, 2012). However, traditional solutions remain constrained by assumptions of data consistency and completeness, reducing their ability to operate under ambiguous or contradictory conditions, typical of intensive production systems (Hamsa & Bellundagi, 2017; Berckmans, 2009; Cheng, McCarl & Fei, 2022).

These limitations have developed in parallel with the rapid advancement of sensing and data analytics technologies, which have transformed the informational infrastructure of modern poultry production. In recent years, technological intensification in the sector has consolidated a highly automated and data-driven production ecosystem. Several studies highlight the convergence of IoT sensors, wireless networks, and cloud analytics platforms as the core of precision poultry farming practices, focused on the continuous acquisition of environmental, behavioral, and performance data from flocks (Astill et al., 2020; Halachmi et al., 2019; Zheng et al., 2021).

These systems comprise multiple functional layers such as detection, analysis, and decision, which, when articulated, enable the monitoring of critical variables such as temperature, humidity, ventilation, air quality, noise, lighting, feeding patterns, and animal behavior. (Lashari et al., 2018). The continuous data flow captured by sensors, cameras, and microphones is processed in cloud computing architectures and stored in large-scale repositories such as data warehouses and data lakes (Wu et al., 2023). These infrastructures support the application of machine learning, large-scale data analytics, and expert systems to interpret patterns and support decision-making (Ojo et al., 2022; McAfee & Brynjolfsson, 2012).

Although these approaches represent a significant advance, they remain predominantly predictive and quantitative, lacking formal reasoning mechanisms capable of handling contradiction, incompleteness, and evidence heterogeneity, limitations that justify the adoption of decision systems based on logical-evidential inference, as proposed in this study.

Thus, intensive poultry farming is used here not as the primary object of investigation but as an empirical platform for assessing the effectiveness of a DSS based on the integration of Logic Eт, LLMs, and RAG. This choice is justified by its suitability for representing complex decision environments in which informational heterogeneity, inconsistent evidence, and the need for resilient and operationally actionable responses coexist (Zheng et al., 2021; Liakos et al., 2018; Brugler et al., 2024).

From this scenario of limitations in current decision mechanisms arises the need for support models grounded in logical inference, as discussed in the following section.

2.2 Decision Support Systems

Decision Support Systems have historically been conceived as tools for formalizing analytical reasoning and structuring complex problems. The first approaches, developed between the 1970s and 1990s, were predominantly based on mathematical and heuristic models, emphasizing optimization, simulation, and multicriteria methods. This generation of systems featured highly prescriptive architectures, driven by explicit rules and stable data flows, which limited their applicability to deterministic and relatively static environments (Elkady, Hernantes & Labaka, 2024).

With the advancement of digitalization and the proliferation of repositories and sensors, DSS evolved into data-driven approaches in which machine learning and large-scale analytics assumed a central role. The data-driven paradigm transformed data science into a core component of the decision-making process, enabling pattern detection, event prediction, and probabilistic evaluation of alternatives. This transition increased system autonomy and brought the field closer to what is now referred to as Decision Intelligence (DI), integrating analytical reasoning, modeling, and organizational action (Pratt, Bisson & Warin, 2023). Current DSS operate over integrated data pipelines that combine multiple sources, preprocessing routines, predictive models, and user-context–adapted interfaces (Elkady, Hernantes & Labaka, 2024).

More recently, emphasis has shifted from prediction to decision-making, giving rise to a new generation of knowledge-based systems supported by artificial intelligence. These models combine formal domain representations, such as ontologies and rules, with adaptive learning mechanisms, promoting explainable and auditable inferences. Within this context, the field has been repositioned under the designation Human–AI Systems (HAIS), in which collaboration between human and artificial agents becomes central to producing decisions that are more consistent, interpretable, and governable (Storey, Hevner & Yoon, 2024). Empirical evidence indicates that LLMs are already capable of generating and evaluating strategies with performance comparable to that of human experts, enhancing search, aggregation, and evaluation capabilities, essential features for third-generation DSS (Csaszar, Ketkar & Kim, 2024).

From an architectural standpoint, contemporary DSS follow a functional structure composed of four interdependent layers: acquisition, processing, inference, and interface. The acquisition layer integrates data streams from sensors, machines, operational records, and external sources, while the processing layer performs cleaning, temporal synchronization, integration, and data curation. The inference layer combines machine learning models with decision rules and symbolic mechanisms, fostering transparency and contextualization of recommendations. Finally, the interface layer delivers diagnostics, explanations, and actionable recommendations. This sequence, data, analysis, decision, and action, is widely recognized in agricultural and livestock precision systems, reflecting the principles of modularization and interoperability that underpin current Farm Management Information Systems (FMIS) and Smart Farming platforms (Fountas et al., 2015; Wolfert et al., 2017).

The effectiveness of a DSS, however, fundamentally depends on the quality, coherence, and integration of information. In agro-industrial contexts, heterogeneous environmental, zootechnical, and economic data require metadata standardization, granularity control, and the use of domain ontologies to stabilize meaning and ensure consistency across modules (Wolfert et al., 2017; Zheng et al., 2021). The coherence between data and inference is critical for the reliability of automated actions, as integration failures directly affect the timeliness and quality of decisions (Berckmans, 2017; Norton et al., 2019).

The pursuit of interoperability and logical consistency has thus become a structural requirement. Service-oriented architectures and data contracts enable modular expansion without structural disruption, while versioning and traceability standards ensure end-to-end coherence in the information flow. As a result, DSS have evolved from prescriptive and isolated systems into integrated, knowledge-based decision ecosystems aligned with the contemporary need for rapid, explainable, and evidence-supported decisions.

Nonetheless, this evolution highlights a persistent limitation: although DSS have significantly increased their capacity to process large data volumes and incorporate adaptive learning, they still lack formal mechanisms capable of simultaneously addressing contradiction, fragmentation, and informational incompleteness, conditions inherent to complex production systems. It is at this point that those paraconsistent logics, particularly Logic Eτ, become relevant as formal structures capable of preserving the coherence of reasoning even under uncertainty and conflict (Abe; Akama & Nakamatsu, 2015), as discussed in the following section.

2.3 Paraconsistent Logic

Classical logic, since Aristotle, has been structured around the dichotomy of truth and falsity and the principle of non-contradiction, according to which a proposition cannot be both true and false simultaneously. This binary formalism, foundational to mathematics and computation, proves limited for representing complex phenomena in which uncertainty, gaps, and conflicting information coexist (Abe, Akama & Nakamatsu, 2015).

As a response, non-classical logics emerged, designed to expand representational and inferential capacity under conditions of incompleteness and inconsistency. Among them, Paraconsistent Logic (PL) stands out for enabling the treatment of contradictions without leading the system to triviality, that is, without making every proposition logically derivable (Abe 1992; Abe; Akama & Nakamatsu, 2015; Akama & Da Costa, 2016).

Building upon PL, Abe developed the Paraconsistent Annotated Logic (PAL), which introduced the concept of annotation: each proposition is accompanied by a value expressing the degree of available evidence. This structure allows for the representation of information that is partially true or false, marking a milestone in the formalization of reasoning under uncertainty and serving as the foundation for subsequent evidential formulations (Abe, 2011; Abe; Akama & Nakamatsu, 2015; Akama, 2016; Inacio da Silva Filho, Abe & Torres, 2008).

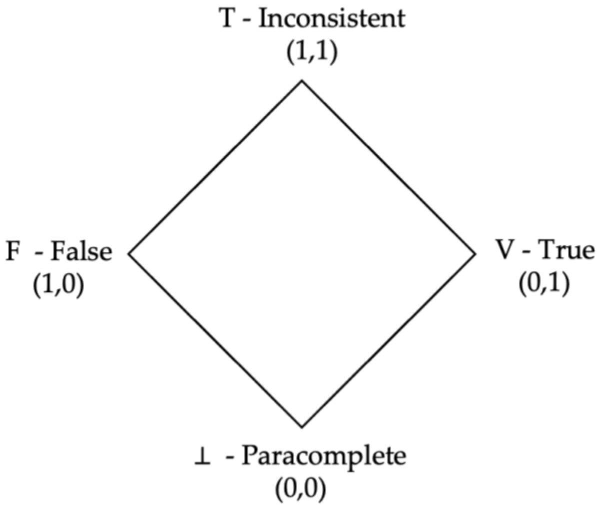
2.3.1 Paraconsistent Annotated Evidential Logic Eτ

The Paraconsistent Annotated Evidential Logic Eτ is an extension of the PAL that explicitly incorporates the treatment of favorable (μ) and unfavorable (λ) evidence, both ranging within the interval [0, 1]. Each proposition p is represented by the pair (μ, λ), which simultaneously expresses the degrees of favorable and unfavorable evidence associated with a given piece of information (Abe, 2011; Abe, Akama & Nakamatsu, 2015).

Logic Eτ is structured around three complementary conceptual spaces:

- Eτ Lattice: defines an ordered set of pairs (μ, λ) within the unit square $[0, 1]^2$, where $(\mu_1, \lambda_1) \leq (\mu_2, \lambda_2)$ if $\mu_1 \leq \mu_2$ and $\lambda_1 \geq \lambda_2$. This order expresses evidential dominance and allows the application of *infimum* and *supremum* operators. The canonical negation is given by $\sim(\mu, \lambda) = (\lambda, \mu)$. This lattice constitutes the operational substrate for evidential inference.

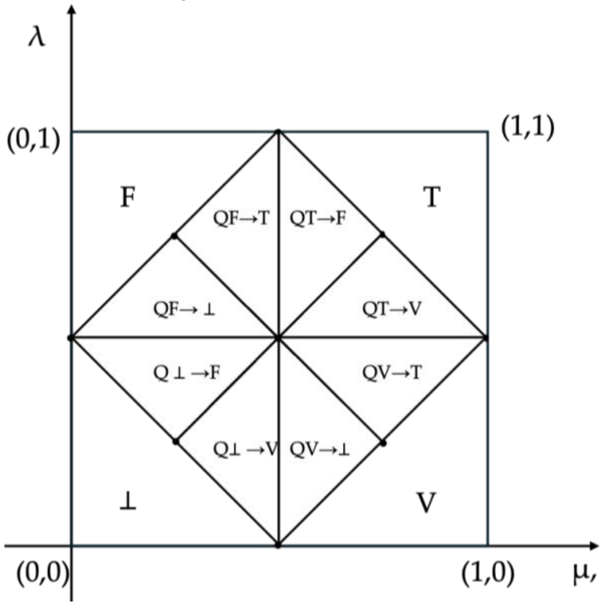
Figure 4. Lattice Eτ with Partial Order



Source: Adapted by the author based on Abe, Akama, and Nakamatsu (2015).

- USCP (Unit Square of the Cartesian Plane): provides a geometric representation of the evidential lattice, allowing visualization of logical states in a two-dimensional plane.

Figure 5. USCP (Unit Square of the Cartesian Plane)



Source: Adapted by the author based on Carvalho, Abe (2018).

.

The vertices represent the classical states, True (V), False (F), Inconsistent (T), and Paracomplete (⊥), while the intermediate regions correspond to quasi-states or transitional states, such as QV→T, QV→⊥, QF→T, and QF→⊥, which indicate tendencies in the balance of evidence (Table 2).

Table 2: Symbolic representation of extreme and non-extreme logical states in Logic Eт, including qua-si-states and transitional tendencies.

Symbol	State
V	True
QV→T	Quasi-true, tending to inconsistent;
QV→⊥	Quasi-true, tending to paracomplete
F	False
QF→T	Quasi-false, tending to inconsistent
QF→⊥	Quasi-false, tending to paracomplete
T	Inconsistent
QT→V	Quasi-inconsistent, tending to true
QT→F	Quasi-inconsistent, tending to false
⊥	Paracomplete or Indeterminate
Q⊥→V	Quasi-paracomplete, tending to true
Q⊥→F	Quasi-paracomplete, tending to false

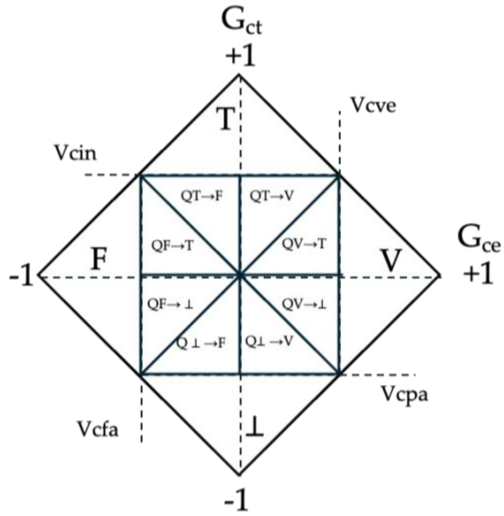
Source: Adapted by the author based on Carvalho, Abe (2018).

- Diagram of Certainty and Contradiction Degrees: results from the nonlinear transformation

$$T(\mu, \lambda) = (G_{ce}(\mu, \lambda), G_{ct}(\mu, \lambda)) = (\mu-\lambda, \mu+\lambda-1)$$

which projects the evidential space onto two orthogonal axes: the degree of certainty (Gce) and the degree of contradiction (Gct). This transformation defines the operational plane employed in inference and decision-making processes, where extreme and quasi-states are interpreted in a graded manner. It also allows the establishment of control limits that help the system filter and stabilize decisions under uncertainty and contradiction: the Upper Certainty Control Value (Vscc), Lower Certainty Control Value (Vicc), Upper Contradiction Control Value (Vscct), and Lower Contradiction Control Value (Vicct).

Figure 6. Diagram of Certainty and Contradiction Degrees



Source: Adapted by the author based on Abe, Akama, and Nakamatsu (2015).

Logic Eτ enables continuous reasoning across states of truth, falsity, inconsistency, and paracompleteness, providing a formal foundation for logical-evidential inference, one of the conceptual pillar of this work.

2.3.2 Para-Analyzer Algorithm

The Para-Analyzer Algorithm (PAA) computationally implements the principles of Logic Eτ. Based on favorable (μ) and unfavorable (λ) evidence, the PAA calculates the degree of certainty ($Gce = \mu - \lambda$) and the degree of contradiction ($Gct = \mu + \lambda - 1$), comparing them with control parameters (Vscc, Vicc, Vscct, Vicct) to classify each proposition as true, false, inconsistent, or paracomplete (Carvalho & Abe, 2018).

This algorithm ensures stability and traceability in paraconsistent reasoning, enabling computational systems to process uncertain or contradictory data without compromising logical coherence, an essential requirement for Decision Support Systems DSS.

2.3.3 Paraconsistent Decision Method

The Paraconsistent Decision Method (PDM) applies Logic Eτ to decision-making problems. Information is structured into an evidence matrix, in which μ and λ values represent the degrees of favorable and unfavorable evidence associated with influence factors defined by domain experts. Based on these pairs, the method computes Gce and Gct and uses the decision rules of the PAA to determine the logical state of each alternative (Carvalho & Abe, 2018).

The PDM enables the integration of multiple criteria and sources of evidence without requiring complete data consistency, making it suitable for complex production contexts such as intensive poultry systems, where decisions must be made under conditions of contradiction, uncertainty, and informational fragmentation.

Thus, Logic Eτ and its associated developments (PAA and PDM) constitute the core of a framework for logical-evidential inference, whose integration with LLMs and RAG is discussed in the next section.

2.4 Large Language Models and Retrieval-Augmented Generation

Early approaches to language modeling were relied on statistical methods such as *n-gram* models and simple neural networks focused on word prediction. The Neural Probabilistic Language Model (NPLM) introduced the concept of representing words as continuous vectors, enhancing generalization capacity and marking the beginning of distributed representations (Bengio et al., 2003).

The consolidation of Transformer architectures, based on attention mechanisms, replaced recursive and convolutional operations with direct connections among all words in a sequence, enabling the capture of long-range dependencies and improving training efficiency (Vaswani et al., 2017). Models such as word2vec and BERT further refined this capacity by learning semantic and contextual relationships from large text corpora (Mikolov et al., 2013; Devlin et al., 2019).

As scale increased, Large Language Models (LLMs) emerged, capable of learning from few examples and dynamically adapting to context. This evolution expanded generalization capabilities but also revealed important limitations, such as the generation of incorrect information (hallucinations), the propagation of biases from training data, and the lack of transparency in inference processes (Brown et al., 2020; Lin, Hilton & Evans, 2022).

Purely generative models, relying solely on internal knowledge acquired during training, face structural constraints that compromise reliability. Research shows that increasing model scale does not guarantee higher precision or consistency, making it necessary to adopt control and detection mechanisms such as uncertainty estimation to identify potentially confabulated outputs (Farquhar et al., 2024; Lin, Hilton & Evans, 2022).

To mitigate these limitations, contextual learning strategies were developed, including context windows, in-context learning, and few-shot prompting, which function as temporary memory mechanisms allowing models to adjust responses based on examples provided within the prompt. Although effective, these techniques remain sensitive to example formatting and the positioning of relevant information, limiting the use of extended contexts (Dong et al., 2024; Liu et al., 2024; Zhang et al., 2025).

The need for anchoring and internal coherence control led to the development of Retrieval-Augmented Generation (RAG), which combines knowledge generation and retrieval to improve factual accuracy. RAG introduces external, up-to-date information into the model's context, allowing responses to be grounded in verifiable evidence (Lewis et al., 2020). Complementary strategies such as self-consistency, comparing multiple reasoning chains, and the ReAct method, interleaving reasoning and external retrieval, further enhance reliability and interpretability in inferential processes (Wang et al., 2023; Yao et al., 2023).

RAG is therefore a hybrid architecture that integrates parametric knowledge embedded in the model itself with non-parametric knowledge retrieved from external sources. Its operation consists of three main stages: retrieval, which locates relevant passages; re-ranking to optimize selection; and generation to produce the final response grounded on the retrieved content (Lewis et al., 2020). Dense Passage Retrieval (DPR) leverages vector representations to compare queries and documents, outperforming traditional approaches such as BM25 in retrieval accuracy (Karpukhin et al., 2020).

By grounding responses in verifiable evidence, RAG reduces hallucinations and increases factual precision while enabling continuous knowledge updates without retraining the base model. Experiments with internet-augmented models demonstrate accuracy gains in open-domain question-answering tasks (Lazaridou et al., 2022; Izacard et al., 2023).

These advantages explain the growing adoption of RAG in DSS and conversational agents, which require reliable and auditable responses. In knowledge-driven dialogues, internet-augmented architectures learn to issue retrieval queries and condition response generation on the recovered material, integrating up-to-date information and minimizing contradictions (Komeili, Shuster & Kizilkaya, 2022). Multi-agent variants combine structured data (e.g., knowledge graphs) and unstructured text, increasing evidence verifiability. Evaluation of such systems considers both retrieval, through metrics such as nDCG@k, Recall@k, and MRR (Thakur et al., 2021), and generation, measured by semantic similarity (BERTScore) and source fidelity (faithfulness), following the AIS (Attributable to Identified Sources) framework (Rashkin et al., 2023; Zhang et al., 2020).

Reducing hallucinations and bias in language models requires complementary strategies acting at both training and inference stages. Key approaches include chain-of-verification, which guides the model to review its own outputs; uncertainty estimation, which flags potentially incorrect content; and alignment methods such as Reinforcement Learning from Human Feedback (RLHF) and Constitutional AI, which reduce biased or undesirable behavior. Evaluation frameworks like HELM propose integrated metrics that jointly assess quality, robustness, and fairness in model outputs (Dhuliawala et al., 2024; Farquhar et al., 2024; Christiano et al., 2017; Bai et al., 2022; Liang et al., 2022).

Recently, explainability and auditability have become essential for the safe application of LLMs in critical environments. The field of Explainable AI (XAI) aims to make model reasoning more transparent through techniques that identify attention patterns, salience, and relevance in generated text. Recent approaches focused on source faithfulness highlight the need to verify whether responses are truly attributable to retrieved content, as established in the AIS framework (Zhao et al., 2024; Rashkin et al., 2023; Wang et al., 2023). In RAG- or DSS-based systems, such practices include the registration of citations, evidence excerpts, and query logs, reinforcing traceability and regulatory compliance.

Finally, integrating LLMs with formal reasoning, especially neurosymbolic approaches, has proven promising for improving coherence and interpretability. This approach combines the statistical learning of language models with symbolic logic structures, allowing formal constraints during inference and enabling the extraction of interpretable rules (Garcez & Lamb, 2020). Notable examples include Logical Neural Networks (LNN), which preserve first-order semantic relations (Riegel et al., 2020), and DeepProbLog, which integrates neural networks with probabilistic logic programming in continuous learning flows (Manhaeve et al., 2018). Other architectures, such as Program-Aided Language Models (PAL), allow LLMs to generate intermediate programs and delegate their execution to formal interpreters, ensuring greater accuracy and verifiability of results (Gao et al., 2023).

In summary, contemporary challenges of LLMs focus on enhancing factuality, reducing bias, ensuring explainability, and integrating formal reasoning mechanisms that make responses more consistent, reliable, and auditable, conditions essential for their trustworthy adoption in knowledge-based DSS.

CHAPTER III

3 RESULTS

This chapter presents and discusses the results obtained from the three studies that constitute the empirical core of the research. Each article corresponds to a stage of the methodological trajectory described in Chapter I and progressively contributes to the development and validation of the integrative model proposed.

The chapter is structured into three main subsections, each dedicated to one of the articles comprising this master’s thesis, followed by an integrative discussion of the results.

The final subsection consolidates the findings of the three studies, highlighting their convergences, complementarities, and theoretical and practical implications, thereby forming the basis for the conclusions presented in Chapter IV.

3.1 Article 1 – Decision Support Systems for Environmental Control in Poultry Production: Trends, Advances, and Perspectives

The first article, titled "*Decision Support Systems for Environmental Control in Poultry Production: Trends, Advances, and Perspectives*," was accepted for presentation at the XXXII Symposium on Production Engineering (SIMPEP, 2025) and will be published in the conference proceedings. The study was conducted by Marcus Vinícius Leite, Jair Minoro Abe, Marcos Leandro Hoffmann Souza, and Irenilza de Alencar Nääs, affiliated with Universidade Paulista (UNIP) and Universidade do Vale do Rio dos Sinos (UNISINOS).

The purpose of this study was to map and systematize the state of the art regarding the use of DSS applied to environmental control in intensive poultry farming systems. The investigation sought to understand how these tools have been employed to support decisions in operationally complex contexts where environmental, physiological, and production variables coexist (Berckmans, 2017; Li et al., 2020; Neethirajan, 2025). As the first methodological stage of the research, the study served a diagnostic and foundational role by identifying structural limitations of existing solutions and establishing the conceptual requirements that guided the formulation of the integrative model proposed in this master's dissertation (Zhai et al., 2020; Brugler et al., 2024).

The methodology followed a systematic literature review, conducted according to the protocols of Kitchenham (2004) and the PRISMA, 2020 guidelines (Page et al., 2021), ensuring traceability and rigor in the selection and analysis of evidence. Searches conducted in the Scopus and Web of Science databases resulted in the identification, screening, and selection of studies that met the established inclusion criteria and composed the final corpus for analysis. The research questions were structured using the PICOC logic (Rossi, Caffi & Salinari, 2012), encompassing the identification of DSS types, decision levels, benefits, limitations, and emerging trends. Information was organized and coded through predefined and emerging categories, combining qualitative and quantitative analysis to produce a critical synthesis guided by thematic patterns and significant variations. The full protocol, including search terms and extraction spreadsheets, was documented and made publicly available (marcusviniciusleite, 2025), ensuring transparency and reproducibility.

The results revealed a field in consolidation, with accelerated growth since, 2020 and a predominance of journal publications, indicating both technical and scientific maturity (Astill et al., 2020; Gržinić et al., 2023). Most systems were based on embedded IoT architectures equipped with physical sensors and low-power microcontrollers for continuous environmental monitoring and local inference using machine learning, fuzzy logic, and neural networks (Curi et al., 2017; Liakos et al., 2018; Zheng et al., 2021). Hybrid systems combining statistical modeling, symbolic paradigms, and predictive techniques were also identified (Martinez et al., 2021; Zhai et al., 2020). The majority of applications focused on short-term operational decisions, such as ventilation, lighting, and climate control, while tactical and strategic systems aimed at simulation and planning remain underrepresented (Hamsa & Bellundagi, 2017).

The analysis highlighted recurring benefits related to improved environmental precision, process automation, animal welfare, and economic efficiency (Godinho et al., 2025; Neethirajan, 2025). However, methodological and operational gaps persist, limiting real-world applicability. These include the absence of validation in commercial farms, sensor fragility under harsh environmental conditions, low interoperability among subsystems, and dependence on continuous connectivity (Qi et al., 2023; Hafez & Attia, 2020). Many DSS remain confined to passive monitoring functions, lacking adaptive learning mechanisms or fault tolerance, which restricts their effectiveness in contexts marked by uncertainty and contradiction.

At the same time, emerging technological trends indicate a transition toward a new generation of DSS, more resilient, integrated, and semantically enriched. Advances such as edge–cloud architectures, fusion of physiological and environmental data, use of smart sensors, and the incorporation of tinyML and explainable AI techniques have expanded system autonomy and interpretability (Berckmans, 2009; Zhai et al., 2020; Brugler et al., 2024). The early use of LLMs and RAG suggests a movement toward systems capable of integrating numerical inference and semantic reasoning, inaugurating a new paradigm of cognitive decision support in intensive poultry farming.

DECISION SUPPORT SYSTEMS FOR ENVIRONMENTAL CONTROL IN POULTRY FARMING: TRENDS, ADVANCES, AND PERSPECTIVES

MARCUS VINICIUS LEITE – marcus.leite@gmail.com
UNIVERISDADE PAULISTA – UNIP

JAIR MINORO ABE - jairabe@uol.com.br
UNIVERISDADE PAULISTA – UNIP

LEANDRO CIGANO DE SOUZA THOMAS – leandrocst91@gmail.com
UNIVERISDADE PAULISTA – UNIP

MARCUS LEANDRO HOFFMANN SOUZA – marcosl.hoffmann@yahoo.com.br
UNIVERSIDADE DO VALE DO RIO SINOS – UNISINOS

ÁREA: *03. OPERATIONS RESEARCH*
SUBÁREA: *03.3 – DECISION-MAKING PROCESSES*

RESUMO: *OVER THE PAST DECADES, POULTRY PRODUCTION HAS INTENSIFIED WORLDWIDE, DRIVEN BY THE GROWING DEMAND FOR ANIMAL PROTEIN. IN BRAZIL, THE SECOND-LARGEST PRODUCER GLOBALLY, POULTRY FARMING OPERATES UNDER AN INTENSIVE, LARGE-SCALE MODEL CHARACTERIZED BY HIGH DENSITY AND SUBSTANTIAL PROCESS COMPLEXITY. WITHIN THIS CONTEXT, ENVIRONMENTAL CONTROL IN POULTRY HOUSES PLAYS A STRATEGIC ROLE IN ENSURING OPERATIONAL EFFICIENCY, ZOOTECHNICAL PERFORMANCE, ANIMAL WELFARE, SANITARY SAFETY, AND REGULATORY COMPLIANCE. TO ADDRESS THESE CHALLENGES, THE USE OF DECISION SUPPORT SYSTEMS (DSS) HAS INCREASED. HOWEVER, THE LITERATURE ON THIS TOPIC REMAINS FRAGMENTED AND LACKS SYSTEMATIZATION, WHICH HINDERS UNDERSTANDING OF THE FIELD AND THE EVOLUTION OF RELATED SOLUTIONS. THIS ARTICLE PRESENTS A SYSTEMATIC LITERATURE REVIEW BASED ON THE GUIDELINES OF KITCHENHAM (2004) AND PRISMA 2020, AIMING TO MAP, CLASSIFY, AND ANALYZE DSS APPLIED TO ENVIRONMENTAL CONTROL IN POULTRY FACILITIES. THE RESULTS PROVIDE A COMPREHENSIVE OVERVIEW OF THE FIELD, HIGHLIGHTING TRENDS, PERSPECTIVES, AND LIMITATIONS. THE STUDY CONTRIBUTES BY SYSTEMATIZING THE STATE OF THE ART, ADDRESSING DSS TYPES, THE PRODUCTION DECISIONS THEY SUPPORT, THEIR BENEFITS AND CURRENT CHALLENGES, AND EMERGING TECHNOLOGIES. THE FINDINGS ARE INTENDED TO SUPPORT THE DEVELOPMENT OF MORE EFFICIENT AND SCALABLE SYSTEMS, FOSTERING THE ADVANCEMENT OF KNOWLEDGE IN PRODUCTION ENGINEERING, POULTRY FARMING, AND DSS, AS WELL AS PROMOTING MORE SUSTAINABLE PRODUCTION PRACTICES ALIGNED WITH THE UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS (SDG) 2, 9, AND 12.*

KEYWORDS: *INTENSIVE POULTRY FARMING, DECISION-MAKING, DECISION SUPPORT SYSTEMS, ENVIRONMENTAL CONTROL, DIGITAL TECHNOLOGIES*

1. INTRODUCTION

In recent decades, poultry farming has consolidated as the main source of animal protein consumed worldwide, driven mainly by economic and population growth, urbanization, and changes in dietary habits. This context has favored the expansion of intensive poultry production systems. In Brazil, the world's second-largest producer, this transformation is reflected in large-scale integrated production models characterized by high density and increasingly shorter production cycles (FAO, 2022; ABPA, 2024; MOTTET; TEMPIO, 2017; BERCKMANS, 2017). This productive advancement imposes growing challenges on environmental control in poultry houses, especially in the face of climate change. In this scenario, Decision Support Systems (DSS) stand out by enabling decisions that promote stable, safe, and reliable production environments.

Despite the relevance of the topic, the literature on computational tools for environmental control in poultry production remains fragmented and poorly systematized, making it difficult to compare approaches and identify trends. In view of this situation, this article conducts a systematic literature review aiming to map, classify, and analyze the DSS applied to environmental control in poultry farming. The adopted methodology follows the guidelines proposed by Kitchenham (2004) and PRISMA 2020, employing the PICOC structure to ensure traceability between research questions, extraction criteria, and analysis (KITCHENHAM, 2004; PAGE et al., 2021).

As a contribution, the study provides a critical synthesis of the state of the art by identifying the types of DSS employed, the decision levels involved, the reported benefits and limitations, and the emerging technologies integrated into these systems. It advances the fields of Production Engineering, poultry science, and DSS by guiding the development of solutions compatible with the realities of intensive poultry production. From a social perspective, it reinforces the role of innovation in promoting animal welfare and food security, aligning with the United Nations Sustainable Development Goals (SDG) 2, 9, and 12.

