8th Conference on Production Systems and Logistics

Reinforcement Learning for Intercropping Sequences: Transferable Decision Policies for Circular Industrial Production Systems

Stefan Sarawinsky, Kai Schaller, Kajan Kandiah
Institute for Industrial Managament (FIR) at RWTH Aachen University, Aachen, Germany

Abstract

Improving the sustainability of production systems requires decision support methods that can solve the problem of optimizing the combination of resources under ecological, operational, and compatibility constraints [1]. This challenge arises in agriculture and manufacturing, particularly in Circular Industrial Production Systems (CIPS), where the creation of value depends on the coordinated use of materials, processes and resources, while minimizing waste and environmental impact [2]. As production environments become more complex and dynamic, traditional rule-based or static optimization approaches struggle to cope with combinatorial decision spaces and evolving system constraints [3]. To address this issue, the authors use intercropping as an example proxy domain with high complexity that mirrors these combinatorial challenges and compatibility constraints. This paper therefore presents a Deep Q-Network (DQN)-based Reinforcement Learning (RL) approach embedded in a domain-specific simulation environment for planning intercropping sequences. The RL-based Decision Support System (DSS) learns to maximize an aggregated reward comprising agronomic quality indicators and binary user preference feedback. Results from a simulation demonstrate a consistent convergence and structured long-term policy learning in complex combinatorial decision spaces. Beyond agriculture, this framework shows how RL-based sequencing policies can help CIPS facing similar optimization problems involving material combinations, process compatibility, and sustainability trade-offs.

Keywords

Reinforcement Learning; Circular Economy; Decision Support Systems; Sustainable Sequence Optimization

1. Introduction

In agriculture, intercropping is a strategy that enhances soil health, resource efficiency and ecological resilience by combining different plant species within a cultivation system [4]. From a decision-theoretical perspective, intercropping constitutes a highly complex optimization problem [5]. Its success depends on non-linear interactions between plant species, soil properties, seasonal planting windows and operational constraints [5]. A similar level of complexity exists in Circular Industrial Production Systems (CIPS), where planners must identify combinations of materials, processing steps and resource allocations that are compatible with sustainability and cost constraints [6].

In practice, intercropping and planning are not yet widely adopted. One reason for this is the complexity of the planning process based on challenges such as the limited mechanization of work steps and complex interactions between plants [7]. Furthermore, mixed and monoculture planning is still largely driven by expert knowledge, look-up tables, and heuristic rules [8,9]. While these approaches encode valuable domain expertise, they are inherently static and limited in their ability to adapt to new conditions or information, or to incorporate feedback from system operators. Similar limitations are evident in circular industrial

DOI: https://doi.org/10.15488/20994
ISSN: 2701-6277

production planning, where predefined rules and optimization heuristics often fail to capture dynamic interactions and long-term effects of sequential decisions [10].

Consequently, there is a growing need for adaptive, learning-based decision-making methods that can optimize sequences of independent choices over time while incorporating user feedback. Although LLMs excel in semantic reasoning, their inherent stochasticity and challenges in handling delayed rewards mean they are less well-suited than Reinforcement Learning (RL) to reproducible, long-term sequencing problems. RL explicitly optimizes cumulative policies [11,12,13,14]. RL addresses sequential decision-making directly by learning policies that maximize cumulative reward through interaction with the environment. It is particularly suited to problems characterized by multi-step dependencies, uncertainty and delayed feedback [15].

This research aims to evaluate the effectiveness of RL in managing the combinatorial complexity of intercropping as a basis for industrial applications. To this end, we address the following research questions:

RQ1: To what extent can a DQN-based agent internalize and optimize intercropping sequences under heterogeneous agronomic constraints without explicit rule-based programming?

RQ2: How can the structural properties of agricultural intercropping be mapped to CIPS to ensure methodological transferability?

Accordingly, the focus lies on analyzing learning convergence and structural transferability rather than statistical performance benchmarking.

