JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

CSARIM 2025

Transfusion guidance based on reinforcement learning

Prof. Dr. Jens Meier

Clinic of Anesthesiology and Intensive Care Medicine

Kepler University Clinic Linz, Austria

Conflict of interest

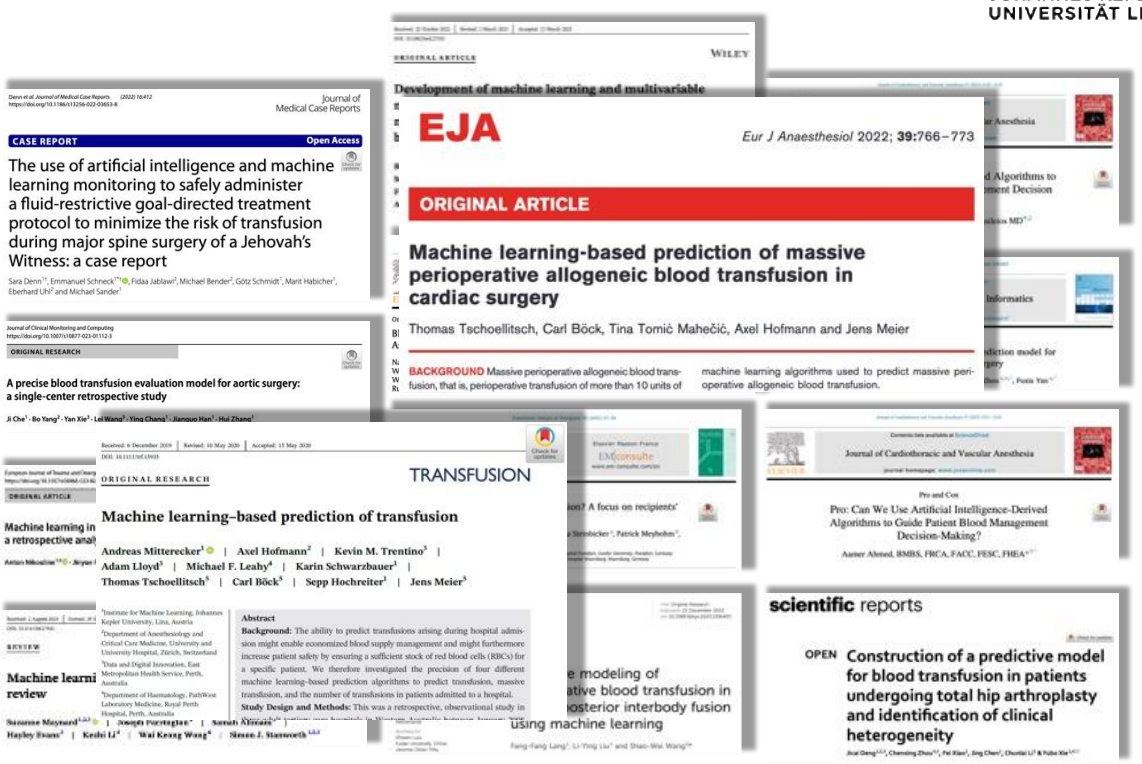
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