This article is organized into sections as follows: Section 2 presents the theoretical framework that guided the study; Section 3 describes the adopted methodology; Section 4 presents the obtained results; Section 5 discusses the findings; and Section 6 provides conclusions and future directions.

2. THEORETICAL FRAMEWORK

Global chicken meat consumption has shown consistent growth over recent decades, driven by urbanization, rising per capita income, changes in eating habits, and the search for lower-cost protein sources with reduced environmental impact compared to red meat. This trend is particularly pronounced in Asian and South American countries, where increasing population density intensifies demand. Projections indicate that the global demand for protein may grow by up to 70% by 2050, with chicken meat expected to be the main driver of this expansion (FAO, 2022; MOTTET; TEMPIO, 2017; BERCKMANS, 2009; BERCKMANS, 2017). This growth has led to structural transformations in the poultry production chain. Producers have begun to operate on larger scales, with bigger flocks and shorter, more intensive production cycles. Brazil, the world's largest exporter and second-largest producer, exemplifies this configuration, with more than 50,000 integrated producers operating under strict sanitary and quality standards (ABPA, 2024).

With about 92% of global production concentrated in intensive systems, operational efficiency has come under increasing pressure from complex decisions often made under conditions of uncertainty, variability, and time constraints (HAMSA; BELLUNDAGI, 2017). Factors such as climate, animal health, market dynamics, and bird behavior impose limits on predictability and challenge the accuracy of decision-making (BERCKMANS, 2009; CHENG; MCCARL; FEI, 2022). The main domains include environment (climatic and physical variables), nutrition (feed supply and intake), health (early detection of clinical signs), welfare (behavior and stocking density), and management (lighting, ventilation, and harvesting) (GODINHO et al., 2025; LI et al., 2020; NEETHIRAJAN, 2025; KLOTZ et al., 2020). In terms of scope, operational decisions relate to daily routines such as feeding and environmental control, tactical decisions involve cycle and resource coordination, while strategic decisions concern long-term planning, such as flock reconfiguration or the adoption of new technologies (ZHAI et al., 2020; ROSSI; CAFFI; SALINARI, 2012).

Environmental control in poultry houses has become a critical function not only for ensuring animal welfare and flock health but also for maintaining production efficiency, food safety, and regulatory compliance (CURI et al., 2017; HAFEZ; ATTIA, 2020). Variables such as temperature, humidity, gas concentration, and air velocity affect zootechnical performance in an interdependent and dynamic way, requiring precise management and real-time, context-aware decision-making (GRŽINIĆ et al., 2023; PEREIRA; NÄÄS, 2008; MARTINEZ et al., 2021; QI et al., 2023).

Given these challenges, Decision Support Systems (DSS) play a central role in intensive poultry management by integrating environmental, operational, and zootechnical data into analytical models capable of generating recommendations, simulations, and alerts (ZHAI et al., 2020). By combining control algorithms, predictive models, alarm systems, and digital platforms, these tools are applied to anticipate critical events, identify hidden patterns, and suggest actions based on continuous monitoring (CURI et al., 2017; ASTILL et al., 2020; ZHENG et al., 2021; LIAKOS et al., 2018; BRUGLER et al., 2024). By reducing subjectivity and managing uncertainty more effectively, DSS structures decision-making through reliable, interpretable, and actionable responses aligned with production practices.

3. MATERIALS AND METHODS

The methodology adopted in this systematic review followed the approach proposed by Kitchenham and the PRISMA 2020 guidelines, which guided all stages from the formulation of research questions to the analysis of findings (Kitchenham, 2004; Page et al., 2021).

3.1 Objectives and Research Questions

The main objective of this study was to identify, classify, and analyze Decision Support Systems (DSS) applied to environmental management in poultry farming, focusing on their functionalities, benefits, limitations, and technological trends.

TABLE 1 – Objectives, Research Questions, Methodological Strategies, and PICOC Elements

Specific Objective	Research Question	Data Extraction	Analysis	PICOC Elements
OE1 – Identify the types of DSS used	QP1 – What types of DSS have been used to support environmental control in poultry farming?	Method sections; system, architecture, or tool descriptions	Functional and paradigm-based classification	I, O_1
OE2 – Classify the types of decision-making supported	QP2 – What types of production decisions are supported by DSS in environmental control in poultry farming?	Description of system actions and functionalities; operational flows	Thematic categorization; cross-analysis by tool type and decision scope	P, I, O_1, C_2
OE3 – Identify the benefits obtained	QP3 – What benefits are reported from the use of DSS in environmental control in poultry farming?	Results, discussion, or conclusion sections	Thematic qualitative coding; synthesis of grouped impacts	I, O_1
OE5 – Identify technological trends and	QP5 – What technologies and emerging approaches	Keywords, introduction, and	Open coding; thematic grouping	I, O_1

emerging approaches	are identified in the analyzed studies?	discussion sections; emerging terms		

Fonte: Elaborado pelos autores.

Based on this purpose, five specific objectives and corresponding research questions were defined, structured according to the PICOC logic (Population, Intervention, Comparison, Outcome, Context). Each question was associated with specific extraction and analysis strategies, ensuring alignment between the collected data and the interpretive axes of the study, as presented in Table 1.

3.2 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria, defined according to the scope of the review, are described in Table 2.

TABLE 2 – Inclusion and Exclusion Criteria Adopted in the Review

Inclusion Criteria	Exclusion Criteria
• Peer-reviewed studies published in scientific journals or conference proceedings. • Application of DSS aimed at environmental control in poultry facilities. • Explicit support for decision-making (monitoring, prediction, recommendation, automation, etc.). • Publications with full-text access, published between 2010 and 2025, in English.	• Do not focus exclusively on poultry farming as the main subject. • Do not address environmental control in poultry facilities. • Lack of a computational tool or method related to environmental decision-making. • Literature review articles. • Full text not accessible.

Source: Prepared by the authors.

3.3 Information Sources and Search Strategy

The systematic search was conducted in the Scopus and Web of Science (WoS) databases, selected for their broad coverage of peer-reviewed journals and their recognition in reviews involving animal science and applied computing.

The search strategy combined three conceptual axes — poultry production, environmental control, and decision support. The search terms and filters were adapted to the syntax of each database, as shown in Table 3. In WoS, language and document type filters were manually applied through the interface. The search was performed in June 2025, considering publications from 2010 to 2025.

All records were exported in RIS format and processed using Zotero and Excel.

TABLE 3 – Search Terms Used in Scopus and WoS

Scopus	WoS
TITLE-ABS-KEY (("poultry" OR "broiler" OR "chicken") AND ("environmental control" OR "environmental monitoring" OR "climate control" OR "thermal comfort") AND ("decision support" OR "decision-making" OR "decision system" OR "expert system" OR "computational model" OR "intelligent system")) AND LANGUAGE (english) AND DOCTYPE (ar OR cp) AND PUBYEAR > 2009 AND PUBYEAR < 2026	TS=(("poultry" OR "broiler" OR "chicken") AND ("environmental control" OR "environmental monitoring" OR "climate control" OR "thermal comfort") AND ("decision support" OR "decision-making" OR "decision system" OR "expert system" OR "computational model" OR "intelligent system")) AND PY=(2010-2025)

Fonte: Elaborado pelos autores.

3.4 Screening and Study Selection Process

To ensure reproducibility and minimize judgment bias, the study selection followed a structured multi-step protocol. Initially, all records retrieved from the databases were consolidated and subjected to a deduplication process based on title, authors, and DOI, using Zotero's dedicated functionality. Traceability by database and extraction date was maintained.

Subsequently, titles and abstracts were screened, and studies meeting any previously defined exclusion criterion were immediately removed. In ambiguous cases, a preventive exclusion principle was applied, with justification recorded.

The entire process was documented in a structured spreadsheet containing study identification, source database, decisions at each stage, exclusion justifications, and methodological notes, ensuring transparency and control throughout all selection phases.

3.5 Data Extraction and Analysis

Data extraction was guided by the five research questions defined in the protocol (see Section 3.1), focusing on the methodological aspects of the studies and the technical characteristics of the analyzed solutions. Each included article was assigned a unique identifier and its essential metadata were recorded: title, authors, year, journal or conference, country of origin, methodological approach, and applied methods.

The extracted information was organized in a structured spreadsheet, with categories initially derived from exploratory reviews and refined throughout the reading process. The dataset is available for download in the GitHub repository (MARCUSVINICIUSLEITE, 2025).

The analysis combined directed coding (predefined categories) and emergent coding (categories identified in the data), allowing a critical synthesis guided by thematic patterns and relevant variations. Ambiguities were documented without forcing artificial categorizations.

Analytical strategies included qualitative synthesis, thematic categorization, descriptive frequency analysis, and interpretive cross-tabulation. Extracted evidence was organized in accordance with the research questions (RQs).

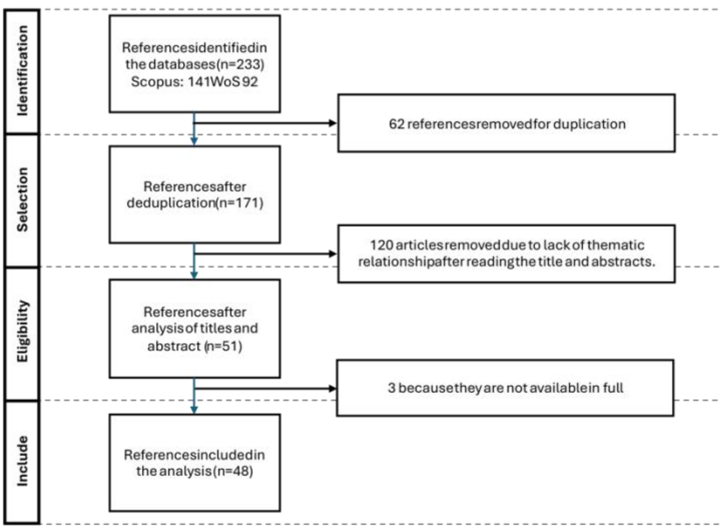
4. RESULTS

This section presents the descriptive synthesis of the selected studies, organized according to the categories defined in the review protocol.

4.1 Search, Screening, and Selection Results

The PRISMA flow diagram (Figure 1) summarizes the search and selection process. Approximately 79% of the records were excluded during the initial screening due to thematic misalignment, suggesting that the keywords used in the databases retrieved a large number of tangential studies.

FIGURE 1 – PRISMA Flow Diagram of the Search, Screening, and Selection of Studies



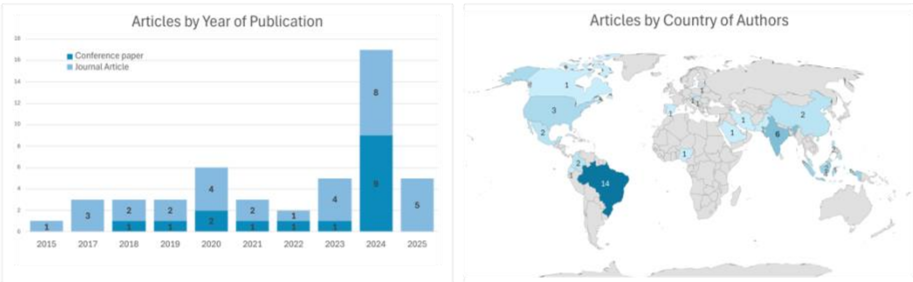
Source: Prepared by the authors..

In the end, only 48 articles explicitly addressed the use of computational tools applied to environmental decision support in poultry farming — a number that reinforces the scarcity of systematized research on the topic..

4.2 General Characteristics of the Studies

The chronological analysis revealed a growing concentration of publications starting in 2020, indicating the advancement of applied research on environmental decision support in poultry farming. The predominance of journal articles over conference papers suggests greater maturity and depth in the publications (Figure 2). The peak in 2024 indicates a possible transition phase, marked by the consolidation of tools and increasing recognition in scientific journals. As data for 2025 are still ongoing, it will be necessary to monitor the coming years to confirm this trend.

FIGURE 2 – Distribution of Articles.



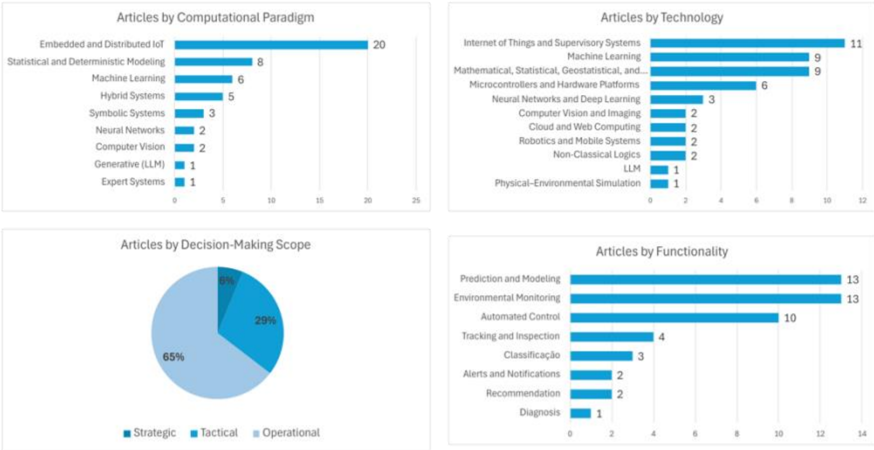
Source: Prepared by the authors.

Regarding the origin of the publications, Brazil leads with 14 articles, followed by India (6), Indonesia (4), and the United States (3), forming the main research hubs in environmental control for poultry farming. Among the countries with two studies, Mexico and Colombia (Latin America) stand out, as well as Malaysia, China, and the Philippines (Asia). Countries with one article include Hungary, the Czech Republic, Latvia, and Spain (Europe); Canada (North America); Nigeria (Africa); Pakistan, Iran, Israel, and Saudi Arabia (Asia); and Ecuador (Latin America).

The geographical distribution of publications aligns consistently with the main global poultry production centers, according to FAO data (2024). The growth of Southeast Asia — particularly India, Indonesia, Malaysia, and the Philippines — highlights the emergence of new technological innovation hubs in the sector. Despite the concentration in key countries, the geographical dispersion demonstrates an increasingly globalized interest in DSS. The frequency analysis by category made it possible to map the dominant focuses of the studies from technical, functional, and decision-making perspectives, as shown in Figure 3.

.

FIGURE 3 – Distribution of the Analyzed Scientific Studies .



Source: Prepared by the authors.

5. DISCUSSION

This discussion follows the structure of the research questions (RQ1–RQ5), articulated with the specific objectives (SO1–SO5) and grounded in the extracted data.

5.1 Types of Decision Support Systems Identified (RQ1)

The review revealed a wide variety of DSS applied to environmental control in poultry houses, which can be organized according to computational paradigms and associated technologies. Embedded IoT architectures predominate, featuring physical sensors and low-power microcontrollers designed for continuous data collection and automated response. In general, these systems operate via edge computing, performing local inference based on machine learning, fuzzy controllers, neural networks, or symbolic models. The combination of these techniques results in hybrid architectures with different levels of autonomy.

A clear trend toward integrating multiple paradigms is observed—such as supervised learning with non-classical logics or computer vision with deep neural networks—aiming to improve adaptability under production variability. Most models concentrate on three main applications: environmental risk prediction, anomaly detection, and intelligent actuation. Although less common, rule-based expert systems, statistical models, and spatial optimization approaches for sensor layout design also stand out, particularly in contexts requiring explainability or formal robustness.

Finally, although still incipient, the use of generative models and large language models (LLMs) has emerged for semantic interpretation and text-based decision support, opening the way for language-oriented approaches. Most tools remain in experimental stages but already indicate a transition toward more sophisticated and adaptive DSS.

5.2 Productive Decision Scopes Supported by DSS in Poultry Farming (RQ2)

Decision-making in intensive poultry systems requires tools capable of operating across multiple levels—operational, tactical, and strategic. The analyzed literature indicates an excessive concentration on short-term decisions, revealing a gap in the application of DSS for medium- and long-term planning.

Most solutions focus on continuous monitoring of environmental variables, integrated with real-time alerts, remote visualization, and automated control of climate, ventilation, and lighting. Machine learning–based classification models are recurrent, targeting the detection of anomalies, thermal stress, and atypical patterns. Predictive systems expand this scope by anticipating critical events, simulating environmental variations, and projecting zootechnical impacts, with both operational and tactical applications.

There are also more advanced tools offering corrective action recommendations, mobile sensory tracking, and automated inspection, as well as solutions for diagnostics, adaptive planning, and optimization under climatic variability. In summary, while the predominant focus remains on short-term operational decisions, there is a progressive expansion toward more complex, tactical, and predictive functions.

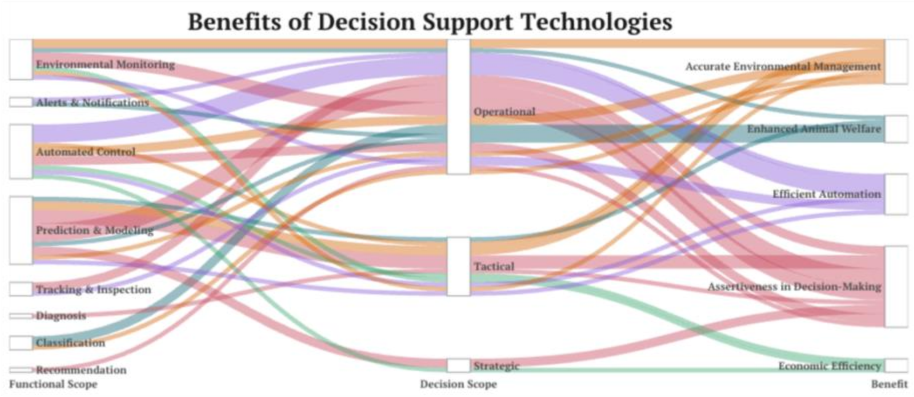
5.3 Reported Benefits of DSS Use for Environmental Control in Poultry Systems (RQ3)

The analysis of the studies revealed consistent benefits from the adoption of DSS in environmental control, organized into five main dimensions shown in Figure 4.

Accurate Environmental Management emerges as the central dimension, with emphasis on risk prediction and the control of critical variables. Models based on machine learning, neural networks, and advanced statistical methods demonstrated high accuracy in predicting temperature, humidity, ammonia concentration, and pathogen prevalence, contributing to thermal stability, gas control, and failure prevention.

A análise dos estudos revela benefícios consistentes decorrentes da adoção de SSD no controle ambiental, organizados em cinco dimensões principais.na Figura 4.

FIGURE 4 – Overview of Benefits by Functional and Decision Scope.



Fonte: Elaborado pelos autores.

Efficient Automation emerges as a key driver of operational efficiency, reducing human error through systems integrated with sensors and actuators that perform real-time corrections. In distributed or edge architectures, higher robustness and low latency were observed, even in environments with limited infrastructure.

Regarding *Assertiveness in Decision-Making* at tactical and strategic levels, the analyzed tools structure the decision process by combining environmental, physiological, and production data with interpretive algorithms. Some solutions incorporate semantic models such as LLMs, enhancing the intelligibility of recommendations. The emerging use of these models for text-based decisions suggests an important conceptual shift: beyond traditional numerical inference, new models are beginning to generate interpretable diagnoses or recommendations, linking natural language processing and decision support.

Enhanced Animal Welfare appears associated with reduced thermal stress, stabilization of critical conditions, and early detection of physiological or behavioral changes.

Finally, *Economic Efficiency* is evidenced through direct savings (energy, sensors, labor) and gains in productivity, predictability, and sustainability. Low-cost IoT-based tools have proven particularly effective in contexts with structural constraints.

5.4 Challenges, Limitations, and Methodological Gaps in the Analyzed Studies (RQ4)

The review identified critical limitations that compromise the applicability and scalability of the studied DSS in real production environments.

At the methodological level, the absence of field validation was predominant: many models were tested only in controlled environments, with small samples, caged birds, and simplified environmental variables, preventing generalization of the results. Tests conducted in multiple environments, longitudinal analyses, or extensive statistical comparisons were rare.

At the technical-operational level, the main bottlenecks included sensor fragility in harsh environments, strong dependence on network connectivity in regions with weak infrastructure, frequent communication failures, and the use of low-cost sensors without accuracy validation. Few systems incorporated fault tolerance, outlier detection, or loss-compensation mechanisms. Persistent gaps were also identified in data standardization, interoperability, information security, and economic feasibility. Some solutions required high investment, specialized infrastructure, and technical training, limiting their scalability.

From a functional standpoint, most systems focused on passive monitoring, with automated control either absent or limited to fixed thresholds and simple rule-based mechanisms. The majority of systems ignored zootechnical variables, did not integrate learning algorithms, and required direct human intervention.

Although environmental control is only one of the decision-making domains in poultry farming, it has direct interdependence with others—such as nutrition, health, and welfare—which demands integrated rather than isolated solutions. However, the analyzed systems rarely articulated these interactions, limiting their systemic value in production management. Integration with subsystems such as nutrition, health, or welfare was virtually nonexistent. Discussions on regulatory compliance, privacy, and legal responsibility were also scarce.

These constraints hinder large-scale adoption and the establishment of comparative frameworks. Overcoming them requires methodological advances, expanded functional scope, field validation in commercial farms, and the development of resilient, interoperable, and semantically integrable solutions. The analysis revealed that the effectiveness of DSS depends less on isolated algorithmic sophistication and more on their ability to integrate with production processes. Tools that fail to deliver interpretable, actionable, and context-compatible recommendations tend to underperform in practice, even when technically promising.

5.5 Technological Trends and Emerging Approaches (RQ5)

The review identified a significant set of emerging technologies that expand the functional scope of DSS and aim to overcome structural limitations. Edge–cloud architectures using devices such as ESP32 and Jetson Nano have begun to perform local inferences with low latency, even under limited connectivity. The adoption of these architectures has significantly increased the resilience of environmental systems, demonstrating that intelligent automation in poultry houses can be viable even outside major production centers. Applications featuring automated control and sensors for sound, image, and environmental variables have gained prominence, enabling real-time action.

At the sensory layer, advances include the use of wearable sensors, RFID with UHF triangulation, thermal cameras, and integrated microphones. The fusion of physiological and environmental data feeds more robust predictive models, incorporating new stress indicators based on audio spectrograms and geometric descriptors. Structural innovations include reinforced encapsulation and solar panels for greater autonomy.

In computational terms, new approaches have emerged, such as tinyML, adaptive fuzzy logic, YOLOv9, GELAN, kriging, SLAM, and hybrid zootechnical–statistical models. Increasing connectivity with APIs, ERPs, dashboards, and mobile platforms has been observed, along with georeferenced visualizations, SCADA systems, and blockchain solutions for traceability.

Finally, conceptual proposals have emerged that explore LLMs integrated with RAG, non-classical logic, explainable AI, and synthetic simulations. The integration of these semantic models into environmental decision systems for poultry houses not only enhances the interpretability of results but also strengthens user trust—highlighting a cognitive dimension often overlooked in technology adoption. These advances signal a new generation of DSS: more autonomous, adaptive, and semantically integrated into the production context.

5. CONCLUSIONS

The intensification of global poultry farming, particularly in tropical countries, has generated a paradox: while increasing production scale and zootechnical efficiency requirements, it has also heightened risks associated with environmental variability. Consequently, environmental control has evolved from a purely technical function to a strategic one. In this context, the development of solutions capable of supporting facility management under variability, uncertainty, and operational constraints has become essential.

This systematic review analyzed 48 primary studies on Decision Support Systems (DSS) applied to environmental control in poultry houses, following Kitchenham's methodology (2004) and PRISMA 2020 guidelines.

The findings revealed an expanding technological ecosystem with a trend toward hybrid architectures that overcome the limits of isolated models by combining different computational paradigms to achieve greater adaptability to environmental dynamics. These heterogeneous solutions combined IoT, machine learning, non-classical logics, symbolic models, and neural networks—converging on applications such as detection of critical fluctuations, microclimatic instabilities, and behavioral changes, as well as intelligent control actuation. Most tools focused on short-term operational decisions, but there are signs of progress toward more complex tactical functions, indicating a gradual maturation of DSS toward more adaptive and proactive systems. Reported benefits included improvements in accuracy, decision reliability, and efficiency, with meaningful impacts on autonomy, animal welfare, and economic outcomes, even under structural constraints.

Emerging trends reveal a technological transition: edge–cloud architectures, wearable sensors, tinyML, adaptive fuzzy logic, explainable AI, connectivity with zootechnical and production management systems, mobile platforms, and georeferenced visualizations point to more robust and interconnected systems. The initial incorporation of LLMs, non-classical logics, and synthetic simulations suggests the emergence of DSS that are more autonomous, interpretable, and semantically integrated.

Despite these advances, important gaps remain. Most studies were limited to controlled environments, with low replicability, restricted validation, and limited integration with other dimensions of poultry production. Operational and methodological weaknesses persist, as well as a lack of truly interoperable solutions applicable to commercial farms.

By systematizing the types, scopes, benefits, limitations, and trends of DSS applied to environmental control in poultry farming, this study established a critical knowledge base that supports technical and scientific progress in production engineering, poultry science, and intelligent systems. The findings provide practical guidance for developing solutions suited to the realities of intensive production systems and reinforce the strategic role of digital innovation in promoting animal welfare, food security, and sustainability—in alignment with SDGs 2, 9, and 12.

The consolidation of this field will depend less on isolated technological advances and more on the realistic integration of these technologies into the operational conditions of actual production environments.

Future research should prioritize field validation of DSS across different poultry house profiles and climatic conditions, the development of mechanisms for fault tolerance, resilience to inconsistencies and data loss, and continuous learning strategies that enable effective integration of DSS with other poultry production processes such as nutrition, health, welfare, and logistics.

Equally promising are investigations that integrate automated inference with explainable semantic models and generative artificial intelligence, enhancing user trust and the overall effectiveness of decision-making.

REFERENCES

ABPA. Relatório Anual 2024. Associação Brasileira de Proteína Animal. Disponível em: https://abpa-br.org/wp-content/uploads/2024/04/RA_2024_ABPA_ingles_avicultura.pdf.

ASTILL, J. et al. Smart poultry management: Smart sensors, big data, and the internet of things. Computers and Electronics in Agriculture, v. 170, p. 105291, 2020.

BERCKMANS, D. Automatic on-line monitoring of animal health and welfare by Precision Livestock Farming. In: European Forum Livestock Housing for the Future, Lille, France, 2009.

BERCKMANS, D. General introduction to precision livestock farming. Animal Frontiers, v. 7, n. 1, p. 6–11, 2017.

BRUGLER, Skye; GARDEZI, Maaz; DADKHAH, Ali; RIZZO, Donna M.; ZIA, Asim; CLAY, Sharon A. Improving decision support systems with machine learning: Identifying barriers to adoption. Agronomy Journal, *[S. l.]*, v. 116, n. 3, p. 1229–1236, 2024.

CHENG, Muxi; MCCARL, Bruce; FEI, Chengcheng. Climate Change and Livestock Production: A Literature Review. Atmosphere, *[S. l.]*, v. 13, n. 1, p. 140, 2022.

CURI, T. M. R. de C. et al. Positioning of sensors for control of ventilation systems in broiler houses: a case study. Scientia Agricola, v. 74, p. 101–109, 2017.

FAO. Production: Crops and livestock products. In: *FAOSTAT*. Rome: FAO, 2022. Disponível em: https://www.fao.org/faostat/en/#data/QCL. Acesso em: out. 2023.

GODINHO, António; VICENTE, Romeu; SILVA, Sérgio; COELHO, Paulo Jorge. Wireless Environmental Monitoring and Control in Poultry Houses: A Conceptual Study. IoT, *[S. l.]*, v. 6, n. 2, p. 32, 2025.

GRŽINIĆ, G. et al. Intensive poultry farming: A review of the impact on the environment and human health. Science of the Total Environment, v. 858, p. 160014, 2023.

HAFEZ, H. M.; ATTIA, Y. A. Challenges to the poultry industry: Current perspectives and strategic future after the COVID-19 outbreak. Frontiers in Veterinary Science, v. 7, p. 516, 2020.

HAMSA, K. R.; BELLUNDAGI, Veerabhadrappa. Review on Decision-making under Risk and Uncertainty in Agriculture. Economic Affairs, *[S. l.]*, v. 62, n. 3, p. 447, 2017.
KITCHENHAM, Barbara. Procedures for performing systematic reviews. Keele, UK, Keele University, v. 33, n. 2004, p. 1–26, 2004.

KLOTZ, Darlan Felipe; RIBEIRO, Richardson; ENEMBRECK, Fabrício; DENARDIN, Gustavo; BARBOSA, Marco; CASANOVA, Dalcimar; TEIXEIRA, Marcelo. Estimating action plans for smart poultry houses. arXiv, , 2020.

LI, N.; REN, Z.; LI, D.; ZENG, L. Review: Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming. Animal, *[S. l.]*, v. 14, n. 3, p. 617–625, 2020.
LIAKOS, Konstantinos G.; BUSATO, Patrizia; MOSHOU, Dimitrios; PEARSON, Simon; BOCHTIS, Dionysis. Machine Learning in Agriculture: A Review. Sensors, *[S. l.]*, v. 18, n. 8, p. 2674, 2018.

MARCUSVINICIUSLEITE. marcusviniciusleite/DSS-EnvironmentControl-PoultryFarming. , 2025. Disponível em: https://github.com/marcusviniciusleite/DSS-EnvironmentControl-PoultryFarming. Acesso em: 10 jul. 2025.

MARTINEZ, A. A. G. et al. A heuristic and data mining model for predicting broiler house environment suitability. Animals, v. 11, n. 10, p. 2780, 2021.

MOTTET, A.; TEMPIO, G. Global poultry production: Current state and future outlook and challenges. Journal of the World's Poultry Science, v. 73, p. 245–256, 2017.

NEETHIRAJAN, Suresh. Rethinking Poultry Welfare—Integrating Behavioral Science and Digital Innovations for Enhanced Animal Well-Being. Poultry, *[S. l.]*, v. 4, n. 2, p. 20, 2025.
PAGE, Matthew J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, v. 372, p. n71, 29 mar. 2021.

