Vol.9 No.1 (2026)

Journal of Applied Learning & Teaching

ISSN: 2591-801X

Proudly owned and sponsored by Kaplan Business School, Australia

Content Available at: https://jalt.open-publishing.org/index.php/jalt/index

Modelling optimal learning pathways: A Markov Decision Process approach to the pedagogy–heutagogy continuum

Dyane Smith[A] — [A]Data Science & Analytics, Kaplan Business School, Australia

Sanjeev Naguleswaran[B] — [B]Data Science & Analytics, Kaplan Business School, Australia

Keywords

Andragogy;
heutagogy;
Markov Decision Process (MDP);
optimal learning;
pedagogy.

Correspondence

dyane.smith@kbs.edu.au[A]

Article Info

Received 14 November 2025
Received in revised form 17 March 2026
Accepted 18 March 2026
Available online 27 March 2026

DOI: https://doi.org/10.37074/jalt.2026.9.1.16

Abstract

This paper presents a novel application of an AI model based on a Markov Decision Process (MDP) and leverages Dynamic Programming and value iteration to model the student learning journey across pedagogical, andragogical, and heutagogical learning paradigms. In contrast to a traditional static educational model, the proposed approach can adapt to changing cohort engagement and progress to provide more personalised and effective learning. Thus, the model can provide deeper insights into how students navigate self-directed learning, stay motivated, and achieve job readiness by considering different learning states, choices, and their associated outcomes. This work offers a theoretical contribution by formalising the pedagogy-heutagogy continuum and a practical framework through integration of analytics systems to optimise the learning process. It establishes a conceptual shift where personalisation moves from a design choice to a mathematically optimised strategy, bridging educational theory with computational decision science. While the current model uses illustrative data, it establishes a scalable foundation for future empirical integration using learning analytics.

Introduction

Higher education institutions are increasingly challenged by student disengagement, diverse learner profiles, and the demand for job-ready graduates. The advent of digital transformation has introduced new methodologies, profoundly reshaping the requirements of teaching and learning. For example, developments in artificial intelligence have emerged with great speed, requiring new critical skills and knowledge, influencing the effectiveness of the learning process, and highlighting the role of continuous learning in modern education.

The rapidly evolving labour markets, along with flexible delivery models, amplify the above-mentioned challenges putting a strain on traditional pedagogical frameworks that tend to be a one-size-fits-all approach. Although new modes of delivery and data-rich learning environments have been introduced, little has been done to improve students' engagement and learning outcomes. Much of the current implementation is around static learner dashboards or predictive models that identify at-risk students. While they provide sound fundamentals, these approaches treat learner engagement as a linear concept, discounting the dynamic and non-linear continuum in the learning journey.

A deeper exploration of curriculum design frameworks can identify possible alternative, complementary pathways for knowledge acquisition and engagement. These pathways, framed within the paradigms of pedagogy, andragogy, and heutagogy, cater to adult learners with distinct needs, fostering differentiated levels of learner engagement and autonomy.

Conceptual and research gap

The growing interest in personalised approaches to education has ignited an interest in individualised learning methods. While there is activity in this area, current models lack the explicit representation of the learner journey as a probabilistic and state-dependent process. Pedagogical theory is rarely integrated with computational modelling, making this framework highly contemporary, addressing the conceptual and current research gap. Most models assume a fixed pathway for learners rather than a continuum and do not adequately capture the transitions between learning paradigms based on learner engagement. Primarily, current models do not adequately compute the transitions between pedagogy, andragogy and heutagogy based on the learner's capabilities.

A Markov process-based conceptualisation offers a framework to mitigate these limitations. This model represents learning as probabilistic state transitions, where future learning is based on current conditions and not static contexts. The model computes learning as a dynamic progression capturing transitions between states and the end goal, in this case, job readiness. This work contributes to a theoretical and computational framework to model learner progression and lays the groundwork for adaptive curriculum design and personalisation in higher education. It integrates behavioural, cognitive, and contextual data into a systematic structure, thereby supporting adaptive curriculum design pathways informed by transitional probabilities. The timeliness of this study can take advantage of the educational systems that generate continuous learner interaction data, making it adaptable at scale.

The conceptualisation of learner progression as a Markov process nested into the pedagogy-andragogy-heutagogy continuum undertaken in this study links educational theory with probabilistic modelling. In addition, this AI-based model lays out the groundwork that informs curriculum personalisation, student autonomy and scalable learning interventions.

The proposed model and its adaptation

We use a Markov Decision Process (MDP) to model the student learning journey across pedagogical, andragogical, and heutagogical learning paradigms. An MDP is a mathematical model used to describe situations where an agent makes decisions in an uncertain environment (Puterman, 2014). In our context, the "agent" is the student. This framework is used to model sequential decision making and has been widely applied in operations research, robotics and reinforcement learning (Sutton & Barto, 2018). Applications of AI in education have been discussed in Rasul et al. (2023). In education research, there have been exploratory studies of modelling adaptive learning pathways and curriculum sequencing, where learner progression can be represented as transitions between knowledge states (Canty & Greyling, 2020; Chi et al., 2011; Clement et al., 2015).

Figure 1. The conceptual gap: Fixed pathways across the learning journey (generated by ChatGPT with guidance from the authors).

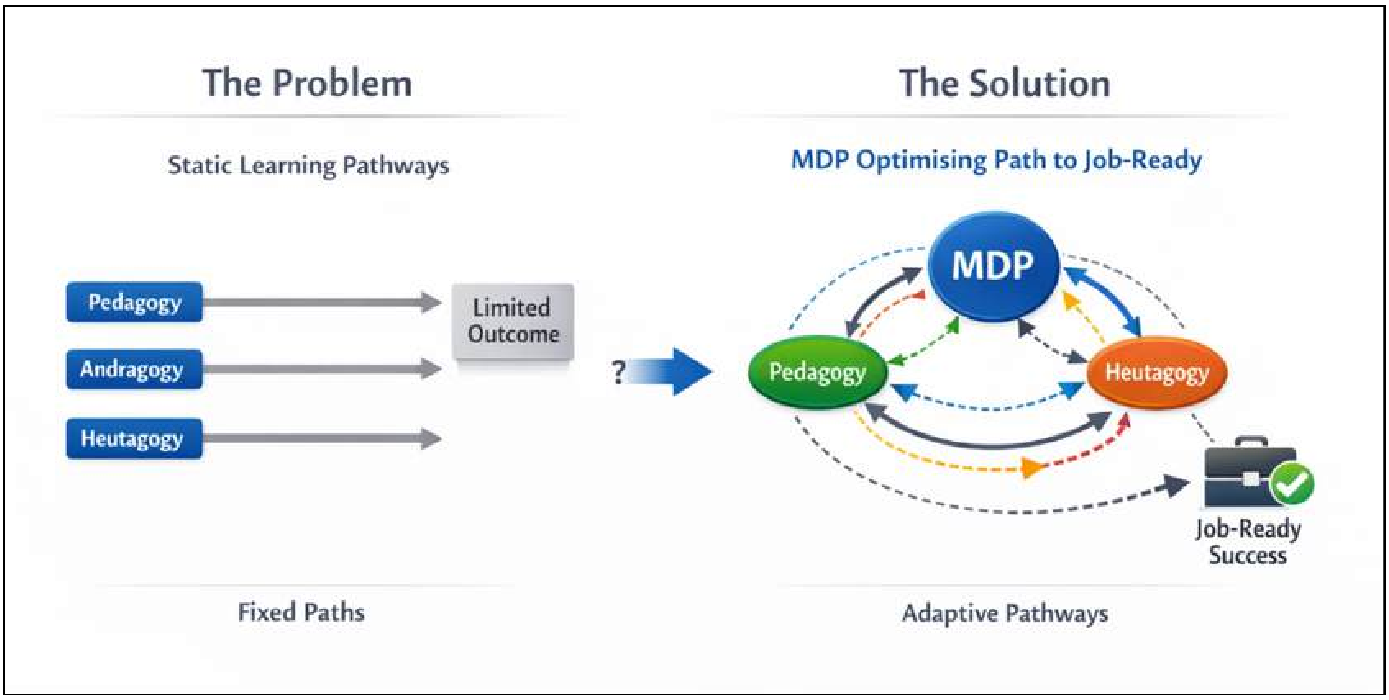

Figure 2. MDP as a solution to the limitations (generated by ChatGPT with guidance from the authors).

