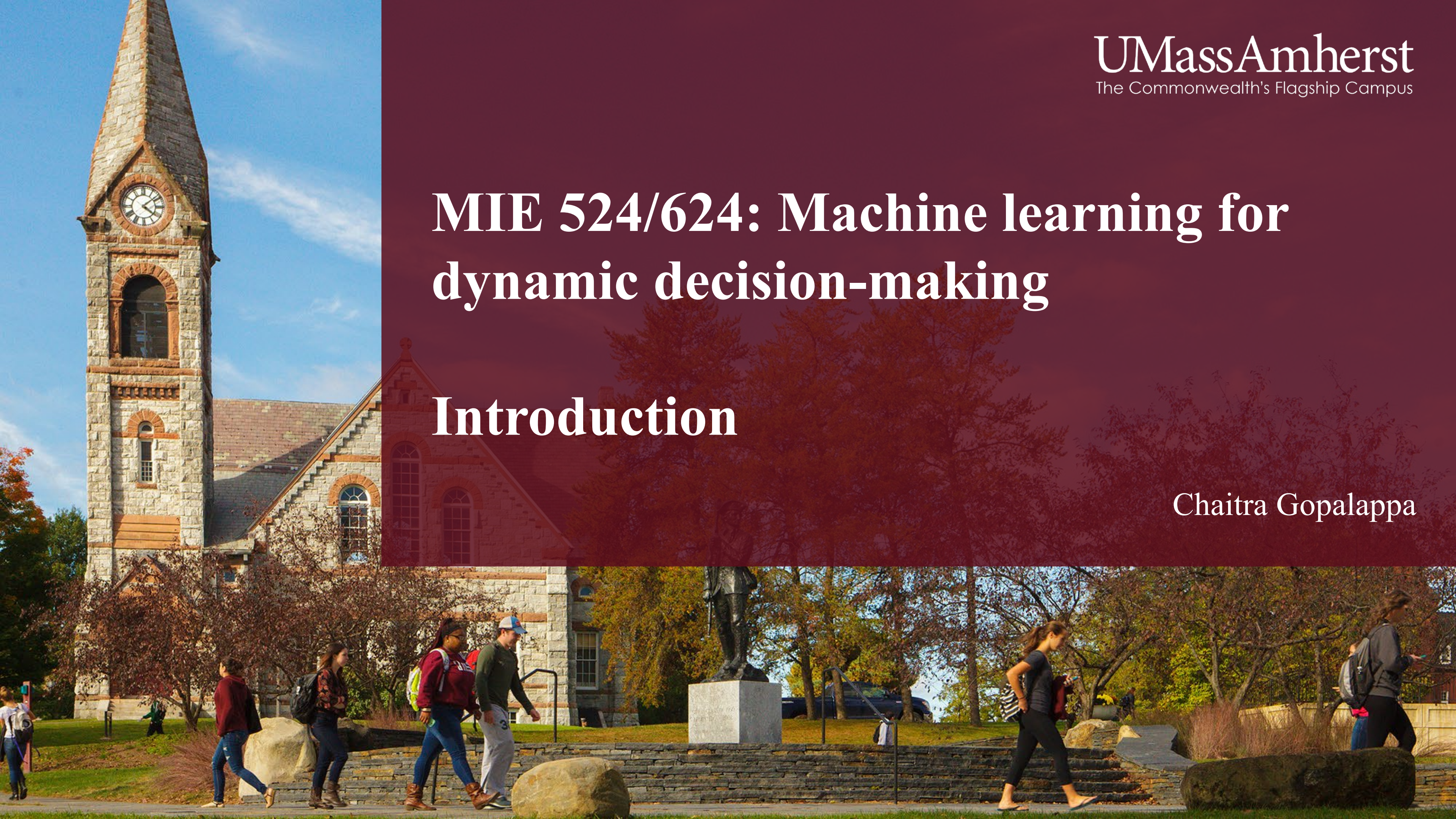


UMassAmherst
The Commonwealth's Flagship Campus

MIE 524/624: Machine learning for dynamic decision-making

Introduction

Chaitra Gopalappa

Applications with dynamic decision-making (DDM)

- Industry
  - Dynamic pricing (airline seat pricing)
  - Inventory control (stochastic demand and supply)
  - Network routing (Amazon delivery routing)
- Policies:
  - Controlled forest fires for wildfire management
  - Fishing/hunting license/limits for sustainable wildlife management
- Investments:
  - Phased implementation/investment into renewable energy to balance economic growth and carbon emissions
  - Quarantine containments/vaccine rollouts for control of new/emerging disease outbreaks

Machine learning (ML) for dynamic decision making (DDM)

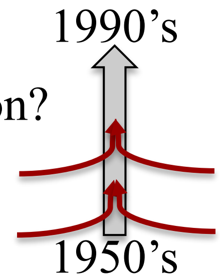
- What is reinforcement learning?
- What is simulation-based optimization?
- What is control optimization?
- What is dynamic decision-making?
- What is sequential decision-making?

- What is deep learning
- What is machine learning?

- What is neural networks?

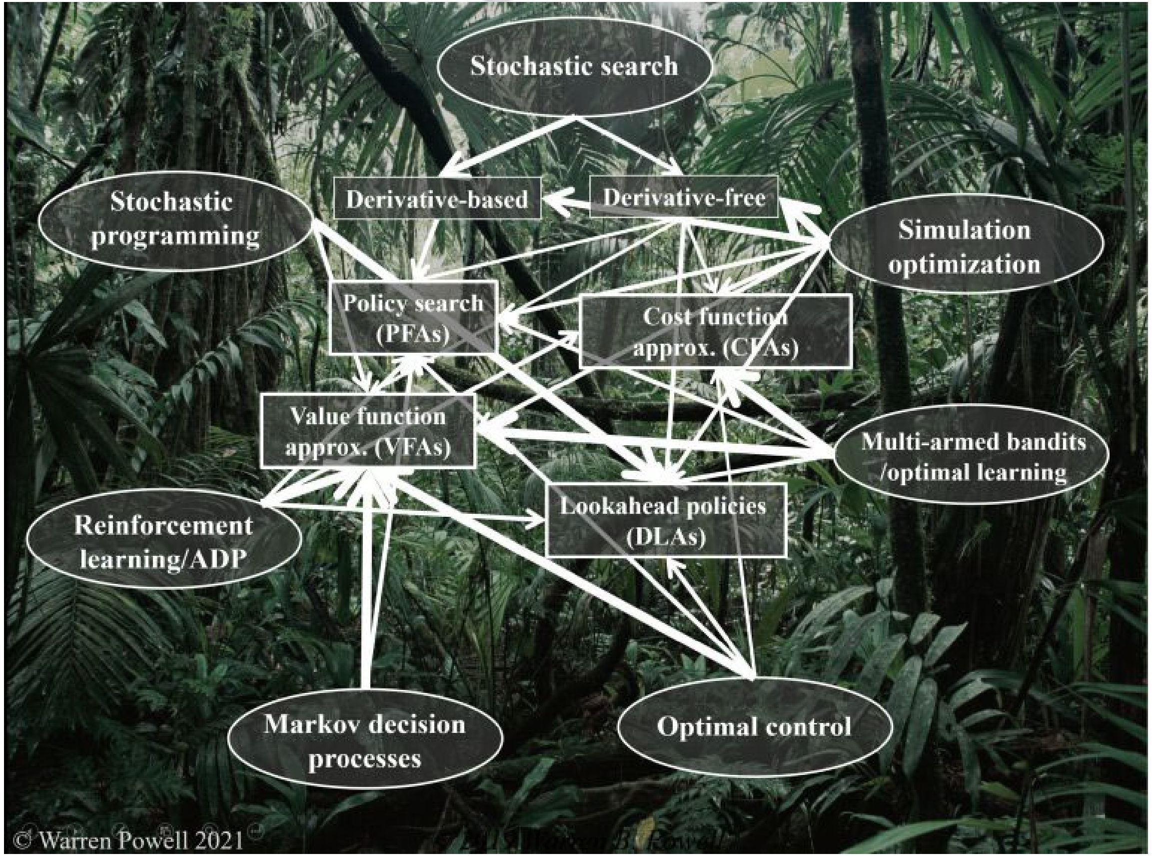
Machine learning (ML) for dynamic decision making (DDM)