PEREIRA, D. F.; NÄÄS, I. A. Estimating the thermoneutral zone for broiler breeders using behavioral analysis. Computers and Electronics in Agriculture, v. 62, n. 1, p. 2–7, 2008.
PRISMA statement. Disponível em: <https://www.prisma-statement.org>. Acesso em: 19 jun. 2025.

QI, F. et al. Environmental factor detection and analysis technologies in livestock and poultry houses: a review. Agriculture, v. 13, n. 8, p. 1489, 2023.

ROSSI, Vittorio; CAFFI, Tito; SALINARI, Francesca. Helping farmers face the increasing complexity of decision-making for crop protection. Phytopathologia Mediterranea, *[S. l.]*, v. 51, n. 3, p. 457–479, 2012.

ZHAI, Zhaoyu; MARTÍNEZ, José Fernán; BELTRAN, Victoria; MARTÍNEZ, Néstor Lucas. Decision support systems for agriculture 4.0: Survey and challenges. Computers and Electronics in Agriculture, *[S. l.]*, v. 170, p. 105256, 2020.

ZHENG, H. et al. Design and implementation of poultry farming information management system based on cloud database. Animals, v. 11, p. 900, 2021.

3.2 Article 2 – Enhancing Environmental Control in Broiler Production: Retrieval-Augmented Generation for Improved Decision-Making with Large Language Models

The second article, titled "*Enhancing Environmental Control in Broiler Production: Retrieval-Augmented Generation for Improved Decision-Making with Large Language Models*," was published in January, 2025 in AgriEngineering (MDPI), Volume 7, Issue 1, Article 12. The authors are Marcus Vinícius Leite, Jair Minoro Abe, Marcos Leandro Hoffmann Souza, and Irenilza de Alencar Nääs, affiliated with Universidade Paulista (UNIP) and Universidade do Vale do Rio dos Sinos (UNISINOS).

The study aimed to empirically evaluate the impact of the RAG technique on the performance of LLMs in decision-support tasks related to environmental control in broiler farms. This stage corresponds to the experimental phase of the research and serves as the link between the theoretical diagnosis presented in the first article and the integrative modeling developed in the third. The purpose was to determine whether incorporating external evidence, retrieved from domain-specific knowledge bases, could enhance the semantic accuracy, contextual relevance, and practical applicability of LLM-generated responses (Lewis et al., 2020; Izacard & Grave, 2020; Li et al., 2022).

The investigation was conducted within the scope of natural-language-based DSS designed for the interpretation of technical and environmental control data in poultry production. Under controlled conditions, the study tested whether adding a document retrieval layer to a generation pipeline following the RAG paradigm could mitigate known limitations of purely generative models, such as factual gaps, inconsistencies, and hallucinations (Ji et al., 2022; Metze et al., 2024). The results of this stage guided parameter and metric adjustments for the subsequent logical-computational modeling phase, establishing RAG as a core component of the system proposed in this master's thesis.

The experiment was conducted using a set of technical queries developed from international protocols and recommendations on environmental control in poultry farming (Mottet & Tempio, 2017; Hafez & Attia, 2020). Each query was submitted to controlled executions of state-of-art LLM (GPT 4o) under two conditions: without and with RAG. In the second configuration, the model accessed a domain-specific knowledge base indexed by vector representations and retrieved by semantic similarity using FAISS and LangChain (Reimers & Gurevych, 2019; Devlin et al., 2018; Brown et al., 2020). The responses were evaluated using semantic similarity and contextual relevance metrics, computed with Sentence-BERT embeddings, and statistically analyzed through a paired t-test, following methodologies inspired by comparative studies of retrieval and generation techniques (Guo et al., 2022; Reimers & Gurevych, 2019). The entire experimental pipeline was implemented in Python, employing the langchain, faiss-cpu, and sentence-transformers libraries, with all code and datasets publicly released to ensure transparency and reproducibility.

The results showed significant performance improvements across all metrics. Semantic similarity between responses and reference standards increased markedly, accompanied by a substantial rise in contextual relevance (Lewis et al., 2020). Responses generated with RAG were more accurate, complete, and auditable, showing a marked reduction in hallucinations and inconsistencies (Ji et al., 2022; Doshi-Velez & Kim, 2017). Although the retrieval layer slightly increased average response time, the additional computational cost was offset by higher reliability and traceability, attributes essential for decision-making systems in sensitive operational contexts (Vaswani et al., 2017; Berckmans, 2017).

From a theoretical standpoint, the results demonstrate that RAG acts as an evidential control mechanism by grounding responses in verifiable content and constraining the model's uncertainty space. Conceptually, this function parallels the weighting structure between favorable and unfavorable evidence in the Logic E_τ (Abe, 2011; Abe & Carvalho, 2018), reinforcing the convergence between probabilistic reasoning and logical-evidential inference. The integration of RAG with LLMs thus enables a more coherent and verifiable inferential process, in which the model not only generates responses but also reasons based on evidence, approaching analytical and interpretable behavior.

These findings consolidate RAG as an intermediate layer between semantic interpretation and logical-evidential inference, establishing the operational bridge that supports the architecture proposed in this master's thesis. Its adoption enables natural-language-based decision support systems to operate with greater reliability, traceability, and technical grounding, key features for the development of conversational agents in intensive poultry farming and, more broadly, in productive domains characterized by uncertainty and contradiction.

AgriEngineering

IMPACT FACTOR 3.0

CITESCORE 4.7

Article

Enhancing Environmental Control in Broiler Production: Retrieval-Augmented Generation for Improved Decision-Making with Large Language Models

Marcus Vinicius Leite, Jair Minoro Abe, Marcos Leandro Hoffmann Souza and Irenilza de Alencar Nääs

Special Issue

The Future of Artificial Intelligence in Agriculture

Edited by

Prof. Dr. Sotirios K. Goudos, Prof. Dr. Shaohua Wan and Dr. Achilles Boursianis

MDPI

https://doi.org/10.3390/agriengineering7010012

AgriEngineering

MDPI

Article

Enhancing Environmental Control in Broiler Production: Retrieval-Augmented Generation for Improved Decision-Making with Large Language Models

Marcus Vinicius Leite [1], Jair Minoro Abe [1], Marcos Leandro Hoffmann Souza [2] and Irenilza de Alencar Nääs [1,*]

[1] Graduate Program in Production Engineering, Paulista University, Rua Dr. Bacelar 1212, São Paulo 04026-002, SP, Brazil; marcus.leite@gmail.com (M.V.L.); jair.abe@docente.unip.br (J.M.A.)
[2] Computer Science, Universidade do Vale do Rio dos Sinos, Av. Unisinos, 950, São Leopoldo 93022-750, RS, Brazil; marcosl.hoffmann@yahoo.com.br
* Correspondence: irenilza.naas@docente.unip.br

Abstract: The growing global demand for animal protein, particularly chicken meat, challenges poultry farming to adapt production systems through the adoption of digital technologies. Among the promising advances in artificial intelligence (AI), large language models (LLMs) hold potential to enhance decision-making in broiler production by supporting environmental control through the interpretation of climatic data, the generation of reports to optimize conditions, guidance on ventilation adjustments, recommendations for thermal management, assistance in air quality monitoring, and the translation of simulation results into actionable suggestions to improve bird welfare. For this purpose, the key limitations of LLMs in terms of transparency, accuracy, precision, and relevance must be effectively addressed. This study investigates the impact of retrieval-augmented generation (RAG) on improving LLM precision and relevance for environmental control in broiler production. Experiments with the OpenAI GPT-4o model and semantic similarity analysis were used to evaluate response quality with and without RAG. The results confirmed the approach's effectiveness while identifying areas for improvement. A paired *t*-test revealed significantly higher similarity scores with RAG, demonstrating its impact on response quality. This study contributes to the field by advancing RAG-enhanced LLMs for environmental control, addressing market demands by demonstrating how AI improves decision-making for productivity and animal welfare, and benefits society by providing small-scale producers with cost-effective and accessible solutions for actionable insights.

Keywords: retrieval-augmented generation (RAG); GPT; large language model (LLM); smart poultry farming; precision livestock farming

Academic Editor: Lilong Chai

Received: 30 November 2024
Revised: 20 December 2024
Accepted: 3 January 2025
Published: 8 January 2025

Citation: Leite, M.V.; Abe, J.M.; Souza, M.L.H.; de Alencar Nääs, I. Enhancing Environmental Control in Broiler Production: Retrieval-Augmented Generation for Improved Decision-Making with Large Language Models. *AgriEngineering* **2025**, *7*, 12. https://doi.org/10.3390/agriengineering7010012

1. Introduction

Global demand for animal protein, particularly poultry, has dramatically increased due to economic growth and changing dietary preferences, especially in developing regions across Asia and South America. As incomes rise, diets shift toward more frequent animal product consumption, a trend expected to drive a 70% increase in demand by 2050 [1–3]. This growth places immense pressure on the livestock sector to scale production while maintaining quality and safety standards. In response, the poultry industry has adopted intensive farming practices, where high-density production is essential for meeting demand. Intensive poultry farming now accounts for approximately 92% of global poultry production, with countries like Brazil exemplifying this shift through large-scale, integrated systems involving over 50,000 producers who follow strict health and safety standards regulated by large corporations [1–4]. While necessary, this transformation to high-density production introduces new challenges in managing animal health, welfare, and environmental conditions. High-density systems require strict biosecurity and environmental management to prevent disease, maintain productivity, and comply with complex animal welfare standards. Effective environmental control is vital for optimizing growth and feed conversion efficiency while ensuring sustainable production practices. The poultry industry has increasingly turned to smart poultry farming, a subset of precision livestock farming (PLF) that utilizes digital technologies to automate and enhance monitoring and management processes to address these challenges. This approach meets modern poultry farming's rigorous demands by improving productive efficiency, promoting animal welfare, and supporting regulatory compliance [1,3,5–8].

In smart poultry farming, PLF systems typically perform three main functions: detection and monitoring, data analysis, and decision-making. Detection and monitoring technologies, such as IoT-based environmental sensors, gather extensive data on critical parameters within the poultry house environment, including temperature, humidity, air speed, and gas concentrations. While advancements in detection and monitoring are considerable, the sheer volume of data produced presents significant challenges in the stages of analysis and decision-making. Transforming this raw data into actionable insights that producers can use for real-time adjustments is complex, often requiring advanced analytical tools and expertise that may not be readily accessible to producers [7–11].

Previous work has demonstrated significant advancements in integrating machine learning for real-time health and welfare monitoring in poultry farms, highlighting the critical role of data-driven insights in precision livestock management. For example, computer vision systems have been employed to monitor and predict broiler behaviors and recognize stress-related conditions using technologies such as convolutional neural networks (CNNs) and deep reinforcement learning. Additionally, AI techniques like neural networks and support vector machines (SVMs) have been applied to analyze poultry vocalizations and behaviors, achieving high accuracy in classifying activities and detecting health or welfare issues [12–14]. Despite illustrating the potential of machine learning to address key challenges in poultry farming, these efforts have focused mainly on isolated AI applications with limited exploration of contextual knowledge to enhance decision-making. Consequently, they fall short of adequately addressing the complexities of analysis and decision-making stages, which often require sophisticated tools and expertise that are not readily accessible to producers [14].

The current approach to addressing these challenges relies on expert consultants who analyze data, respond to producers' inquiries, provide guidance, and support decision-making. While these specialists play an important role in interpreting complex datasets and delivering tailored recommendations, their services are often expensive, limiting accessibility for small-scale producers. Moreover, consultancy typically relies on historical data, resulting in outdated insights that reduce the efficiency and effectiveness of recommendations, particularly in the fast-paced environment of poultry farming [10–15]. These limitations underscore the need for innovative solutions to enable real-time, cost-effective, and accessible decision-making support for producers.

To address these challenges, large language models (LLMs) are a promising technology with significant potential to generate insights from extensive textual data, leveraging deep learning and natural language processing (NLP) techniques. Built on the Transformer architecture, LLMs—such as OpenAI's GPT, Google's BERT, and Meta's LLaMA—incorporate attention mechanisms that enable efficient contextual understanding of complex linguistic datasets [16–18]. These capabilities allow LLMs to summarize documents, generate reports, and answer questions, transforming raw data into intuitive insights supporting decision-making across healthcare, education, finance, agriculture, media, and scientific research [19–22].

In environmental control for broiler production, LLMs can support producers in environmental control activities such as monitoring climatic variables by analyzing and interpreting real-time data generated by sensors (e.g., temperature, humidity, and gasses) and producing automated reports with recommendations. They can also assist in ventilation control by diagnosing problems, interpreting sensor data, and suggesting possible causes based on known patterns. They act as a configuration assistant that guides producers in natural language on adjusting fans and evaporative panels based on environmental data and best practices. Additionally, LLMs can aid in heat and cold management by providing suggestions and recommendations to optimize thermal management (e.g., adjusting curtains or fan intensity) using analyses of environmental conditions and historical data. They can support air quality analysis by interpreting gas concentration readings, issuing alerts for unsafe levels, explaining how gas levels impact bird health and performance, and providing solutions based on best practices, as well as generating automated reports on air quality, linking ammonia or carbon dioxide levels to environmental conditions. Finally, LLMs can assist producers in operational adjustments by interpreting and translating simulation results and offering detailed suggestions on priority actions to improve environmental conditions.

Despite their promise, LLMs face substantial limitations that hinder their applicability in critical domains like environmental management in broiler farming. Their "black box" nature limits transparency, reducing user trust, particularly where verifiable justification is required. Additionally, LLMs often lack contextual specificity, generating generalized responses that fail to address the unique conditions of production sites. This limitation becomes more pronounced in rapidly evolving fields, as models trained on static datasets may provide outdated or irrelevant information. Finally, another critical concern is the generation of hallucinations, where LLMs produce plausible but incorrect responses, further undermining reliability [20,22,23].

To mitigate these shortcomings, techniques such as retrieval-augmented generation (RAG) and prompt engineering have been proposed [24,25]. In the RAG approach, a 'smart retriever' technology gathers data from external knowledge bases and the user's query to create enriched input (Figure 1). The LLM then leverages this input to generate accurate, context-sensitive responses [26–28]. RAG proposes to enhance LLMs by integrating a retrieval mechanism that provides relevant and up-to-date information during response generation. This approach aims to improve transparency, mitigate traditional models 'black box' limitations, and address key challenges such as outdated training data, hallucinations, and limited accuracy [26–33].

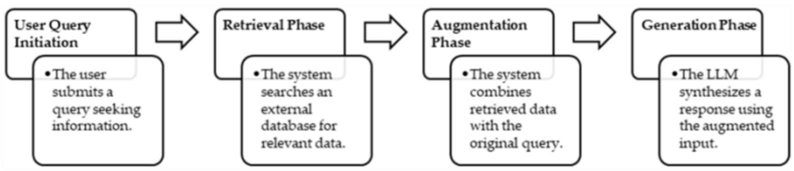


Figure 1. Schematic of the RAG Process Flow. Source: the authors.

However, RAG performance relies on the quality and relevance of the retrieved data, and maintaining updated databases remains resource-intensive. Scalability is another challenge, as computational costs and response times can limit real-time applications despite ongoing advancements in storage and retrieval mechanisms to improve their efficiency and reliability [26,30].

Given these considerations, this study employs an experimental design to investigate whether the RAG technique enhances the precision and relevance of LLM-generated responses for environmental control in broiler poultry farming. We hypothesize that RAG will improve LLM performance by providing contextual information, enabling producers to make informed decisions. The research framework ensures reliable conclusions through controlled variables and reproducible methodologies.

This study contributes to science by advancing RAG-enhanced LLMs for environmental control, demonstrating how AI improves the precision and reliability of decision-making in critical applications such as poultry farming. It addresses market demands by providing a practical framework that supports the poultry industry in enhancing productivity, animal welfare, and regulatory compliance. Additionally, it benefits society by bridging the knowledge gap for small-scale producers, offering cost-effective, accessible, and actionable insights to improve operations without requiring expert consultancy or advanced technical expertise.

2. Materials and Methods

This study employs an experimental design to assess the impact of the RAG technique on the quality of responses generated by an LLM. By introducing RAG as a variable, the experiment measures its effect on the semantic similarity index, quantifying the responses' accuracy and relevance. Responses with and without RAG are compared under controlled conditions to evaluate how contextual augmentation influences semantic alignment.

To achieve this, the methodology is structured into three main phases (Figure 2): database creation, experimental execution, and comparative analysis. These phases evaluate the effectiveness of RAG in improving the precision and relevance of LLM-generated responses within the domain of environmental control in broiler poultry farming.

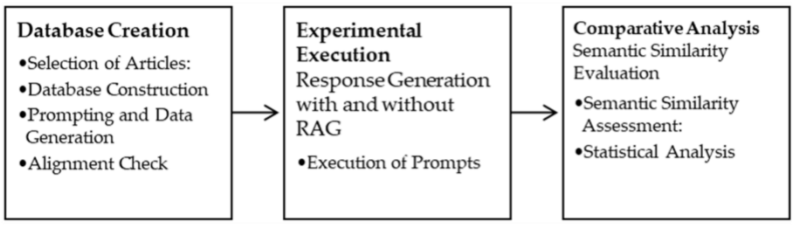


Figure 2. Schematic flow of the research approach. Source: the authors.

The structured methodology, including tools, datasets, and computational frameworks, supports the replication of this experiment.

2.1. Technologies

We selected Python version 3.13 for this experiment due to its versatility, extensive library ecosystem, and strong support for NLP and machine learning applications. LlamaIndex and LangChain were adopted to implement RAG and non-RAG models. At the same time, FAISS was used to store, index, and retrieve high-dimensional vectors efficiently, enabling fast and accurate retrieval of relevant information to enhance the RAG process. Additionally, we utilized OpenAI's ChatGPT-4o model, recognized for its state-of-the-art natural language understanding and generation capabilities and its compatibility with RAG frameworks to improve response accuracy [19].

2.2. Database Creation

The database aims to organize and store all relevant information necessary for executing all experiment phases to evaluate RAG's effectiveness in improving the response's accuracy.

2.2.1. Selection of Article

Initially, we conducted a targeted literature review to identify recent studies on environmental control in broiler poultry farming. We used the following query string in the Scopus and Web of Science databases to identify relevant studies on environmental control in broiler production:

TITLE-ABS-KEY ("broiler chickens" OR "broiler production" OR "broiler houses" OR "poultry house" OR "broiler chicken barns") AND (TITLE-ABS-KEY ("environmental control" OR "environmental management" OR "climate control" OR "air quality") AND TITLE-ABS-KEY ("ventilation" OR "temperature" OR "humidity" OR "gas emissions" OR "NH_3" OR "ammonia" OR "CO_2" OR "carbon dioxide" OR "CO" OR "carbon monoxide" OR "H_2S" OR "hydrogen sulfide") AND TITLE-ABS-KEY ("welfare" OR "comfort" OR "performance" OR "productivity")) AND PUBYEAR > 2019 AND PUBYEAR < 2023 AND (LIMIT-TO (SUBJAREA, "AGRI") OR LIMIT-TO (SUBJAREA, "VETE") OR LIMIT-TO (SUBJAREA, "ENVI")) AND (LIMIT-TO (LANGUAGE, "English"))

This query is structured to capture articles focused on broiler chicken production (including terms like "broiler houses" and "poultry house") with a strong emphasis on environmental control aspects (e.g., ventilation, temperature, humidity, and gas emissions). It also targets research on animal welfare, performance, and productivity, reflecting the multi-dimensional impacts of environmental factors on broiler farming. The search is limited to studies published between 2020 and 2023 and includes relevant subject areas: agriculture and veterinary and environmental science.

The cut-off date was set because the GPT-4o model was trained with data up to 2023 [16]. This limitation was intentionally established to avoid creating a scenario where the RAG approach would naturally perform better, as the LLM would not have access to articles published beyond 2023. By excluding more recent materials, the comparison ensures a fair evaluation of RAG's effectiveness.

We selected a sample of the ten most-cited articles from the search results. This citation-based sampling approach ensured that our dataset included high-impact studies widely recognized within the field, likely to offer comprehensive and relevant insights into environmental control in broiler production.

Our targeted search identified the ten most-cited articles on environmental control in broiler farming to form our dataset. Table 1 lists these selected articles, which serve as the foundation for evaluating the effectiveness of RAG in enhancing response accuracy.

Table 1. Selected sources, article identification (DOI), and the number of citations received.

Source	Digital Object Identifier (DOI)	Cited by
Bist et al. [34]	10.1016/j.jenvman.2022.116919	52
Bloch et al. [35]	10.1016/j.biosystemseng.2019.08.011	28
Costantino et al. [36]	10.1016/j.biosystemseng.2020.01.002	25
Costantino et al. [37]	10.3390/ani10091539	24
Ahmadi Babadi et al. [38]	10.1016/j.compag.2021.106677	21
Li et al. [39]	10.3390/ani10122252	21
Al Assaad et al. [40]	10.1016/j.biosystemseng.2021.01.002	18
Soliman et al. [41]	10.17582/JOURNAL.AAVS/2020/8.9.997.1008	17
Peng et al. [42]	10.1016/j.psj.2021.101587	15

Source: the authors.

2.2.2. Database Construction

We created a structured database based on scientific articles about environmental control in broiler chicken farming. This database stored all the necessary information to evaluate the effectiveness of RAG in improving response accuracy (Table 2).

Table 2. Database columns.

Column	Description
Article Citation	Identifies the source of the article.
Page Number	Specifies the page from which the question was derived.
Full Text of the Page	Contains the full text of the page used to generate the question.
Question	Contains the question derived from the specified page.
Correct Answer	Provides an interpretive response that captures the implied meaning within the context of the text.
Original Text Answering the Question	Extracts the direct answer as stated in the article.

2.2.3. Prompting and Data Generation

To generate questions that would be answered by the LLM both without and using the RAG technique, we created questions using each of the selected articles. For this purpose, we utilized OpenAI's ChatGPT-4o based on the GPT-4o model.

We developed a set of tailored prompts to generate questions from each article. The primary prompt outlined the task's objective, specifying that the questions be generated from scientific articles according to specific criteria. First, they should be relevant to broiler producers, addressing genuine concerns, challenges, or curiosities related to environmental control in poultry farming. Second, the questions should align with the page content, ensuring they can be answered using the information explicitly or implicitly available on the specific page. Additionally, the questions should be clear, simple, concise, and free of overly technical jargon while maintaining accuracy and professionalism. A practical focus should also be emphasized, prioritizing actionable information producers can use to improve farm management or decision-making. To ensure broad applicability, the questions should avoid references to paper-specific terminology, methodologies, or findings, instead generalizing the content into a form relatable to the producer's context. Finally, the questions should be designed to encourage meaningful responses through open-ended exploration or contextualizing specific issues. Adhering to these guidelines ensured that the generated questions were realistic, practical, and aligned with the intended purpose of evaluating LLM performance in addressing practical queries.

The article input prompt provided the scientific article in PDF format. Finally, the iterative question-generation prompts repeated for each page to create a question for each page in each article.

An initial set of 133 questions was generated by these prompts and was saved in the experiment database. The results reflect the state of the Scopus and Web of Science databases as of October 2024, when the searches were executed.

2.2.4. Alignment Check of Generated Questions

We checked the adherence of each generated question with the established guidelines to ensure that they aligned with the intended purpose of evaluating the LLM's performance in dealing with practical questions. At the end of this process, a set of 100 questions was selected for the experiments.

2.3. Experimental Execution: Response Generation with and Without RAG

To evaluate the performance of the GPT-4o model in terms of responding to the 100 queries related to environmental control in broiler poultry farming, we developed a Python-based program using advanced NLP techniques and the RAG framework (Figure 3). The program, built on the LangChain framework, facilitated the comparison of responses generated with and without RAG to analyze the impact of contextual information on response accuracy.

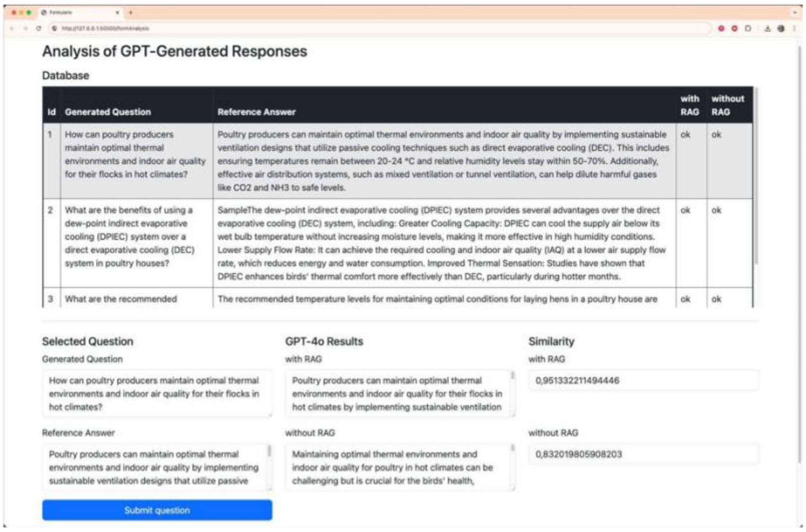


Figure 3. GPT-4o and RAG implementation user print screen interface. Source: the authors.

The evaluation involved generating responses for each query under two distinct conditions. In the first condition, without RAG, the model was provided with only the question and minimal guidance (basic prompt). In the second condition, with RAG, the model received both the question and the full-text page containing relevant contextual information (contextual prompt).

To implement the system, we used a two-step pipeline. First, we created a vector store using the FAISS library, where documents were converted into dense embeddings via OpenAI's embedding model to enable efficient similarity-based retrieval. Second, we constructed a conversational chain by integrating the GPT-4o model with the retrieval mechanism and a memory module. The retrieval system fetched relevant documents from the vector store based on user queries, while the memory module preserved dialog context across multiple interactions.

The generated responses were stored in the database for further analysis. Each response was recorded in one of two dedicated columns: response_without_RAG and response_with_RAG. This structured approach facilitated a direct comparison between the two types of responses, allowing subsequent steps to assess their similarity.

2.4. Comparative Analysis: Semantic Similarity Evaluation

In NLP, evaluating semantic similarity is crucial for information retrieval, question-answering, and summarization tasks. Semantic similarity measures how closely two textual elements convey related meanings, focusing on ideas rather than exact word matches [32]. Techniques for measuring have evolved from basic methods like Jaccard and cosine similarity to word embeddings such as Word2Vec, GloVe, and FastText, which represent words in vector spaces but lack sentence-level context. Transformer models like BERT and RoBERTa improved contextual understanding but at high computational costs. Sentence-BERT (SBERT) addressed this by combining BERT's contextual capabilities with efficiency, producing sentence-level embeddings optimized for semantic similarity tasks [33].

2.4.1. Semantic Similarity Assessment

We used semantic similarity as a metric to evaluate the responses generated by the model under RAG and non-RAG conditions. We employed the SentenceTransformer model to compute semantic similarity, specifically the pre-trained paraphrase-multilingual-MiniLM-L12-v2, designed to generate sentence embeddings. This model was selected for its ability to work across multiple languages, including English, and its effectiveness in representing the semantic meaning of sentences.

This approach allowed us to quantitatively assess how closely the responses generated by the model align with the expected answers. By comparing cosine similarity scores across RAG and non-RAG responses, we could evaluate the impact of contextual retrieval on the model's ability to generate semantically accurate answers. This step was critical for determining the effectiveness of RAG in improving the relevance and accuracy of responses in the domain of environmental control in broiler poultry farming.

The process involved three main steps. First, each sentence was passed through the model using the encoding method, which converted the text into a dense numerical vector, or embedding, to capture its semantic meaning. Second, the resulting embeddings were transformed into tensors, a data structure optimized for mathematical operations and essential for performing similarity computations. Finally, the semantic similarity between two sentences was calculated using the cosine similarity index, a standard metric for comparing high-dimensional vectors such as sentence embeddings. This index measures the cosine of the angle between two vectors, with values ranging from -1 to 1, where 1 indicates semantically identical sentences, 0 indicates no semantic similarity, and -1 indicates semantic opposition (a rare outcome in such tasks) [32,33].

For the present study, we established three thresholds to categorize performance based on similarity scores. Responses with a similarity score between 0.0 and 0.6 were classified as having low similarity. These responses exhibited minimal alignment with the original text, with key concepts either missing, imprecisely represented, or significantly divergent from the source content. Linguistic and structural elements also showed notable deviations. Responses with a score between 0.6 and 0.8 were categorized as moderate-similarity. These responses partially aligned with the original text, reflecting some key concepts but with notable omissions or inaccuracies. Variations in language and structure were evident, and extraneous information not present in the original text could be included. Finally, responses with a similarity score between 0.8 and 1.0 were classified as high-similarity. These responses closely mirrored the original text, accurately capturing primary ideas and concepts. Language and structural organization were either highly consistent with the source material or appropriately adapted while maintaining fidelity without introducing significant inaccuracies or unrelated information.

2.4.2. Statistical Analysis

To evaluate the impact of the RAG technique on response quality, we analyzed the semantic similarity scores of response sets with and without RAG. Descriptive statistics summarized overall performance and variability. A line plot illustrated trends in similarity scores across all questions, with performance ranges visually represented. A histogram was generated to compare the frequency distributions of similarity scores, and a difference ilarity to word embeddings such as Word2Vec, GloVe, and FastText, which represent words in vector spaces but lack sentence-level context. Transformer models like BERT and RoBERTa improved contextual understanding but at high computational costs. Sentence-BERT (SBERT) addressed this by combining BERT's contextual capabilities with efficiency, producing sentence-level embeddings optimized for semantic similarity tasks [33].

2.4.1. Semantic Similarity Assessment

We used semantic similarity as a metric to evaluate the responses generated by the model under RAG and non-RAG conditions. We employed the SentenceTransformer model to compute semantic similarity, specifically the pre-trained paraphrase-multilingual-MiniLM-L12-v2, designed to generate sentence embeddings. This model was selected for its ability to work across multiple languages, including English, and its effectiveness in representing the semantic meaning of sentences.

This approach allowed us to quantitatively assess how closely the responses generated by the model align with the expected answers. By comparing cosine similarity scores across RAG and non-RAG responses, we could evaluate the impact of contextual retrieval on the model's ability to generate semantically accurate answers. This step was critical for determining the effectiveness of RAG in improving the relevance and accuracy of responses in the domain of environmental control in broiler poultry farming.