2. Related Work

2.1 Reinforcement Learning in Agriculture

RL has gained increasing attention in the field of agriculture decision support [16,17], particularly in areas such as irrigation management [18,19], fertilization, and crop scheduling [18]. Sutton and Barto established the foundational theoretical framework for RL, setting out the principles of value-based learning, policy optimization, and temporal difference methods that underpin modern RL algorithms [20]. Chen et al. demonstrated the effectiveness of Deep Q-Network (DQN) in irrigation decision-making under uncertain weather conditions, achieving better results than rule-based methods [21]. Gautron et al. provide a thorough overview of RL applications in crop management, emphasizing the suitability for handling delayed rewards and complex interactions while highlighting challenges related to data availability and interpretability [22].

Despite the growing body of work, RL research in agriculture has predominantly focused on single-variable decision problems (e.g. irrigation timing) and monoculture-based systems [18,19]. The optimization of intercropping sequences involving discrete combinatorial decisions and compatibility constraints remains under-explored. Furthermore, there are no studies that explicitly address the transferability of agricultural RL methods for intercrop sequence planning to the context of circular industrial production.

2.2 Decision Support in Circular Industrial Production Systems

In manufacturing, DSS traditionally rely on mathematical programming, heuristics, or simulation-based optimization to address production planning and control problems [23]. However, with the increasing importance of sustainability and circular economy principles, production systems are becoming more complex due to material recyclability, process compatibility, and lifecycle constraints [24,25]. These challenges have motivated research into artificial intelligence (AI)-based decision support, including simulation-driven optimization and machine learning approaches such as RL [26].

Many manufacturing decision-support solutions focus on static optimization or supervised learning methods, which require labeled datasets and struggle to adapt to new operating conditions (e.g. supply chain

disruptions) [26,27]. Sequential decision problems, such as process routing or material sequence optimization, are often addressed through heuristics (e.g. First-Come-First-Served) rather than learning-based approaches [28,29]. This creates a methodological gap that RL is well-positioned to fill, particularly when combined with simulation environments analogous to digital twins.

2.3 Explainability and Human-in-the-Loop Learning

Explainability is a critical requirement for DSS within intercropping practices and circular industrial production. Explainable AI (XAI) methods aim to make model decisions transparent to users. In RL, interpretability can be enhanced through explicit reward decomposition or policy visualization [30]. Compared to end-to-end deep learning, value-based RL approaches such as DQN offer greater opportunities for stepwise explanation of decisions, which is essential for user acceptance [31].

Human-in-the-loop (HITL) learning, where user feedback is incorporated during training, has been shown to improve the alignment of automated decisions with operational preferences [31]. In DQN, reward shaping provides a structured mechanism to embed such feedback into RL systems without compromising learning stability. Despite its potential, this approach remains underutilized in joint agricultural and manufacturing decision support research.

3. Research Methodology

3.1 Problem Abstraction

The intercropping planning problem is formulated as a Markov Decision Process (MDP) defined by the tuple (S, A, R, γ) where S denotes the state space, A the action space, R the reward function, and γ the discount factor. In detail the model abstraction components are:

- **State Space (S):** A state s_t represents a partially filled planting row, encoded by the sequence of crops already placed, the soil type, planting month, and row length.
- **Action Space (A):** An action a_t corresponds to selecting the next crop to be placed at the current position from the available catalog.
- **Implicit Time-Step $(t, t+1)$:** The time index t represents an implicit decision step. Each step corresponds to the placement of one crop in the planting row. $t+1$ denotes the subsequent decision state following this placement.
- **Episode:** An episode ends when the planting row has been completely assigned.