- What is reinforcement learning?
- What is simulation-based optimization?
- What is control optimization?
- What is dynamic decision-making?
- What is sequential decision-making?



- What are attention mechanisms/transformers
- What is deep learning?
- What is machine learning?
- What is neural networks?

The jungle
Foundations of DDM algorithms



Source: Powell 2021, https://castlelab.princeton.edu/sda

The jungle- Evolution

Source: Powell 2021, https://castlelab.princeton.edu/sda

- *"Stochastic search traces its roots to two papers in 1951 – one for derivative-based stochastic search, and one for derivative-free stochastic search. Derivative-based stochastic search can be presented as a sequential decision problem where the decision is the stepsize, almost always determined with a parameterized stepsize rule (a PFA). However, derivative-free stochastic search has quickly evolved to using all four classes of policies.*

- *Simulation-optimization – This community originally focused on evaluating competing configurations for manufacturing systems using a DLA called optimal computing budget allocation (OCBA). A few years ago Michael Fu came out with "Handbook of Simulation Optimization" that covers many of the major topics in stochastic optimization using all four classes, but never formalizes the process of searching over all four classes.*

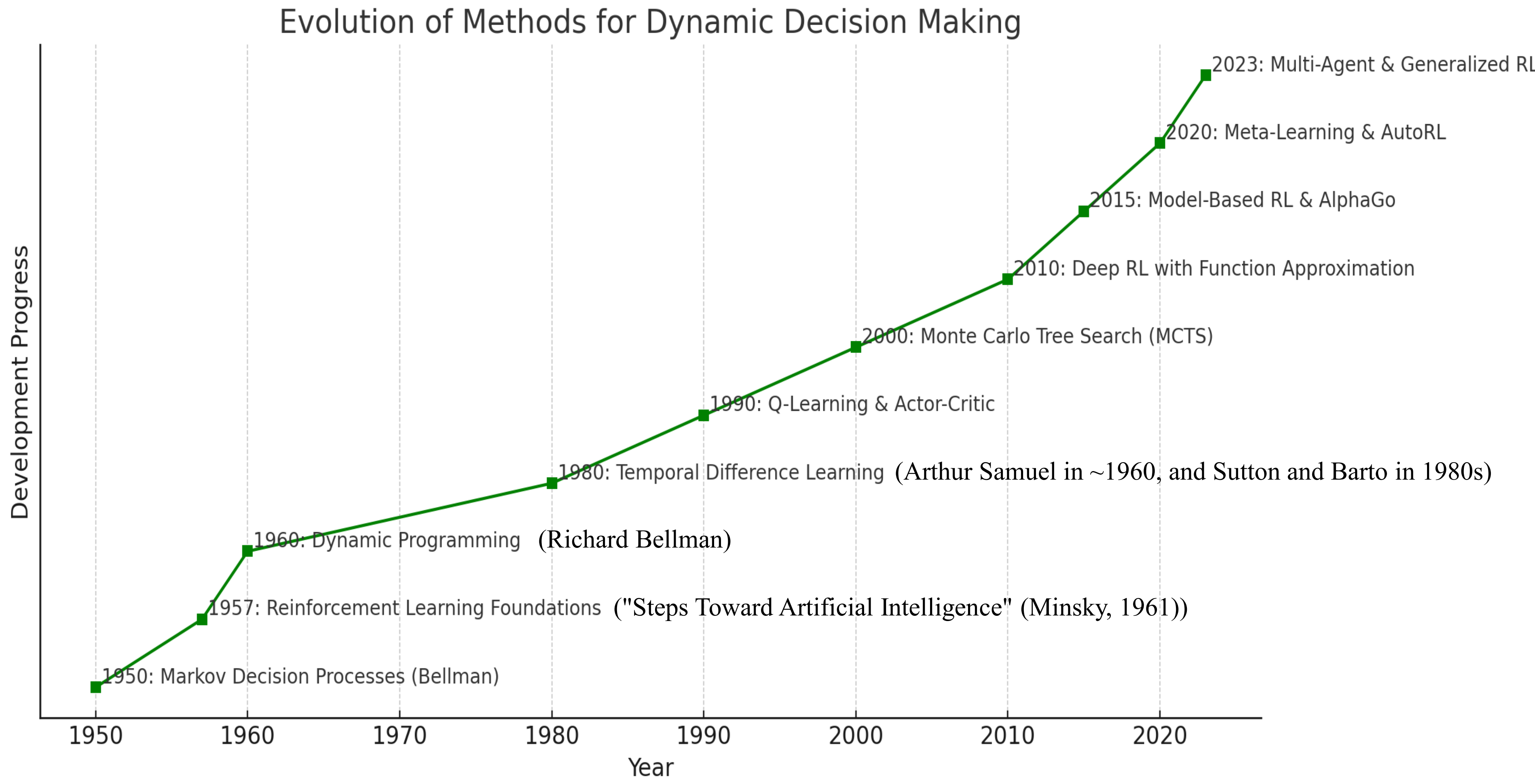
- *Multi-armed bandit problems – These are active learning problems where we try discrete alternatives to improve our understanding of how well it works. It started in the 1970s with a breakthrough known as "Gittins indices" which is based on solving Bellman's equation (a VFA), then transitioned to upper confidence bounding in the 1980s (a CFA). Today, all four classes of policies are used. [Multi-armed bandit problems are all a form of derivative-free stochastic search.]*