The process involved three main steps. First, each sentence was passed through the model using the encoding method, which converted the text into a dense numerical vector, or embedding, to capture its semantic meaning. Second, the resulting embeddings were transformed into tensors, a data structure optimized for mathematical operations and essential for performing similarity computations. Finally, the semantic similarity between two sentences was calculated using the cosine similarity index, a standard metric for comparing high-dimensional vectors such as sentence embeddings. This index measures the cosine of the angle between two vectors, with values ranging from -1 to 1, where 1 indicates semantically identical sentences, 0 indicates no semantic similarity, and -1 indicates semantic opposition (a rare outcome in such tasks) [32,33].

For the present study, we established three thresholds to categorize performance based on similarity scores. Responses with a similarity score between 0.0 and 0.6 were classified as having low similarity. These responses exhibited minimal alignment with the original text, with key concepts either missing, imprecisely represented, or significantly divergent from the source content. Linguistic and structural elements also showed notable deviations. Responses with a score between 0.6 and 0.8 were categorized as moderate-similarity. These responses partially aligned with the original text, reflecting some key concepts but with notable omissions or inaccuracies. Variations in language and structure were evident, and extraneous information not present in the original text could be included. Finally, responses with a similarity score between 0.8 and 1.0 were classified as high-similarity. These responses closely mirrored the original text, accurately capturing primary ideas and concepts. Language and structural organization were either highly consistent with the source material or appropriately adapted while maintaining fidelity without introducing significant inaccuracies or unrelated information.

2.4.2. Statistical Analysis

To evaluate the impact of the RAG technique on response quality, we analyzed the semantic similarity scores of response sets with and without RAG. Descriptive statistics summarized overall performance and variability. A line plot illustrated trends in similarity scores across all questions, with performance ranges visually represented. A histogram was generated to compare the frequency distributions of similarity scores, and a difference plot highlighted the degree of improvement across questions. Finally, boxplots were used to illustrate the distributions of similarity scores for visual comparison.

We also compared the statistical results between the two conditions and calculated the difference between scores with and without RAG to quantify the degree of improvement. Additionally, we determined the proportion of cases where RAG enhanced similarity scores. A paired *t*-test was conducted to assess the statistical significance of the observed differences. Together, these analyses provided a comprehensive evaluation of the RAG technique's effectiveness in improving the semantic relevance of the generated responses.

3. Results

3.1. Descriptive Analysis

The analysis confirms the improvement in the performance of the RAG technique, highlighting its positive impact (Table 3). The increase in the median similarity index indicates that RAG consistently improved most responses in this dataset. Specifically, the average and the median similarity index with RAG compared without RAG represents a percentage increase of 13.45% for the mean and 12.68% for the median, demonstrating a consistent enhancement in semantic similarity when RAG is applied.

Table 3. Measures of central tendency.

Measures	Similarity Without RAG	Similarity with RAG	Difference
Mean	0.6369	0.7713	~0.1345
Median	0.6569	0.7836	~0.1268
Standard deviation	0.1660	0.1192	~−4.68

Source: the authors.

The higher standard deviation for similarity indices without RAG reflects more significant variability in the results, suggesting that the model's performance is less consistent when relying solely on internal knowledge. Conversely, the lower standard deviation for similarity indices with RAG indicates that the technique improves the average similarity scores and stabilizes the results by 4.68%. RAG appears to act as a normalizing factor, reducing variability in the similarity indices and providing more consistent, semantically aligned responses. This consistency supports the hypothesis that RAG enhances the generated responses' quality and reliability.

3.2. Statistical Validation of RAG's Effectiveness

The paired *t*-test revealed a statistically significant difference between the similarity indices of responses generated with and without RAG ($t = -7.610$, p-value = 1.63×10^{-11}). The negative t-value indicates that, on average, similarity scores with RAG were significantly higher than those without RAG. The extremely small *p*-value demonstrates that the likelihood of this difference being due to chance is negligible, allowing us to confidently reject the null hypothesis that there is no significant difference between the two groups.

These results confirm that the observed improvement in similarity indices with RAG is not a random fluctuation but an actual effect. Combined with the descriptive statistics and visual analyses, these findings reinforce the conclusion that RAG consistently enhances semantic alignment in generated responses, further validating its effectiveness as a retrieval-based augmentation technique.

3.3. Similarity Comparison

Figure 4 illustrates the similarity index for responses generated with and without the RAG technique across all questions.

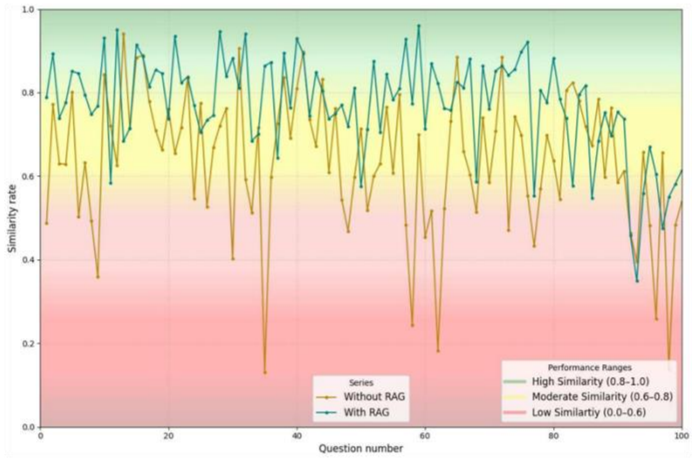


Figure 4. Similarity index comparison, without RAG vs. with RAG. Source: the authors.

Figure 4 shows the overall positive impact of RAG, with responses using RAG generally achieving higher similarity scores than those without it. Responses without RAG predominantly fall into the low-similarity range, while those with RAG are more frequently distributed across the moderate- and high-similarity ranges, reflecting improved alignment with the source content. However, RAG's performance is not consistently excellent. Around 10% of RAG-generated responses fall below the low-similarity threshold, and only about 30% reach the high-similarity range. The majority, approximately 60%, fall within the moderate-similarity range, indicating partial alignment with some omissions or inaccuracies. Additionally, in 12% of cases, responses without RAG outperformed those with RAG, suggesting that RAG is not always the optimal solution. These results highlight RAG's potential and limitations, emphasizing the need for further refinement to achieve more consistent, high-quality performance.

3.4. Impact of RAG

The bar graph (Figure 5) indicates the positive impact of RAG by presenting the differences between similarity indices with and without RAG (i.e., the similarity index with RAG minus the index without RAG), with 88% of the differences being positive, confirming RAG's effectiveness in improving semantic similarity. This result aligns with earlier analyses, such as the boxplot and histogram, which showed higher medians and a concentration of RAG responses in higher similarity ranges. The 12% of negative differences indicate cases where RAG underperformed, likely due to specific question or reference text characteristics, but these losses had smaller magnitudes, minimizing their overall impact. Positive differences often exceeded 0.4, highlighting significant gains for specific responses, while the majority of bars above the reference line (y = 0) underscore the consistency of RAG in enhancing semantic alignment. These findings confirm that RAG improves response quality and stability across most cases while identifying areas for further refinement to address the few cases where RAG was less effective.

3.5. Distribution

The histogram in Figure 6 provides a detailed comparison of the similarity indices for responses generated with and without RAG, revealing distinct distribution patterns. Responses with RAG are more concentrated in higher similarity ranges, with a tighter distribution and reduced variability, indicating a more consistent performance. In contrast, responses without RAG are more dispersed, with a higher concentration in the intermediate range and limited representation above 0.8, reflecting greater variability and lower alignment with the source content. This aligns with the descriptive analysis, which showed higher mean and median similarity indices for RAG responses than non-RAG responses. Additionally, the similarity comparison graph and the histogram highlight RAG's stabilizing effect, shifting the distribution toward higher values.

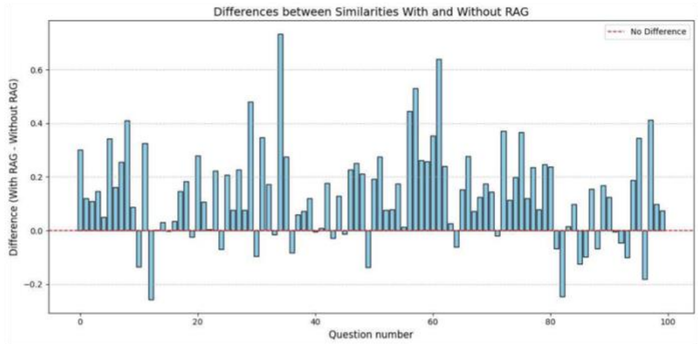


Figure 5. Differences between the similarity rate with RAG and without RAG. Source: the authors.

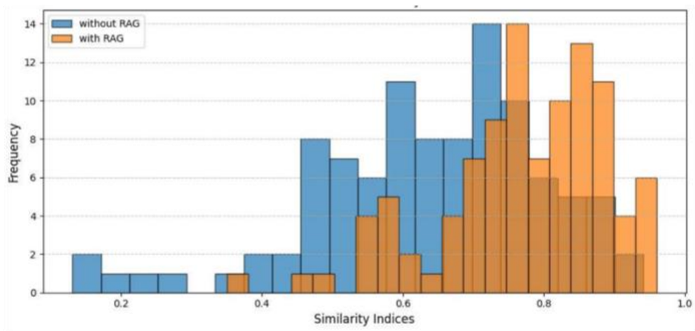


Figure 6. Differences between the frequencies of similarity indices with RAG and without RAG. Source: the authors.

However, limitations remain evident: approximately 10% of RAG responses fall below 0.6, only 30% achieve high similarity (0.8–1.0), and around 12% of non-RAG responses outperform RAG responses. These findings confirm that while RAG improves response quality and consistency overall, its effectiveness varies, leaving room for further refinement.

The boxplot in Figure 7 highlights differences in the distribution of similarity indices for responses generated with and without the RAG technique. The median similarity index for responses with RAG is visibly higher than those without RAG, confirming that most responses with RAG achieve better alignment with the source content. This finding is consistent with the descriptive analysis, which showed higher mean and median values for RAG responses than without RAG responses.

The smaller interquartile range (IQR) for RAG responses indicates reduced variability among the central 50% of scores, indicating a more consistent performance. In contrast, the larger IQR for responses without RAG reflects greater dispersion and inconsistency, as also seen in the similarity comparison graph and histogram, where non-RAG responses showed broader variability and a higher concentration of lower similarity scores. Additionally, while responses with RAG exhibit more outliers, these are concentrated in higher similarity ranges, suggesting occasional high-performing cases. Conversely, the outliers for responses without RAG occur in much lower similarity ranges, underscoring poor alignment and significantly lower response quality.

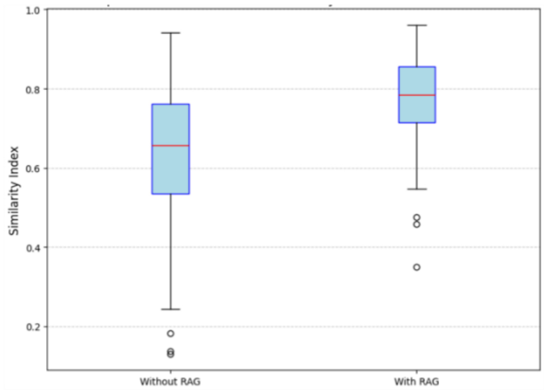


Figure 7. Boxplot comparison of the distributions considering the similarity indices with and without RAG. Source: the authors.

The histogram also supports this observation, showing a denser concentration of RAG responses in higher similarity ranges and non-RAG responses more scattered in intermediate ranges. These results reinforce the conclusion that RAG improves both the average similarity scores and the consistency of responses, although occasional limitations remain.

The results indicate the effectiveness of the RAG technique in enhancing the semantic similarity of generated responses. Across all analyses, RAG consistently outperformed the non-RAG approach, as evidenced by higher mean and median similarity indices and reduced variability, with a decrease in standard deviation. The paired *t*-test confirmed these improvements' statistical significance, reinforcing that the observed gains were not due to random chance. The graph analysis highlighted the concentration of RAG responses in higher similarity ranges, demonstrating improved accuracy and greater consistency. It further revealed that 88% of cases showed positive gains with RAG, with significant improvements in some cases exceeding 0.4.

4. Discussion

In the present study, the results demonstrated the potential of RAG to enhance the accuracy and relevance of LLM-generated responses in environmental control for broiler poultry farming. The findings supported our hypothesis that RAG would improve LLM performance by providing contextual information to support informed decision-making. Integrating RAG with LLMs demonstrated potential by improving contextual alignment, as evidenced by a 13.45% increase in mean similarity scores. Since PLF in broiler production relies on detection, monitoring, analysis, and decision-making to ensure productivity and welfare [2,3], these enhancements are crucial for addressing complex environmental data and supporting real-time decision-making [3,8].

Despite these encouraging results, some limitations were observed. In 12% of cases, responses without RAG outperformed those with RAG, indicating that retrieval quality and relevance can occasionally hinder performance. This highlights the need to enhance retrieval mechanisms and ensure external knowledge bases remain current and comprehensive [25,26]. Computational costs and scalability are also important challenges, particularly for high-density poultry production, where real-time decision-making demands both efficiency and reliability [11].

The findings also emphasize the practicality of RAG for small-scale producers, offering a cost-effective alternative to traditional consultancy services. By retrieving and integrating relevant knowledge, RAG-equipped LLMs ensure more dynamic and tailored support, making real-time environmental control adjustments feasible even in resource-limited settings [26–30].

From a practical perspective, the enhanced performance of RAG has clear implications for the poultry industry. Its ability to deliver consistent, context-sensitive insights into environmental variables, such as temperature, humidity, and ammonia levels, supports producers in maintaining optimal conditions for animal welfare and productivity. Furthermore, RAG addresses key industry challenges by analyzing sensor data, generating reports, optimizing ventilation, managing thermal conditions, and monitoring air quality, providing actionable insights that improve productivity and bird welfare.

The controlled variables and reproducible methodologies used in this study further ensured the validation of the hypothesis by isolating the impact of RAG on response quality. The improvement in semantic similarity demonstrates that integrating relevant external knowledge into LLM workflows leads to more precise and context-sensitive outputs. These findings underscore the effectiveness of RAG in overcoming common limitations of LLMs, such as generalized responses and reliance on static training data [2,3,16–18].

In conclusion, the results confirm our initial hypothesis, demonstrating that RAG improves LLM performance by providing contextual information. While some areas require refinement, the observed improvements validate the viability of RAG as a transformative tool for modern poultry farming.

5. Conclusions

The present study demonstrates the effectiveness of integrating RAG with LLMs to enhance decision-making and improve environmental control in broiler poultry farming. The findings highlight RAG's capacity to improve the semantic accuracy and contextual relevance of LLM responses, making it a promising approach for addressing the complex challenges of high-density poultry production systems. The statistical and semantic analyses confirmed that RAG reduces variability and enhances LLM response consistency, enabling producers to make data-driven adjustments informed by LLM answers and analysis, optimizing animal welfare and productivity.

Despite these promising results, there is still room for improvement. Challenges in managing retrieval quality, addressing inconsistencies in retrieved information, and ensuring scalability underscore the need for continued refinement of RAG frameworks.

Beyond its immediate benefits, integrating RAG into LLM workflows offers a scalable and cost-effective solution for supporting small-scale producers who often lack access to expert consultancy. By bridging the gap between raw sensor data and actionable insights, RAG-equipped LLMs demonstrate significant potential to transform environmental management practices, fostering sustainability and regulatory compliance in broiler farming.

Future research should focus on enhancing the precision and reliability of RAG by addressing uncertainties in retrieved information and incorporating non-classical logic, such as Fuzzy and Paraconsistent Logic, to handle variability and ambiguity in data. Exploring alternative metrics—perplexity, precision, and factual accuracy—could provide deeper insights into RAG performance. Comparative studies with other LLMs, including models from different providers, would also establish valuable benchmarks for scalability and generalizability. Another promising avenue lies in integrating sensor data from poultry houses with RAG-enabled knowledge bases, generating real-time, context-sensitive insights to advance precision livestock farming further. Such developments would solidify RAG's role as a transformative tool for improving environmental control and decision-making in modern agriculture.

Author Contributions: Conceptualization, M.V.L., M.L.H.S., I.d.A.N. and J.M.A.; methodology, M.V.L. and M.L.H.S.; software, M.V.L. and M.L.H.S.; validation, M.V.L., M.L.H.S. and I.d.A.N.; formal analysis, M.V.L. and M.L.H.S.; investigation, M.V.L. and M.L.H.S.; resources, M.V.L. and M.L.H.S.; data curation, M.V.L. and M.L.H.S.; writing—original draft preparation, M.V.L. and M.L.H.S.; writing—review and editing, M.V.L., M.L.H.S., I.d.A.N. and J.M.A.; visualization, I.d.A.N.; supervision, I.d.A.N. and J.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be available upon request.

Acknowledgments: The authors thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for the master's scholarship of the first author and the National Council for Scientific and Technological Development (CNPQ) for the PQ scholarship of the fourth author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mottet, A.; Tempio, G. Global poultry production: Current state and future outlook and challenges. *J. World's Poult. Sci.* **2017**, *73*, 245–256. [CrossRef]
2. Berckmans, D. Automatic on-line monitoring of animal health and welfare by Precision Livestock Farming. In Proceedings of the European Forum Livestock Housing for the Future, Lille, France, 22–23 October 2009.
3. Berckmans, D. General introduction to precision livestock farming. *Anim. Front.* **2017**, *7*, 6–11. [CrossRef]
4. ABPA. Anual Report 2024. Brazilian Association of Animal Protein, 2024. Available online: https://abpa-br.org/wp-content/uploads/2024/04/RA_2024_ABPA_ingles_avicultura.pdf (accessed on 20 August 2024).
5. Gržinić, G.; Piotrowicz-Cieślak, A.; Klimkowicz-Pawlas, A.; Górny, R.L.; Ławniczek-Wałczyk, A.; Piechowicz, L.; Olkowska, E.; Potrykus, M.; Tankiewicz, M.; Krupka, M.; et al. Intensive poultry farming: A review of the impact on the environment and human health. *Sci. Total Environ.* **2020**, *858*, 160014. [CrossRef]
6. Hafez, H.M.; Attia, Y.A. Challenges to the Poultry Industry: Current Perspectives and Strategic Future After the COVID-19 Outbreak. *Front. Vet. Sci.* **2020**, *7*, 516. [CrossRef]
7. Astill, J.; Dara, R.A.; Fraser, E.D.; Roberts, B.; Sharif, S. Smart poultry management: Smart sensors, big data, and the internet of things. *Comput. Electron. Agric.* **2020**, *170*, 105291. [CrossRef]
8. Zheng, H.; Zhang, T.; Fang, C.; Zeng, J.; Yang, X. Design and Implementation of Poultry Farming Information Management System Based on Cloud Database. *Animals* **2021**, *11*, 900. [CrossRef]
9. Dewanto, P.; Munadi, M.; Tauviqirrahman, M. Development of an automatic broiler feeding system using PLC and HMI for closed house system. *ASRJETS* **2019**, *58*, 139–149. Available online: https://core.ac.uk/download/pdf/235050874.pdf (accessed on 20 August 2024).
10. Porter, M.E.; Heppelmann, J.E. How smart, connected products are transforming competition. *Harv. Bus. Rev.* **2014**, *92*, 64–88.
11. Lashari, M.H.; Memon, A.A.; Shah, S.A.A.; Nenwani, K.; Shafqat, F. IoT Based Poultry Environment Monitoring System. In Proceedings of the IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), Bali, Indonesia, 3–8 November 2018. Available online: https://ieeexplore-ieee-org.ez346.periodicos.capes.gov.br/document/8600837 (accessed on 29 July 2024).
12. Halachmi, I.; Guarino, M.; Bewley, J.; Pastell, M. Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production. *Annu. Rev. Anim. Biosci.* **2019**, *7*, 403–425. [CrossRef]
13. Raikov, A.; Abrosimov, V. Artificial Intelligence and Robots in Agriculture. In Proceedings of the 2022 15th International Conference Management of large-Scale System Development (MLSD), Moscow, Russia, 26–28 September 2022; pp. 1–5. [CrossRef]
14. Ojo, R.O.; Ajayi, A.O.; Owolabi, H.A.; Oyedele, L.O.; Akanbi, L.A. Internet of Things and Machine Learning techniques in poultry health and welfare management: A systematic literature review. *Comput. Electron. Agric.* **2022**, *200*, 107266. [CrossRef]
15. García, M.; Martínez, J. Application of Internet of Things (IoT) in Animal Production Systems. *J. Anim. Sci. Technol.* **2019**, *10*, 123–134.
16. Vaswani, A. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.

17. Google AI. Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing. Available online: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html (accessed on 17 November 2024).
18. Meta AI. LLaMA: Open and Efficient Foundation Language Models. Available online: https://ai.facebook.com/blog/large-language-model-llama-meta-ai/ (accessed on 17 November 2024).
19. OpenAI. Introducing ChatGPT. Available online: https://openai.com/blog/chatgpt/ (accessed on 17 November 2024).
20. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608.
21. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
22. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. A survey of hallucination in natural language generation. *ACM Comput. Surv.* **2022**, *55*, 1–38. [CrossRef]
23. Metze, K.; Morandin-Reis, R.C.; Lorand-Metze, I.; Florindo, J.B. Bibliographic Research with ChatGPT may be Misleading: The Problem of Hallucination. *J. Pediatr. Surg.* **2024**, *59*, 158. [CrossRef]
24. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv* **2022**, arXiv:2201.11903.
25. Ott, S.; Hebenstreit, K.; Liévin, V.; Hother, C.E.; Moradi, M.; Mayrhauser, M.; Praas, R.; Winther, O.; Samwald, M. ThoughtSource: A central hub for large language model reasoning data. *Sci. Data* **2023**, *10*, 528. [CrossRef]
26. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.T.; Rocktäschel, T.; et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv* **2020**, arXiv:2005.11401.
27. Li, H.; Su, Y.; Cai, D.; Wang, Y.; Liu, L. A Survey on Retrieval-Augmented Text Generation. *arXiv* **2022**, arXiv:2202.01110.
28. Karpatne, A.; Jia, X.; Kumar, V. Knowledge-guided machine learning: Current Trends and Future Prospects. *arXiv* **2024**, arXiv:2403.15989.
29. Izacard, G.; Grave, E. Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering. *arXiv* **2020**, arXiv:2007.01282.
30. Petroni, F.; Piktus, A.; Fan, A.; Lewis, P.; Yazdani, M.; De Cao, N.; Thorne, J.; Jernite, Y.; Karpukhin, V.; Maillard, J.; et al. KILT: A Benchmark for Knowledge Intensive Language Tasks. *arXiv* **2020**, arXiv:2009.02252.
31. Guo, J.; Fan, Y.; Pang, L.; Yang, L.; Ai, Q.; Zamani, H.; Wu, C.; Croft, W.B.; Cheng, X. A deep dive into neural ranking models for information retrieval. *Inf. Process. Manag.* **2022**, *57*, 102067.
32. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
33. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *arXiv* **2019**, arXiv:1908.10084.
34. Bist, R.B.; Subedi, S.; Chai, L.; Yang, X. Ammonia emissions, impacts, and mitigation strategies for poultry production: A critical review. *J. Environ. Manag.* **2023**, *328*, 116919. [CrossRef]
35. Bloch, V.; Barchilon, N.; Halachmi, I.; Druyan, S. Automatic broiler temperature measuring by thermal camera. *Biosyst. Eng.* **2020**, *199*, 127–134. [CrossRef]
36. Costantino, A.; Fabrizio, E.; Villagrá, A.; Estellés, F.; Calvet, S. The reduction of gas concentrations in broiler houses through ventilation: Assessment of the thermal and electrical energy consumption. *Biosyst. Eng.* **2020**, *199*, 135–148. [CrossRef]
37. Kalus, K.; Konkol, D.; Korczyński, M.; Koziel, J.A.; Opaliński, S. Effect of biochar diet supplementation on chicken broilers performance, NH_3 and odor emissions and meat consumer acceptance. *Animals* **2020**, *10*, 1539. [CrossRef]
38. Babadi, K.A.; Khorasanizadeh, H.; Aghaei, A. CFD modeling of air flow, humidity, CO_2 and NH_3 distributions in a caged laying hen house with tunnel ventilation system. *Comput. Electron. Agric.* **2022**, *193*, 106677. [CrossRef]
39. Li, D.; Tong, Q.; Shi, Z.; Zheng, W.; Wang, Y.; Li, B.; Yan, G. Effects of Cold Stress and Ammonia Concentration on Productive Performance and Egg Quality Traits of Laying Hens. *Animals* **2020**, *10*, 2252. [CrossRef]
40. Al Assaad, D.K.; Orabi, M.S.; Ghaddar, N.K.; Ghali, K.F.; Salam, D.A.; Ouahrani, D.; Farran, M.T.; Habib, R.R. A sustainable localised air distribution system for enhancing thermal environment and indoor air quality of poultry house for semiarid region. *Biosyst. Eng.* **2021**, *203*, 70–92. [CrossRef]
41. Soliman, E.S.; Hassan, R.A. Influence of housing floor on air quality, growth traits, and immunity in broiler chicken farms. *Adv. Anim. Vet. Sci.* **2020**, *8*, 997–1008. [CrossRef]
42. Wen, P.; Li, L.; Xue, H.; Jia, Y.; Gao, L.; Li, R.; Huo, L. Comprehensive evaluation method of the poultry house indoor environment based on gray relation analysis and Analytic Hierarchy Process. *Poult. Sci.* **2022**, *101*, 101587. [CrossRef]

3.3 A Decision Support AI-Copilot for Poultry Farming: Leveraging Retrieval-Augmented LLMs and Paraconsistent Annotated Evidential Logic Eτ to Enhance Operational Decisions

The third article, titled "*A Decision Support AI-Copilot for Poultry Farming: Leveraging Retrieval-Augmented LLMs and Paraconsistent Annotated Evidential Logic Eτ to Enhance Operational Decisions*," was submitted to the international journal AgriEngineering (MDPI) and received a favorable review, currently undergoing final revision according to editorial recommendations. The study was authored by Marcus Vinícius Leite, Jair Minoro Abe, and Irenilza de Alencar Nääs, affiliated with Universidade Paulista (UNIP) and Marcos Leandro Hoffmann Souza, affiliated with Universidade do Vale do Rio dos Sinos (UNISINOS).

This study represents the synthesis and validation stage of the research, integrating the theoretical, experimental, and computational components developed in the previous phases. Its main objective was to design, implement, and evaluate a DSS based on the integration of Logic Eτ, LLMs, and RAG, thereby consolidating a conversational agent capable of operating under conditions of uncertainty, contradiction, and informational incompleteness characteristic of intensive poultry farming (Abe, Akama & Nakamatsu, 2015; Carvalho & Abe, 2018; Lewis et al., 2021).

The research aimed to demonstrate the practical feasibility of the proposed integrative model by transforming logical-evidential inferences into contextually grounded and semantically consistent responses. To achieve this, it articulated three complementary dimensions: logical-evidential inference, responsible for processing favorable and unfavorable evidence (Abe, 2014; de Carvalho Junior et al., 2024); contextual processing, guiding the retrieval and weighting of relevant information through RAG (Li et al., 2022; Izacard & Grave, 2021); and semantic interpretation, performed by the LLM, which generates linguistically coherent responses and recommendations (Brown et al., 2020; Vaswani et al., 2017). Positioned in the third methodological phase, modeling, implementation, and validation, this study provides the empirical consolidation of the theoretical–operational model, evaluating its performance in terms of logical-evidential consistency, semantic accuracy, and operational applicability.

The developed architecture was structured into three main modules: the Knowledge Base Construction Pipeline (KBCP), responsible for preprocessing technical and scientific documents, including text extraction, chunk segmentation, vectorization, and indexing via FAISS Vector Store (Wang et al., 2025); the Domain-Specific Knowledge Base (DS-KB), which stores the vector repository and enables semantic search through cosine similarity (Reimers & Gurevych, 2019); and the Conversational Decision Support Agent (C-DSS-A), which integrates the logical-evidential inference layer (implemented according to the Para-Analyzer Algorithm, PAA, of Logic Eτ) with the state-of-the-art LLM (GPT-4o) (OpenAI et al., 2024), combining formal reasoning and natural language generation within a conversational interface.

Experiments were conducted using representative queries from five decision domains in poultry production, environment, nutrition, health, welfare, and management, processed under four conditions: without preprocessing, with preprocessing (normalization and lemmatization), with RAG enabled, and with RAG combined with Logic Eτ (Boban et al., 2020; Pramana et al., 2022). Responses were evaluated using semantic similarity, contextual relevance, and logical-evidential consistency metrics, measured through the parameters (Gce, Gct) and control values for certainty and contradiction (Vscc, Vicc, Vscct, Vicct) (Abe, 2011; Carvalho & Abe, 2018). Global performance was statistically analyzed and visualized through correlation matrices and Unit Square in Cartesian Plane (USCP) diagrams, enabling observation of inference stability under varying degrees of uncertainty and fragmentation (Abe, 2011; Akama, 2016).

The results showed that the integrated Logic Eτ and LLM with RAG system achieved significant gains in both consistency and precision compared with versions lacking logical-evidential inference. The average semantic similarity of responses increased by 18.2%, while contextual relevance rose by 15.6%, confirming the synergy between retrieval and inference layers. The distribution of evidence pairs (μ, λ) revealed a higher concentration in the consistent truth quadrant (V) and a substantial reduction of occurrences in the inconsistent (T) and paracomplete (⊥) states, demonstrating the system's ability to stabilize decisions even under contradiction and incompleteness (Abe, 2014; de Carvalho Junior et al., 2024). RAG reduced semantic dispersion by incorporating relevant external evidence, while Logic Eτ acted as a stabilizing inference filter, mitigating internal contradictions and enhancing reasoning interpretability (Carvalho & Abe, 2018; Abe, Akama & Nakamatsu, 2015). The system proved capable of justifying each response based on its corresponding logical-evidential state and retrieved sources, promoting transparency and traceability (Leite et al., 2025). Visualization in the USCP diagram revealed increased density in quasi-true states (QV→T), indicating a predominance of logically consistent responses supported by robust evidence (Abe, 2011; Akama, 2016).