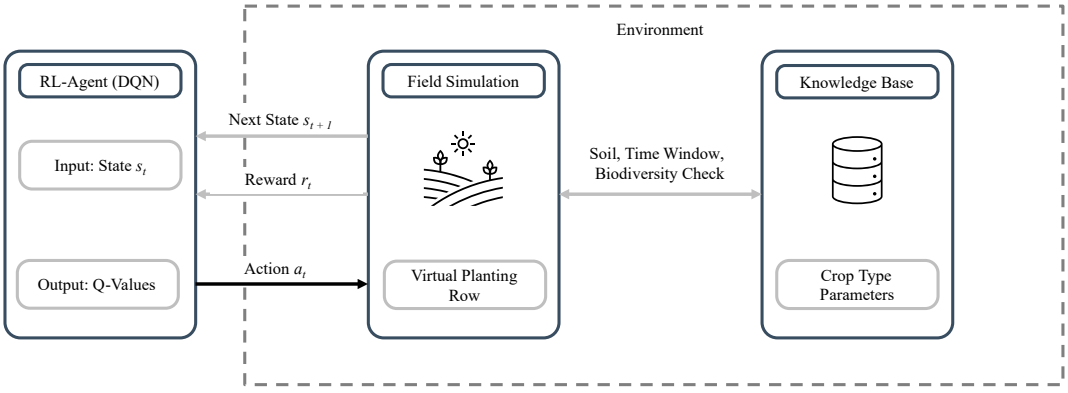

Figure 1: Intercropping approach based on MDP

Figure 1 illustrates the RL loop, in which a DQN agent makes a sequential planting decision based on the current field state s_t. The selected action (e.g. crop index 55: zucchini) is passed to the field simulation in

the environment, where it is validated against an expert knowledge base derived from the agricultural dataset. Within this simulation, specialized filter modules evaluate specific data attributes, such as soil compatibility, seasonal planting windows, and biological synergies, such as nutrient demand. The environment then returns a composite reward r_t, which aggregates agronomic scores for soil and neighborhood effects. It also returns an updated state vector s_{t+1}, which represents the growing planting row. This iterative process enables the agent to learn an optimized sequencing policy (Q-Values) that balances complex operational and ecological constraints effectively.

Simulation Environment

The environment incorporates deterministic expert knowledge through structured data attributes such as soil suitability, planting windows, pH tolerance, nutrient requirements, light demand, pesticide infestation and crop family affiliation. These attributes are used to compute interpretable sub-rewards reflecting agronomic feasibility and diversity.

Reward Structure

The reward function is defined as an additive composition:

$$R_{\text{total}} = R_{\text{agronomic}} + \alpha \cdot R_{\text{user}} \quad (1)$$

where $R_{\text{agronomic}}$ aggregates domain-specific sub-scores (soil compatibility, temporal suitability, nutrient/ light diversity), and R_{user} represents a learned user preference signal derived from binary feedback. The scaling factor α controls the influence of user preferences on the overall optimization objective.

Deep Q-Network Architecture

A DQN is used to approximate the action-value function $Q(s, a)$. The network takes the state vector as input and outputs Q-values for all admissible actions. Training follows standard DQN techniques including experience replay, ε-greedy exploration and temporal-difference learning, as described by Mnih et al. [32].

User Feedback and Reward Shaping

Binary user feedback (acceptable/not acceptable) is collected after each proposed sequence and used to train a lightweight preference model. The output of this model is mapped to a bounded reward adjustment and integrated via online reward shaping. This enables continuous adaptation of the policy to farm specific operational preferences while preserving the underlying agronomic constraints.

3.2 Transferability to Circular Industrial Production Systems

From a methodological perspective, the defined MDP maps directly to CIPS:

Table 1: MDP to CIPS mapping

Intercropping	Circular Industry	Mapping
Crops	Materials	Biological constraints with material compatibility constraints (e.g. in recycling or disassembly)
Planting Rows	Production Sequences	Sequence of crops effects e.g. soil health; the sequence of manufacturing steps affects energy usage and waste generation
Agronomic Constraints	Process Feasibility	pH/Light requirements map to machine capabilities and temperature process windows

As shown in Table 1, crops correspond to material options, planting rows to production sequences, and agronomic constraints to process feasibility in CIPS, regarding compatibility and sustainability requirements. The learned RL policy thus provides a transferable optimization paradigm for sequential decision problems, particularly where combinatorial complexity and sustainability criteria dominate.

4. Experimental Setup

The proposed approach was implemented in Python using a custom OpenAI Gym environment which models a single planting row as a sequential decision problem. The state of the environment is represented as an integer vector of 10 (the maximum length of row), where each position either encodes an empty slot or the index of a plant species from the dataset. The environment models intercropping planning as a finite-horizon decision task, in which a RL agent selects crops incrementally to form a planting sequence, subject to agronomic and operational constraints.

The agronomic knowledge base is stored in a lightweight SQL database containing information on vegetable crops and their attributes. This includes soil suitability scores for three soil types (sand, loam and clay), planting windows (month from/to), pH tolerance ranges, light and nutrient requirements, botanical family and indices for sowing and harvesting techniques. These attributes are used to compute an additive reward comprising (i) soil suitability, (ii) temporal alignment with the planting windows, and (iii) neighborhood effects with respect to the previous plant in the row (light and nutrient complementarity, pH interval overlap, botanical diversity, and process compatibility for sowing and harvesting). Global and local duplicate penalties discourage the repeated use of the same crop.