An MDP consists of states represented by stages in the learning paradigm, actions that represent learner choices or instructor interventions facilitating state transitions, transition probabilities which quantify the likelihood of moving from one state to another after taking a specific action, and the reward function which assigns a value (positive or negative) to each state or state-action pair. This element represents the value of achieving a state or performing an action. For instance, successfully transitioning to the "Job-Ready Graduate" state yields the highest reward. Dynamic programming is used to optimise the learning journey towards job readiness, where the cumulative rewards are maximised. In the context of this paper, the model is applied to a student's learning journey to optimise the pathway with the highest rewards. This is explained in more detail in the sections titled "Conceptual framework and model development" and "Learning states and transitions: A conceptual view".

Learning paradigms

Pedagogy is a teacher-centred approach, where the instructor transmits knowledge and students passively receive it. It's often associated with childhood education, where learners are seen as dependent on the teacher for guidance (Kumar Shah, 2021).

Andragogy focuses on adult learners who are self-directed and bring their own experiences to the learning process. The instructor acts as a facilitator, guiding the learning process and creating opportunities for discussion and application of knowledge (Halupa, 2015).

Heutagogy takes self-directed learning a step further. Learners become self-managed, identifying their learning needs, finding resources, and evaluating their progress independently. The instructor acts more as a mentor or coach, providing support and guidance as learners navigate their own learning journeys (Hase & Kenyon, 2013).

These paradigms form a progression of learner autonomy, which this study proposes to model computationally as a decision-making process. The interplay between these approaches can be seen as a continuum in a tertiary educational setting where pedagogy is often a starting point, especially for learners who lack background knowledge or skills to be fully self-directed. Andragogy becomes more prominent as learners gain experience and confidence (Knowles, 1980). Heutagogy represents the stage where learners take complete ownership of their learning process. However, it's important to note that this continuum isn't always linear. Effective education may use elements of all three approaches depending on the learning context and the needs of students. The pr-

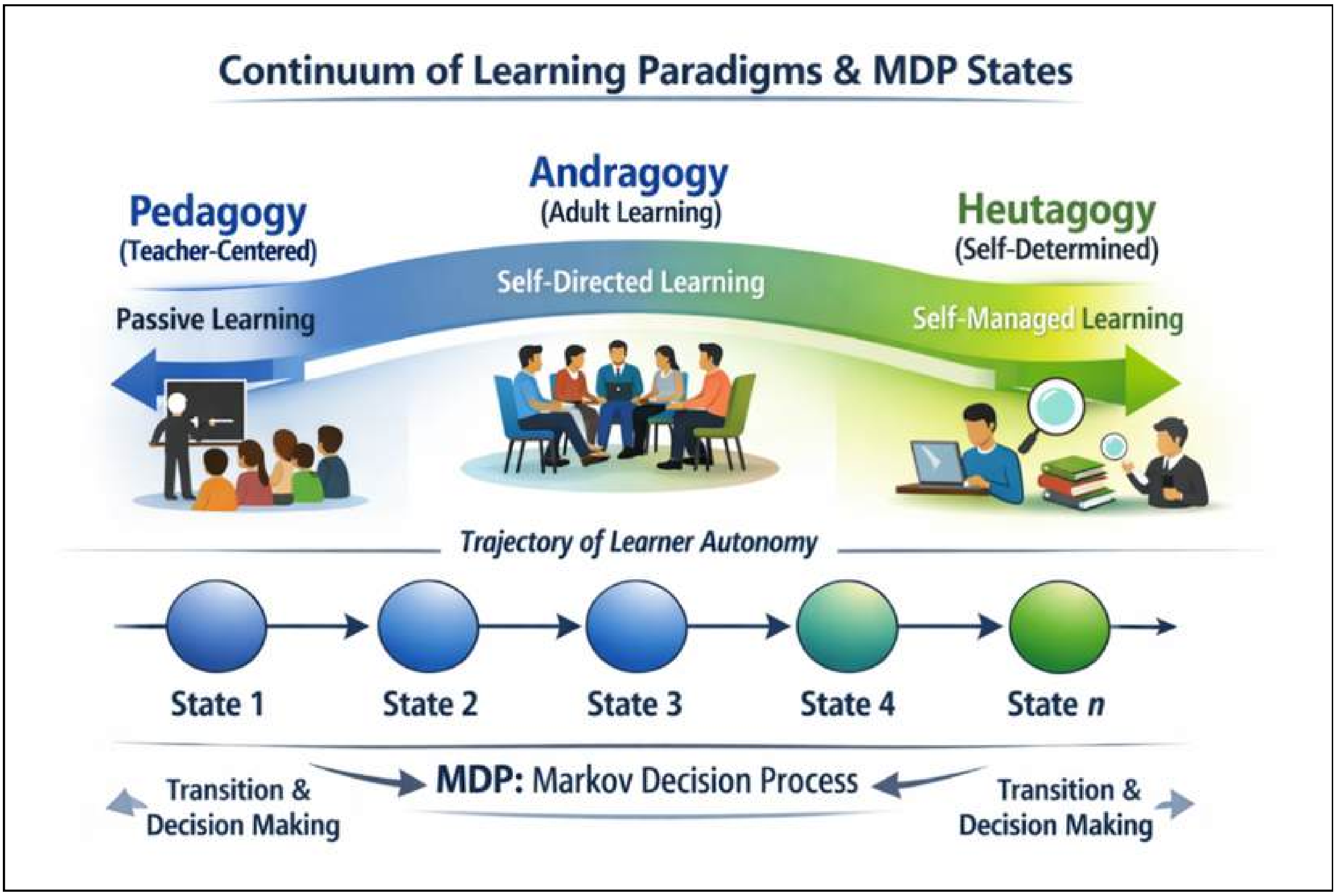

Figure 3. A conceptual representation of the continuum of learning paradigms and MDP states (generated by ChatGPT with guidance from the authors).

-ogression to heutagogy is important in the context of rapid advancements that would make pedagogical learning obsolete within a short time frame (Kaushal et al., 2022).

In this paper, we propose that it is possible to approach the decision problem about learning methodology with more mathematical sophistication by modelling discrete states of the continuum between pedagogy and heutagogy. We propose a model based on an MDP that can be optimised using Dynamic Programming.

Conceptual framework and model development

This section introduces a conceptual MDP framework as an illustrative and theoretical tool for how learners move through different modes of learning on their journey to being job-ready. The purpose of the framework is to support curriculum designers and educators to understand how different teaching approaches can be effective based on state transitions, learner autonomy and study conditions.

The model is developed using the Value Iteration algorithm to determine optimal policies for transitioning between learning states. The MDP framework was selected due to its suitability for modelling sequential decision making under uncertainty, conditions commonly found in educational environments where learners' choices evolve based on feedback and outcomes. Its formal structure also allows for integration with real-time analytics in digital learning systems.

Learning as a journey through a country

A student's learning progression can be understood through the analogy of guiding a student through the roads of a country toward a shared destination of the capital of the country, in this case, Job readiness.

Consider the three learning paradigms as their own districts: the Pedagogy district, the Andragogy district and the Heutagogy district. Pedagogical districts are structured, guided and well-signposted. Andragogical districts al-allow for choice, reflection and negotiated pathways, while heutagogical districts are open, exploratory and self-directed.

Each district will have neighbourhoods with streets and intersections. Some neighbourhoods are closer to the capital, and others are further away. Well-lit structured roads are like Pedagogical approaches, while andragogy roads encourage exploration. Heutagogy roads are self-directed backstreets and shortcuts. Traffic conditions represent uncertainty where not every road will lead to the capital, some routes may take longer, and some may require detours and additional support.

As a learner moves through each neighbourhood using various roads, their autonomy and circumstances can dictate how they navigate through the district across the different neighbourhoods via various roads that lead them to the capital. Some learners may start with a well-lit road, and as they gain confidence and skills in navigation, they may explore backstreets and shortcuts, while other learners may have the confidence and know-how of the shortcuts to get to the capital sooner than the others. While on route, learners could face uncertainty by choosing a route that took longer or road work resulting in detours.

Figure 4. Learner progression is shown through navigating to the capital of a country (generated by ChatGPT with guidance from the authors).

This dynamic navigation to the capital can be computed as probabilistic state transitions. Imagine guiding a learner through a country with multiple routes to the same destination: “Job Readiness”. Each road (learning state) offers different experiences, time durations, and levels of difficulty. At each intersection (decision point), the learner chooses which way to go with some chance of success or failure. Our model helps identify the most efficient, rewarding path based on how learners typically travel, adjusting for their choices and progress.