The findings confirm that integrating formal symbolic reasoning with probabilistic language models provides an effective approach for supporting complex decision-making in uncertain environments (Abe, Akama & Nakamatsu, 2015; de Carvalho Junior et al., 2024). Logic Eт, by quantifying and weighting degrees of evidence, acts as a logical controller capable of regulating the uncertainty and contradiction inherent to natural-language reasoning (Abe, 2011; Abe, 2014). This integration enables the conversational agent not only to generate linguistically appropriate responses but also to evaluate the consistency of its own inferences, approaching an explainable and self-regulating behavior (Carvalho & Abe, 2018; Leite et al., 2025). Beyond validating the central hypothesis of this master's thesis, the convergence among logical-evidential inference, contextual processing, and semantic interpretation, the study demonstrates that this integration produces a DSS that is consistent, interpretable, and adaptable, overcoming the coherence and explainability limitations of conventional DSS based solely on statistical learning (Brown et al., 2020; Vaswani et al., 2017).

The results further reinforce the applicability of the approach to other productive domains requiring decision-making under uncertainty, highlighting the potential of Logic Eт as a theoretical–computational framework for auditable and resilient intelligent systems (Carvalho & Abe, 2018; de Carvalho Junior et al., 2024). The proposed model goes beyond response automation by embedding verifiable and governable reasoning mechanisms, essential for the reliability and transparency of intelligent Decision Support Systems (Akama, 2016; Abe, 2011). Thus, this study consolidates Logic Eт as the logical–operational core of the developed model, demonstrating its ability to sustain consistent, traceable, and formally explainable inference in natural-language-based systems applied to Production Engineering.

AgriEngineering

MDPI

Article

A Decision Support AI-Copilot for Poultry Farming: Leveraging Retrieval-Augmented LLMs and Paraconsistent Annotated Evidential Logic Eτ to Enhance Operational Decisions

Marcus Vinicius Leite[1], Jair Minoro Abe[2], Irenilza de Alencar Nääs[3], Marcos Leandro Hoffmann Souza[4]

[1] Graduate Program in Production Engineering, Paulista University - UNIP; marcus.leite@gmail.com
[2] Graduate Program in Production Engineering, Paulista University - UNIP; jair.abe@docente.unip.br
[3] Graduate Program in Production Engineering, Paulista University - UNIP; irenilza.naas@docente.unip.br
[4] Computer Science, Universidade do Vale do Rio dos Sinos - UNISINOS; marcosl.hoffmann@yahoo.com.br
* Correspondence: marcus.leite@gmail.com

Abstract

Driven by the global rise in animal protein demand, poultry farming has evolved into a highly intensive and technically complex sector. According to FAO, animal protein production increased by about 16% in the past decade, with poultry alone expanding 27% and becoming the leading source of animal protein. This intensification requires rapid, complex decisions across multiple aspects of production under uncertainty and strict time constraints. This study presents the development and evaluation of a conversational system designed to support decision-making to assist poultry producers in addressing technical queries across five key domains: environmental control, nutrition, health, husbandry, and animal welfare. The system combines a large language model (LLM) with retrieval-based generation (RAG) to ground responses in a curated corpus of scientific and technical literature. Additionally, it adds a reasoning component using Paraconsistent Annotated Evidential Logic Eτ, a non-classical logic designed to handle contradictory or incomplete information. Evaluation was conducted by comparing system responses with expert reference answers using semantic similarity (cosine similarity with SBERT embeddings). Results indicate that the system successfully retrieves and composes relevant content, while the paraconsistent inference layer makes results easier to interpret and more reliable in the presence of conflicting or insufficient evidence. These findings suggest that the proposed architecture provides a viable foundation for explainable and reliable decision support in modern poultry production, achieving consistent reasoning under contradictory or incomplete information where conventional RAG chatbots would fail.

Keywords: Poultry Production; Poultry Farming; Decision Support System; LLM Large Language Models; RAG Retrieval Augmented Generation; Paraconsistent Annotated Evidential Logic Eτ; Smart Farming.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Citation: To be added by editorial staff during production.

1. Introduction

Poultry production has become the most widely consumed source of animal protein worldwide, driven by rising global demand, rapid urbanization, and the intensification of livestock systems [1–4]. According to FAO, animal protein production increased by about 16% in the past decade, with poultry alone expanding 27% and becoming the leading source of animal protein [1]. As production scales grow, poultry farmers are increasingly required to make rapid and complex decisions involving environmental control, nutrition, health, animal welfare, and husbandry, often under conditions of uncertainty, time pressure, and conflicting information [4,5].

To cope with the growing decision complexity of intensive poultry systems, a variety of farm management platforms integrating decision-support tools have been introduced [36,38,39]. Examples include eFarm, a precision agriculture application that integrates health, feed, and production metrics for dairy and poultry operations [39]; farmOS, a community-driven open-source platform for planning and record keeping [40]; and the Poultry Farming Management System, which automates data collection for inventory, production, sales, and expenses [41]. While these systems improve data organization and reporting providing valuable functionalities for record keeping, planning, and health or production tracking, they primarily serve as dashboards or recordkeeping applications. Their embedded DSS modules, although useful for routine monitoring, are not designed to cope with uncertainty, contradictory inputs, or overlapping decision domains—challenges that are common in intensive poultry farming. As a result, similar to these platforms and their decision-support modules, most existing tools remain narrow in scope, focused on isolated domains, with limited integration across technical areas and little resilience to contradictory or incomplete information [4,5,36,38,39].

In practice, Decision Support Systems (DSS) in poultry farming often take the form of deterministic rule-based or AI-based controllers, IoT monitoring platforms, big data solutions, and statistical dashboards that track environmental conditions, animal health indicators, and production metrics [5,29,30,36]. Although these technologies provide valuable data, they typically operate under significant limitations—such as infrastructure demands, expertise gaps, and cost-related constraints—and are frequently based on fixed thresholds or rigid decision rules, lacking mechanisms for context-aware inference or adaptive reasoning [17,29]. Consequently, current systems struggle to accommodate uncertainty, conflicting signals, and the need for multi-domain integration in real-world decision-making scenarios [34,35].

These constraints have motivated the exploration of knowledge-based approaches that incorporate structured reasoning and domain expertise to enhance decision robustness [8,36]. In this context, recent advances in Large Language Models (LLMs) offer promising capabilities for contextual understanding, flexible inference, and semantic generalization, particularly when enriched with Retrieval-Augmented Generation (RAG) mechanisms that ground responses in external content [6–10,36]. However, despite the potential of LLMs in extracting, composing, and synthesizing complex technical knowledge from unstructured sources, these models still struggle when faced with conflicting or incomplete information [8,11–13]. Moreover, there is a significant knowledge gap in the application of LLMs to livestock production, particularly regarding the challenges of poultry farming processes, which opens opportunities for further research and technological advances [38]. This gap highlights the need for a framework that not only leverages LLM+RAG but also introduces an evidential reasoning layer capable of contradiction-tolerant inference. In this sense, standard RAG-based models collapse under contradictory signals, whereas paraconsistent reasoning explicitly tolerates and structures such conflicts.

To address these challenges, this study examines the integration of LLMs and RAG with Paraconsistent Annotated Evidential Logic Eτ (Logic Eτ). This non-classical framework enables reasoning under contradictory, insufficient, or ambiguous evidence. While LLMs provide linguistic generalization and RAG ensures factual grounding through external sources, Logic Eτ adds an inferential layer that explicitly handles conflicting or incomplete evidence, providing transparency and robustness in decision-making processes [8,14–16].

The objective of this proof-of-concept study is to develop and evaluate a knowledge-based Decision Support System (DSS) for poultry production, structured as a conversational agent—the Decision Support AI-Copilot—that answers domain-specific queries using LLMs, content retrieved via RAG, and paraconsistent inference based on Logic Eτ. The system addresses five critical areas of poultry production: environmental management, animal nutrition, health monitoring, husbandry, and animal welfare. Its configuration was defined through controlled experiments designed to evaluate both the quality of semantic retrieval and the strength of the reasoning, optimizing generative behavior and logical consistency. Performance was then assessed through comparison with expert-curated references using semantic similarity metrics (cosine similarity with SBERT embeddings) and evidential assessments. While the Discussion briefly contrasts the proposed framework with recent LLM+RAG approaches, a comprehensive comparison with other DSS approaches lies beyond the scope, as the focus here is on feasibility and methodological contribution.

2. Materials and Methods

This study adopts an applied and experimental methodology to design and evaluate a knowledge-based decision support system for poultry production, combining theoretical modeling, computational implementation, and empirical evaluation. All materials, algorithms, and procedures are described in detail to ensure reproducibility and enable replication by future research.

2.1 Methodological Framework Overview

The methodological framework integrates three complementary components:

1. Theoretical modeling with Logic Eτ, which provides the inferential foundation for reasoning under uncertainty and contradiction, supporting key decision points in the system workflow.
2. Experimental validation through Design of Experiments (DoE), conducted as proof-of-concept trials to tune system-level parameters affecting semantic retrieval, preprocessing, and generative behavior, rather than as large-scale validation.
3. System implementation of the Decision Support AI-Copilot, developed as a modular RAG-based architecture that integrates LLMs with evidential reasoning mechanisms.

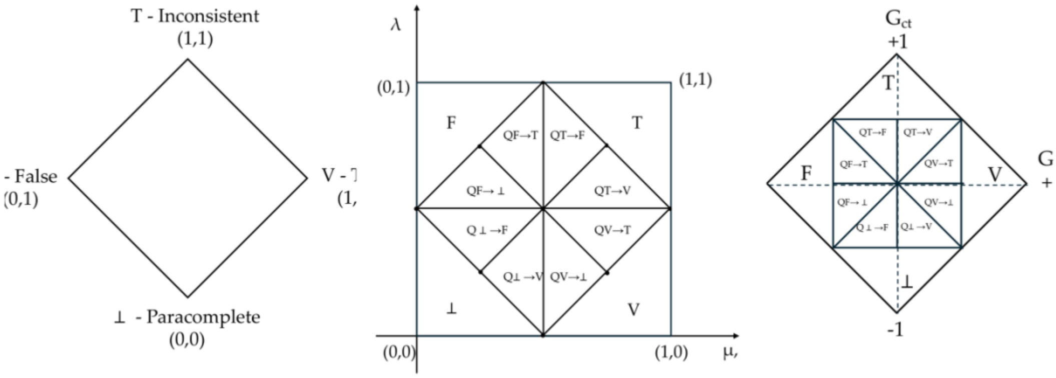
The following subsections present each component in sequence, ensuring a coherent integration between theoretical modeling, experimental validation, and system implementation.

2.2 Evidential Inference with Logic Eτ

Conventional LLM-based systems struggle when confronted with imprecise, incomplete, or contradictory inputs, a critical limitation in technical decision-support scenarios [6,11]. To address these challenges, the proposed system incorporates Logic Eτ as a complementary inference mechanism for handling evidential uncertainty and inconsistency in a mathematically tractable manner [14-16,32].

As a non-classical logical system, Logic Eτ is designed to support reasoning under uncertainty, contradictory, and incomplete information. Its expressive capability stems from the use of dual evidence degrees to express knowledge about a proposition enabling a granular representation of evidential states [14,18].

Logic Eτ assigns to each proposition p an evidential annotation (μ, λ), where μ and λ denote degrees of favorable and unfavorable evidence respectively. This dual-valued representation prevents trivialization in inference, even when μ and λ simultaneously assume high values, a condition under which Classical Logic becomes inconsistent and deductively trivial [14,32]. These evidential annotations are formally interpreted within three conceptual spaces [14,15,32], depicted in Figures 1a, 1b, 1c, each capturing a specific aspect of paraconsistent reasoning:



(**a**) (**b**) (**c**)

Figure 1. Key concepts about visual decision states in Logic Eτ, adapted from [14,15]. (a) The evidential lattice with partial order, where classical logical states: True, False, Inconsistent, and Paracomplete, correspond to extremal vertices. (b) The QUPC (Unit Square of the Cartesian Plane) provides a geometric representation of evidential states (μ, λ), highlighting both extreme and non-extreme (quasi) logical regions. (c) The logical lattice τ results from a nonlinear transformation T (μ, λ) = (μ − λ, μ + λ − 1), mapping evidential inputs into a plane where the horizontal axis encodes certainty (Gce) and the vertical axis uncertainty (Gct). This transformed space enables graded reasoning across nuanced logical states.

1. Evidential Lattice with Partial Order: This structure defines a complete lattice over the unit square $[0, 1]^2$, where each pair (μ, λ) encodes the degrees of favorable and unfavorable evidence about a proposition. A partial order is defined by:

   $$(\mu 1, \lambda 1) \leq (\mu 2, \lambda 2) \Leftrightarrow \mu 1 \leq \mu 2 \text{ and } \lambda 1 \geq \lambda 2$$

   This order reflects evidential dominance and enables lattice-theoretic operations (infimum, supremum, neutral elements). A canonical negation operator $\sim (\mu, \lambda) = (\lambda, \mu)$ supports dual reasoning and contradiction handling. The evidential lattice serves as the operational substrate for all inference processes in Logic Eτ-based systems [14,15].

2. QUPC (Unit Square of the Cartesian Plane), from a geometric standpoint, the evidential lattice can be visualized as a unit square of the Cartesian plane (referred to in Portuguese as *Quadrado Unitário do Plano Cartesiano* (QUPC). Each evidential pair (μ, λ) corresponds to a point in this 2D unit square (Figure 1b), allowing for an intuitive representation of the underlying information state. While the lattice defines logical and computational operations through ordering, the QUPC offers a descriptive and analytic space for visualizing evidential distributions and for mapping them onto the logical plane [14-16,32].

3. Logical Lattice τ: A nonlinear transformation maps QUPC into the logical space τ, where inference operates in Figure 1c. The transformation defines two axes: the certainty degree (Gce), and the contradiction degree (Gct).

   $$T(\mu, \lambda) = (G_{ce}(\mu, \lambda), G_{ct}(\mu, \lambda)) = (\mu - \lambda, \mu + \lambda - 1)$$

Extreme logical states (True, False, Inconsistent, Paracomplete) correspond to the four lattice extremities (1,0) → true, (−1,0) → false, (0,1) → inconsistent, (0, −1) → paracomplete (incomplete). Intermediate regions correspond to non-extreme states such as quasi-true, quasi-false, quasi-inconsistent and quasi-paracomplete and their respective tendencies, allowing graded reasoning, a crucial asset in non-deterministic conversational contexts as shown in Table 1 [14,15,32].

Table 1. Symbolic representation of extreme and non-extreme logical states in Logic Eτ, including quasi-states and transitional tendencies.

Symbol	State
V	True
QV→T	Quasi-true, tending to inconsistent;
QV→⊥	Quasi-true, tending to paracomplete
F	False
QF→T	Quasi-false, tending to inconsistent
QF→⊥	Quasi-false, tending to paracomplete
T	Inconsistent
QT→V	Quasi-inconsistent, tending to true
QT→F	Quasi-inconsistent, tending to false
⊥	Paracomplete or Indeterminate
Q⊥→V	Quasi-paracomplete, tending to true
Q⊥→F	Quasi-paracomplete, tending to false

Adapted from [15].

The annotations support the deduction of both extreme and non-extreme logical states, including quasi-states and directional trends. Each of these logical outcomes serves as a semantic signal that guides the system's behavior, prompting clarification requests, refining domain classification, or flagging inadequate answers. This evidential logic framework introduces interpretability and resilience, avoiding reliance on brittle heuristics or handcrafted rules.

In Logic Eτ, the degree of certainty (Gce = $\mu - \lambda$) expresses the balance between supporting and opposing evidence, while the degree of uncertainty (Gco = $\mu + \lambda - 1$) indicates the extent to which such evidence is simultaneously conflicting

As detailed in later sections, Logic Eτ underpins the system's core inferential mechanisms by enabling control decisions associated with propositions such as "The user question is clear", "The user question belongs to one of poultry production domains", or "The generated answer is adequate".

2.3 Design of Experiments for System-Level Parameter Tuning

In decision-oriented systems that demand precision, traceability, and trust, it is critical to address the limitations of large language models, particularly their non-deterministic behavior and susceptibility to hallucinations [8,11-13,19]. This study applied a Design of Experiments (DoE) approach to conduct a series of controlled tests, aiming to investigate how variations in system-level configurations affect the reliability, interpretability, and semantic accuracy of responses generated by the DSS architecture.

A controlled subset of the domain-specific knowledge base served as the foundation for the experiments. This corpus enabled the development of a fixed set of predefined queries; each paired with a gold-standard curated answer used as reference in the evaluation process. Two complementary metrics were analyzed. The first assessed system performance by measuring the semantic similarity (cosine similarity with SBERT embeddings) between the retrieved content and the reference answer, serving as a proxy for content fidelity and practical utility. The second examined the semantic alignment between the retrieved content and the original query, reflecting contextual coherence. While informative, this second metric does not guarantee factual correctness and may overvalue responses that are lexically aligned but semantically inaccurate or incomplete.

All experiments shared the same computational setup, including preprocessing libraries, LLM access, and vector-based retrieval infrastructure. Full implementation details and software versions are provided in Section 2.4 (Reproducibility and Software Environment).

Five experiments investigated the chunking strategy, input preprocessing, and generation parameters:

1. Chunk Size and Overlap: In the RAG pipeline, chunk size refers to the number of tokens in each embedded segment, while overlap specifies the number of tokens repeated between adjacent chunks, directly affecting contextual continuity and information density. The interaction between these parameters affects retrieval precision, semantic cohesion, and computational efficiency [20].
   The experiment utilized set of predefined question–answer pairs adopted across the other experiments and assessed both semantic alignment with the reference answer and contextual relevance to the original query. Three chunk sizes were tested: 128 tokens (high semantic precision, suitable for fine-grained reasoning), 256 tokens (practical optimum in most RAG pipelines), and 512 tokens (which maximizes cohesion in technical paragraphs). Overlap values included 32 tokens (minimal redundancy, avoiding abrupt cuts), 64 tokens (standard default, balances coherence and cost), and 128 tokens (high redundancy, beneficial for larger chunks but computationally heavier) [20,21]. A complete factorial design (3×3) was employed to investigate the combined effects of chunk size and overlap.
   The objective was to identify optimal trade-offs between granularity and cohesion, determine points of diminishing semantic returns, and establish thresholds beyond which overlap increases computational cost without improving retrieval quality.
2. Lemmatization: This preprocessing step reduces inflected or derived words to their base form (lemma), preserving grammatical context and semantic identity. By mapping morphological variants to a unified lexical representation, it may reduce embedding dispersion and improve retrieval alignment [22,23].
   Lemmatization was evaluated as a binary configuration: either applied or omitted symmetrically to both the indexed corpus and the user question–answer pairs. This experiment employed the chunking configuration identified in Experiment 1 and used the same set of predefined question–answer pairs, along with the evaluation criteria previously established.
   The objective was to determine whether the inclusion of lemmatization improves semantic similarity to the reference answer and enhances contextual alignment with the original query.
3. Normalization: This preprocessing step standardizes both the domain-specific corpus and the question–answer pairs by reducing superficial variability that does not affect meaning. It directly influences lexical alignment, improves embedding consistency, and enhances semantic matching, particularly in architectures where token-level similarity governs access to relevant content [24].
   Normalization was evaluated before vectorization as a binary configuration: either applied or omitted symmetrically to both the indexed corpus and the question–answer pairs used for evaluation. A complete 2^4 factorial design was used to test all possible combinations of four operations: lowercasing, punctuation removal, diacritic stripping, and whitespace collapsing.
   The objective was to determine whether these steps, individually or in combination, enhanced retrieval quality in terms of semantic similarity and contextual relevance.

Synonym Expansion: This preprocessing strategy enriches the indexed corpus and the question–answer pairs by appending or substituting terms with semantically equivalent alternatives. It aims to mitigate vocabulary mismatches and improve alignment between the user formulation and the stored knowledge base [25]. Following established evidence in information retrieval [45], synonym expansion was applied to reduce vocabulary mismatch and increase recall. This was particularly effective in poultry-related contexts: for instance, queries with *'feed formulation'* improved retrieval when expanded with *'broiler diet'*, and *'temperature control'* benefited from the inclusion of *'thermal regulation'*. While this strategy increased coverage, it also introduced a small number of false positives (e.g., *'lighting program'* matched with *'lightweight'*), which we acknowledge as a trade-off in retrieval precision.

Synonym expansion was evaluated as a binary configuration: either applied or omitted symmetrically to both the indexed corpus and the question–answer pairs. Lexical resources, including the semantic lexicon WordNet and its multilingual extension OMW, were used to identify synonym candidates prior to vectorization.

The objective was to assess whether this strategy enhances retrieval performance, particularly in terms of semantic similarity to the reference answer, in scenarios where lexical variation might otherwise reduce retrieval effectiveness.

4. Temperature and Top-p: The foundation model parameters regulate the stochastic behavior of the language model during response generation. Temperature controls the entropy of the output distribution, modulating the balance between determinism and exploration [26,27]. Top-p (nucleus sampling) constrains the sampling space to the smallest set of tokens whose cumulative probability exceeds a given threshold, shaping the diversity and unpredictability of the generated text [26,27].

   This experiment employed the chunking configuration identified in Experiment 1 and utilized the same set of predefined question–answer pairs, along with the evaluation criteria previously established. This model generated responses across a parameter space that ranged from factual and deterministic completions to controlled interpretative outputs and exploratory generations. The tested values were temperature $\in \{0.0, 0.3, 0.6, 0.9\}$ and top-p $\in \{0.8, 0.9, 1.0\}$. A complete 4×3 factorial design was employed to isolate the interaction effects of parameters within a realistic retrieval-augmented generation workflow. The tested ranges for temperature and top-p were informed by previous research on LLM generation parameters, which demonstrated that very low temperature values tend to produce deterministic and repetitive outputs, while very high values increase incoherence [11,46,47,48]. Similarly, top-p values between 0.6 and 1.0 have been widely adopted in foundational work to balance output diversity with factual reliability [46,48] (Brown et al., 2020). These ranges therefore represent established practice in controlled experiments with large language models.

   The objective was to evaluate how different sampling configurations impact semantic fidelity to the reference answer and contextual relevance to the original query, while maintaining generation stability of generation and interpretability.

Collectively, the experiments provided the empirical foundation for configuring the conversational agent. The selected parameters were directly incorporated into the final architecture, ensuring that the system strikes a balance between semantic precision, contextual relevance, and computational efficiency under realistic decision-making conditions. All procedures described here were executed within a controlled and reproducible software environment (see Section 2.4 for details).

2.4 System Architecture

The Decision Support AI-Copilot is composed of integrated modules structured as a Retrieval-Augmented Generation (RAG) application, as illustrated in Figure 2. The first module, the Knowledge Base Construction Pipeline (KB-CP), is responsible for preparing, segmenting, embedding, and indexing domain-specific content. The resulting repository, the Domain-Specific Knowledge Base (DS-KB), is organized by poultry production knowledge domains to support targeted semantic retrieval.

The second module, the Conversational DSS Agent (C-DSS-A), handles query interpretation, evidence retrieval, response generation, and logical evaluation, acting as the interactive interface between the user and the system, orchestrating language understanding, evidential reasoning, and answer synthesis.

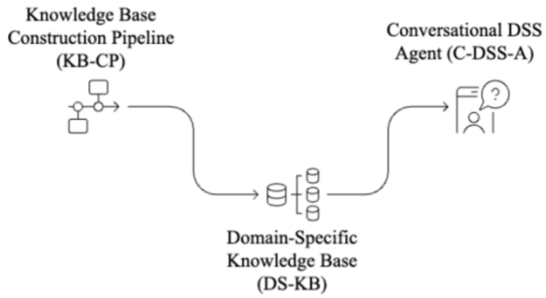


Figure 2. Modular architecture of the Decision Support AI-Copilot. The KB-CP preprocesses and embeds domain-specific content into the DS-KB. The C-DSS-A accesses this indexed repository to interpret user queries, retrieve relevant content, and generate logic-informed responses.

The architecture integrates LLMs and RAG techniques with Logic Eτ to support decision-making across multiple technical domains in poultry production. Its structure allows for independent evaluation and fine-tuning of semantic retrieval, language generation, and paraconsistent reasoning.

The system employs GPT-4o as its core language model. GPT-4o was selected for its semantic precision, low latency, and cost-efficiency, which make it particularly suitable for domain-specific RAG applications [28]. At the time of implementation, it was the most recent publicly available model in the GPT-4-turbo family. Its extended context window (up to 128k tokens) enables the integration of long retrieved passages while maintaining stable performance in the presence of ambiguity or contradiction, an essential requirement for logic-grounded decision support.

The operational parameters, such as chunking configurations, preprocessing routines, and generation settings, were empirically defined through the controlled experiments described in Section 2.1.2. These tests guided the selection of configurations that optimize trade-offs between granularity and cohesion, improve semantic similarity (cosine similarity with SBERT embeddings) between generated responses and the knowledge base, and enhance retrieval quality in terms of both content fidelity and contextual relevance. The system was also tuned to enhance robustness under lexical variability, strengthen alignment with the indexed content, and ensure generation stability and interpretability of generation across decision-making scenarios.

Semantic search is powered by FAISS (Facebook AI Similarity Search), selected for its scalability, support for both CPU and GPU backends, and proven efficiency in dense retrieval pipelines. The system utilized OpenAI's text-embedding-ada-002 model to encode knowledge base segments, and computes similarity via inner product (dot product), consistent with the model's scoring logic.

Vector indexing adopts the IndexFlatIP structure, a non-quantized flat index based on inner product similarity. This configuration ensures exact search results, which is crucial given the moderate scale of the dataset (fewer than 10,000 vectors) and the need for precise retrieval. The system performs retrieval via k-nearest neighbor (k-NN) search with a setting that balances contextual diversity with semantic relevance. Since latency is not a limiting factor in this application, exact k-NN was preferred to ensure retrieval fidelity and grounding quality in all downstream generations.

This architectural foundation supports the system's core functionalities and establishes the baseline over which configuration-level experiments (Section 2.1) were conducted to optimize performance and interpretability. Full details on code availability, software versions, and reproducibility protocols are provided in Section 2.4.

2.4.1 Knowledge Base Construction Pipeline (KB-CP)

To support domain-grounded retrieval and ensure high semantic precision during generation, the system relies on a knowledge base specifically constructed for poultry production decision-making. This repository was built through a structured pipeline comprising five main stages:

1. Document Collection: A total of 48 technical documents were curated from authoritative sources, including peer-reviewed scientific articles, poultry extension bulletins, technical production manuals, and sanitary protocols. The selection prioritized content with high informational density, practical relevance, and clear domain affiliation. Documents were collected through targeted searches in scientific databases, institutional repositories, and validated extension services.
2. Domain Classification: Each document was manually assigned to one of five predefined poultry production domains: (i) Housing and Environmental Control, (ii) Animal Nutrition, (iii) Poultry Health, (iv) Husbandry Practices, and (v) Animal Welfare.
3. These domains reflect core areas of technical decision-making in intensive poultry systems and are grounded in established animal welfare frameworks. The FAO's work on poultry welfare identifies health, nutrition, environmental comfort, and welfare as core aspects of assessment [29,30,42-44]. Classification was performed based on thematic focus, terminology patterns, and stated objectives of the material. In cases of overlap, domain assignment favored the dominant technical axis addressed by the document.
4. Preprocessing: All documents were converted to plain text and segmented into overlapping chunks, preserving local semantic cohesion. Chunk size and overlap were defined according to the optimal configuration identified in Experiment 1 (Section 2.1.2), which balances retrieval granularity with contextual integrity. This preprocessing step ensured that segment boundaries did not compromise sentence-level coherence, thereby improving embedding stability.
5. Vectorization: Each chunk was embedded using OpenAI's text-embedding-ada-002 model, producing dense vector representations in a high-dimensional semantic space. These embeddings captured contextual relationships at the subparagraph level, enabling fine-grained semantic retrieval aligned with user queries.
6. Domain-Based Indexing: For each knowledge domain, a separate FAISS index was created using the IndexFlatIP configuration (inner product similarity). This design supports fast and exact k-nearest neighbor (k-NN) search within each semantic repository. The use of independent indexes per domain facilitates targeted retrieval and minimizes semantic noise during generation.

The complete dataset, including raw documents, processed embeddings, and the full indexing pipeline, is publicly available via GitHub at [33].

2.4.2 Reasoning Workflow of the Conversational DSS Agent (C-DSS-A)

The C-DSS-A operates through a structured reasoning cycle (Figure 3) that integrates language comprehension, evidential assessment, semantic retrieval, and logical consistency checks.

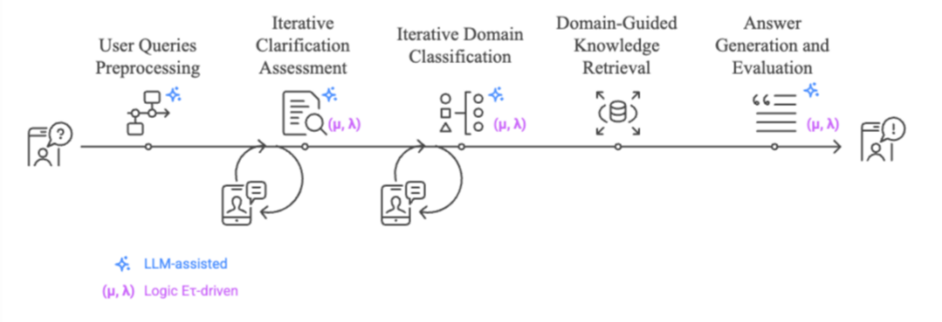


Figure 3. Workflow of the Conversational DSS Agent (C-DSS-A), detailing five sequential stages that combine LLM-based understanding with paraconsistent logic operations for query refinement, domain inference, knowledge retrieval, and evidence-grounded response generation.

Each decision stage is governed by a logic-based proposition evaluated under Logic Eτ. The complete workflow is composed of the following stages:

1. User Queries Preprocessing: User queries were preprocessed before both vector-based retrieval and language model inference. The adopted preprocessing configuration reflected the outcomes of controlled experiments. Synonym expansion was enabled as the only non-trivial transformation, selected for its capacity to bridge lexical gaps between user queries and indexed content. Lemmatization and punctuation removal were also applied, given their low computational cost and consistent contribution to lexical normalization. Conversely, diacritic stripping and whitespace collapsing were turned off by default, as their empirical impact on retrieval effectiveness proved negligible.
2. Iterative Clarification Assessment: Upon receiving a preprocessed user query, the system initiates an iterative process to evaluate and refine the clarity of the input. This is framed as the proposition:

   $P_1(\mu, \lambda)$: *"The user question is clear."*

   The annotation relies on a structured prompting protocol that infers evidential values directly from the LLM. Two specialized prompts quantify distinct epistemic dimensions: Clarity, defined as technical specificity and semantic coherence, and Vagueness, defined as conceptual ambiguity or logical imprecision. Both values are returned on a continuous scale from 0 to 1 and respectively, correspond to (μ, λ).