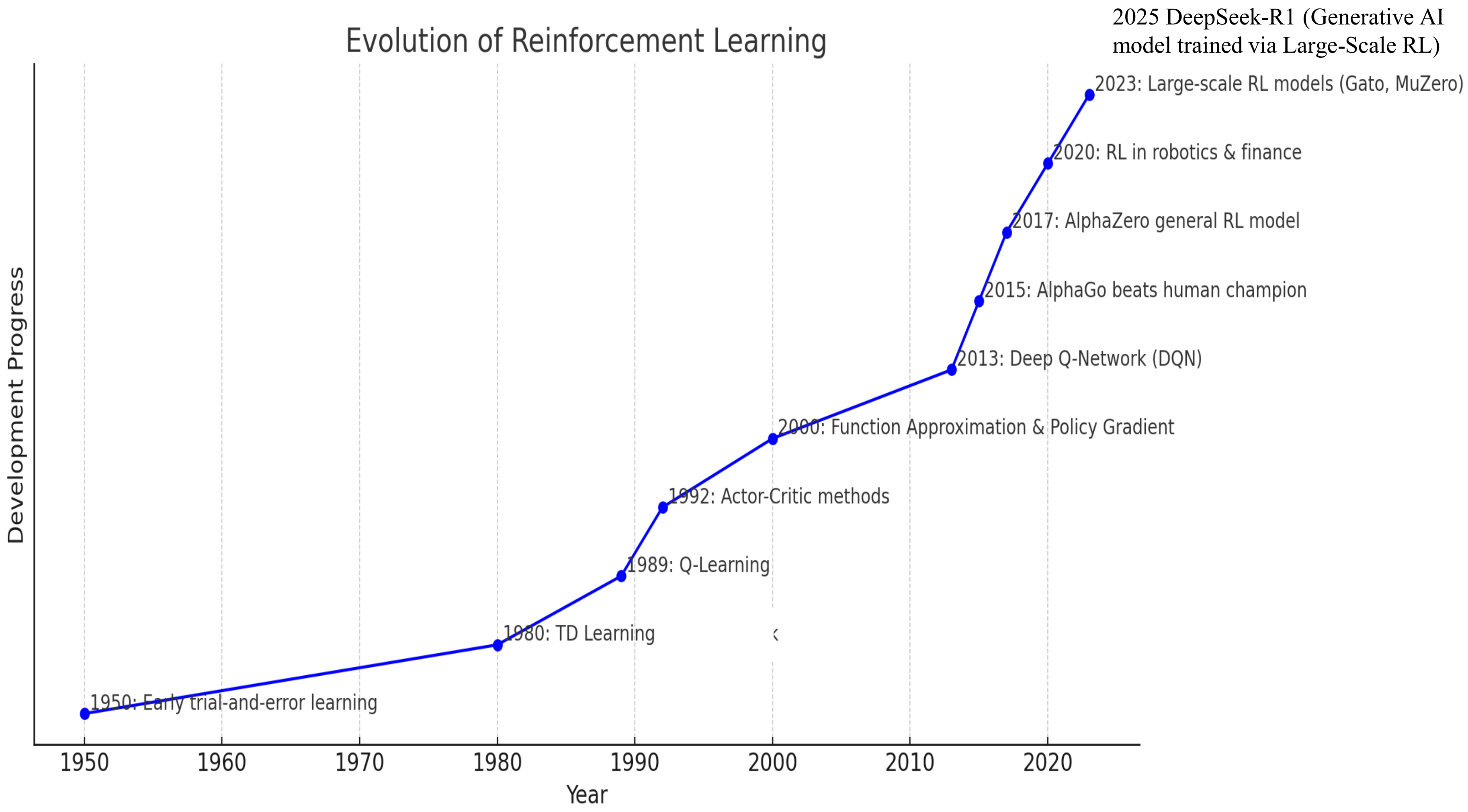
- *Optimal control started with Hamilton-Jacobi (that is, Bellman) equations, which is a VFA, but was able to show for a class of problems known as "linear quadratic regulation" that the policy reduced to a linear function (a PFA). In the 1970s Paul Werbos learned how to use backpropagation to approximate value functions (a VFA). Finally, there is an entire subfield in optimal control called model predictive control (MPC), which is a deterministic lookahead (a DLA).*

- *Markov decision processes evolved originally using Bellman's equation using discrete states (lookup table VFAs), which is a VFA-based policy, although special cases were found that produced analytical solutions to policies (PFAs). However, the curse of dimensionality that arises with lookup table representations motivated the use of approximate dynamic programming, which is all VFA-based, but the research literature also explored the use of parameterized policy (PFAs).*

- *Reinforcement learning in the 1990s and early 2000s was limited to Q-learning (see the discussion on what is reinforcement learning here), as presented in the first edition of Sutton and Barto's now classic Reinforcement Learning: An Introduction. But if you compare this edition to the second edition (2018), the range of methods has expanded to include a variety of other algorithmic strategies (parameterized policies (PFAs), upper confidence bounding (a CFA), Q-learning (a VFA) and Monte Carlo tree search (a stochastic DLA). These are samples drawn from each of the four classes, but there the field has never formalized the process of searching over all four classes to find which one is best.*

- *Stochastic programming was initiated by George Dantzig in the 1950s using scenario trees, which is a form of stochastic DLA, but in the 1990s the research community developed "stochastic dual dynamic programming" (SDDP) which is a form of approximate dynamic programming using Benders cuts to approximate convex value functions (a VFA-based policy). SDDP may be a stochastic lookahead – it depends on how the model is being used (this is true of all ADP-based applications)."*