Each episode represents the planning of a single planting row between 2 and 10 positions long. For each episode, the planting month and soil type are randomly selected. The state is defined as a fixed-length vector encoding the partially filled row, where unassigned positions are padded. The action space is therefore discrete, corresponding to the selection of a crop species from those available at each position.

As mentioned, the reward function is additive and combines multiple agronomic criteria. Duplicate crops within a row incur a penalty. All reward components are logged separately to enable post-hoc explainability.

In addition to expert-based rewards, a lightweight preference model can be activated to incorporate binary user feedback. This model predicts a preference score for individual placement decisions based on contextual features (crop, neighbor, soil and month position), which is mapped to a bounded bonus and added via reward shaping. For the core experiments reported here, the focus is on the base agronomic reward to analyze fundamental learning behavior.

This is how the RL agent and training were formed. The decision policy is learnt using a DQN with two fully connected hidden layers (64 neurons each). Experience replay and a target network with soft updates are employed for stabilization. Training was performed over 15.000 episodes using an ε-greedy exploration strategy that gradually shifted towards exploitation. The total episodic reward is recorded after each episode and serves as the primary performance indicator.

The performance of the RL agent is evaluated based on its behavior as it converges within the complex decision space. The initial training phase (exploration) establishes a foundation for uninformed decision-making. The resulting policy is analyzed in terms of its long-term reward structure and decision-making patterns across different scenarios.

5. Results

The learning behavior of the RL agent based on DQN was evaluated by analyzing the evolution of the total episodic reward over training time.

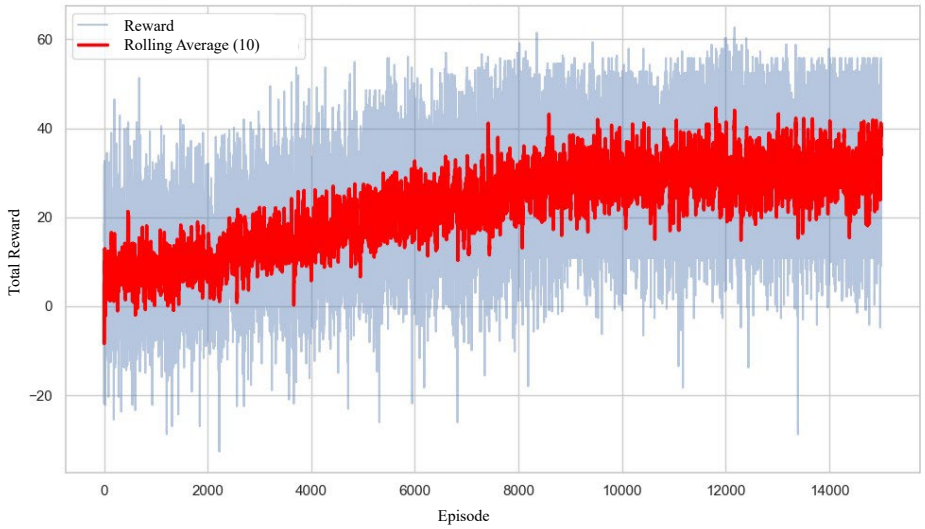

Figure 2: Evolution of episodic rewards during DQN training.

Figure 2 shows the progression of rewards across episodes, displaying both raw episodic rewards and a rolling average over 10 episodes. The rolling mean highlights convergence behavior. Initially, the agent exhibits highly variable rewards, reflecting random exploration and the absence of an informed policy. As training progresses, the rolling average steadily increases, indicating that the agent is learning to exploit the structured agronomic signals embedded in the reward function. After approximately 8.000-10.000 episodes, the reward trend stabilizes, suggesting convergence towards a consistent policy.

Qualitative inspection of sequences generated confirms that the agent learns to avoid planting duplicate crops, to respect planting windows and optimize neighbor compatibility (e.g. by pairing deep-rooting plants with shallow-rooting plants). However, the variance in raw rewards remains relatively high even in later episodes, due to the stochastic sampling of months, soil types, and row lengths. Nevertheless, the monotonic increase in the rolling average demonstrates robust learning progress across heterogeneous scenarios rather than overfitting to a fixed configuration.