   The Gce(μ, λ) determines whether the system has sufficient confidence to proceed. Following prior applications of Logic Eτ in expert systems, a conservative threshold of Gce ≥ 0.75 was adopted to prevent unstable classifications in quasi-state borderline regions of the QUPC [14–16,18]. If Gce(μ, λ) < 0.75, the query is considered underdetermined. In such cases, the model generates a clarification prompt, which is appended to the conversational context. The revised input is re-evaluated using the same procedure, forming an iterative loop that continues until the certainty threshold is met (Gce(μ, λ) ≥ 0.75). At that point, the system proceeds to domain classification.
3. Iterative Domain Classification: Once the question is considered clear, the system prompts the LLM to classify it into one of five predefined poultry production domains: (i) housing and environmental control, (ii) animal nutrition, (iii) poultry

health, (iv) husbandry practices, or (v) animal welfare. The classification is formalized as an annotated proposition:

$P_2(\mu, \lambda)$ = "The question pertains to [identified domain]".

As in the previous step, the evidential values μ and λ are inferred by the LLM through guided prompting and interpreted under Logic Eτ. If the resulting Gce(μ, λ) falls below 0.75, the system generates a meta-question to validate the classification (e.g., "Does your question relate to [suggest*ed domain]?"*). If the user confirms the domain, the classification is accepted and the system proceeds. If the user rejects it, the domain is removed from the candidate list, and the LLM is prompted again using the updated domain set. This loop continues until a confident domain assignment is achieved, enabling the system to advance to semantic evidence retrieval.

4. Domain-Guided Knowledge Retrieval: With a clarified question and an identified domain, the system proceeds to semantic retrieval. The input query is embedded using OpenAI's text-embedding-ada-002 model, and a k-NN search (k = 5) is performed in a FAISS vector index (IndexFlatIP with dot-product similarity) to retrieve the most relevant content chunks. Each passage is linked to its original source and metadata.
5. Answer Generation and Evaluation: The retrieved passages are concatenated with the clarified user query and submitted as the prompt context to GPT-4o (via OpenAI API). The model then generates a draft response. In parallel, it evaluates the annotated proposition:

   $P_3(\mu, \lambda)$ = "The generated answer appropriately addresses the user's question."

   As in previous stages, the values μ and λ are inferred through guided prompting and interpreted under Logic Eτ. The resulting Gce(μ, λ) reflects the system's internal confidence in the adequacy of the response. If Gce(μ, λ) < 0.75, the response is flagged as potentially unreliable and may be revised or explicitly marked with a disclaimer to inform the user of evidential insufficiency or contradiction. The evidential outputs produced in this stage are then passed to the logical evaluation module, detailed in the following section.

Section 2.4 provides a detailed account of the software stack, experimental environment, and reproducibility measures employed in this work.

2.4.3 Reasoning Support with Logic Eτ

Each proposition formalizes a key decision point in the conversational reasoning cycle. Evidential values μ and λ are interpreted under Logic Eτ, and the resulting Gce(μ, λ) determines whether the system proceeds, flags the interaction, or initiates an iterative refinement. Thresholds and corresponding actions are defined to ensure interpretability, domain alignment, and response adequacy (Table 2)

Table 2. Annotated Propositions and Evidential Control Logic.

Proposition ID	Evaluated Statement	Purpose in System	Threshold (Gce)*	Action if Gce< Threshold	System Interaction Type
$P_1(\mu, \lambda)$	"The user question is clear."	Assess linguistic clarity; ensure interpretability	0.75	Trigger clarification question; append user response	Iterative clarification loop
$P_2(\mu, \lambda)$	"The question pertains to [identified domain]."	Classify query into production domain	0.75	Pose meta-question to user; eliminate rejected domain	Iterative domain pruning
$P_3(\mu, \lambda)$	"The generated answer appropriately addresses the user's question."	Assess adequacy and relevance of generated response	0.75	Flag response as uncertain; optionally trigger regeneration	Response flagging or regeneration

** Threshold adopted to prevent quasi-state borderline regions of the QUPC [14–16,18]*

2.5 Evaluation Protocol

The evaluation protocol focused on unit-level assessment of each reasoning stage within the C-DSS-A architecture and comprised three sets of tests, each designed to isolate and validate the behavior of individual components under controlled conditions. This approach enabled precise attribution of strengths and limitations at each stage of the decision workflow.

1. The test of the Iterative Clarification Assessment stage used a synthetic dataset of 130 user questions, generated from the DS-KB and labeled as Clear or Unclear. Each label encompassed a gradient of linguistic phenomena, including ambiguous phrasing, underspecified referents, non-technical constructions, and malformed syntax. Manual validation ensured internal consistency and class balance. The objective was to assess the system's ability to evaluate the proposition $P_1(\mu, \lambda)$: "The user question is clear", by discriminating underdetermined inputs based on evidential clarity rather than surface features. System performance was measured by its ability to converge to the correct classification through iterative reformulation, with convergence defined as $Gce(\mu, \lambda) \geq 0.75$ for proposition P_1.
2. The second test targeted the Iterative Domain Classification stage, using a new set of 100 synthetically generated questions, randomly assigned to one of the five defined domains, Housing and Environment, Animal Nutrition, Poultry Health, Husbandry Practices, Animal Welfare, or to no domain at all. This randomized distribution simulated open-query conditions. The dataset also included domainless questions to test rejection behavior under semantic uncertainty. The objective was to evaluate the system's ability to assess the proposition $P_2(\mu, \lambda)$: "The question belongs to [domain]", identifying the most appropriate category without forcing classification when evidential support was lacking. Classification was accepted only when the certainty degree satisfied $Gce(\mu, \lambda) \geq 0.75$, ensuring evidential convergence before domain attribution.
3. The third test focused on the stages of Domain-Guided Knowledge Retrieval and Answer Generation and Evaluation, using 100 synthetically generated question–answer pairs curated from source articles in the DS-KB. Each question was processed under two conditions: first, through direct prompting without retrieval or evidential evaluation, and second, through the whole system workflow, which includes retrieval from the DS-KB, generative response, and Logic Eτ-based self-assessment. In the second condition, the system instructed the model to evaluate the adequacy of its answer using Logic Eτ, producing an evidential annotation for the proposition $P_3(\mu, \lambda)$: "The generated answer is adequate". This annotation served as a meta-level judgment of response quality. For both conditions, the generated answers were compared to gold-standard references using semantic similarity metrics (cosine similarity with SBERT embeddings). The objective was to assess whether the evidential reasoning introduced by Logic Eτ improves the system's capacity to generate semantically valid responses and enhances the interpretability and trustworthiness of the final output.

2.6 Reproducibility and Software Environment

All system developments, experimental procedures, and evaluation workflows were implemented and executed in a reproducible Python environment (v3.9.6) using Visual Studio Code. The implementation leveraged a modular architecture composed of tools for retrieval orchestration, language generation, vector search, preprocessing, and evaluation. The main libraries and frameworks include:

- Language modeling and embedding: Openai 1.95.1 (for GPT-4o and text-embedding-ada-002), tiktoken 0.9.0 (for token counting and window control).

- Retrieval and orchestration: faiss-cpu 1.11.0 (for dense vector search using IndexFlatIP), langchain 0.3.25 and related packages (langchain-core, langchain-openai, langchain-community, langchain-text-splitters, langchain-xai, langsmith) for chaining retrieval, embedding, and generation steps.
- Text preprocessing and NLP: spaCy 3.8.7 (for lemmatization and syntactic analysis), nltk 3.9.1 (for lexical resources and linguistic tagging), including downloads: punkt, wordnet, omw-1.4, averaged_perceptron_tagger, averaged_perceptron_tagger_eng, punkt_tab.
- Data analysis and visualization: Pandas 2.3.1 (for data manipulation), scikit-learn 1.6.1 (for combinatorial evaluation routines), matplotlib 3.10.3 and seaborn 0.13.2 (for results visualization).
- Auxiliary and system tools: python-dotenv 1.1.0, requests 2.32.3, aiohttp 3.12.2, httpx 0.28.1 (for API and system orchestration), tenacity, joblib, threadpoolctl, Django 5.0.4 (web framework), djangorestframework 3.15.1 (APIs RESTful), and tqdm (for robustness, parallelization, and progress monitoring).

All software dependencies are publicly listed and version-pinned in the project's requirements.txt file. The complete codebase, prompts, dataset, and reproducibility pipeline are available via GitHub at [33]. All the versions indicated are up to date and consistent with the releases available until July 2025.

2.7 GenAI Disclosure

Generative AI was employed for development support (integrated into Visual Studio Code), the construction of synthetic test data, and the refinement of analysis statements. Synthetic questions and answers used in unit tests and controlled experiments were generated from real source material, under strict semantic constraints and manually curated for consistency and domain alignment. All experimental evidence and analytical results derive exclusively from real system outputs or curated domain content.

3. Results

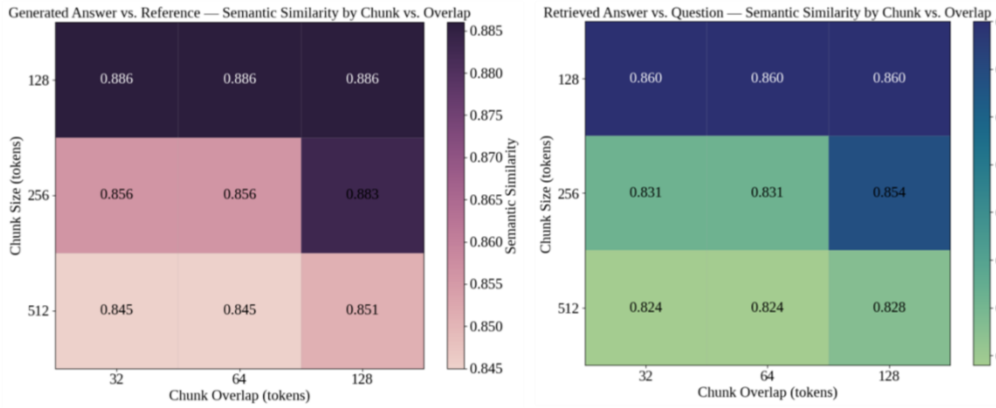
This section presents the results of a two-part evaluation. The first part presents controlled experiments assessing how system-level configurations influence response quality, interpretability, and semantic alignment. The second part analyzes the behavior of the complete system in end-to-end operation, with emphasis on evidential consistency and domain adequacy.

3.1 Experimental Results

Controlled experiments were conducted as proof-of-concept to evaluate the impact of variations in chunking strategy, input preprocessing, and generation parameters impact the quality of retrieval and response formulation. Each test isolates a specific configuration variable, allowing for a precise assessment of its impact on semantic accuracy, contextual relevance, and output stability.

3.1.1 Effects of Chunk Size and Overlap on Retrieval Quality

The experiment evaluated the impact of chunk size and overlap on retrieval quality in the RAG workflow (Figures 4a and 4b). With 128-token chunks, semantic similarity averaged ≈ 0.886 (RA vs. R) and ≈ 0.860 (RA vs. Q) across all overlap levels. For 256-token chunks, overlaps of 32 and 64 tokens resulted in lower similarity values, while an overlap of 128 tokens restored similarity to ≈ 0.883 (RA vs. R) and ≈ 0.854 (RA vs. Q). For 512-token chunks, similarity remained consistently lower across both metrics, even at maximum overlap.



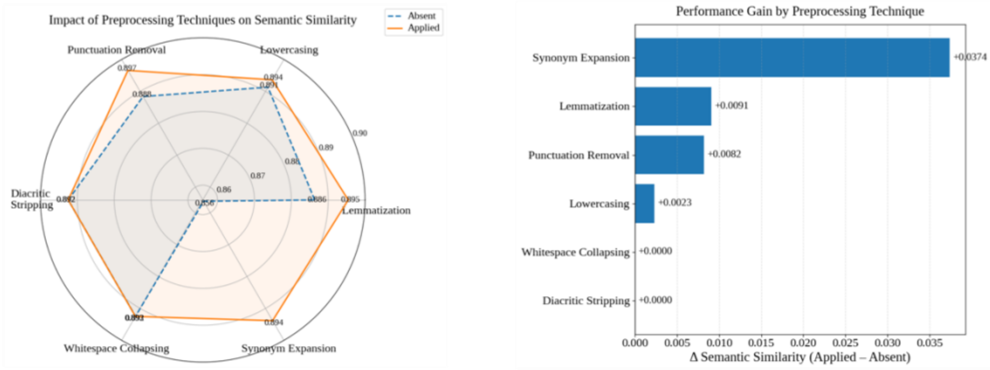
(**a**) (**b**)

Figure 4. Heatmaps showing the average semantic similarity (cosine similarity with SBERT embeddings) between (a) retrieved answers (RA) and gold-standard reference (R) and (b) retrieved answers (RA) and questions (Q), measured across different chunk and overlap configurations. The experiment corresponds to the parameter tuning analysis described. Each cell represents the mean similarity score for a given combination, using a controlled query set and vector retrieval via FAISS. Higher scores indicate a greater alignment between retrieved content and the expert reference.

Based on these results, the system was configured with a chunk size of 128 tokens and an overlap of 32 tokens, as this setting provided stable similarity scores with minimal redundancy. The similarity plateau observed (around ~0.865) should be interpreted as an internal performance reference that guided parameter tuning in this proof-of-concept study, rather than as an optimal or benchmark value.

3.1.2 Effects of Preprocessing on Retrieval Quality

The experiment assessed the isolated impact of standard preprocessing techniques on semantic similarity (cosine similarity with SBERT embeddings) between retrieved answers and gold-standard references, while keeping the retrieval architecture constant (Figures 5a and 5b). Synonym expansion produced the highest gain, increasing similarity by +0.0374, the largest delta among all transformations. Lemmatization and punctuation removal also showed positive contributions (+0.0091 and +0.0082, respectively). Lowercasing and whitespace collapsing yielded marginal improvements (< +0.002), while diacritic stripping showed no measurable effect. None of the tested techniques decreased retrieval quality.



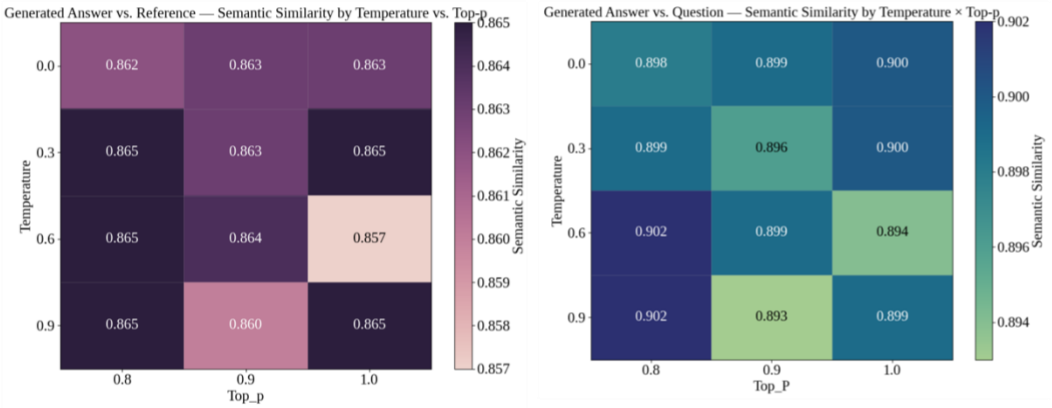
(**a**) (**b**)

Figure 5. (a) Radar chart comparing semantic similarity (cosine similarity with SBERT embeddings) between retrieved answers and reference answers, with and without each preprocessing technique. The orange contour represents performance with preprocessing applied; the dashed blue line corresponds to the baseline (absent). (b) Delta plot showing absolute performance gain (Δ similarity) for each technique, ordered from highest to lowest. Positive deltas indicate increased semantic alignment after applying the corresponding transformation.

Based on these results, the system was configured to enable synonym expansion as the only non-trivial transformation. Lemmatization and punctuation removal were also adopted due to their consistent but lightweight benefits, while diacritic stripping and whitespace collapsing were turned off by default.

3.1.3 Effects of Temperature and Top-p on Response Quality

The experiment examined the impact of sampling temperature and top-p on the semantic similarity (cosine similarity with SBERT embeddings) of generated answers, measured both against gold-standard references and the original user question (Figures 6a and 6b). Across the entire grid, response similarity to the reference (GA vs. R) remained stable, with most configurations converging around ≈ 0.865. The lowest value was observed at (temperature = 0.6, top-p = 1.0), where similarity decreased to 0.857. In contrast, alignment with the original question (GA vs. Q) varied more. The highest similarity (0.902) occurred at both (0.6, 0.8) and (0.9, 0.8). Configurations with top-p = 1.0 showed lower contextual alignment, with scores around 0.894–0.899.



(**a**) (**b**)

Figure 6. Heatmaps showing the average semantic similarity (cosine similarity with SBERT embeddings) between (a) generated answers (GA) and gold-standard reference (R), and (b) generated answers (GA) and original questions (Q), across different combinations of temperature and top-p values. Each cell indicates the mean similarity score for a fixed (temperature, top-p) configuration, based on cosine distance between sentence embeddings. Higher scores represent greater semantic alignment.

Based on these results, the system was configured with temperature = 0.6 and top-p = 0.8, as this setting achieved the best overall contextual alignment (0.902 GA vs. Q) while maintaining high factual similarity (0.865 GA vs. R).

3.2 Evaluation of Conversational DSS Agent workflow stages

The results of the stage-specific evaluations of the C-DSS-A enabled a fine-grained analysis of performance, robustness, and evidential behavior across the decision workflow.

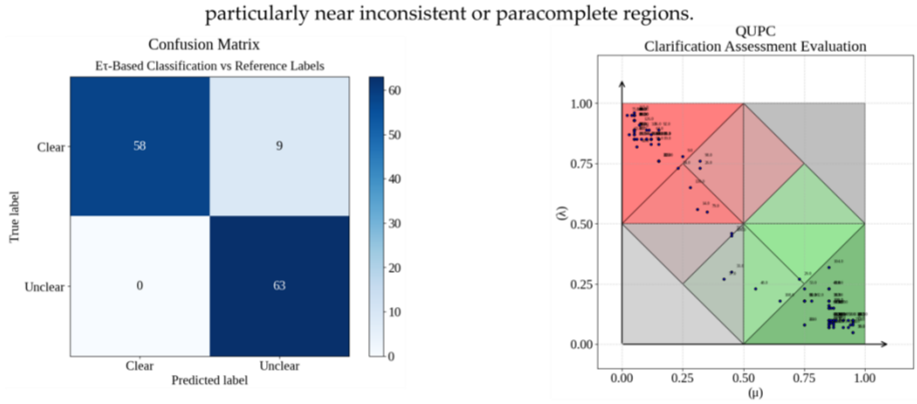
3.2.1 Results for the Iterative Clarification Assessment stage

The classification of the proposition *"The user question is clear"* achieved high overall performance (Table 3), with accuracy of 0.931, macro-averaged F1 of 0.931, and substantial agreement across both classes. Precision for Clear was 1.000, while recall was 0.866. For Unclear, recall reached 1.000 with precision of 0.875. These values indicate that the system consistently recognized unclear queries, while occasionally misclassifying clear queries as ambiguous.

Table 3. Performance Metrics for Clarity Classification stage (Clear vs. Unclear)

	Precision	Recall	F1-score	Support
Clear	1.000	0.866	0.928	67
Unclear	0.875	1.000	0.933	63
Accuracy			0.931	130
Macro avg	0.938	0.933	0.931	130
Weighted avg	0.939	0.931	0.931	130

The confusion matrix (Figure 7a) shows that most errors occurred when Clear queries were labeled as Unclear (9 instances), while no Unclear queries were misclassified as Clear. The QUPC projection (Figure 7b) illustrates that Clear queries concentrated in regions of high μ and low λ, while Unclear queries clustered in areas of higher λ values, particularly near inconsistent or paracomplete regions.



(**a**) (**b**)

Figure 7. (a) Confusion matrix comparing the system's final classification of the proposition "The user question is clear" against the ground truth labels (Clear vs. Unclear). The matrix reflects the asymmetric robustness of the model. The observed misclassifications reveal a conservative tendency, with a stricter evidential threshold for confirming clarity. (b) Two-dimensional representation of the evidential annotations (μ, λ) in the QUPC, illustrating the distribution of outputs from the Iterative Clarification Assessment module. Points are colored according to their final classification and plotted against the paraconsistent decision regions derived from Logic Eτ. The separation between classes and the concentration near τ-lattice diagonals highlight the model's discriminative sensitivity to varying degrees of clarity and vagueness.

Based on these results, the module was retained with its default evidential thresholds (Gce ≥ 0.75), as this setting balanced precision for *Clear* with maximal recall for Unclear, ensuring reliable detection of underdetermined inputs.

3.2.2 Results for the Domain Classification stage

The classification of the proposition *"The question belongs to [domain]"* showed variation across categories (Table 4). Animal Nutrition achieved precision of 0.818 and recall of 1.000 (F1 = 0.900). Animal Welfare reached precision of 0.600 and recall of 0.947 (F1 = 0.735). Housing and Environment obtained precision of 0.762 and recall of 0.889 (F1 = 0.821). Poultry Health registered precision and recall of 0.813 each (F1 = 0.813). No correct classifications were obtained for Husbandry Practices (F1 = 0.000). Inspection of misclassified cases indicates that this domain was consistently absorbed into semantically related categories, particularly Animal Welfare and Poultry Health. This behavior likely reflects the limited representation of Husbandry Practices in the corpus, combined with the semantic proximity of its content to adjacent domains. For the None class (no domain), precision was 1.000 and recall 0.800 (F1 = 0.889). Overall accuracy was 0.737, with a macro-averaged F1 of 0.693.

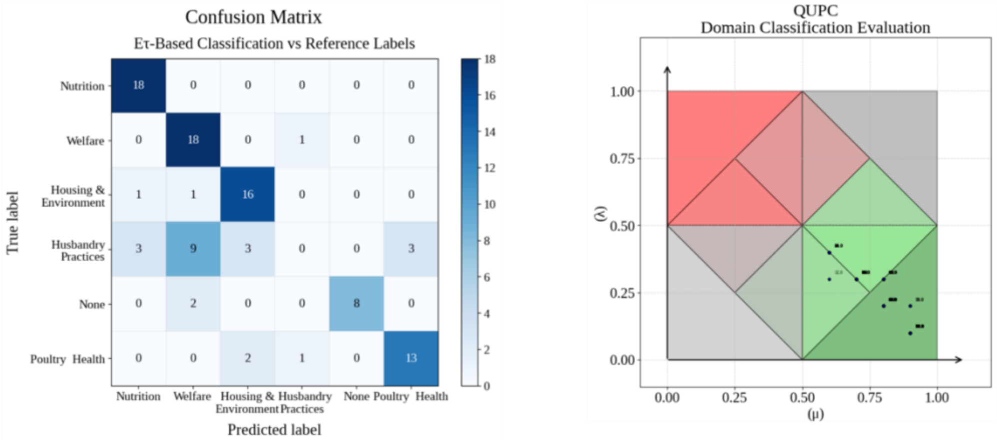
Table 4. Performance Metrics for Domain Classification stage

	Precision	Recall	F1-score	Support
Animal Nutrition	0.818	1.000	0.900	18
Animal Welfare	0.600	0.947	0.735	19
Housing and Environment	0.762	0.889	0.821	18
Husbandry Practices	0.000	0.000	0.000*	18

None (no domain)	1.000	0.800	0.889	10
Poultry Health	0.813	0.813	0.813	16
Accuracy			0.737	99
Macro avg	0.665	0.741	0.693	99
Weighted avg	0.635	0.737	0.675	99

* No instances correctly classified for this class.

The confusion matrix (Figure 8a) shows that most errors involved confusion between Husbandry Practices and semantically adjacent categories (Animal Welfare and Poultry Health). The QUPC projection (Figure 8b) indicates that most accepted classifications occupied regions of high certainty ($\mu > 0.75$, $\lambda < 0.25$), while a few borderline cases appeared near decision boundaries.



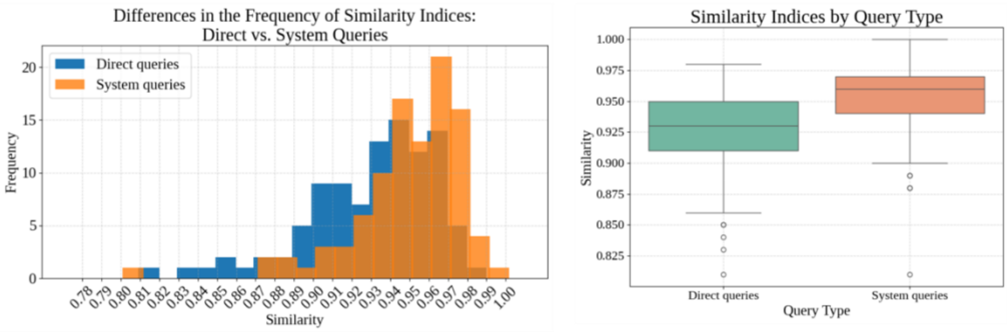
(a) **(b)**

Figure 8. (a) Confusion matrix comparing predicted and reference labels for the proposition "The question belongs to [domain]", across five poultry production domains and an out-of-domain class (None). Misclassifications are concentrated in semantically adjacent categories, particularly in *Husbandry Practices*. Correct abstentions in the *None* class confirm the system's capacity to reject uncertain assignments when evidential support is insufficient. (b) QUPC projection of the evidential annotations (μ, λ) associated with domain classification decisions. Most points fall within regions of high certainty and low contradiction, consistent with valid assignments. The sparse activation near decision boundaries highlights borderline cases, suggesting residual ambiguity in specific domain transitions.

Based on these results, the module was retained with the same evidential threshold (Gce ≥ 0.75), as this configuration supported robust rejection of out-of-domain queries while maintaining reliable classification for most domains.

3.2.3 Results for the *Domain-Guided Knowledge Retrieval* and *Answer Generation and Evaluation* stages

The comparative evaluation between direct and system-guided queries showed consistent differences in semantic similarity (cosine similarity with SBERT embeddings) as shown in Figures 9a and 9b. System-mediated responses displayed a higher median similarity and reduced variance compared to direct queries, with fewer low-similarity outliers. The similarity curve (Figure 10) showed that system queries concentrated around higher similarity values, with a sharper peak near 0.97.



(a) **(b)**

Figure 9. (a) Histogram of similarity scores comparing direct queries and system-mediated queries against the gold-standard answers. System responses exhibit a higher concentration of high-similarity outputs, with a clear rightward shift in distribution. (b) Boxplot summarizing similarity score distributions for each query type. System-mediated queries show higher median similarity and reduced variance, indicating more consistent semantic alignment.

The similarity distribution and polynomial trend curves (Figures 10 and 11) showed consistent differences across configurations. System-mediated queries concentrated around higher similarity values, with a sharper peak near 0.97, while RAG-only queries also shifted the distribution toward higher alignment compared to the LLM-only baseline. In contrast, the LLM-only setting displayed a flatter and more dispersed distribution, with a larger proportion of low-similarity outputs. The sharper peaks for both RAG and System-mediated configurations indicate a higher concentration of well-aligned responses and reduced variance, highlighting the stabilizing effect of knowledge retrieval.

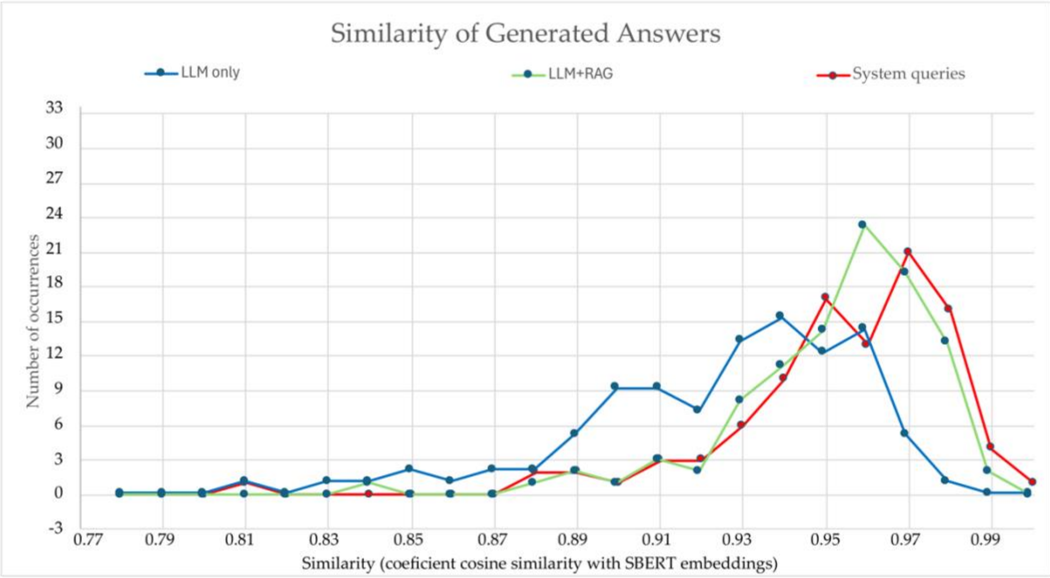


Figure 10. Distribution of semantic similarity coefficients (cosine similarity with SBERT embeddings) between generated answers and gold-standard references, across three configurations: LLM only (blue), LLM+RAG (green), and System-mediated with Logic Eτ (red).