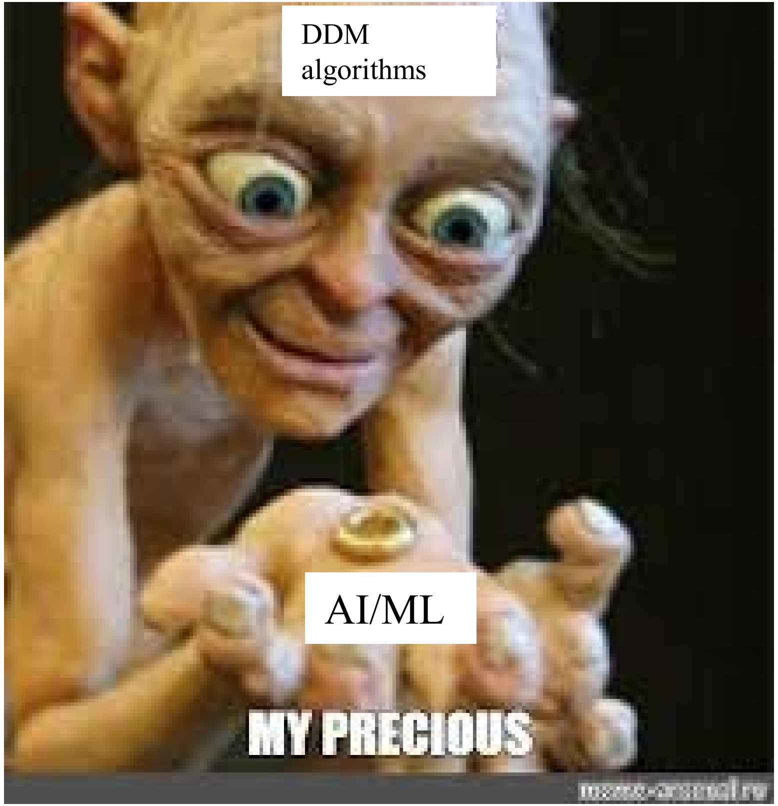

Evolution of Methods for Dynamic Decision Making
2023: Multi-Agent & Generalized RL
2020: Meta-Learning & AutoRL
2015: Model-Based RL & AlphaGo
2010: Deep RL with Function Approximation
2000: Monte Carlo Tree Search (MCTS)
1990: Q-Learning & Actor-Critic
1980: Temporal Difference Learning
(Arthur Samuel in ~1960, and Sutton and Barto in 1980s)
1960: Dynamic Programming
(Richard Bellman)
1957: Reinforcement Learning Foundations
("Steps Toward Artificial Intelligence" (Minsky, 1961))
1950: Markov Decision Processes (Bellman)
Development Progress
1950
1960
1970
1980
1990
2000
2010
2020
Year


Evolution of Reinforcement Learning
2025 DeepSeek-R1 (Generative AI model trained via Large-Scale RL)
2023: Large-scale RL models (Gato, MuZero)
2020: RL in robotics & finance
2017: AlphaZero general RL model
2015: AlphaGo beats human champion
2013: Deep Q-Network (DQN)
2000: Function Approximation & Policy Gradient
1992: Actor-Critic methods
1989: Q-Learning
1980: TD Learning
1950: Early trial-and-error learning
Development Progress
Year
1950
1960
1970
1980
1990
2000
2010
2020

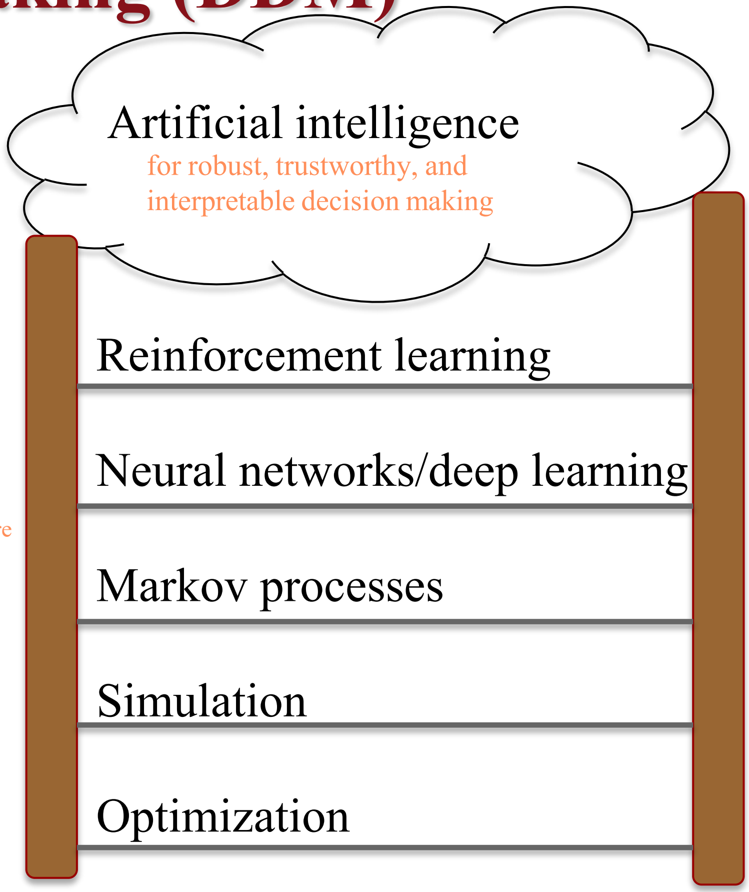
Why the sudden boom in DDM

Machine learning + large compute capacity → model-free intelligent decision-making systems that overcome curse of model-based methods

- 2017 Self-attention mechanisms /transformers
- 1980s Deep learning (model-free)
  - Backpropagation for multi-layer neural networks Geoffrey Hinton and John Hopfield the 2024 Nobel Prize in Physics for *"foundational discoveries and inventions that enable machine learning with artificial neural networks"*.
- Statistical machine learning (model-based)

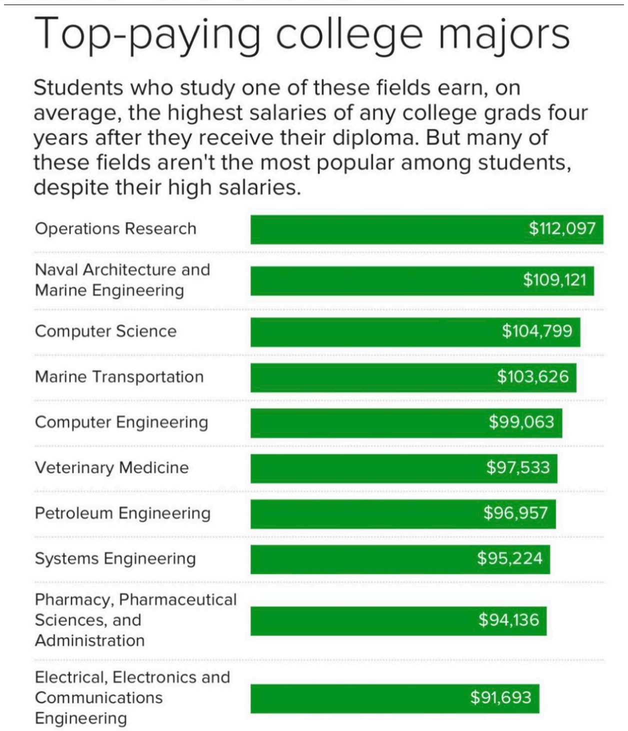


Machine learning (ML) for dynamic decision making (DDM)



Strengthen core math foundations

Applied math + CS for practical engineering applications (for robust, trustworthy, and interpretable decision making)