The results demonstrate that the DQN agent successfully internalizes multi-criteria agronomic constraints. The steady increase from an initial low reward (representing random exploration) to a stable plateau confirms that the agent learns a policy that demonstrates structured policy learning beyond random exploration. Qualitative analysis confirms that the agent considers long-term neighbor compatibility, which would be impossible for a purely random or short-sighted approach.

6. Discussion

The aim of this study was to determine whether RL with DQN could be an effective method for optimizing complex sequencing problems under heterogeneous constraints. From a methodological perspective, RL differs from greedy heuristics in that it optimizes cumulative reward over a finite planning horizon rather than immediate local effectiveness. This structural property enables the RL agent to consider long-term compatibility effects within the sequence. The following sections discuss the performance of the proposed approach, its methodological transferability to CIPS and the limitations of the current study.

6.1 Performance and Contributions

The experimental results indicate that the RL approach successfully internalizes multi-criteria constraints without relying on explicit rule-based programming. The convergence of the cumulative reward (as shown in Figure 2) suggests that the RL agent learns to balance conflicting objectives, such as maximizing diversity while ensuring soil compatibility, by exploring the structure of the reward function. In contrast to static rule-

based heuristics that optimize based on predefined local criteria, the RL agent optimizes cumulative reward over the full planning horizon.

A key strength of the proposed approach lies in its interpretability. By decomposing the reward into individual agronomic components and logging them for every placement decision, the system provides transparent explanations for its recommendations. This is particularly important in practitioner-facing DSS, where trust and domain acceptance are critical.

A critical requirement for DSS in both agriculture and industry is explainability. The proposed framework addresses this by utilizing a decomposed reward structure. Since every decision score is an aggregation of transparent sub-rewards, the system can provide post-hoc explanations for its recommendations. This transparency is essential for building trust with domain experts, who are often skeptical of black-box optimization methods.

6.2 Transfer to Circular Industrial Production Systems

A central contribution of this work is the validation of intercropping systems as a high-complexity proxy domain for CIPS. Although the semantic differs, the underlying mathematical structure of the decision problem is isomorphic between the two domains. Optimizing a sequence of discrete entities under strict compatibility and sustainability constraints applies for both approaches.

6.2.1 Structural Isomorphism

The methodological transferability rests on the following mapping between the intercropping agriculture approach and CIPS. Crops correspond to material options, planting rows to production sequences, and agronomic constraints to process feasibility (see Table 1). Sustainability objectives, such as soil fertility preservation in agriculture, are mirrored by circularity metrics in industry, including material reuse, energy consumption and waste reduction.

Just as crops have biological attributes (pH, nutrients), industrial components possess material attributes (alloy type, recyclability, toxicity) that dictate compatibility [33]. The order of crops affects soil health and pest resilience, analogously, the sequence of manufacturing or disassembly steps determines energy consumption, tool changes, and material purity [6]. Constraints such as planting windows map to temporal production restrictions (e.g. curing times, shift schedules) or machine capabilities [10].

6.2.2 Sustainability Trade-offs

Both domains face the challenge of multi-objective optimization where economic efficiency conflicts with ecological goals. In intercropping, the RL agent balances yield against soil regeneration. In CIPS, planners must balance throughput against circularity metrics, such as maximizing the reuse of components or minimizing hazardous waste [2]. The results of this paper demonstrate that RL is particularly suited for this problem because it enables sustainability as a cumulative reward signal rather than a hard constraint, allowing the system to find flexible trade-offs that rigid rule-based system often misses.

6.2.3 Human-in-the-Loop with Reward Shaping

The integration of user feedback highlights a crucial mechanism for industrial transferability. Production planners, like farmers, possess tacit knowledge [31]. The proposed reward shaping mechanism allows the RL agent to adapt these latent preferences over time. This capability is necessary for the deployment of CIPS, where soft factors like worker experience or varying material quality often deviate from the theoretical model.