Learning states and transitions: A conceptual view

An MDP is a mathematical model used to describe situations where an agent makes decisions in an uncertain environment. This is what the MDP captures: it simulates the learner’s learning journey using mathematics to identify optimal steps, balancing structure, autonomy, and goal-orientation. In other words, the framework models learning as a sequence of states and decisions under uncertainty. The MDP framework enables us to estimate which sequence of decisions is likely to get the student to job readiness most efficiently. It consists of four key elements (Hull, 2021; Russel, 2016):

- **States** represent stages within pedagogical, andragogical, and heutagogical pathways. In the case of the example, the various neighbourhoods would be the states.
- **Actions** or the roads chosen represent movement between these states as a result of learner choices and instructor interventions.
- **Transition probabilities** quantify the likelihood of moving from one state to another after taking a specific action.
- **Reward function** assigns a value (positive or negative) to each state or state-action pair.

If the action leads to the destination more quickly, then a higher reward is assigned, while actions that lead away from the destination, such as risk or uncertainty, would yield a negative value. Successfully transitioning to the “Job-Ready Graduate” state yields the highest reward. The rewards here are not considered as performance metrics; they reflect educational desirability, such as the development of skills, increased learner autonomy, and professional practice readiness.

The model assumes that where the learner goes next depends on where their current state is and is not a function of their entire history. This ties in well with common educational practice, where educators and curriculum designers respond to learners based on their current state and not every previous action they may or may not have taken. Curriculum designers have a balancing act of structure vs. autonomy, feedback vs. application, and the value of exploration vs. disengagement. Educators must articulate the kind of learning experience that is most likely to support progress toward job readiness. This framework provides reasoning to systematically consider these questions rather than relying on a static state of a curriculum or intuition.

The optimal pathway taken by learners is interpreted as curriculum design signals, indicating the transitions that need to be made for the learner to be job-ready. It suggests that while a structured pedagogical approach is vital at the start of the learning journey, the curriculum should fade pedagogical structure and conditionally introduce andragogical and heutagogical elements to the curriculum. Such teaching practices shape the likelihood of moving from one state to another (transition probabilities) through assessment and learning design. Therefore, this conceptual model situates the learning paradigms on a continuum of increasing learner autonomy, mediated by curriculum design levers, and culminating in job readiness as an emergent outcome. The learning design choices are considered probabilistic transitions and rewards in the MDP process formalising education theory. It bridges teaching science and computational decision science.

Value Iteration as a planning metaphor

Value Iteration is an algorithm used in Markov Decision Processes (MDPs) to find the optimal policy (Russel, 2016). Value Iteration can be used to systematically assess the long-term rewards of each learning approach. Starting with an initial estimate of each learning state’s value, the algorithm repeatedly updates its understanding of each state until convergence to a best strategy based on final state values. This is analogous to planning each step of a student’s educational journey to determine the best strategy for job readiness.

While Value Iteration is a mathematical construct, it is best understood as a planning metaphor where it conceptually asks *which learning approach would consistently lead to outcomes that are of higher value over time?* For educators and curriculum designers, value iteration can be interpreted as an iterative curriculum refinement, starting with assumptions about which states are valuable, examining how different transitions influence long-term outcomes, and adjusting learning design to support the desirable transitions and outcomes, mirroring reflective teaching practice that is articulated in a formal conceptual structure.

The technical explanations are provided in the Appendix.

Assumptions and limitations

The model currently assumes:

- Learning pathways are independent of external disruptions such as illness, changes in personal motivation, or institutional constraints.
- Rewards and transition probabilities are static and do not vary over time, cohort, or individual learner profiles.

That learners behave rationally to maximise long-term rewards, and that all transition probabilities remain static over time. In practice, learner decisions may be influenced by emotional, social, or contextual factors not captured in this framework.

Representation of learning

Five high-level states corresponding to the pedagogical, andragogical, and heutagogical paths are first identified. While these learning paradigms were mentioned in the Introduction section, this section focuses on how their associated learning states are represented within the MDP.

Table 2. Comparative analysis of synchronous and asynchronous formats in doctoral writing groups.

Paradigm	Stage	Description
Pedagogy	No Knowledge and Skills	Learners depend on the teacher to deliver fundamental knowledge and guidance.
	Basic to Advanced Knowledge	Structured progression through curriculum content.
	Validation of Knowledge	Demonstration of learning via assessments.
Andragogy	Personalised Goals	Learners set personal goals through self-assessment.
	Skill Development	Independent study through projects, reading, and feedback loops.
	Growth Mindset	Learners adopt reflective practices and focus on continual improvement.
Heutagogy	Self-Directed Exploration	Learners initiate learning, challenge assumptions, and co-create knowledge.
	Meta-Cognitive Reflection	Focus on lifelong learning, adaptability, and knowledge transfer.
	Self-Actualisation	Learners achieve personal and professional transformation.

Pedagogical path

In the pedagogical pathway of a learning environment, states could represent different stages or levels of progress within the instructional process. Exemplar states in the pedagogical pathway:

1. No knowledge and skills: This is the initial stage where learners explore course objectives, curriculum, and basic concepts, starting with enrolment.
2. Basic Knowledge: Learners progress to this state after completing the introductory concepts and topics.
3. Intermediate Knowledge: Learners reach this state after mastering intermediate topics, delving into advanced concepts and applications within the field of study.
4. Advanced Knowledge: This state signifies advancement to more complex topics and theories, building upon the fundamentals learned in the previous module.
5. Validation of knowledge: This indicates successful completion of the course requirements and attainment of learning objectives, which covers assessments and feedback on performance.

For the agent to progress from one state to another, the choices they make are considered actions. Exemplar actions in the above pedagogical pathway can be listed as follows:

1. Introduction/enrolment to course.
2. Completing Introduction and Fundamentals.
3. Completing Module 2: Intermediate Topics.
4. Completing Module 3: Advanced Concepts.
5. Complete Assessment and Evaluation.

For the agent to progress from one state to another, the choices they make are considered actions. Exemplar actions in the above pedagogical pathway can be listed as follows:

Andragogical path

1. Personalised learning preference: Learners conduct a self-assessment and then set personalised learning goals based on their aspirations and professional development needs.
2. Enhanced skills through self-directed learning, literature review, and projects: Learners engage in self-directed study, immersing themselves in chosen topics and acquiring new skills through a combination of reading, practice exercises, and hands-on projects.
3. Awareness of progress and improvement areas: Learners regularly reflect on their learning experiences, assessing goals and identifying areas for improvement. They seek feedback to gain insights and refine their approach.
4. Mastered skills through self-directed projects: Learners apply acquired knowledge and skills to real-world challenges, integrating new learnings with existing expertise to deepen understanding.
5. Growth mindset: The final stage emphasises lifelong learning where learners cultivate a growth mindset and stay current on advancements in their field.

Exemplar actions in the above andragogical pathway can be listed as follows:

1. Self-Assessment and Goal Setting
2. Independent Study and Skill Development
3. Reflection and Feedback
4. Application and Integration
5. Continual Learning and Growth

Heutagogical path

1. New knowledge: Learners explore by posing questions, challenging assumptions, and seeking answers through independent inquiry.
2. Ownership of the learning process: Learners take ownership of their learning process by developing personalised learning plans, identifying interests, setting objectives, and strategies for acquiring and applying knowledge.

3. Co-create knowledge and enhance new knowledge: Learners participate in communities, engage with peers, mentors, and experts to share insights, and co-create knowledge collaboratively. They embrace experiential learning, experimenting with new concepts, technologies, and approaches in real-world contexts, embracing failure as an essential part of the learning process.

4. Meta-cognitive skills: Learners engage in reflection, evaluate their learning journey and strategies, and adjust their approach as needed. They develop meta-cognitive skills, becoming aware of their own thinking processes and learning habits.

5. Lifelong learners: The final stage represents the culmination of the heutagogical learning journey, where learners achieve self-actualisation and personal transformation. They emerge as lifelong learners, capable of navigating complex challenges, pursuing passions, and contributing to society.

Exemplar actions in the above heutagogical pathway can be listed as follows:

1. Curiosity and Inquiry
2. Autonomous Learning Planning
3. Collaboration and Experimentation
4. Reflective Practice and Meta-Cognition
5. Self-Actualisation and Transformation

Job-ready graduate

The final state across all pathways is the "Job-Ready Graduate." This represents a learner who has not only acquired the necessary content knowledge and skills but also exhibits lifelong learning capacity, adaptability, and professional readiness.