Pic credit and source: https://www.cbsnews.com/news/college-major-top-and-lowest-earning-majors-impact-on-income-pay/

Course outline

Course outline- congruence of subfields

- Non-linear optimization (introduction)
  - Stochastic search (static decision making) – **introduction**
    - Steepest descent, stochastic approximation
- Machine learning (static)- **introduction**
  - Neural networks/deep learning
- Markov processes (refresher)
  - Markov decision process - **introduction**
  - Dynamic programming – **introduction**
    - Value iteration, policy iteration
- Systems models/ Simulation (refresher)
- Reinforcement learning– **introduction**
  - Q-learning, DQN (Deep Q network), Policy gradient methods

Course outline- congruence of subfields

- Non-linear optimization (introduction)
  - Stochastic search (static decision making) – **introduction**
    - Steepest descent, stochastic approximation
- Machine learning (static)- **introduction**
  - Neural networks/deep learning
- Markov processes (refresher)
  - Markov decision process - **introduction**
  - Dynamic programming – **introduction**
    - Value iteration, policy iteration
- Systems models/ Simulation (refresher)
- (ML for DDM)Reinforcement learning– **introduction**
  - Q-learning, DQN (Deep Q network), Policy gradient methods

Course outline- congruence of subfields

- Non-linear optimization (introduction) (static modeling)
  - Stochastic search (static decision making) – **introduction**
    - Steepest descent, stochastic approximation
- Machine learning (static)- **introduction** (static)
  - Neural networks/deep learning
- Markov processes (refresher) (dynamic modeling)
  - Markov decision process - **introduction**
  - Dynamic programming – **introduction**
    - Value iteration, policy iteration
- Systems models/ Simulation (refresher) (dynamic)
- (ML for DDM)Reinforcement learning– **introduction** (dynamic)
  - Q-learning, DQN (Deep Q network), Policy gradient methods

Course outline- congruence of subfields

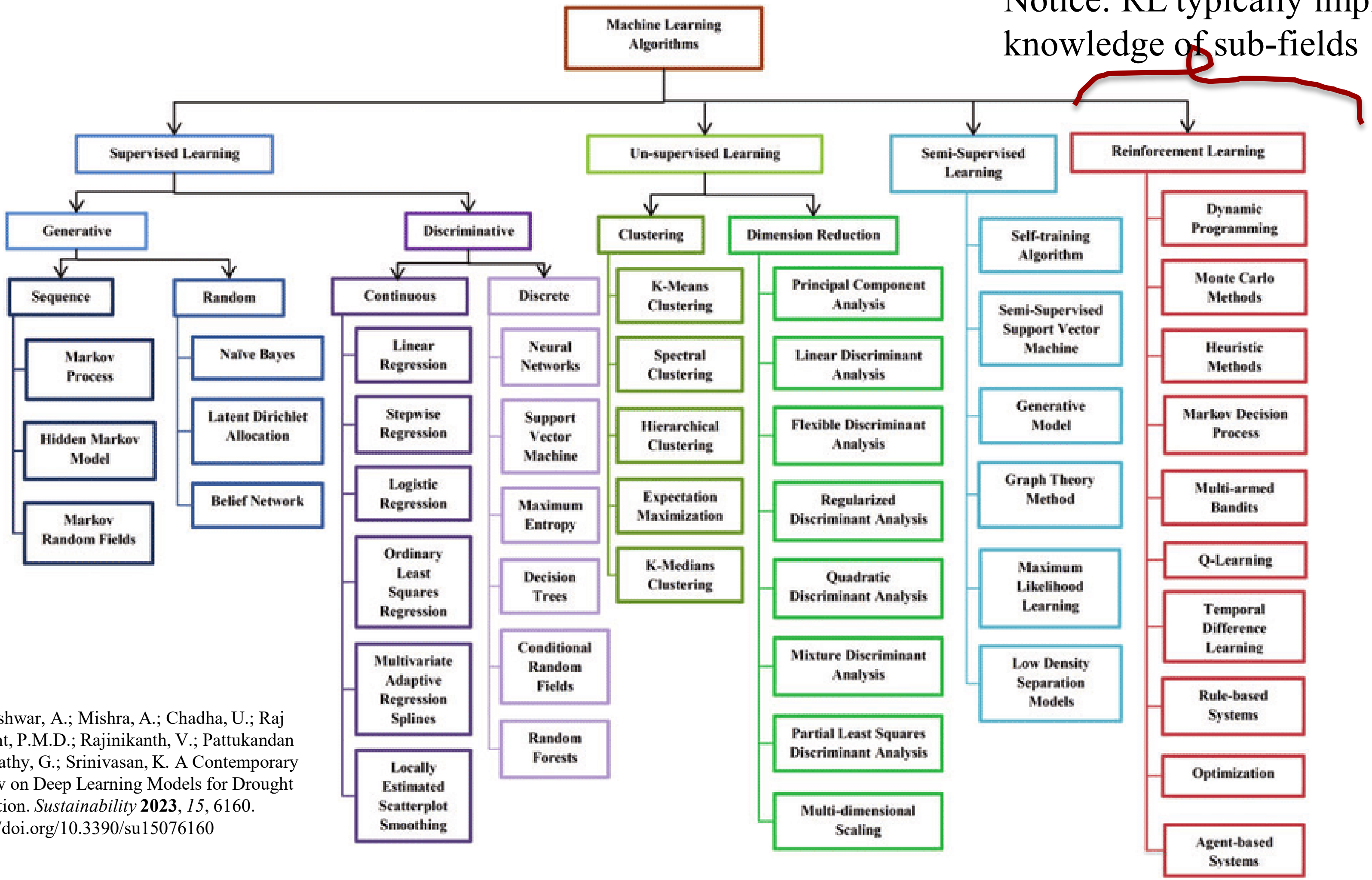
- Optimization (introduction)
  - Stochastic search (static decision making) – introduction
    - Steepest descent
- Machine learning (static)- introduction
  - Neural networks/deep learning
- Markov processes (refresher)
  - Markov decision process - **introduction**
  - Dynamic programming (model-based method for dynamic decision making) – introduction
    - Value iteration, policy iteration
- Systems models/ Simulation (refresher)
- Reinforcement learning (model-free method for dynamic decision making) – introduction
  - Q-learning, DQN (Deep Q network), Policy gradient