6.3 Limitations

While the results are promising, several limitations must be acknowledged to contextualize the findings:

- **Simulation and Reality (L1):** The evaluation relies on a simulation environment with deterministic biological interactions. Real-world biological and industrial systems exhibit a higher degree of stochasticity (e.g. weather anomalies or machine breakdowns) that were not fully modeled.
- **Planning Horizon (L2):** The current optimization focuses on single-row sequences within a single planting season. It does not yet account for multi-period crop rotations or the complex interdependencies between parallel production lines, which would exponentially increase the state space.
- **Feedback Data (L3):** The user preference module was tested with limited binary feedback data. In a real-world deployment, the agent would lack sufficient feedback in the start phase, which would result in not aligned decisions. This remains a challenge that requires more robust initialization strategies, such as imitation learning from historical logs.

Despite these limitations, the study confirms that RL-based sequencing policies provide a robust alternative to static heuristics. The successful application in the intercropping domain serves as a proof-of-concept for the broader class of sustainable sequencing problems, offering a methodological blueprint for intelligent decision support in circular economy.

7. Conclusion & Outlook

This paper investigated the application of RL to optimize intercropping sequences, using the agricultural context as a methodological proxy domain for CIPS to address complex operational constraints. A DQN was embedded in a domain-specific simulation environment and trained using an interpretable, multi-criteria reward function. The experimental results show that the agent learns stable sequencing policies that internalize non-trivial compatibility relationships between crops and exhibit structured long-term optimization behavior.

Beyond its agricultural contribution, the work demonstrates that intercropping constitutes a valuable high-complexity proxy domain for sustainability-oriented decision support. The methodological principles underlying the proposed RL framework by sequential policy learning, reward decomposition and HITL adaptation. Therefore, it's directly transferable to circular industrial production systems facing similar optimization challenges.

The paper highlights several strengths of RL-based decision support: the ability to balance conflicting objectives over time, adaptability to changing constraints, and inherent support for explainable decision-making through structured reward signals. At the same time, limitations related to simulation fidelity (L1), limited feedback data (L3) and simplified planning horizons (L2) underline the need for further research.

Future work will extend the framework along three main directions. First, the environment will be expanded to more complex multi-row and multi-period planning scenarios to capture longer-term interactions and rotations. Second, RL will be systematically compared with non-deterministic LLM-based decision approaches, investigating also hybrid architectures that combine symbolic reasoning with sequential optimization. Third, the transfer potential will be empirically validated through an industrial case study in a circular production context, such as remanufacturing or disassembly sequencing.

To summarize, this work provides both a concrete RL-based solution for intercropping planning and a transferable methodological contribution for adaptive, sustainability-oriented decision support. As production systems increasingly adopt circular economy principles, RL offers a promising foundation for next-generation intelligent planning and control systems.

Acknowledgements

Funded as DLR project no. 01MN23012B by the Federal Ministry of Research, Technology and Space (BMFTR) due to a decision of the German Parliament.

References

[1] Ghadimi, P., Wang, C., Lim, M.K., 2019. Sustainable supply chain modeling and analysis: Past debate, present problems and future challenges. Resources, Conservation and Recycling 140, 72–84.

[2] Geissdoerfer, M., Savaget, P., Bocken, N.M., Hultink, E.J., 2017. The Circular Economy – A new sustainability paradigm? Journal of Cleaner Production 143, 757–768.

[3] Helo, P., Hao, Y., 2022. Artificial intelligence in operations management and supply chain management: an exploratory case study. Production Planning & Control 33 (16), 1573–1590

[4] Malézieux, E., Crozat, Y., Dupraz, C., Laurans, M., Makowski, D., Ozier-Lafontaine, H., Rapidel, B., Tourdonnet, S., Valantin-Morison, M., 2009. Mixing plant species in cropping systems: concepts, tools and models. A review. Agron. Sustain. Dev. 29 (1), 43–62

[5] Dury, J., Schaller, N., Garcia, F., Reynaud, A., Bergez, J.E., 2012. Models to support cropping plan and crop rotation decisions. A review. Agron. Sustain. Dev. 32 (2), 567–580

[6] Lieder, M., Rashid, A., 2016. Towards circular economy implementation: a comprehensive review in context of manufacturing industry. Journal of Cleaner Production 115, 36–51.