A job-ready graduate is considered to possess the subject matter knowledge and technical and soft skills needed to effectively enter and perform in the workforce. Additionally, such graduates need an attitude of continuous learning to thrive in the workforce and facilitate growth in their careers.

Students incorporate various learning practices to be job-ready, navigating combinations of different learning practices. The proposed model, based on a Markov Decision Process (MDP), will present the optimal pathway across Pedagogical, Andragogical, and Heutagogical high-level states for a learner to achieve their policy of being job-ready.

The five high-level states for each teaching and learning practice led to the long-term goal of being job-ready. We established that learning may not be linear and could contain transitioning pathways that lead to the end goal. These pathways can be mapped with informed estimates of success probabilities and rewards.

Each of these states is treated as a node in the MDP graph. Actions taken by the learner (or instructor) serve as transitions between nodes. Transition probabilities and rewards are attached based on expected learner progression, motivation, and instructional impact. This allows for comparative analysis of learning trajectories through each paradigm.

Simulation of the conceptual framework

This section simulates the contextual framework discussed thus far. While the estimates for the probabilities and rewards provide a proof-of-concept foundation, we acknowledge that they are not drawn from a formal empirical dataset. The primary aim here is to demonstrate the feasibility and utility of applying MDPs to model learning journeys. Future work will focus on calibrating the model using anonymised academic performance records, LMS engagement logs, and survey-based learner self-assessments to strengthen empirical validity. The total rewards from one learning practice to another are the cumulative rewards of all the states in that learning practice. Pedagogical pathways may take longer and yield fewer rewards compared to andragogical and heutagogical approaches to achieve a job-ready state. Higher rewards were assigned to heutagogical states to reflect their alignment with long-term employability skills such as adaptability, metacognition, and lifelong learning, all valued outcomes in modern workforce environments.

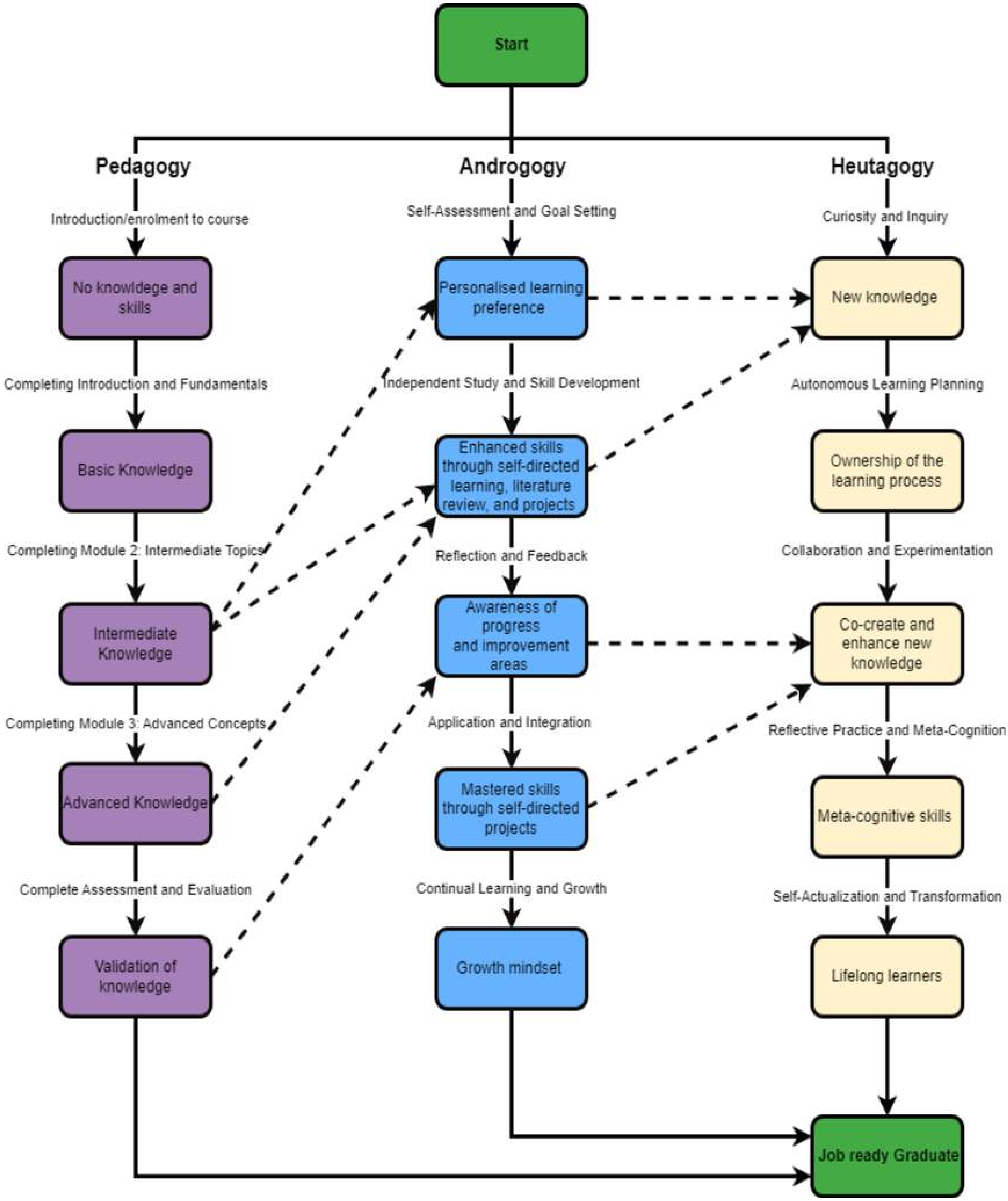

Figure 5. Learning pathways map representing states, actions, and transitions toward achieving job readiness.