Course outline- congruence of subfields

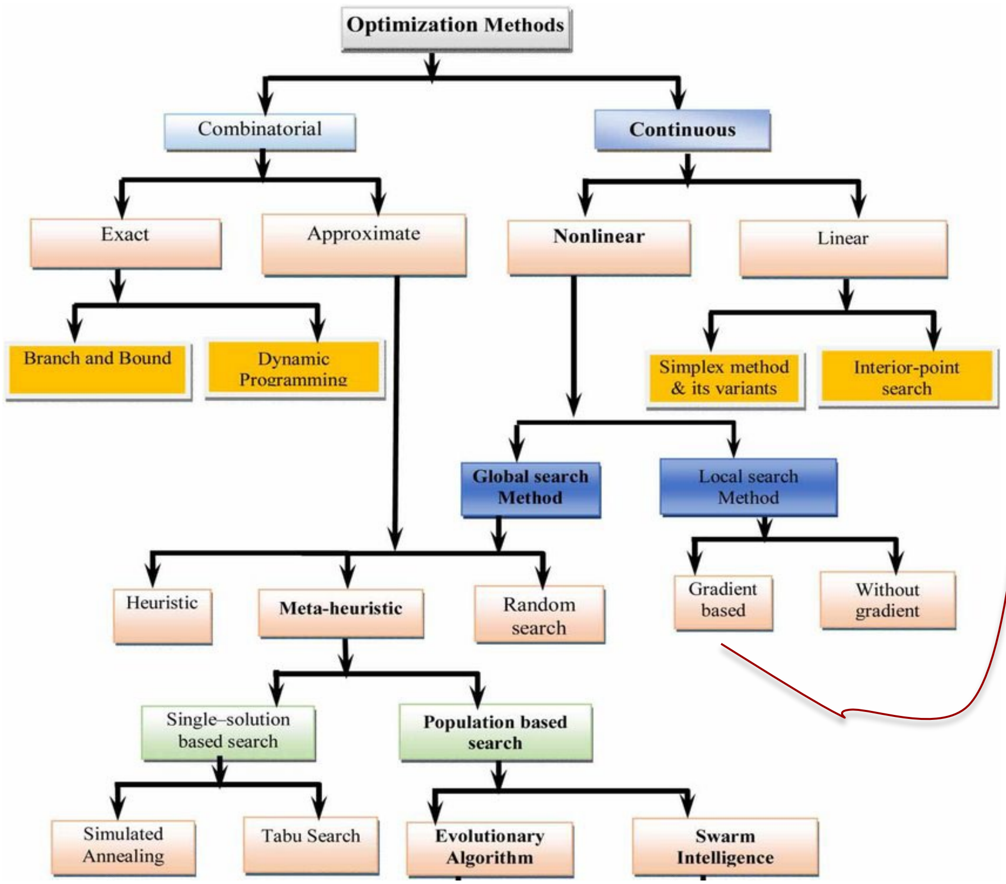
- Optimization (introduction)
  - Stochastic search (static decision making) – introduction (Solution methods for the 'learning' process)
    - Steepest descent
- Machine learning (static)- introduction
  - Neural networks/deep learning (Mechanism to 'learn' decisions/ Sometimes also mechanism to learn environment/simulation)
- Markov processes (refresher)
  - Dynamic programming (model-based method for dynamic decision making) – introduction
    - Value iteration, policy iteration
- Systems models/ Simulation (refresher) (Serves as 'environment' to 'train' neural net to 'learn' decisions)
- ML for DDM (RL algorithms) (model-free method for dynamic decision making) – introduction
  - Q-learning
  - DQN (Deep Q network), Policy gradient

Where do these methods lie in the sub-field taxonomy?

Notice: RL typically implies knowledge of sub-fields



Gyaneshwar, A.; Mishra, A.; Chadha, U.; Raj Vincent, P.M.D.; Rajinikanth, V.; Pattukandan Ganapathy, G.; Srinivasan, K. A Contemporary Review on Deep Learning Models for Drought Prediction. *Sustainability* **2023**, *15*, 6160. https://doi.org/10.3390/su15076160



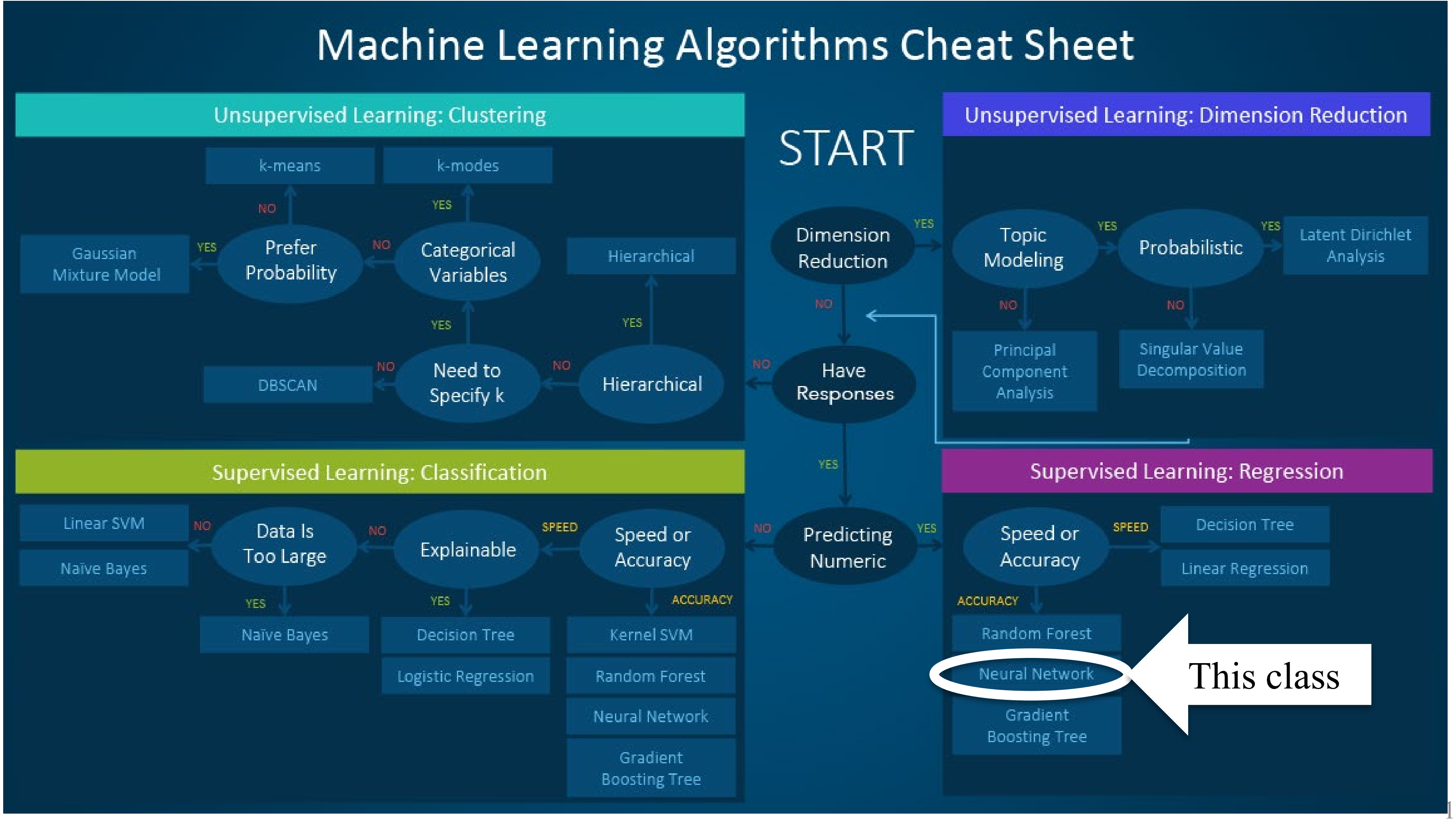
M. Janga Reddy, D. Nagesh Kumar; Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: a state-of-the-art review. *H2Open Journal* 1 January 2020; 3 (1): 135–188. doi: https://doi.org/10.2166/h2oj.2020.128