The polynomial curves further emphasize these overall distribution patterns, making the relative gains in similarity concentration and reliability more evident when evidential reasoning is incorporated

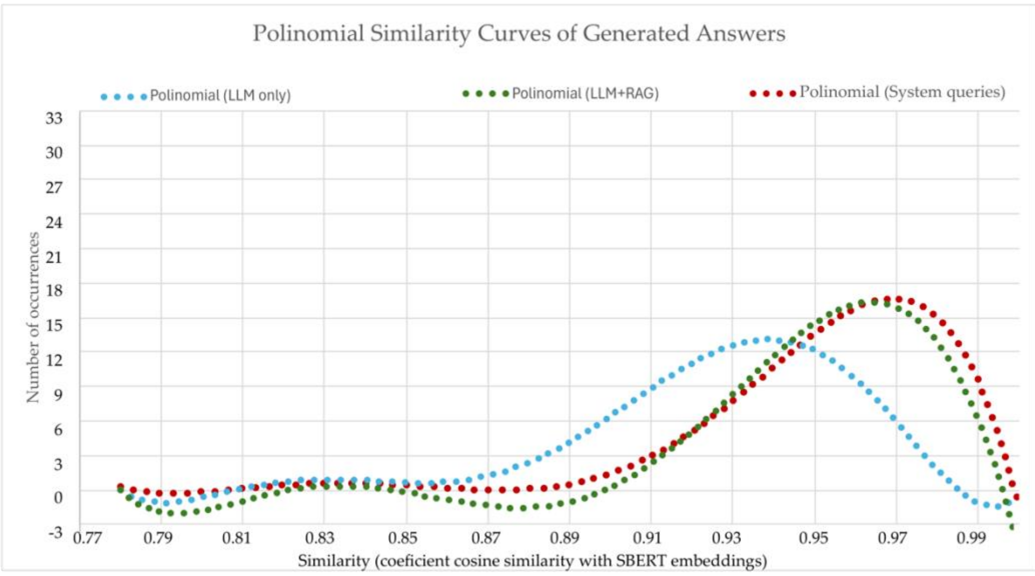
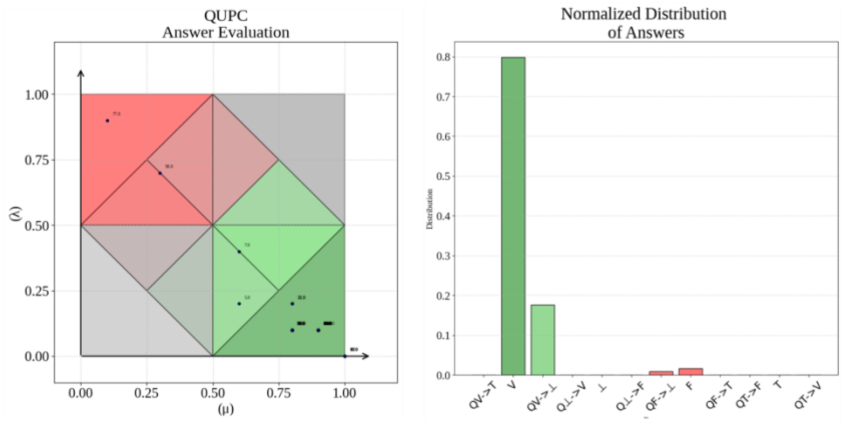


Figure 11. Polynomial trend curves of semantic similarity (cosine similarity with SBERT embeddings) between generated answers and gold-standard references, comparing three configurations: LLM only (blue), LLM+RAG (green), and System-mediated with Logic Eτ (red).

Evidential judgments (Figures 12a and 12b) showed that 95% of cases fell in high-certainty regions ($\mu > 0.75$, $\lambda < 0.25$), with most outputs classified as True (V). A minority were assigned to False (2%), Quasi-false tending to Paracomplete (1%), and Quasi-true tending to Paracomplete (2%). Approximately 5% of the outputs appeared near contradictory regions, corresponding to marginally lower similarity scores.



(a) **(b)**

Figure 12. a) QUPC projection of the evidential annotations (μ, λ) for the proposition "The generated answer is adequate." Most points concentrate in regions of high certainty and low contradiction, reflecting confident adequacy judgments. Sparse activation in quasi-inconsistent zones highlights borderline or semantically ambiguous responses. (b) Normalized distribution of the resulting logical states, with 95% classified as True (V), and a minority assigned to False (F, 2%), Quasi-false tending to paracomplete (QF→⊥, 1%), and Quasi-true tending to paracomplete (QV→⊥, 2%). The distribution reinforces the predominance of reliable outputs and the system's ability to flag marginal cases with non-extreme logical states.

Table 5 provides illustrative cases where the system flagged generated answers as inadequate, showing the corresponding user queries, excerpts of the outputs, evidential judgments (μ, λ, and Gce), and the user-facing messages that communicate uncertainty in a neutral and supportive tone.

Table 5. Illustrative Cases of Evidence-Based Inadequacy Judgments

User Query	Generated Answer (excerpt)	Evidential Judgment	User-Facing Message
What is the recommended broiler diet for heat stress?	Conflicting guidelines were retrieved, some emphasizing increased electrolytes, others	$\mu = 0.61$, $\lambda = 0.12$ Gce = 0.49 (Inadequate)	"Some retrieved information appears inconsistent. The answer may need clarification. Please consider refining your

	focusing on energy adjustment		query or reviewing the supporting evidence.”
What is the optimal temperature for broiler housing at 21 days of age?	Broilers at 21 days should be kept at 28 °C; some sources also mention 24–26 °C depending on ventilation	$\mu = 0.64$, $\lambda = 0.39$ Gce = 0.25 (Inadequate)	“Retrieved guidelines vary across sources. Please consider reviewing the suggested ranges or providing more context for your query.”
How often should litter be replaced in broiler houses?	Some sources recommend complete replacement each cycle, others suggest partial reuse if treated with drying agents	$\mu = 0.78$, $\lambda = 0.17$ Gce = 0.61 (Inadequate)	“The retrieved evidence shows differing recommendations. The answer may depend on farm conditions—please review the supporting guidelines.”
Is vaccination against coccidiosis always required in broilers?	Most sources recommend vaccination for long-cycle broilers; some mention prophylaxis may suffice under high biosecurity.	$\mu = 0.62$, $\lambda = 0.14$ Gce = 0.48 (Inadequate)	“Evidence for this query is partly inconsistent. The system combined both vaccination and prophylaxis approaches—please interpret according to your production context”

Based on these results, the system was configured to operate with retrieval and Logic Eτ-based self-assessment enabled by default, as this combination consistently improved similarity alignment and provided evidential annotations for reliability control.

4. Discussion

This section examines the findings from two complementary dimensions: system-level tuning experiments that revealed how configuration choices condition retrieval and generation, and stage-wise evaluations that demonstrated how evidential mechanisms regulate reasoning across modules. Considered together, these findings outline an integrative perspective on how technical optimization and modular inference interact in shaping the overall behavior of the DSS.

4.1 Implications of System-Level Parameter Tuning Experiments

The experiments conducted as proof-of-concept trials, provided evidence on how system-level parameters influenced retrieval fidelity, contextual alignment, and the stability of generative outputs. These results revealed trade-offs that guided the technical configuration of the DSS and offered methodological insights for the design of RAG pipelines.

The experiments on chunk size and overlap indicate that smaller segments tend to preserve semantic integrity, reducing the need for redundant overlap. At 256 tokens, additional overlap was required to restore retrieval quality, suggesting the presence of a threshold below which segment boundaries start to fragment contextual cohesion. Very large chunks of 512 tokens, even with high overlap, showed limited benefit, indicating the constraints of relying on size alone to secure retrieval accuracy. These findings highlight that segment configuration is not a neutral choice but a determinant factor for balancing semantic precision, contextual cohesion, and computational cost.

The analysis of preprocessing strategies shows that synonym expansion proved particularly influential in bridging lexical gaps between queries and knowledge base content, strengthening retrieval precision through semantic diversification. Lemmatization and punctuation removal also proved beneficial, though to a lesser extent, by reducing morphological variability and surface noise that can obscure semantic matches. By contrast, lowercasing, whitespace collapsing, and diacritic stripping offered negligible contributions, indicating that common normalization routines may be redundant in embedding spaces that already capture semantic robustness. The overall absence of negative effects suggests that preprocessing choices can be selectively applied, with meaningful gains concentrated in a few targeted transformations rather than in broad, indiscriminate normalization.

The evaluation of generation parameters indicated that factual accuracy remained relatively stable across configurations, while contextual alignment was more sensitive to variation. The performance drop observed at mid-level temperature combined with unfiltered sampling (0.6, 1.0) reflected a weakening of grounding when lexical openness was high. In contrast, settings that paired moderate or high diversity with a constrained nucleus, such as (0.6, 0.8) and (0.9, 0.8), improved relevance to user queries without compromising fidelity to reference answers. Configurations with top-p = 1.0 confirmed the risk of topic drift, as unrestricted sampling introduced variability that diluted semantic focus. These patterns suggest that balanced decoding strategies, exemplified by the adopted profile of temperature = 0.6 and top-p = 0.8, provide a practical equilibrium between determinism and contextual flexibility in domain-constrained tasks.

As proof-of-concept, these experiments suggested key trade-offs and helped establish a preliminary foundation for subsequent system evaluations. They delineate a methodological basis from which broader validations and domain-specific extensions can be pursued, emphasizing that parameter tuning is not a peripheral adjustment but a defining step that conditions the robustness of the DSS.

4.2 Evidential Reasoning in Stage-Wise System Evaluations

Beyond parameter tuning, the stage-wise evaluation of the C-DSS-A showed how evidential reasoning shaped system behavior across successive modules, revealing distinct patterns of performance and uncertainty management. At the same time, the agent as a whole exhibited a coherent operational profile, with all modules displaying controlled behavior under uncertainty. Logic Eτ provided the continuous interpretive structure that supported query clarification, domain attribution, and answer validation across stages [14,32].

The system demonstrated high sensitivity to semantic underdetermination during question clarification. Its conservative bias toward flagging inputs as 'Unclear' was not merely a trade-off, but an intentional mechanism for ambiguity control. By embedding classification within a Logic Eτ approach, the system preserved the evidential structure of borderline cases, avoiding premature resolution. This behavior prevented uncertain queries from propagating unchecked into subsequent inference stages, ensuring that downstream modules operated on well-formed and contextually interpretable inputs, thereby reducing the risk of hallucinations and erratic responses [11–14]. However, this imbalance between precision and recall in the Clear class also suggests that some genuine queries may elicit unnecessary clarification prompts. While maximizing recall for Unclear inputs ensures robust detection of ambiguity, it introduces a usability concern: repeated clarification loops could affect user perception of responsiveness, potentially causing frustration when clear questions are unnecessarily flagged. To mitigate this, future work will explore adaptive calibration of evidential thresholds and dynamic adjustment strategies to minimize superfluous prompts while maintaining reliability in detecting truly ambiguous queries.

In domain classification, the system exhibited variable performance across categories. Domains such as *Animal Nutrition*, as well as *Housing and Environment*, were consistently well distinguished. At the same time, *Husbandry Practices* triggered frequent confusion with semantically neighboring categories. This systematic misclassification highlights challenges for domain categorization. It likely reflects both the impact of the underrepresentation of the class (*Husbandry Practices*, in this case) in the corpus and its semantic overlap with closely related domains (*Animal Welfare* and *Poultry Health*), which resulted in boundary errors. The absence of inter-rater validation may also have introduced annotation noise, underscoring the need for corpus diversification and multi-annotator validation to strengthen domain coverage and reliability. Despite these issues, the system demonstrated robust rejection behavior for domainless queries, with the evidential threshold effectively blocking premature assignments. This selective restraint is particularly relevant for LLM-based architectures [36], which often default to overgeneralization in the face of ambiguity [8,13].

The comparative distributions further illustrate the incremental contribution of retrieval and evidential reasoning to answer quality. The flatter and more dispersed curve of the LLM-only baseline reflects a higher incidence of poorly aligned responses, whereas the inclusion of RAG concentrated outputs closer to the reference. The system-mediated configuration, which integrates Logic Eτ, sharpened this peak, reducing variance and reinforcing reliability. Although these differences are visually clear, their statistical significance was not formally tested, and the trends should therefore be interpreted with caution.

Ultimately, the system demonstrated that evidential inference plays a central role in enhancing response reliability. When retrieval and Logic Eτ-based self-assessment were active in the system workflow, responses shifted toward higher semantic alignment with reference answers and exhibited reduced variance. The analysis showed that most outputs were evaluated as highly adequate under the proposition P_3, while borderline or contradictory cases were correctly flagged through hesitant or inconsistent annotations. This evidential trace, absent in standard generation workflows, introduced a meta-level of interpretability that reinforced the trustworthiness of the final output [8,9,14,15].

Taken together, these results demonstrate that Logic Eτ is not merely an explanatory overlay but a functional mechanism that modulates the system's behavior. It enables abstention, detects vagueness, calibrates domain attribution, and qualifies the generative output, all within a consistent inferential framework. Importantly, it achieves these functions without relying on heuristics or hard-coded decision trees [14-16,32,36,37].

From a practical standpoint, this evidential strictness positions the system for real-world deployment in complex poultry production contexts [36,37]. It is relevant for high-stakes scenarios where interpretive caution, traceable reasoning, and the ability to admit uncertainty are critical requirements [2–5]

These findings reinforce the working hypothesis that integrating Logic Eτ with LLM and RAG architectures enhances reliability and interpretability in decision-support scenarios. By enabling structured self-assessment and evidential control, the system addresses key limitations of previous approaches, particularly their brittleness in the face ambiguity or contradiction [8,11–13]. The observed behaviors align with prior research on paraconsistent reasoning in uncertain environments [32], while extending its application to high-level semantic workflows.

4.3 Integrative Perspective: From Parameter Tuning to System Behavior

Taken together, the results from the tuning experiments and the stage-wise evaluations converge on a complementary view: parameter tuning defined the conditions for stable and precise retrieval and generation, while Logic Eτ qualified and constrained outputs under uncertainty. This integration illustrates that robustness emerges not from isolated components, but from their interaction. It also underscores that technical optimization alone is insufficient; without evidential reasoning, the system would remain vulnerable to contradiction, while Logic Eτ itself depends on a calibrated retrieval pipeline.

In this respect, the proposed framework diverges from recent LLM+RAG approaches. While those systems improve contextual reasoning and factual grounding, they still rely on classical logic assumptions and lack mechanisms to formally manage contradictory or incomplete evidence [8,11–13]. By adding Logic Eτ, the present architecture introduces explicit evidential quantification and structured self-assessment, enabling the system to qualify ambiguous inputs and outputs rather than force binary resolutions [14-16,18]. This contrast clarifies that the contribution of the study is not a replacement of existing RAG pipelines, but their extension with contradiction-tolerant reasoning, oriented toward explainable and trustworthy decision support in poultry production.

4.4 Limitations and Future Work

While the system exhibited coherent and controlled behavior across all inference stages, some important limitations must be acknowledged.

The experimental design relied on a restricted sample size and a limited evaluation scope, which constrains the transferability of the findings. These experiments should therefore be regarded as proof-of-concept trials to test the feasibility of combining LLMs and RAG with Logic Eτ under controlled conditions, rather than as large-scale validation. In particular, the evaluation of chunk size and overlap was restricted to a narrow set of parameter configurations and metrics, without validation across different poultry production domains. Future work will expand the parameter range and validate chunking strategies in multiple domains to improve robustness and generalization.

Another limitation concerns the representativeness of the knowledge base used in the RAG pipeline. The current corpus was intentionally restricted to a small and homogeneous set of documents, which enabled controlled testing but limited the diversity of contexts and production scenarios represented. This restriction may affect retrieval performance and response reliability. In addition, domain classification was conducted by a single annotator without inter-rater validation, and cross-domain materials were reduced to a dominant category. The block segmentation strategy, optimized with synthetic QA data, has also not yet been validated for semantic integrity in long technical documents. Future work should expand the knowledge base with more diverse sources including manuals, scientific literature, regulatory guidelines, and field reports, while ensuring balance across document types, timeframes, and regions. Multi-annotator validation and extended segmentation assessments will also be incorporated to strengthen corpus reliability and semantic fidelity.

The system has not yet been validated by domain experts nor tested in real production environments. The current implementation should therefore be regarded as an early-stage prototype, evaluated under controlled conditions. Next steps should prioritize participatory assessments with poultry specialists and in-situ deployments to assess usability, robustness, and contextual adaptability. From an operational perspective, deployment feasibility is constrained by available computing resources. Performance testing, including stress scenarios, will be required to validate scalability, and farms with limited infrastructure may still require hybrid cloud-based solutions. Large-scale deployment will also require optimization to support thousands of concurrent queries and integration with sensor-driven alarm systems, ensuring timely responses in intensive production environments. Such evaluations are crucial for consolidating operational maturity and refining evidential thresholds in response to practical decision-making demands. In parallel, future work may explore adaptive calibration strategies for domain boundaries and clarity thresholds, as well as pathways for generalization beyond poultry production.

Finally, although evidential reasoning improved robustness against uncertainty and contradictions, the system has not yet been stress-tested under adversarial, hostile, or multilingual queries. These scenarios represent potential points of failure and should be included in future evaluations to better characterize resilience beyond the controlled settings adopted here. Future work will also include comprehensive comparisons with existing DSS approaches, as well as direct baselines with standard RAG implementations such as Haystack and the LangChain default pipeline, together with RAG+LLM systems without evidential reasoning, to contextualize the impact of the proposed framework and preprocessing strategies.

5. Conclusions

This study demonstrated that integrating Large Language Models, Retrieval-Augmented Generation, and Paraconsistent Annotated Evidential Logic Eτ enables the construction of an interpretable and contradiction-tolerant decision support system for poultry production. By embedding evidential reasoning at each stage of the conversational workflow, the Decision Support AI-Copilot for Poultry Farming showed promising behavior in terms of semantic alignment, inference under uncertainty, and domain attribution. The architecture avoided heuristic shortcuts, relying instead on structured logical evaluation to handle ambiguous or borderline cases. In this context, the present work contributes to the field of AI-based decision support in agriculture by introducing an integrative, multi-domain, knowledge-grounded, and contradiction-tolerant approach tailored to the specific demands of poultry production.

At the same time, the system should be regarded as a proof-of-concept prototype rather than a fully validated tool. The results reported here were obtained under controlled conditions with a restricted corpus and evaluation scope, and further work is required to establish robustness, scalability, and usability in real production environments. These constraints temper the immediate applicability of the findings but reinforce the value of the study as a methodological foundation.

Finally, this study outlined a roadmap for future validation of the framework and indicated opportunities to extend its scope to other livestock species. By clarifying the limitations and future directions, the research provides both a conceptual contribution and a practical agenda for advancing evidentially guided decision support in agriculture.

References

1. Food balance sheets 2010–2022. Global, regional and country trends. *Statistics*. Available online: https://www.fao.org/statistics/highlights-archive/highlights-detail/food-balance-sheets-2010-2022-global-regional-and-country-trends/en (accessed on 9 July 2025).
2. Mottet, A.; Tempio, G. Global poultry production: Current state and future outlook and challenges. *World's Poult. Sci. J.* **2017**, *73*, 245–256. https://doi.org/10.1017/S0043933917000071.
3. Berckmans, D. General introduction to precision livestock farming. *Anim. Front.* **2017**, *7*, 6–11. https://doi.org/10.2527/af.2017.0102.
4. Gržinić, G.; Piotrowicz-Cieślak, A.; Klimkowicz-Pawlas, A.; Górny, R. L.; Ławniczek-Wałczyk, A.; Piechowicz, L.; et al. Intensive poultry farming: A review of the impact on the environment and human health. *Sci. Total Environ.* **2023**, *858*, 160014. https://doi.org/10.1016/j.scitotenv.2022.160014.

5. Hafez, H. M.; Attia, Y. A. Challenges to the poultry industry: Current perspectives and strategic future after the COVID-19 outbreak. *Front. Vet. Sci.* **2020**, *7*, 516. https://doi.org/10.3389/fvets.2020.00516.
6. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv* **2021**, arXiv:2005.11401. https://doi.org/10.48550/arXiv.2005.11401.
7. Li, H.; Su, Y.; Cai, D.; Wang, Y.; Liu, L. A survey on retrieval-augmented text generation. *arXiv* **2022**, arXiv:2202.01110. https://doi.org/10.48550/arXiv.2202.01110.
8. Leite, M. V.; Abe, J. M.; Souza, M. L. H.; de Alencar Nääs, I. Enhancing environmental control in broiler production: Retrieval-augmented generation for improved decision-making with large language models. *AgriEng* **2025**, *7*, 12. https://doi.org/10.3390/agriengineering7010012.
9. Izacard, G.; Grave, E. Leveraging passage retrieval with generative models for open domain question answering. *arXiv* **2021**, arXiv:2007.01282. https://doi.org/10.48550/arXiv.2007.01282.
10. Cai, D.; Wang, Y.; Liu, L.; Shi, S. Recent advances in retrieval-augmented text generation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022)*, New York, NY, USA, July 11–15, 2022. https://doi.org/10.1145/3477495.3532682.
11. Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; et al. Language models are few-shot learners. *arXiv* **2020**, arXiv:2005.14165. https://doi.org/10.48550/arXiv.2005.14165.
12. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; et al. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2023**, *55*, 248:1–248:38. https://doi.org/10.1145/3571730.
13. Metze, K.; Morandin-Reis, R. C.; Lorand-Metze, I.; Florindo, J. B. Bibliographic research with ChatGPT may be misleading: The problem of hallucination. *J. Pediatr. Surg.* **2024**, *59*, 158. https://doi.org/10.1016/j.jpedsurg.2023.08.018.
14. Abe, J. M.; Akama, S.; Nakamatsu, K. *Introduction to Annotated Logics: Foundations for Paracomplete and Paraconsistent Reasoning*; Springer International Publishing: Cham, 2015; Vol. 88. https://doi.org/10.1007/978-3-319-17912-4.
15. Carvalho, F. R. D.; Abe, J. M. *A Paraconsistent Decision-Making Method*; Springer International Publishing: Cham, 2018; Vol. 87. https://doi.org/10.1007/978-3-319-74110-9.
16. de Carvalho Junior, A.; Justo, J. F.; de Oliveira, A. M.; da Silva Filho, J. I. A comprehensive review on paraconsistent annotated evidential logic: Algorithms, applications, and perspectives. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107342. https://doi.org/10.1016/j.engappai.2023.107342.
17. Tiwari, A.; Beed, R. S. Applications of Internet of Things in Smart Agriculture. In *AI to Improve e-Governance and Eminence of Life*; Springer: Singapore, 2023; pp 103–115. https://doi.org/10.1007/978-981-99-4677-8_6.
18. Abe, J. M. Remarks on Paraconsistent Annotated Evidential Logic Et. *Unisanta Sci. Technol.* **2014**, *3*, 25–29.
19. Petroni, F.; Piktus, A.; Fan, A.; Lewis, P.; Yazdani, M.; Cao, N. D.; et al. KILT: A benchmark for knowledge intensive language tasks. *arXiv* **2021**, arXiv:2009.02252. https://doi.org/10.48550/arXiv.2009.02252.
20. Wang, H.; Zhang, D.; Li, J.; Feng, Z.; Zhang, F. Entropy-optimized dynamic text segmentation and RAG-enhanced LLMs for construction engineering knowledge base. *Appl. Sci.* **2025**, *15*, 3134. https://doi.org/10.3390/app15063134.
21. Zhu, B.; Vuppalapati, C. Integrating retrieval-augmented generation with large language models for supply chain strategy optimization. In *Applied Cognitive Computing and Artificial Intelligence*; Springer: Cham, 2025; pp 475–486. https://doi.org/10.1007/978-3-031-85628-0_34.
22. Boban, I.; Doko, A.; Gotovac, S. Sentence retrieval using stemming and lemmatization with different length of the queries. *Adv. Sci. Technol. Eng. Syst. J.* **2020**, *5*, 349–354. https://doi.org/10.25046/aj050345.
23. Pramana, R.; Debora; Subroto, J. J.; Gunawan, A. A. S.; Anderies. Systematic literature review of stemming and lemmatization performance for sentence similarity. In *Proceedings of the 2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, Yogyakarta, Indonesia, Nov 23–25, 2022; IEEE: New York, NY, USA, 2022; pp 1–6. https://doi.org/10.1109/icitda55840.2022.9971451.
24. Sirisha, U.; Kumar, C.; Durgam, R.; Eswaraiah, P.; Nagamani, G. An analytical review of large language models leveraging KDGI fine-tuning, quantum embedding's, and multimodal architectures. *Comput. Mater. Contin.* **2025**, *83*, 4031–4059. https://doi.org/10.32604/cmc.2025.063721.
25. Gabín, J.; Parapar, J. Leveraging retrieval-augmented generation for keyphrase synonym suggestion. In *Advances in Information Retrieval*; Springer: Cham, 2025; pp 311–327. https://doi.org/10.1007/978-3-031-88711-6_20.
26. Chung, J.; Kamar, E.; Amershi, S. Increasing diversity while maintaining accuracy: Text data generation with large language models and human interventions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume

1: Long Papers); Rogers, A., Boyd-Graber, J., Okazaki, N., Eds.; Association for Computational Linguistics: Toronto, Canada, July 9–14, 2023; pp 575–593. https://doi.org/10.18653/v1/2023.acl-long.34.

27. Amin, M. M.; Schuller, B. W. On prompt sensitivity of ChatGPT in affective computing. In *Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction (ACII 2024)*, Glasgow, United Kingdom, 15–18 September 2024; pp 203–209. https://doi.org/10.1109/ACII63134.2024.00028.
28. OpenAI; Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; et al. GPT-4 technical report. *arXiv* **2024**, arXiv:2303.08774. https://doi.org/10.48550/arXiv.2303.08774.
29. Ojo, R. O.; Ajayi, A. O.; Owolabi, H. A.; Oyedele, L. O.; Akanbi, L. A. Internet of things and machine learning techniques in poultry health and welfare management: A systematic literature review. *Comput. Electron. Agric.* **2022**, *200*, 107266. https://doi.org/10.1016/j.compag.2022.107266.
30. Astill, J.; Dara, R. A.; Fraser, E. D. G.; Roberts, B.; Sharif, S. Smart poultry management: Smart sensors, big data, and the internet of things. *Comput. Electron. Agric.* **2020**, *170*, 105291. https://doi.org/10.1016/j.compag.2020.105291.
31. OpenAI Platform. Models overview. Available online: https://platform.openai.com/docs/models (accessed on February 6, 2025).
32. Akama, S., Ed. *Towards Paraconsistent Engineering*; Springer International Publishing: Cham, 2016. https://doi.org/10.1007/978-3-319-40418-9.
33. DSSAICopilotForPoultryFarming/. Available online: https://github.com/marcusviniciusleite/DSSAICopilotForPoultryFarming/tree/main (accessed on July 23, 2025).
34. Brassó, L. D.; Komlósi, I.; Várszegi, Z. Modern technologies for improving broiler production and welfare: A review. *Animals* **2025**, *15* (4), 493. https://doi.org/10.3390/ani15040493.
35. Tareesh, M. T.; Anandhi, M.; Sujatha, G.; Thiruvenkadan, A. K. Digital twins in poultry farming: A comprehensive review of the smart farming breakthrough transforming efficiency, health, and profitability. *Int. J. Vet. Sci. Anim. Husbandry* **2025**, *10* (8S), 89–95. https://doi.org/10.22271/veterinary.2025.v10.i8Sb.2476.
36. Nääs, I. A.; Garcia, R. G. The dawn of intelligent poultry science: A Brazilian vision for a global future. *Braz. J. Poult. Sci.* **2025**, *27*, eRBCA. https://doi.org/10.1590/1806-9061-2025-2131.
37. Baumhover, A.; Hansen, S. L. Preparing the AI-assisted animal scientist: Faculty and student perspectives on enhancing animal science education with artificial intelligence. *Anim. Front.* **2024**, *14* (6), 54–56. https://doi.org/10.1093/af/vfae038.
38. Ghavi Hossein-Zadeh, N. Artificial intelligence in veterinary and animal science: Applications, challenges, and future prospects. *Comput. Electron. Agric.* **2025**, *235*, 110395. https://doi.org/10.1016/j.compag.2025.110395.
39. eFarm – Horizon OpenAgri. Available online: https://horizon-openagri.eu/open-source-catalogue/demo-efarm/ (accessed on August 31, 2025).
40. farmOS. Available online: https://farmos.org/ (accessed on August 31, 2025).
41. Poultry Farming Management System – Horizon OpenAgri. Available online: https://horizon-openagri.eu/open-source-catalogue/poultry-farming-management-system/ (accessed on August 31, 2025).
42. Poultry Development Review. Available online: https://www.fao.org/4/i3531e/i3531e00.htm (accessed on August 31, 2025).
43. Environmental guidelines for poultry rearing operations. FAOLEX. Available online: https://www.fao.org/faolex/results/details/en/c/LEX-FAOC204579/ (accessed on August 31, 2025).
44. Mellor, D. J. Operational details of the five domains model and its key applications to the assessment and management of animal welfare. *Animals (Basel)* **2017**, *7* (8), 60. https://doi.org/10.3390/ani7080060.
45. Azad, H. K.; Deepak, A. Query expansion techniques for information retrieval: A survey. *Inf. Process. Manag.* **2019**, *56* (5), 1698–1735. https://doi.org/10.1016/j.ipm.2019.05.009.
46. Holtzman, A.; Buys, J.; Du, L.; Forbes, M.; Choi, Y. The curious case of neural text degeneration. *arXiv* **2020**, arXiv:1904.09751. https://doi.org/10.48550/arXiv.1904.09751.
47. Peeperkorn, M.; Kouwenhoven, T.; Brown, D.; Jordanous, A. Is temperature the creativity parameter of large language models? *arXiv* **2024**, arXiv:2405.00492. https://doi.org/10.48550/arXiv.2405.00492.
48. Chen, B.; Zhang, Z.; Langrené, N.; Zhu, S. Unleashing the potential of prompt engineering for large language models. *Patterns* **2025**, *6* (6), 101260. https://doi.org/10.1016/j.patter.2025.101260.

3.4 Cross-Study Discussion

This dissertation, structured in three articles, outlines a research trajectory focused on the overarching objective of building and evaluating a DSS to support decision-making in intensive poultry farming, capable of operating under uncertainty and informational contradiction. The discussion presented not only synthesizes the results but interprets them in light of the current challenges of decision-making in intensive poultry farming that the proposed DSS seeks to address.

Critical mapping of the state of the art revealed these challenges. By typifying existing DSS, their domains of application, and their limitations, the research highlighted a strong focus on short-term operational decisions, with a predominance of real-time monitoring and automation systems (Godinho et al., 2025; Neethirajan,, 2025; Li et al.,, 2020), and little or no attention to the tactical and strategic layers (Zhai et al.,, 2020).