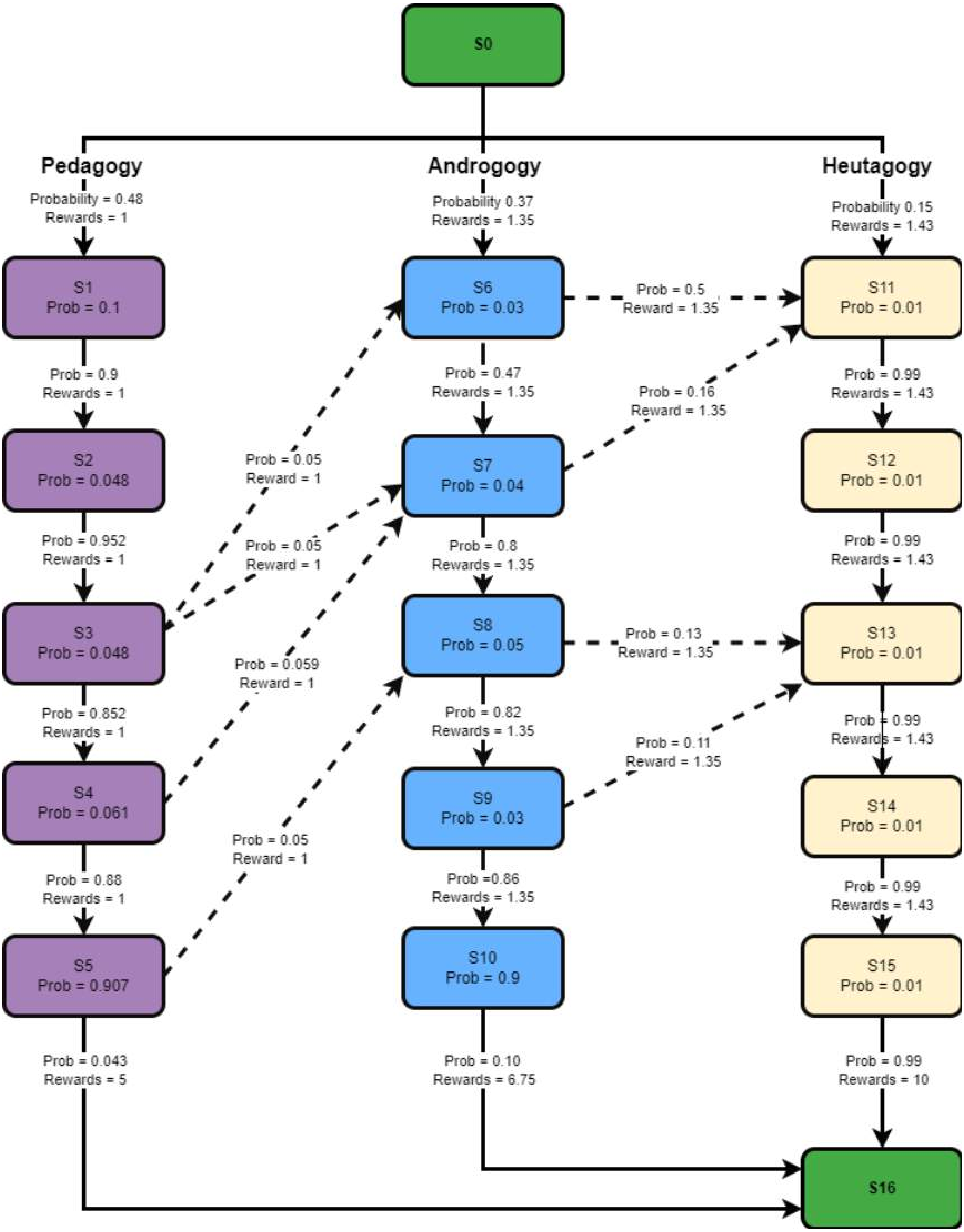

Figure 6. Learning pathways are mapped with associated probabilities and rewards for each transition, informed by an overarching view as given in the Selected Higher Education Statistics 2023 and the authors' tenure in the academic industry.

This transition information was represented as an adjacency matrix with the following columns:

The adjacency matrix includes the following columns:

1. **Current State:** Represents the current state in the MDP.
2. **Current State Title:** Describes the current state.
3. **Next State:** Represents the next possible state.
4. **Next State Title:** Describes the next state.
5. **Probability:** The probability of transitioning from the current state to the next state.
6. **Reward:** The reward for transitioning from the current state to the next state.

The states are listed in the following table.

Table 2. Table of all states sequentially leading to being job-ready.

State	Current State Title
S0	Start
S1	No knowledge and skills
S2	Basic Knowledge
S3	Intermediate Knowledge
S4	Advanced Knowledge
S5	Validation of knowledge
S6	Personalised learning preference
S7	Enhanced skills through self-directed learning, literature review, and projects
S8	Awareness of progress and improvement areas
S9	Mastered skills through self-directed projects
S10	Growth mindset
S11	New knowledge
S12	Ownership of the learning process
S13	Co-create and enhance new knowledge
S14	Meta-cognitive skills
S15	Lifelong learners
S16	Job-ready Graduate

Results of the simulation

The Value Iteration algorithm yielded insights into the optimal pathways for student learning.

Optimal pathway identification

The policy derived from the Value Iteration algorithm highlighted the following:

- **Pedagogical Pathway:** A longer path with moderate cumulative rewards. This pathway benefits learners who need structured instruction and guidance.
- **Andragogical Pathway:** A medium-length pathway offering greater rewards due to self-directed efforts and practical application of skills.
- **Heutagogical Pathway:** The shortest path to the "Job-Ready Graduate" state, characterised by high rewards for learners who exhibit strong autonomy and reflective practices.
- The optimal policy often favours non-linear transitions (e.g., from early andragogical to late heutagogical states) because these pathways yield higher long-term rewards. This reflects the fluidity of adult learning, where experience and self-efficacy can accelerate progression beyond rigid sequences.

Transition probabilities and rewards

- A matrix of transition probabilities and rewards revealed that heutagogical states (e.g., "Meta-Cognitive Skills") offer higher rewards and smoother transitions to advanced states.
- Pedagogical states (e.g., "Basic Knowledge") required more transitions, with lower rewards per action.

Cumulative value function

- The cumulative value function indicated the superiority of the heutagogical pathway in maximising long-term rewards.
- A graph of state values showed a steeper growth curve for self-directed learning approaches compared to instructor-led methodologies.

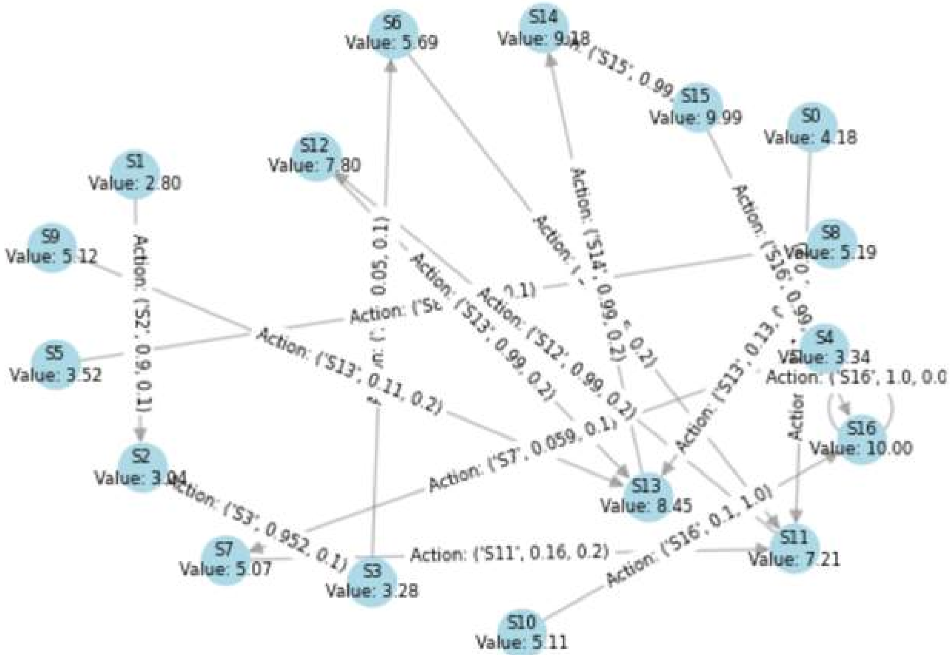

Figure 7. Optimal Policy Map - Highlighting recommended actions for each state.