Gradient (or approximate gradient) -based search methods

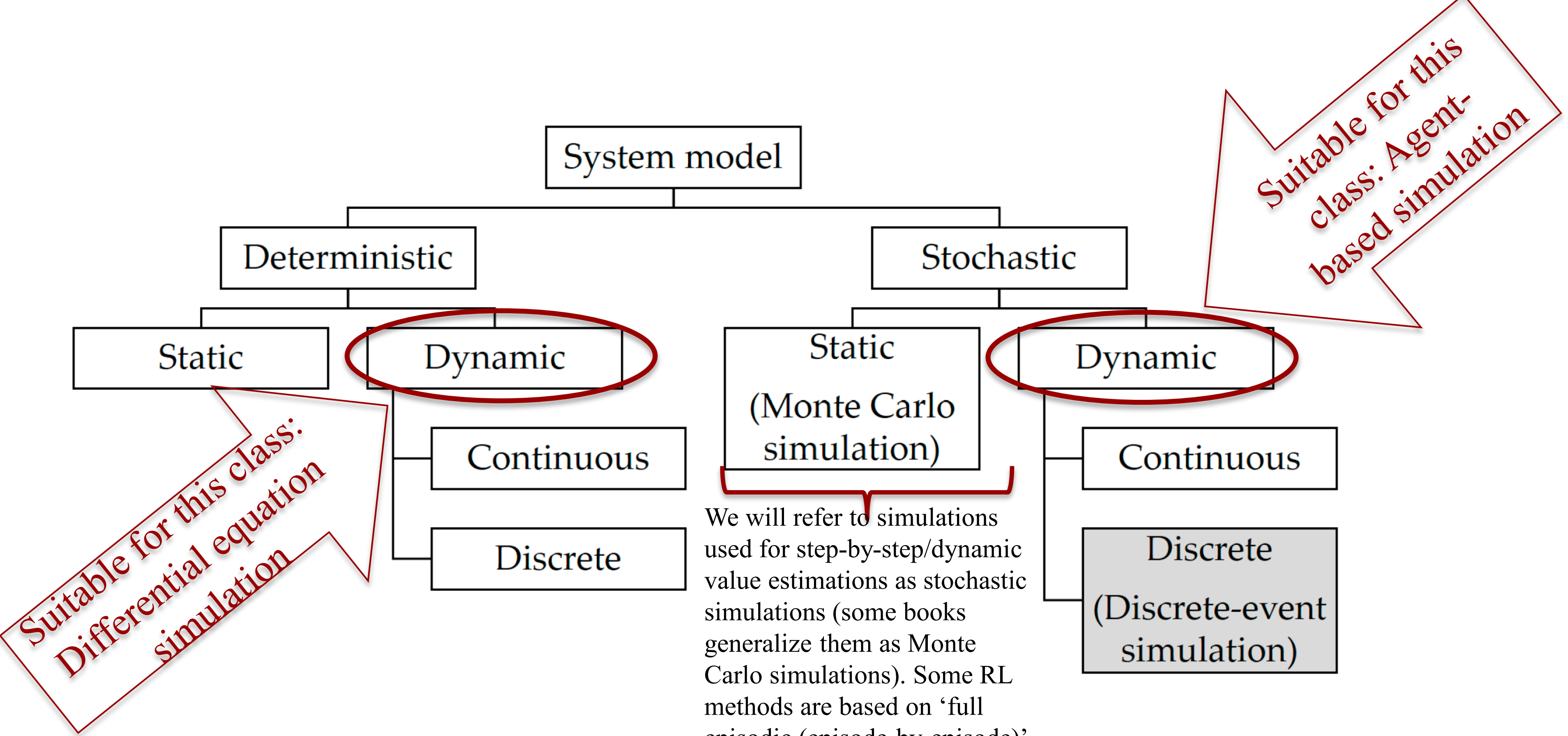
- Exact
- Line search methods
  - Steepest descent
  - Newton's
  - Quasi-Newton's
- Trust-region methods
  - Levenberg-Marquardt