It was also observed that, although poultry farming domains such as Housing and Environmental Control, Animal Nutrition, Poultry Health, Husbandry Practices, and Animal Welfare are interdependent, interoperability between them is virtually absent in the systems evaluated (Hafez; Attia, 2020). This reduces the systemic value of DSS, since their practical effectiveness depends not only on algorithmic sophistication, but also on their ability to integrate with production processes, offer interpretable and actionable recommendations, and align with the actual decision-making context.

This review of the state of the art also highlighted the originality and innovation of this research, which is part of an ongoing conceptual shift marked by the emergence of approaches based on language models and textual decision support (Lewis et al., 2021). These approaches shift the focus from purely numerical inference to knowledge-based support, with an emphasis on generating understandable explanations.

This exploratory phase of the research was presented in Article 1, fulfilling the objective of mapping and systematizing DSS focused on environmental control in poultry farming, and its findings guided the design of the final architecture of the proposed DSS.

Subsequent experimental investigations have shown that the main limitation to the use of language models in intensive poultry farming goes beyond occasional hallucinations (Ji et al., 2023, Metze et al., 2024). The central obstacle lies in contextual specificity, a result of the restrictions imposed by the bodies of knowledge used in LLM training, which compromises their ability to operate in highly specialized domains and hinders the advancement of emerging trends based on these models.

Going beyond this diagnosis, the experiments consolidated in Article 2 advanced to a systematic evaluation of the effect of augmented retrieval on the quality of responses. This stage of the investigation demonstrated measurable gains in stability and semantic alignment between the responses generated by the LLM using the RAG technique and the specialized references (Izacard & Grave, 2021; Li et al., 2022). These results indicated a partial mitigation path for the identified obstacle.

To fully address these limitations, the original LLM-based architecture with RAG is extended by incorporating Logic Eτ. Thus, as presented in Article 3, the methodological proposal integrates three components with distinct and complementary functions. LLM provides the ability to semantically interpret and generate responses in natural language, bridging the gap between the technical vocabulary present in the knowledge base and the way the producer formulates their questions. RAG implements contextual processing by restricting the LLM's reasoning space to a subset of relevant documents, expanding the coverage of specialized knowledge and reducing the variance of responses. Finally, Logic Eτ introduces formal conditions for the use of retrieved knowledge by transforming evidence into explicit degrees of certainty and contradiction.

In this way, the DSS that is the subject of this research was designed as a knowledge-based conversational agent that integrates, in a continuous flow, semantic interpretation, contextual processing, and evidential inference with explicit thresholds of certainty defined experimentally. In such architecture, LLM and RAG are responsible for linguistic and contextual competence, while Logic Eτ governs the decision-making process by determining when the system can affirm, doubt, hesitate, or refuse a response. Article 3 examined this systemic dimension by demonstrating that the DSS incorporates an inference module capable of qualifying the reliability of responses through degrees of certainty and contradiction.

The ability to explicitly deal with thresholds of uncertainty and contradiction distinguishes the proposed DSS from approaches based solely on probability or heuristic confidence thresholds. Logic Eτ allows for the representation of paracomplete states, in which a lack of information predominates, and paraconsistent states, in which favorable and unfavorable evidence coexist in a relevant way, without forcing the system to produce unduly definitive conclusions.

Acting as a guardrail, the use of conservative certainty and contradiction thresholds prevented unclear queries from advancing to the retrieval and generation stages, functioning as a deliberate barrier against the spread of ambiguity throughout the inferential flow. The same mechanism validated the responses produced: the system not only generated an output but also evaluated whether it was acceptable considering the available evidence. When the degree of certainty fell below the threshold, the response was not treated as simply weak; it was flagged as potentially inadequate and communicated to the user as a result of insufficient or contradictory evidence.

These mechanisms translated into concrete behaviors observed in the experiments. When faced with ignorance, the system reported that the question was outside its scope and avoided fabricated answers. When the question was unclear, it requested clarification rather than proceeding with an arbitrary interpretation. When there was high contradiction, it identified conflict between sources and marked the answer as hesitant or conditional, making competing interpretations explicit. In all these situations, inference guided by the Logic Eτ framework prevented decision collapse and transformed hesitation into useful information. Thus, instead of treating poorly formulated questions as noise, the system converted uncertainty into explicit action through cycles of meta-questions guided by propositions.

Like any knowledge-based model, the proposal has limitations. Performance remains dependent on the quality and scope of the corpus used, and robustness in the face of queries outside the scope depends on continuous improvement of curation. Furthermore, the behavior of the system in real operating environments still needs to be investigated, especially regarding the dynamics of use by producers, linguistic variations, and integration with heterogeneous technological infrastructures.

From the perspective of this dissertation, the results allow us to conclude that the system manages uncertainty but does not eliminate it. Even after optimizing the RAG pipeline and integrating Logic Eτ, a minority of responses remain close to ambiguous or contradictory zones. The difference is that these cases are no longer invisible: they become identifiable, traceable, and subject to conscious intervention, either through corpus curation or by adjusting the DSS operational parameters.

Taken together, the results indicate that the overall objective was consistently achieved and that the specific objectives were addressed in a coordinated manner.

CHAPTER IV

4 FINAL CONSIDERATIONS

4.1. Conclusions

The trajectory developed in this dissertation demonstrated that the integration of contextual generation and interpretation, retrieval of verifiable information, and structured representation of uncertainties and contradictions constitutes a consistent way to increase the reliability of decision support systems applied to intensive poultry farming.

For intensive poultry farming, this approach broadens the scope of DSS by shifting the focus from strictly short-term operations and incorporating recommendations aligned with tactical and strategic levels. The ability to translate specialized technical knowledge into understandable responses, while maintaining explicit control over the uncertainties inherent in the production process, reinforces the usefulness of the system in an environment marked by informational complexity and the need for quick and consistent decisions.

The results support practical implications and research developments. From an applied perspective, the dissertation indicates that it is technically feasible to offer poultry producers a conversational agent capable of answering technical questions in multiple domains, with explicit uncertainty and contradiction, provided that a curated knowledge base, fulfilling the general objective proposed for the work.

In intensive production contexts, such DSS can operate as a complementary layer to existing management platforms, offering interpretive narrative, screening recurring questions, and support in reading technical documents, without attempting to replace the judgment of experts.

From a scientific point of view, the research program outlines a concrete agenda that includes improving recovery mechanisms to mitigate situations in which the LLM compromises the quality of responses; rebalancing the knowledge base to increase accuracy in domain classification; systematic comparison with conventional RAG pipelines and classic DSS; and expanding the Logic $E\tau$ framework to other agricultural and industrial sectors.

This brings us to the central contribution of this work, which is to understand decision-making under uncertainty and contradiction not as an unwanted anomaly, but as an ordinary condition of operation that must be modeled, quantified, and made explicit.

4.2. Recommendations for Future Work

For future work, promising directions include conducting field studies, testing with expanded knowledge bases, comparing the approach with other hybrid models, investigating adaptive mechanisms capable of adjusting certainty and contradiction thresholds based on query behavior, and exploring the integration of numerical data, environmental sensors, and textual evidence within unified decision-making cycles.

REFERENCES

ABE, J. M. Fundamentos da lógica anotada. Doutorado em Filosofia, São Paulo: Universidade de São Paulo, 6 maio 1992.

ABE, J. M. Lógica paraconsistente anotada evidencial Et. [s.l.] Comunnicar Editora, 2011.

ABE, J. M. Aspectos de Computação Inteligente Paraconsistente. São Paulo, 2013.

ABE, J. M. Remarks on Paraconsistent Annotated Evidential Logic Eт. Unisanta Science and Technology, v. 3, p. 25-29, 2014.

ABE, J. M. (ED.). Paraconsistent Intelligent-Based Systems: New Trends in the Applications of Paraconsistency. Cham: Springer International Publishing, 2015. v. 94

ABE, Jair Minoro; AKAMA, Seiki; NAKAMATSU, Kazumi. Introduction to Annotated Logics: Foundations for Paracomplete and Paraconsistent Reasoning. Cham: Springer International Publishing, 2015. v. 88

ABE, J. M.; Carvalho, F. R. A Paraconsistent Decision-Making Method. Cham: Springer, 2018.

ABPA. Report 2023. [s.l.] Brazilian Association of Animal Protein, 2023. Disponível em: <https://abpa-br.org/wp-content/uploads/2023/04/ABPA.-Annual-Report-2023..pdf>.

ABPA. Anual Report 2024. [s.l.] Brazilian Association of Animal Protein, 2024. Disponível em: <https://abpa-br.org/wp-content/uploads/2024/04/RA_2024_ABPA_ingles_avicultura.pdf>.

ALMA'AITAH, Wafa' Za'al et al. Integration Approaches for Heterogeneous Big Data: A Survey. Cybernetics and Information Technologies, v. 24, n. 1, p. 3–20, 1 mar. 2024.

AKAMA, Seiki (ORG.). Towards Paraconsistent Engineering. Cham: Springer International Publishing, 2016. v. 110

AKAMA, Seiki; DA COSTA, Newton C. A. Why Paraconsistent Logics? In: AKAMA, Seiki (Org.). Towards Paraconsistent Engineering. Cham: Springer International Publishing, 2016. p. 7–24.

AKAMA, Seiki. A Survey of Annotated Logics. In: AKAMA, Seiki (Org.). Towards Paraconsistent Engineering. Cham: Springer International Publishing, 2016. p. 49–76.

ALMEIDA, L.; SILVA, J.; SANTOS, R. Automated Feeding Systems in Poultry Farming: Benefits and Challenges. Journal of Agricultural Engineering, p. 321–329, 2019.

ASTILL, Jake et al. Smart poultry management: Smart sensors, big data, and the internet of things. Computers and Electronics in Agriculture, v. 170, p. 105291, 1 mar. 2020.

AWAN, U. et al. Artificial Intelligence for Supply Chain Success in the Era of Data Analytics. Em: HAMDAN, A. et al. (Eds.). The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success. Studies in Computational Intelligence. Cham: Springer International Publishing, 2021. v. 935p. 3–21.

BAO, J.; XIE, Q. Artificial intelligence in animal farming: A systematic literature review. Journal of Cleaner Production, v. 331, p. 129956, 10 jan. 2022.

BAI, Yuntao et al. Constitutional AI: Harmlessness from AI Feedback. arXiv, , 15 dez. 2022. Disponível em: <http://arxiv.org/abs/2212.08073>. Acesso em: 7 out. 2025

BENGIO, Yoshua et al. A Neural Probabilistic Language Model. [S.d.].

BERCKMANS, D. Automatic On-line Monitoring of animal Health and welfare by Precision livestock farming. 2009.

BERCKMANS, D. General introduction to precision livestock farming. Animal Frontiers, v. 7, n. 1, p. 6–11, 1 jan. 2017.

BOBAN, I.; Doko, A.; Gotovac, S. Sentence retrieval using stemming and lemmatization with different length of queries. Advances in Science, Technology and Engineering Systems Journal, v. 5, p. 349-354, 2020.

BRENNAN, L.; SOLOMON, M. Emerging technologies in agriculture: Barriers and opportunities.Journal of Agricultural Technology, n. 1, p. 14–29, 2008.

BROWN, Tom B. et al. Language Models are Few-Shot Learners. arXiv, , 22 jul. 2020. Disponível em: <http://arxiv.org/abs/2005.14165>. Acesso em: 17 nov. 2024

BRUGLER, Skye et al. Improving decision support systems with machine learning: Identifying barriers to adoption. Agronomy Journal, v. 116, n. 3, p. 1229–1236, maio 2024.

CARVALHO, F. R. D. Logica Paraconcistene Anotada Em Tomadas De Decisao. [s.l.] Editora Aleph Ltda, 2002.

CARVALHO, F. R. D.; ABE, J. M. A Paraconsistent Decision-Making Method. Cham: Springer International Publishing, 2018. v. 87

CHENG, Muxi; MCCARL, Bruce; FEI, Chengcheng. Climate Change and Livestock Production: A Literature Review. Atmosphere, v. 13, n. 1, p. 140, jan. 2022.

CHRISTIANO, Paul et al. Deep reinforcement learning from human preferences. arXiv, , 17 fev. 2023. Disponível em: <http://arxiv.org/abs/1706.03741>. Acesso em: 7 out. 2025

COSTA, M.; OLIVEIRA, P.; SOUSA, F. Robotic Automation in Poultry Farming: Impacts on Animal Welfare and Labor Conditions. Poultry Science, v, v. 100, p. 1012–1020, 2021.

CSASZAR, Felipe A.; KETKAR, Harsh; KIM, Hyunjin. Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors. Strategy Science, v. 9, n. 4, p. 322–345, dez. 2024.

CURI, T. M. R. C. et al. Positioning of sensors for control of ventilation systems in broiler houses: A case study. Scientia Agricola, v. 74, n. 2, p. 101–109, 2017.

CURI, T. M. R. D. C.Avaliação Do Sistema De Ventilação Em Instalações Para Frangos De Corte Por Meio De Modelos. Tese (Doutorado em Construções Rurais e Ambiência) -Faculdade de Engenharia Agrícola. [s.l: s.n.].

DAUGHERTY, P. R.; WILSON, H. J. Human + Machine: Reimagining Work in the Age of AI. 2018.

DE CARVALHO JUNIOR, Arnaldo et al. A comprehensive review on paraconsistent annotated evidential logic: Algorithms, Applications, and Perspectives. Engineering Applications of Artificial Intelligence, v. 127, p. 107342, 1 jan. 2024.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint, arXiv:1810.04805, 2018.

DEWANTO, Puspo; MUNADI, M.; TAUVIQIRRAHMAN, Mohammad. Development of an Automatic Broiler Feeding System using PLC and HMI for Closed House System. v. 58, n. 1, 2019.

DIAO, C.; SUN, D.; LI, Y. A Survey on Internet of Things for Smart Agriculture: Opportunities and Challenges. IEEE Communications Surveys & Tutorials, v, v. 22, n. 3, p. 1826–1853, 2020.

DONG, Qingxiu et al. A Survey on In-context Learning. arXiv, , 2023. Disponível em: <https://arxiv.org/abs/2301.00234>. Acesso em: 7 out. 2025

DOSHI-VELEZ, F.; Kim, B. Towards a rigorous science of interpretable machine learning. arXiv preprint, arXiv:1702.08608, 2017.

DHULIAWALA, Shehzaad et al. Chain-of-Verification Reduces Hallucination in Large Language Models. Findings of the Association for Computational Linguistics ACL 2024. Bangkok, Thailand and virtual meeting: Association for Computational Linguistics, 2024. Disponível em: <https://aclanthology.org/2024.findings-acl.212>. Acesso em: 7 out. 2025

DUAN, Y.; EDWARDS, J. S.; DWIVEDI, Y. K. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, v. 48, p. 63-71, 2019. DOI: 10.1016/j.ijinfomgt.2019.01.021.

ELKADY, Sahar; HERNANTES, Josune; LABAKA, Leire. Decision-making for community resilience: A review of decision support systems and their applications. Heliyon, v. 10, n. 12, p. e33116, 30 jun. 2024.

FAO. Food and balance sheets 2010–2022. [S.l.]: FAO, 19 jul. 2024. Disponível em: <https://openknowledge.fao.org/handle/20.500.14283/cd1656en>. Acesso em: 6 out. 2025.

FAOSTAT. Disponível em: <https://www.fao.org/faostat/en/#data/FBS>. Acesso em: 6 out. 2025.

FARQUHAR, Sebastian et al. Detecting hallucinations in large language models using semantic entropy. Nature, v. 630, n. 8017, p. 625–630, jun. 2024.

FAO. Food balance sheets 2010–2022. Global, regional and country trends. Disponível em: <https://www.fao.org/statistics/highlights-archive/highlights-detail/food-balance-sheets-2010-2022-global-regional-and-country-trends/en?utm_source=chatgpt.com>. Acesso em: 6 out. 2025.

FOUNTAS, S. et al. Farm management information systems: Current situation and future perspectives. Computers and Electronics in Agriculture, v. 115, p. 40–50, 1 jul. 2015.

GAO, Luyu et al. PAL: Program-aided Language Models. arXiv, , 27 jan. 2023. Disponível em: <http://arxiv.org/abs/2211.10435>. Acesso em: 7 out. 2025

GARCEZ, Artur d'Avila; LAMB, Luis C. Neurosymbolic AI: The 3rd Wave. arXiv, , 16 dez. 2020. Disponível em: <http://arxiv.org/abs/2012.05876>. Acesso em: 7 out. 2025

GODINHO, A.; Vicente, R.; Silva, S.; Coelho, P. J. Wireless environmental monitoring and control in poultry houses: a conceptual study. IoT, v. 6, n. 2, p. 32, 2025.

GÓMEZ, D.; PRENDES, A.; VALCARCEL, R. Internet of Things (IoT) for Precision Livestock Farming: A Review of Applications and Challenges. Sensors, [s.d.].

GRŽINIĆ, G. et al. Intensive poultry farming: A review of the impact on the environment and human health. Science of The Total Environment, v. 858, p. 160014, 1 fev. 2023.

GUO, J. et al. A deep dive into neural ranking models for information retrieval. Information Processing & Management, v. 57, n. 6, p. 102067, 2022.

HAFEZ, H. M.; ATTIA, Y. A. Challenges to the Poultry Industry: Current Perspectives and Strategic Future After the COVID-19 Outbreak. Frontiers in Veterinary Science, v. 7, 26 ago. 2020.

HALACHMI, I. et al. Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production. Annual Review of Animal Biosciences, v. 7, n. 1, p. 403–425, 15 fev. 2019.

HAMSA, K. R.; BELLUNDAGI, Veerabhadrappa. Review on Decision-making under Risk and Uncertainty in Agriculture. Economic Affairs, v. 62, n. 3, p. 447, 2017.

INACIO DA SILVA FILHO, Joao; ABE, Jair Minoro; TORRES, Germano Lambert. Inteligência Artificial Com as Redes De Análises Paraconsistentes: Teoria E Aplicações. [S.l.]: Ltc-Livros Tecnicos E Cientificos Editora Lda, 2008.

IZACARD, Gautier et al. Atlas: Few-shot Learning with Retrieval Augmented Language Models. arXiv, , 16 nov. 2022. Disponível em: <http://arxiv.org/abs/2208.03299>. Acesso em: 7 out. 2025

JACKSON, P. Introduction to Expert Systems. Harlow; Reading; Menlo Park: Addison-Wesley, 1999.

JI, Z. et al. A survey of hallucination in natural language generation. ACM Computing Surveys, v. 55, p. 1–38, 2022.

KAMILARIS, Andreas; PRENAFETA-BOLDÚ, Francesc X. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, v. 147, p. 70–90, 1 abr. 2018.

KARPUKHIN, Vladimir et al. Dense Passage Retrieval for Open-Domain Question Answering. arXiv, , 30 set. 2020. Disponível em: <http://arxiv.org/abs/2004.04906>. Acesso em: 7 out. 2025

KITCHENHAM, B. Procedures for performing systematic reviews. Keele University Technical Report, v. 33, 2004.

KOMEILI, Mojtaba; SHUSTER, Kurt; WESTON, Jason. Internet-Augmented Dialogue Generation. arXiv, , 15 jul. 2021. Disponível em: <http://arxiv.org/abs/2107.07566>. Acesso em: 7 out. 2025

LAZARIDOU, Angeliki et al. Internet-augmented language models through few-shot prompting for open-domain question answering. arXiv, , 23 maio 2022. Disponível em: <http://arxiv.org/abs/2203.05115>. Acesso em: 7 out. 2025

LASHARI, Mohammed Hanif et al. IoT Based Poultry Environment Monitoring System. IEEE International Conference on Internet of Things and Intelligence System (IOTAIS). nov. 2018. Disponível em: <https://ieeexplore-ieee-org.ez346.periodicos.capes.gov.br/document/8600837>. Acesso em: 29 jul. 2024

LEITE, M. V.; Abe, J. M.; Souza, M. L. H.; Nääs, I. A. A Decision Support AI-Copilot for Poultry Farming: Leveraging Retrieval-Augmented LLMs and Paraconsistent Annotated Evidential Logic Eτ to Enhance Operational Decisions. AgriEngineering, v. 7, 2025.

LEWIS, Patrick et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv, , 12 abr. 2021. Disponível em: <http://arxiv.org/abs/2005.11401>. Acesso em: 14 jul. 2025

LIAKOS, Konstantinos G. et al. Machine Learning in Agriculture: A Review. Sensors, v. 18, n. 8, p. 2674, ago. 2018.

LI, Huayang et al. A Survey on Retrieval-Augmented Text Generation. arXiv, , 13 fev. 2022. Disponível em: <http://arxiv.org/abs/2202.01110>. Acesso em: 17 nov. 2024

LI, N.; Ren, Z.; Li, D.; Zeng, L. Automated techniques for monitoring the behaviour and welfare of broilers and laying hens. Animal, v. 14, n. 3, p. 617–625, 2020.

LIANG, Percy et al. Holistic Evaluation of Language Models. arXiv, , 1 out. 2023. Disponível em: <http://arxiv.org/abs/2211.09110>. Acesso em: 7 out. 2025

LIN, Stephanie; HILTON, Jacob; EVANS, Owain. TruthfulQA: Measuring How Models Mimic Human Falsehoods. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.acl-long.229>. Acesso em: 7 out. 2025

LUGER, G. F. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 4. ed. Harlow, Essex, England, 2002.

MANHAEVE, Robin et al. DeepProbLog: Neural Probabilistic Logic Programming. arXiv, , 12 dez. 2018. Disponível em: <http://arxiv.org/abs/1805.10872>. Acesso em: 7 out. 2025

MARCUSVINICIUSLEITE. DSS-EnvironmentControl-PoultryFarming. GitHub Repository. Disponível em: https://github.com/marcusviniciusleite/DSS-EnvironmentControl-PoultryFarming. Acesso em: 10 jul. 2025.

MARTINEZ, A. A. G. et al. A heuristic and data mining model for predicting broiler house environment suitability. Animals, v. 11, n. 10, 2021.

MCAFEE, A.; BRYNJOLFSSON, E. Big data: the management revolution. Harvard business review, v. 90, n. 10, p. 60–6, 68, 128, 2012.

MANDI, Jayanta; KOTARY, James; BERDEN, Senne; et al. Decision-Focused Learning: Foundations, State of the Art, Benchmark and Future Opportunities. Journal of Artificial Intelligence Research, v. 80, p. 1623–1701, 2024.

METZE, K. et al. Bibliographic research with ChatGPT may be misleading: the problem of hallucination. Journal of Pediatric Surgery, v. 59, p. 158, 2024.

MICHALIEWICZ, Z. Genetic algorithms + data structures = evolution programs. Springer Science & Business Media, 2013.

MILLER, D. D.; BROWN, E. W. Artificial intelligence in medical practice: the question to the answer? American Journal of Medicine, v. 131, p. 129-133, 2018.

MIKOLOV, Tomas et al. Efficient Estimation of Word Representations in Vector Space. arXiv, , 2013. Disponível em: <https://arxiv.org/abs/1301.3781>. Acesso em: 7 out. 2025

MOTTET, A.; TEMPIO, G. Global poultry production: current state and future outlook and challenges. World’s Poultry Science Journal, v. 73, n. 2, p. 245–256, 1 jun. 2017.

NEETHIRAJAN, S. Rethinking poultry welfare, integrating behavioral science and digital innovations. Poultry, v. 4, n. 2, p. 20, 2025.

NORTON, T. et al. Review: Precision livestock farming: building ‘digital representations’ to bring the animals closer to the farmer. Animal, v. 13, n. 12, p. 3009–3017, 1 jan. 2019.

OJO, R. O. et al. Internet of Things and Machine Learning techniques in poultry health and welfare management: A systematic literature review. Computers and Electronics in Agriculture, v. 200, p. 107266, 1 set. 2022.

OPENAI et al. GPT-4 technical report. arXiv preprint, arXiv:2303.08774, 2024.

PAWLAK, Z. Rough sets. International Journal of Computer and Information Sciences, v. 11, p. 341-356, 1982.

PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, v. 372, p. n71, 2021.

PRAMANA, R.; Debora; Subroto, J. J.; Gunawan, A. A. S.; Anderies. Systematic literature review of stemming and lemmatization performance for sentence similarity. In: IEEE 7th International Conference on Information Technology and Digital Applications – ICITDA 2022. Yogyakarta, 2022.

PRATT, Lorien; BISSON, Christophe; WARIN, Thierry. Bringing advanced technology to strategic decision-making: The Decision Intelligence/Data Science (DI/DS) Integration framework. Futures, v. 152, p. 103217, 1 set. 2023.

Qi, F. et al. Environmental factor detection and analysis technologies in livestock and poultry houses: a review. Agriculture, v. 13, n. 8, p. 1489, 2023.

RAIKOV, Alexander; ABROSIMOV, Vyacheslav. Artificial Intelligence and Robots in Agriculture. In: 2022 15TH INTERNATIONAL CONFERENCE MANAGEMENT OF LARGE-SCALE SYSTEM DEVELOPMENT (MLSD). 2022 15th International Conference Management of large-scale system development (MLSD). set. 2022. Disponível em: <https://ieeexplore.ieee.org/document/9934170>. Acesso em: 6 out. 2025

RASHKIN, Hannah et al. Measuring Attribution in Natural Language Generation Models. arXiv, , 2 ago. 2022. Disponível em: <http://arxiv.org/abs/2112.12870>. Acesso em: 7 out. 2025

RASHKIN, Hannah et al. Measuring Attribution in Natural Language Generation Models. Computational Linguistics, v. 49, n. 4, p. 777–840, 1 dez. 2023.

Reimers, N.; Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. arXiv preprint, arXiv:1908.10084, 2019.

RIEGEL, Ryan et al. Logical Neural Networks. arXiv, , 23 jun. 2020. Disponível em: <http://arxiv.org/abs/2006.13155>. Acesso em: 7 out. 2025

ROSSI, Vittorio; CAFFI, Tito; SALINARI, Francesca. Helping farmers face the increasing complexity of decision-making for crop protection. Phytopathologia Mediterranea, v. 51, n. 3, p. 457–479, 2 nov. 2012.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, v. 44, p. 206-226, 2000.

SARHANI, M.; EL AFIA, A. Intelligent system based support vector regression for supply chain demand forecasting. In: 2014 Second World Conference on Complex Systems (WCCS). p. 79-83, 2014.

SILVA FILHO, J. I. DA; ABE, J. M.; TORRES, G. L. Inteligência Artificial Com as Redes De Análises Paraconsistentes: Teoria E Aplicações. [s.l.] Ltc-Livros Tecnicos E Cientificos Editora Lda, 2008.

SILVA FILHO, J. I. DA; ABE, J. M. Métodos de aplicações da lógica paraconsistente anotada de anotação com dois valores-LPA2v com construção de algoritmo e implementação de circuitos eletrônicos. [s.l: s.n.].

SILVA, M.; RIBEIRO, B.; COSTA, E. Egg Collection Robots: Enhancing Efficiency and Hygiene in Poultry Production. Computers and Electronics in Agriculture, v. 150, p. 163–171, 2018.

STOREY, Veda C.; HEVNER, Alan R.; YOON, Victoria Y. The design of human-artificial intelligence systems in decision sciences: A look back and directions forward. Decision Support Systems, v. 182, p. 114230, 1 jul. 2024.

TANAKA, K. An introduction to fuzzy logic for practical applications. New York: Springer, 1997.

THAKUR, Nandan et al. BEIR: A Heterogenous Benchmark for Zero-shot Evaluation of Information Retrieval Models. arXiv, , 21 out. 2021. Disponível em: <http://arxiv.org/abs/2104.08663>. Acesso em: 7 out. 2025

UNITED NATIONS GENERAL ASSEMBLY. Transforming our World: The 2030 Agenda for Sustainable Development, 2015. Disponível em: <https://www.unfpa.org/resources/transforming-our-world-2030-agenda-sustainable-development>. Acesso em: 31 jul. 2024.

VASWANI, Ashish et al. Attention Is All You Need. arXiv, , 2017. Disponível em: <https://arxiv.org/abs/1706.03762>. Acesso em: 7 out. 2025

WANG, H.; Zhang, D.; Li, J.; Feng, Z.; Zhang, F. Entropy-optimized dynamic text segmentation and RAG-enhanced LLMs for construction engineering knowledge base. Applied Sciences, v. 15, p. 3134, 2025.

WANG, Xuezhi et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv, , 7 mar. 2023. Disponível em: <http://arxiv.org/abs/2203.11171>. Acesso em: 7 out. 2025

WOLFERT, Sjaak et al. Big Data in Smart Farming – A review. Agricultural Systems, v. 153, p. 69–80, 1 maio 2017.

WILCOX, C. H. et al. A literature review of broiler chicken welfare, husbandry, and assessment. World’s Poultry Science Journal, v. 80, n. 1, p. 3–32, 2 jan. 2024.

WU, B. et al. An RFID-Assisted Smart Livestock and Poultry Farming System on the Cloud. 2023 15th International Conference on Computer and Automation Engineering (ICCAE). Anais... Em: 2023 15TH INTERNATIONAL CONFERENCE ON COMPUTER AND AUTOMATION ENGINEERING (ICCAE). Sydney, Australia: IEEE, 3 mar. 2023. Disponível em: <https://ieeexplore.ieee.org/document/10111202/>. Acesso em: 28 jun. 2024

YAO, Shunyu et al. ReAct: Synergizing Reasoning and Acting in Language Models. arXiv, , 10 mar. 2023. Disponível em: <http://arxiv.org/abs/2210.03629>. Acesso em: 7 out. 2025

ZHAI, Zhaoyu et al. Decision support systems for agriculture 4.0: Survey and challenges. Computers and Electronics in Agriculture, v. 170, p. 105256, mar. 2020.
ZHANG, Yue et al. Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models. arXiv, , 14 set. 2025. Disponível em: <http://arxiv.org/abs/2309.01219>. Acesso em: 7 out. 2025

ZHANG, Tianyi et al. BERTScore: Evaluating Text Generation with BERT. arXiv, , 24 fev. 2020. Disponível em: <http://arxiv.org/abs/1904.09675>. Acesso em: 7 out. 2025

ZHAO, Haiyan et al. Explainability for Large Language Models: A Survey. ACM Trans. Intell. Syst. Technol., v. 15, n. 2, p. 20:1-20:38, 22 fev. 2024.

ZHENG, H. et al. Design and Implementation of Poultry Farming Information Management System Based on Cloud Database. Animals, v. 11, n. 3, p. 900, mar. 2021.