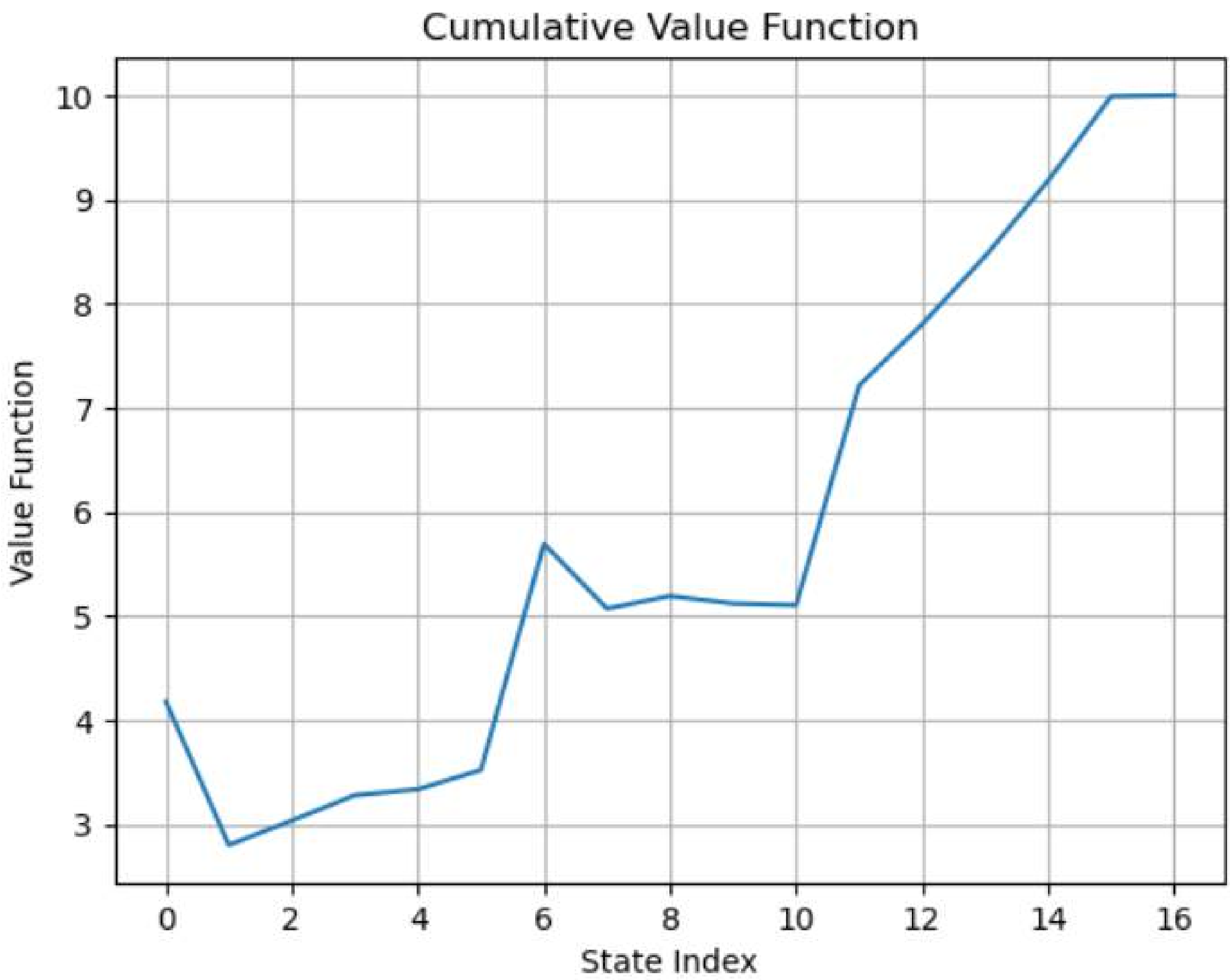

Figure 8. Cumulative value function depicting the rewards of each pathway.

Validation and sensitivity analysis

The robustness of the results was confirmed through sensitivity analysis. Variations in rewards and probabilities had minimal impact on the relative efficiency of pathways, reinforcing the reliability of the findings.

Implications for education

The findings suggest that educational institutions should incorporate elements of self-directed and experiential learning into their curricula to better prepare students for real-world challenges.

In practice, the proposed model can serve as a decision-support tool for curriculum designers, learning technologists, or academic advisors. It can be integrated into Learning Management Systems (LMS) to suggest tailored learning pathways for students based on prior engagement or assessment performance. Consider the following example where students enter a postgraduate subject in Marketing and Social Media Analytics. The students come from diverse backgrounds. Some of the students have a solid understanding of the technical capability needed for this subject, and others have a strategic understanding due to their industry knowledge. The subject consists of technical applications of social media and marketing analytics while also evaluating the strategic implications of social media and marketing on the organisation, amalgamating the technical knowledge with the strategic foresight. For instance, if the subject contains assessments such as quizzes, case study analysis, predictive analytics, and a capstone project, students may do well in combinations of these assessments based on their engagement, autonomy and learning styles. Those who are strategic and grounded in industry knowledge may achieve better grades for the case study analysis and the capstone project, while having room for improvement in technical quizzes and predictive analytics assessments, while tech-savvy students may achieve an inverse outcome. The LMS system could suggest learning activities based on the results of the students. Perhaps the students who need to improve in the technical aspect of the subject may get tailored activities that concentrate on data cleaning, analysis and model creation, while those who need support in strategy may be exposed to more case study analysis and alignment with business goals and decision making (Rudolph et al., 2023).

Figure 9. An example of using the conceptual framework in LMS integration (generated by ChatGPT with guidance from the authors).

Educators could use the model to forecast student progression across learning paradigms and proactively design interventions that accelerate self-directed learning. Learners who display self-directed learning have a higher probability of moving quickly through the learning journey to achieve job readiness, while those who require structured support may need scaffolding that leads to self-directed learning and then job readiness. Currently, analytics in student progression is shown after performance declines, where the optimal next step would need to be deliberated over a period. The delays incurred while deliberating can be a disadvantage to student progression. In addressing this situation, this conceptual framework not only optimises the learning pathway but can also predict the next step in the learner journey. It reduces reaction time and pre-empts a personalised action plan for the learner to reach job readiness fast. It is imperative to mention that it does not replace academic judgement; it augments it by providing the optimal pathways and next steps to achieve job readiness.

Additionally, institutions may use this model to audit existing courses and programs, evaluating whether they disproportionately favour pedagogical approaches or facilitate transitions toward heutagogy. The model's output can inform course redesign efforts that emphasise autonomy, metacognitive skills, and reflective practice. For instance, for courses going through a reaccreditation process, the framework can be used to evaluate the extent of traditional teaching methods vs meaningful progression towards learner autonomy.

The model can be used to analyse the assessments, subject outlines, content for each subject in the course as well as factors external to the course, such as student feedback, academic and facilitator feedback, and employer data. An audit with the use of this framework could reveal an over-reliance on pedagogical or traditional methods with a limited exposure to self-directed learning and autonomy, thereby signalling a scaffolded change in the course to employ meaningful ways of integrating metacognition, autonomy, and practice in lifelong learning. Recommendations could include redefining capstone subjects and projects to involve real-world problems defined by the student, progressively increasing student choice in case study and data analysis and embedding more reflective checkpoints to improve self-directed learning. As mentioned above, rather than replacing academic decision making and judgment, the framework provides structured recommendations to assist scaled decision making in the education institution.

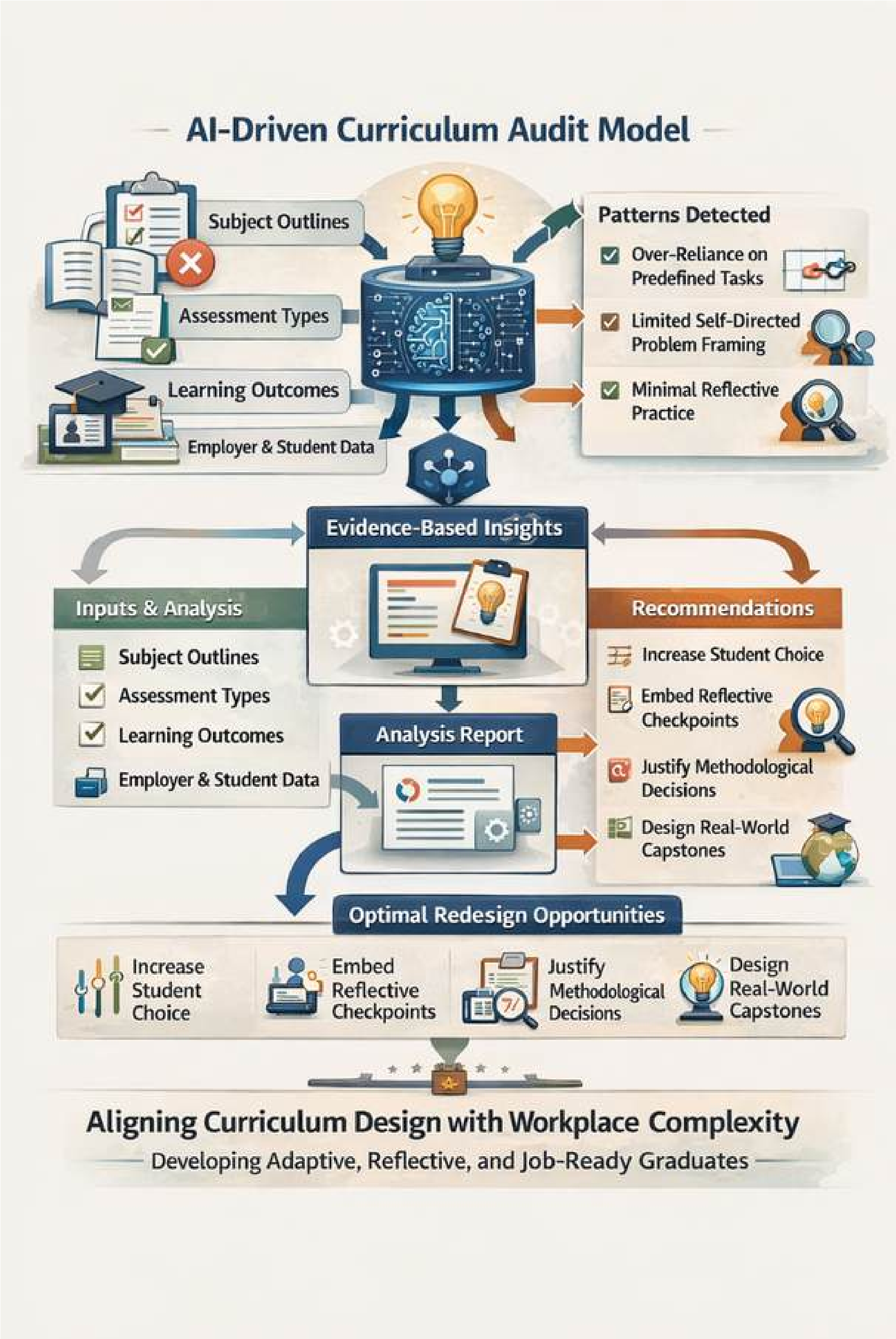

Figure 10. An example of using the conceptual framework in a course audit or reaccreditation (generated by ChatGPT with guidance from the authors).