[7] Raseduzzaman, M., Jensen, E.S., 2017. Does intercropping enhance yield stability in arable crop production? A meta-analysis. European Journal of Agronomy 91, 25–33.

[8] Memmah, M.-M., Lescourret, F., Yao, X., Lavigne, C., 2015. Metaheuristics for agricultural land use optimization. A review. Agron. Sustain. Dev. 35 (3), 975–998.

[9] Saranya, S., Amudha, T., 2016. Crop planning optimization research — A detailed investigation, in: 2016 IEEE International Conference on Advances in Computer Applications (ICACA). 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, India. 24.10.2016 - 24.10.2016. IEEE, pp. 202–208.

[10] Dolgui, A., Ivanov, D., Sokolov, B., 2018. Ripple effect in the supply chain: an analysis and recent literature. International Journal of Production Research 56 (1-2), 414–430.

[11] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D., 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.

[12] Valmeekam, K., Marquez, M., Olmo, A., Sreedharan, S., Kambhampati, S., 2022. PlanBench: An Extensible Benchmark for Evaluating Large Language Models on Planning and Reasoning about Change.

[13] Herrera-Poyatos, D., Peláez-González, C., Zuheros, C., Herrera-Poyatos, A., Tejedor, V., Herrera, F., Montes, R., 2025. An overview of model uncertainty and variability in LLM-based sentiment analysis: challenges, mitigation strategies, and the role of explainability. Frontiers in artificial intelligence 8, 1609097.

[14] Wang, H., Zhu, Z., Shi, F., 2025. Distribution Prompting: Understanding the Expressivity of Language Models Through the Next-Token Distributions They Can Produce.

[15] Kaelbling, L.P., Littman, M.L., Moore, A.W., 1996. Reinforcement Learning: A Survey. jair 4, 237–285.

[16] Bharman, P., Ahmad Saad, S., Khan, S., Jahan, I., Ray, M., Biswas, M., 2022. Deep Learning in Agriculture: A Review. AJRCoS, 28–47.

[17] Chen, D., Huang, Y., 2025. Integrating reinforcement learning and large language models for crop production process management optimization and control through a new knowledge-based deep learning paradigm. Computers and Electronics in Agriculture 232, 110028.

[18] Saikai, Y., Peake, A., Chenu, K., 2023. Deep reinforcement learning for irrigation scheduling using high-dimensional sensor feedback.

[19] Yu, J., Qu, Q., Peng, S., Wei, X., Li, Y., Sun, C., 2025. Deep learning for intelligent irrigation decision-making: A review. Agricultural Water Management 320, 109836.

[20] Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. IEEE Trans. Neural Netw. 9 (5), 1054.

[21] Chen, M., Cui, Y., Wang, X., Xie, H., Liu, F., Luo, T., Zheng, S., Luo, Y., 2021. A reinforcement learning approach to irrigation decision-making for rice using weather forecasts. Agricultural Water Management 250, 106838.

[22] Gautron, R., Maillard, O.-A., Preux, P., Corbeels, M., Sabbadin, R., 2022. Reinforcement learning for crop management support: Review, prospects and challenges. Computers and Electronics in Agriculture 200, 107182.

[23] Pinedo, M.L., 2022. Scheduling. Springer International Publishing, Cham.

[24] Khan, I.S., Ahmad, M.O., Majava, J., 2021. Industry 4.0 and sustainable development: A systematic mapping of triple bottom line, Circular Economy and Sustainable Business Models perspectives. Journal of Cleaner Production 297, 126655.

[25] Machado, C.G., Winroth, M.P., Da Ribeiro Silva, E.H.D., 2020. Sustainable manufacturing in Industry 4.0: an emerging research agenda. International Journal of Production Research 58 (5), 1462–1484.

[26] Li, C., Zheng, P., Yin, Y., Wang, B., Wang, L., 2023. Deep reinforcement learning in smart manufacturing: A review and prospects. CIRP Journal of Manufacturing Science and Technology 40, 75–101.

[27] Usuga Cadavid, J.P., Lamouri, S., Grabot, B., Pellerin, R., Fortin, A., 2020. Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. J Intell Manuf 31 (6), 1531–1558.

[28] Hubbs, C.D., Li, C., Sahinidis, N.V., Grossmann, I.E., Wassick, J.M., 2020. A deep reinforcement learning approach for chemical production scheduling. Computers & Chemical Engineering 141, 106982.