This class

Nocedal and Wright,
Numerical optimization, Springer, 2000

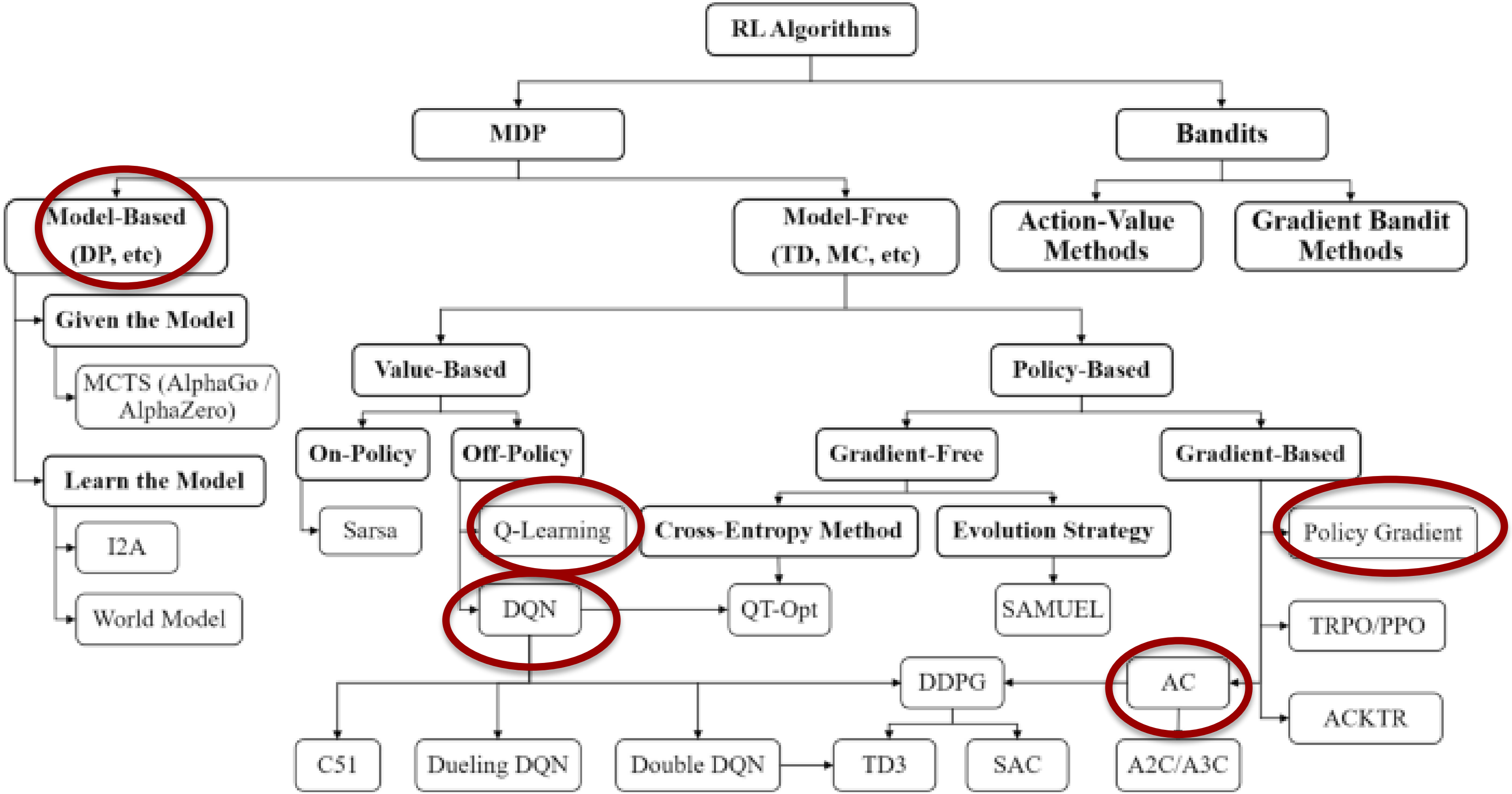


Source: https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use/



We will refer to simulations used for step-by-step/dynamic value estimations as stochastic simulations (some books generalize them as Monte Carlo simulations). Some RL methods are based on 'full episodic (episode-by-episode)' returns and may use Monte Carlo.

Shishvan, M.S.; Benndorf, J. Operational Decision Support for Material Management in Continuous Mining Systems: From Simulation Concept to Practical Full-Scale Implementations. *Minerals* **2017**, *7*, 116. https://doi.org/10.3390/min7070116



Zhang, H., Yu, T. (2020). Taxonomy of Reinforcement Learning Algorithms. In: Dong, H., Ding, Z., Zhang, S. (eds) Deep Reinforcement Learning. Springer, Singapore. https://doi.org/10.1007/978-981-15-4095-0_3

When you have a question

- Course at the congruence of multiple subfields
  - Classes you have taken in UG + new topics
- When you have a question
  - What is your responsibility?
  - What is my responsibility?

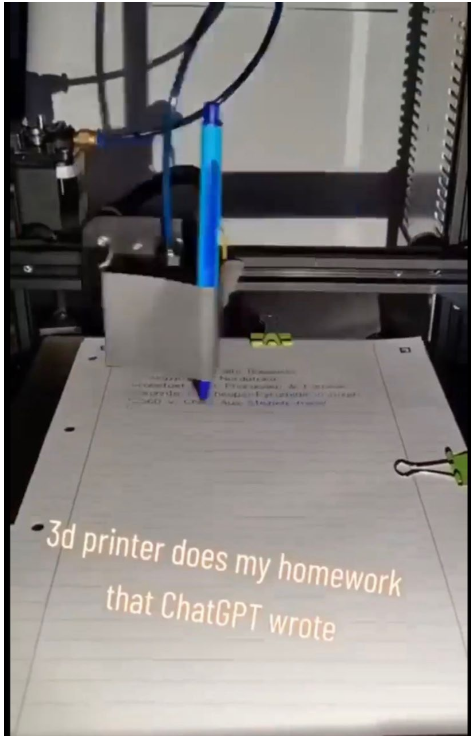
When you have a question

- What is your responsibility?
  - **ASK**
- What is my responsibility?
  - Let you figure it out
  - But **important you ask** so I know and may help as needed

Syllabus

- Assignments: On Canvas
- Grading (most assignments having a Python coding element)
  - ~5 **graded** assignments: 70 points
  - ~5 **ungraded** assignments: 10 points
    - Only 2 points for all this work?
    - Don't be tempted to miss, you will fall behind
  - **Project**: 20 points

Integrity on assignments – "Reclaim, reframe, but revolutionize."



Pic credits: not my own

- "Embrace the wisdom of those who came before you,.."
- But "let the brilliance of your achievements illuminate the world, for borrowed light will always dim."
  - ChatGPT

- Hand in the work of your own creation, built on the knowledge gained from this class.

Integrity on assignments

- I will note on the assignment if/how you can use AI
- But no cheating (asking someone else to do your assignment)
- If you use ChatGPT, Gemini or any other AI apps → cite
  - e.g., I formulated the problem and used ChatGPT for help in writing a code

Reference books (Methods for DDM are perishable)

- Probabilistic Machine Learning: An Introduction (Adaptive Computation and Machine Learning series), Kevin Murphy, 2022 https://probml.github.io/pml-book/book1.html
  - Optimization
  - neural networks
- Probabilistic Machine Learning: Advanced Topics (Adaptive Computation and Machine Learning series), Kevin Murphy, 2022 https://probml.github.io/pml-book/book2.html
  - Reinforcement learning (RL) and DeepRL
- Reinforcement Learning, An Introduction, by Sutton and Barto, second edition, 2018 https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf
  - RL
- Simulation-based Optimization: Parametric Optimization Techniques and Reinforcement Learning, Second edition, by Abhijit Gosavi, Springer, 2014.
  - DP and RL
- Deep reinforcement learning Hands-on, by Maxim Lapan https://github.com/packtpublishing/deep-reinforcement-learning-hands-on
- Reinforcement Learning and Stochastic Optimization, Warren Powell https://castlelab.princeton.edu/RLSO/ https://castlelab.princeton.edu/wp-content/uploads/2019/03/Powell-StochasticOptimizationandLearningMarch212019.pdf
  - RL
  - Stochastic search optimization

Theory and practice

- Class lectures will focus on the theoretical foundations
- Assignments will focus on the application and involves coding

Methods for DDM are perishable – strengthen foundations

Questions?

UMass Amherst

The Commonwealth's Flagship Campus