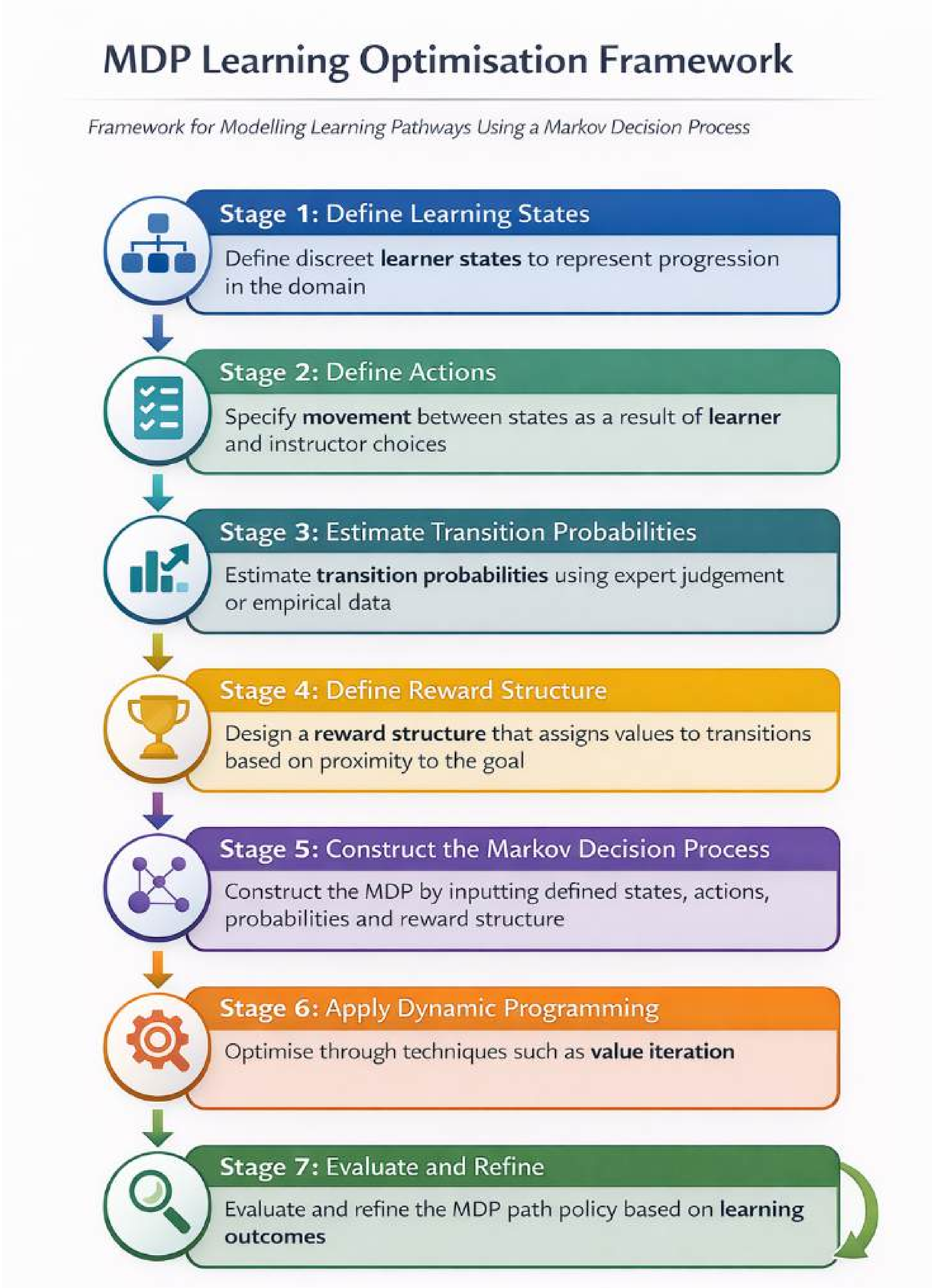

Figure 11. MDP Learning Optimisation Framework (generated by ChatGPT with guidance from the authors).

Application of the model in an education setting

The model is generally feasible across a range of educational contexts, but its successful application depends on several contextual factors. These include the learners' level of maturity and readiness for self-directed learning, the complexity and structure of the subject matter, the institutional culture and assessment requirements, and the availability of supportive learning technologies and resources. In practice, the model is often implemented as a continuum, with more pedagogical approaches used in foundational learning and more self-directed approaches applied as learners gain experience and autonomy. Its success is therefore facilitated by careful instructional design, appropriate scaffolding, and organisational support that enables flexible, learner-centred learning environments.

The following procedure was used when applying the proposed model:

1. **Define the states** that represent stages within the learning paradigm based on discrete learner knowledge. The states should represent the progression in the domain.
2. **Define the actions** that represent movement between the established states due to learner decisions and instructor interventions.
3. **Estimate the transition probabilities** by using expert judgement or empirical data to quantify the likelihood of moving from one state to another after taking a specific action.
4. **Define a reward structure** that assigns a value (positive or negative) to each state or state-action pair. If the action leads to the destination more quickly, then a higher reward is assigned, while actions that lead away from the destination, such as risk or uncertainty, would yield a negative value.
5. **Construct the MDP** by inputting the states, actions, probabilities and reward function into the framework.
6. **Apply dynamic programming** by using optimisation techniques such as value iteration to determine the optimal policy.
7. **Evaluate and refine the results** to validate against the learning outcomes and adjust the model where necessary.

The MDP framework offers a scalable method to evaluate and refine learning strategies across diverse contexts. It is possible to increase the modelling sophistication of MDPs. Partially Observable MDPs (POMDPs), Hierarchical MDPs, and Factored MDPs are examples of extensions that may be useful.

Conceptual reflection

AI modelling helps understand learner self-direction not just by improving pathways, but by grounding the idea of "being self-directed". When concepts like autonomy or agency are translated into modelled choices, constraints, and feedback, there is clarity about what is meant by them.

Limitations and future work

Future research should refine transition probabilities with real-world dynamic data, drawing from academic performance records, online learning analytics, and workplace competency assessments (Kolodner et al., 2003), drawing from academic performance records, online learning analytics, and workplace competency assessments. Challenges in data collection include ensuring privacy, managing data variability across institutions, and addressing biases in self-reported learning outcomes.

Additionally, the assumption of rational learner behaviour does not account for bounded rationality, cognitive biases, or situational constraints. Exploring agent-based simulations or reinforcement learning frameworks with heterogeneous agent profiles could address these nuances.

The current simulation uses illustrative data informed by expert judgment rather than large-scale empirical datasets. This limitation restricts the model's predictive reliability for specific institutional settings. Future research should involve collaboration with universities or online learning platforms to access de-identified learnil-ng analytics, allowing for data-driven tuning of transition probabilities and reward structures. This will improve model generalisability and enhance its potential as a decision-support tool for curriculum designers.

The current model does not account for regressions or skipped stages. In reality, learners may oscillate between states or bypass traditional sequences. A future extension could incorporate stochastic backtracking or multi-path transitions to better reflect such non-linear behaviour.

Implementation of this model in institutional settings may be constrained by ethical considerations, including data privacy, learner autonomy, and algorithmic transparency. Consent protocols and bias mitigation strategies would be required for real-world deployment, particularly when integrating with LMS or advising systems.

Summary and conclusions

This study demonstrated that heutagogical approaches provide the most efficient pathway to achieving job readiness, with significant cumulative rewards and shorter transition times. In comparison, pedagogical and andragogical pathways offer structured and moderately self-directed options, respectively, but with longer trajectories. This signals that learner autonomy is state dependent and a probabilistic capability rather than a fixed trait. Therefore, learner autonomy can be influenced by curriculum design where the increased presence of heutagogical elements in the curriculum will inevitably bring the learner closer and sooner to a state of being job-ready.

These findings suggest a need for strategic integration of self-directed learning frameworks into curriculum design. Educators and programme leaders should consider scaffolding learners' progression through these paradigms, gradually building autonomy and reflection capacity through intentionally designed interventions.