[29] Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., Kyek, A., 2018. Optimization of global production scheduling with deep reinforcement learning. Procedia CIRP 72, 1264–1269.

[30] Ruggeri, F., Russo, A., Inam, R., Johansson, K.H., 2025. Explainable Reinforcement Learning via Temporal Policy Decomposition.

[31] Christiano, P., Leike, J., Brown, T.B., Martic, M., Legg, S., Amodei, D., 2017. Deep reinforcement learning from human preferences.

[32] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I.,

King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533.

[33] Vongbunyong, S., Chen, W.H., 2015. Disassembly Automation. Springer International Publishing, Cham.

Biography

Stefan Sarawinsky

Stefan Sarawinsky is a researcher at the Institute for Industrial Management (FIR) at RWTH Aachen University. His research focuses on machine learning, agent-based AI, and the design and management of scalable IT architectures. Since joining FIR in 2024, he has been actively involved in research on reinforcement learning in the agronomic domain, particularly within the SAAT project. He holds a master's degree in computer science.

Kai Schaller

Kai Schaller is a research assistant at the Institute for Industrial Management (FIR) at RWTH Aachen University. His research activities primarily focus on artificial intelligence, with particular emphasis on machine learning methods. Since joining FIR in 2025, he has been engaged in research on reinforcement learning applications in the agriculture domain. He holds a bachelor's degree in industrial engineering from RWTH Aachen University.

Kajan Kandiah

Kajan Kandiah has been a researcher at the Institute for Industrial Management (FIR) at RWTH Aachen University since 2020. In his current role within the Information Management department, he is part of the technical group IT-complexity management. His expertise includes IT architecture management, smart grid technologies and transformation strategies.

8th Conference on Production Systems and Logistics

A Digital Framework For Structured Meeting Preparation In Kaizen Projects

Kevin Zhao, Yuxi Wang, Joachim Metternich
PTW/ TU Darmstadt, Darmstadt, Germany

Abstract

Unstructured meetings are a major source of efficiency-related losses in organizations. To address this, a role-based, multi-agent meeting support framework powered by large language models (LLMs) is introduced to enhance structured preparation for cross-functional collaboration. In this framework, each agent simulates a domain-specific role (e.g., logistics, assembly, machining), contributing specialized knowledge to generate targeted agendas and prepare contextual materials aligned with standard operating procedures (SOPs). The approach was evaluated in a learning factory environment using a simulated Kaizen case study, demonstrating its ability to generate goal-driven agendas. Initial results show that well-structured agenda creation is possible, and that the framework can support goal-oriented, cross-functional meetings by generating actionable agendas and providing relevant, explainable information for stakeholder selection. By improving agenda relevance and reducing coordination overhead, this framework provides a scalable method to integrate intelligent meeting support into continuous manufacturing improvement processes, offering a novel application of LLMs to support cross-functional continuous improvement initiatives in manufacturing.

Keywords

Lean Manufacturing; Continuous Improvement; Large Language Models; Agenda Automation; Cross-functional collaboration

1. Introduction

Of the many inefficiencies that exist in organizations, meetings are especially prominent due to their significant time consumption and clear potential for enhancement. They represent a significant cost factor, especially for managerial staff [1]. For instance, in one case study, 300 managers collectively spend an estimated 60,000 hours annually in meetings. At an average personnel cost of €45 per hour, this results in a total annual meeting cost of approximately €2.7 million [1]. In addition to the financial impact, meeting quality influences employee job satisfaction [2], which has been shown to correlate with customer satisfaction [3]. Beyond these human and financial implications, effective meetings are also challenged by organizational structures. Manufacturing organizations typically consist of production and support departments, each with distinct objectives and specialized domain knowledge. However, this division can hinder effective collaboration across departments due to limited mutual understanding and a heightened risk of miscommunication [4]. A lack of coordination and integration of different departments in the internal supply chain can lead to lower performance and numerous non-value-adding activities, including delays due to waiting, time spent searching for information, and participation in non-goal-oriented meetings [5,6,4,7]. These inefficiencies contribute to increased transactional costs that are rarely captured in commercial accounting systems [6].

DOI: https://doi.org/10.15488/20995
ISSN: 2701-6277