Heutagogical pathways encourage autonomy and meta-cognitive skill development, which align with modern workforce demands. Pedagogical and andragogical methods remain essential for learners at different starting points or those requiring varied levels of guidance. These findings align with Knowles’ adult learning theory, which posits that greater autonomy and self-direction result in more effective and sustainable learning outcomes. The model confirms that heutagogical learners reach job-readiness faster, consistent with experiential and constructivist learning principles.

In order to translate these findings into practice, institutions could begin by identifying courses where learners typically transition toward higher autonomy (e.g., capstone projects, internships). The MDP model can then be piloted in these settings to recommend next-step learning activities and track progress toward job-readiness.

Recommendations

To maximise learning outcomes, curricula should adopt a phased approach that transitions learners from pedagogy to heutagogy. By fostering a growth mindset and autonomy early, institutions can better equip students for lifelong learning and professional success.

In future phases, this modelling framework can be integrated into Learning Management Systems (LMS) to provide real-time, data-informed guidance to students and feedback loops to instructors. By aligning instructional design with optimal learning pathways, the model holds potential to enhance both learner agency and institutional effectiveness in preparing job-ready graduates.

Acknowledgements

An earlier version of this paper was circulated as a pre-print via Smith, D., & Naguleswaran, S. (2025, March 31). Artificial Intelligence for Optimal Learning. https://doi.org/10.31219/osf.io/56eby_v1.

Generative AI tools were used to assist in generating conceptual diagrams and subsequently edited and verified by the authors.

The authors thank colleagues who provided feedback during the development of this work.

References

Canty, A., & Greyling, J. (2020). Addressing student attrition within higher education online programs through a collaborative community of practice. *Journal of Applied Learning & Teaching, 3*(2), 1–13. https://doi.org/10.37074/jalt.2020.3.s1.3

Chi, M., VanLehn, K., Litman, D., & Jordan, P. (2011). An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach. *International Journal of Artificial Intelligence in Education, 21*(1–2), 83–113.

Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2015). Multi-armed bandits for intelligent tutoring systems. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 408–415). International Educational Data Mining Society (IEDMS).

Department of Education. (2023). Selected higher education statistics 2023: *Attrition, retention and success rates*. Australian Government. https://www.education.gov.au/higher-education-statistics/student-data/selected-higher-education-statistics-2023-student-data

Halupa, C. (2015). Pedagogy, andragogy, and heutagogy. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (3rd ed., pp. 143–158). IGI Global. https://doi.org/10.4018/978-1-4666-8571-0.ch005

Hase, S., & Kenyon, C. (2013). Heutagogy and the role of self-determined learning in higher education. *International Journal of Self-Directed Learning, 10*(2), 1–17.

Hull, J. C. (2021). *Machine learning in business: An introduction to the world of data science* (3rd ed.). Amazon Fulfillment Poland Sp.

Kaushal, B., Singh, D., Magoon, R., & Kashav, R. (2022). Pedagogy–andragogy–heutagogy: Towards transformative educational epistemologies. *Journal of Anaesthesiology Clinical Pharmacology, 38*(3), 497–498. https://doi.org/10.4103/joacp.JOACP_270_20

Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy* (Rev. ed.). Cambridge Adult Education.

Kolodner, J. L., Gray, J., & Fasse, B. B. (2003). Promoting transfer through case-based reasoning: Rituals and practices in learning by design classrooms. *Cognitive Science Quarterly, 3*(2), 183–232.

Kumar Shah, R. (2021). Conceptualizing and defining pedagogy. *IOSR Journal of Research and Method in Education, 11*(1), 6–29.

Puterman, M. L. (2014). *Markov decision processes: Discrete stochastic dynamic programming* (2nd ed.). Wiley.

Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F. D. O., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning & Teaching, 6*(1), 1–14. https://doi.org/10.37074/jalt.2023.6.1.29

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Appendix

Markov Decision Process

An MDP is a mathematical model used to describe situations where an agent makes decisions in an uncertain environment. In our context, the "agent" is the student. The "states" represent stages in their learning journey (e.g., beginner, skilled, job-ready), and the "actions" refer to decisions such as taking a course module, self-studying, or seeking feedback. The MDP framework enables us to estimate, based on historical data or expert input, which sequence of decisions is likely to get the student to job readiness most efficiently. It consists of four key elements (Hull, 2021; Russell, 2016),

The agent's goal is to learn a **policy**, which dictates the best action to take in each state to maximise the long-term reward.

More precisely, an MDP is a discrete time control process characterised by a set of states; in each state, there are several actions from which the decision maker must choose. For a state s and an action a, a state transition function Pa(s,s') determines the transition probabilities to the next state. There is a cost associated with each action. The states of an MDP possess the Markov property implying that if the current state of the MDP at time t is known, transitions to a new state at time t1 are independent of all previous states.

The solution to an MDP can be expressed as a policy p, which gives the action to take for a given state, regardless of prior history.

It is our goal to represent the learning process as a shortest-path problem where the deterministic problem is defined as:

Assuming a discrete finite-state system, the deterministic shortest path (DSP) problem consists of finding the path formed by a minimum-cost sequence of successor states starting at the initial state i and terminating at a special cost-free goal state G. The system can be represented on a directed graph consisting of nodes 1, 2, . . ., N, where node G is a goal state, and non-negative costs are assigned to each edge. At each node s, s' is a successor node such that (s, s') is an edge, which is selected.

The Stochastic Shortest Path (SSP) problem generalises the DSP problem by introducing transition probabilities with the goal of choosing an optimal action at each state to minimise cost in reaching the goal state.

Value Iteration

Value Iteration is an algorithm used in Markov Decision Processes (MDPs) to find the optimal policy (Russel, 2016). It can be used to systematically assess the long-term rewards of each learning approach. Starting with an initial estimate of each learning state's value, the algorithm repeatedly updates its understanding of each state until convergence to a best strategy based on final state values. This is analogous to planning each step of a student's educational journey to determine the best strategy for job readiness and is accomplished by solving the Bellman equation:

Vk+1(s) = max_a ∑_{s'} P(s'|s,a)[R(s,a,s') + γVk(s')] where, *Vk+1(s): Value of state s at iteration k+1.*

a: Action.
P(s′ | s,a): Probability of transitioning to state s′ from state s by taking action a.
R(s,a,s′): Reward received after transitioning from state s to s′ via action a.
γ: Discount factor (between 0 and 1).

The process of Value Iteration involves:

1. Initialisation: Start with an arbitrary value function V_0(s) for all states.
2. Iteration: Update the value function using the key equation until convergence.
3. Policy Extraction: Derive the optimal policy from the converged value function.

Implementation and simulation

1. Dynamic Programming Approach:

- Value Iteration was implemented to compute the optimal policy.
- The Bellman equation was iteratively solved until convergence.
- The adjacency matrix of state transitions and rewards was directly integrated into the algorithm.
- The algorithm was iterated until the maximum change in state values between successive iterations fell below a convergence threshold ($\varepsilon = 0.001$), ensuring stability of the derived policy.

2. Data Modelling:

- The transition probabilities and rewards were based on an analysis of academic performance data.
- It is important to note that all probabilities and rewards in this model are illustrative, derived from expert judgement and observed learner behaviours rather than empirical datasets. These estimates were intended to demonstrate the conceptual feasibility of the MDP framework, with future versions designed for calibration using institutional learning analytics.
- Probabilities were adjusted to reflect the observed tendencies of learners in progressing through states.

3. Simulation Setup:

- Python libraries such as NumPy were employed for matrix computations.
- The state space was implemented as a directed graph, with nodes representing learner states and edges encoding possible transitions. An adjacency matrix captured these transitions and associated rewards, allowing efficient computation of value functions and policy updates.
- States were encoded as nodes in a directed graph, with edges representing feasible transitions.

4. Pathway Analysis:

- Pathways through pedagogical, andragogical, and heutagogical states were compared based on time to job readiness and cumulative rewards.
- Metrics included the number of transitions and the rewards accumulated in the process of achieving the "Job-Ready Graduate" state.

Transitioning probabilities and rewards

Each action taken to achieve a defined state has an associated probability and reward value. For this initial model, reward and transition probability values were estimated based on aggregate trends observed in university-level progression data, combined with domain expertise from instructors and programme designers. While not empirically derived, these values reflect typical learner behaviours and were calibrated to distinguish between structured, self-directed, and autonomous learning patterns. These values were normalised so that cumulative transition probabilities from each state sum to 1, and reward values reflect the perceived learning impact of each action.