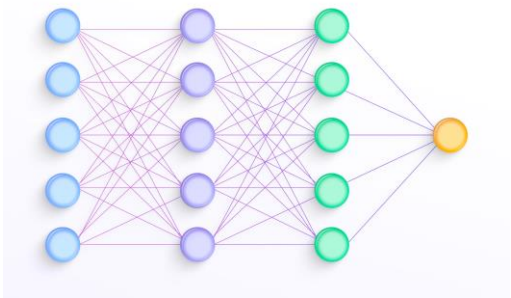
What we already know…

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum



clinical data



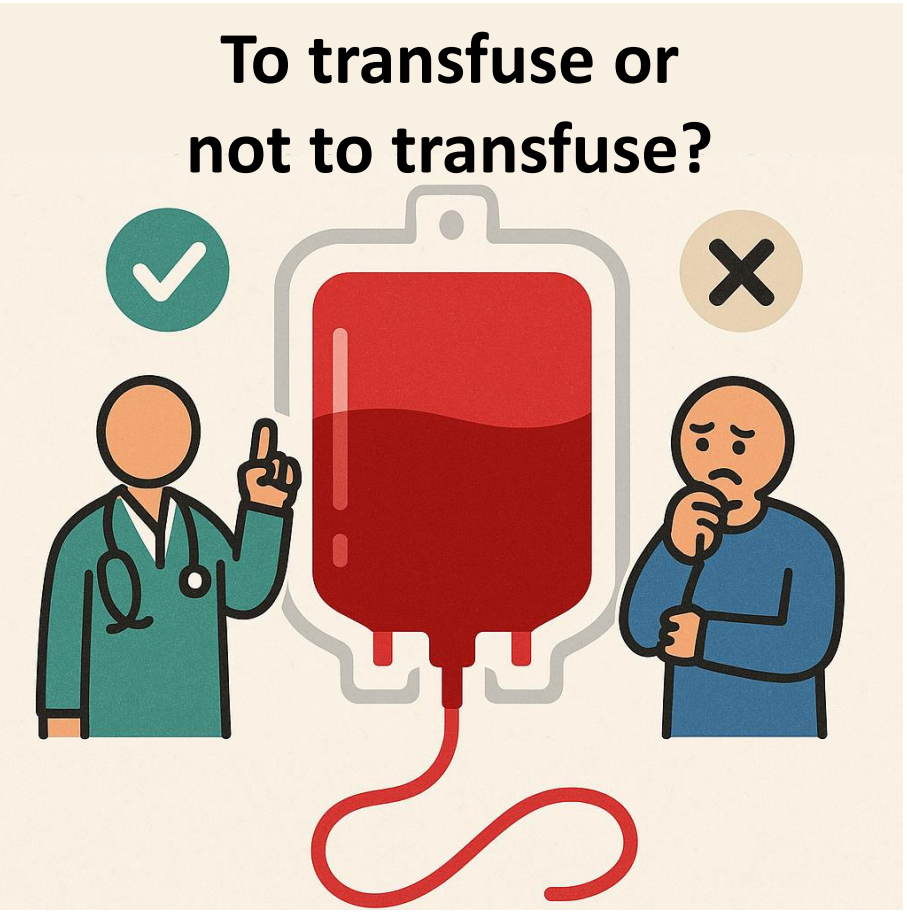
algorithm mimics supervisor

OUTCOME (transfusion)

The elephant in the room

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum



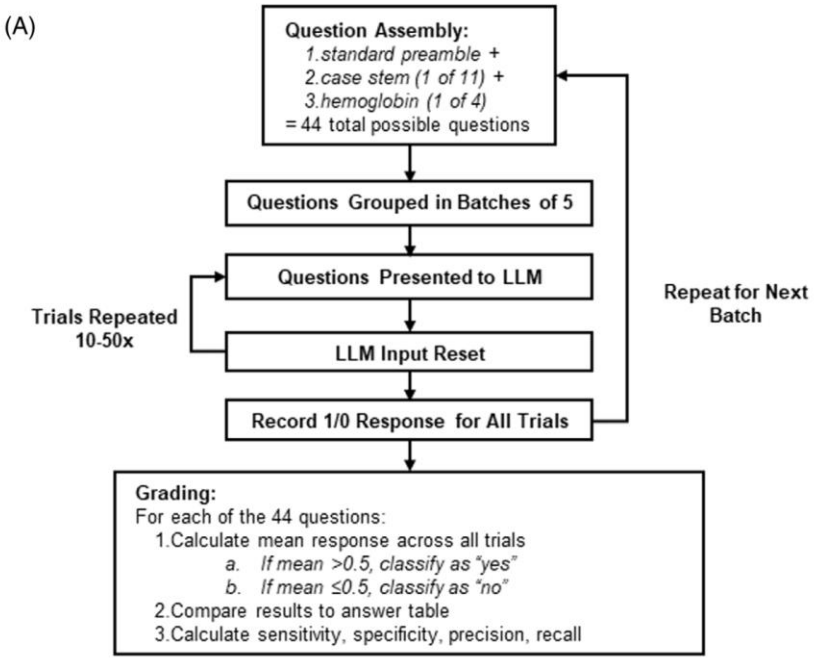
The solution everyone is talking about…

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

ChatGPT

Google Bard AI

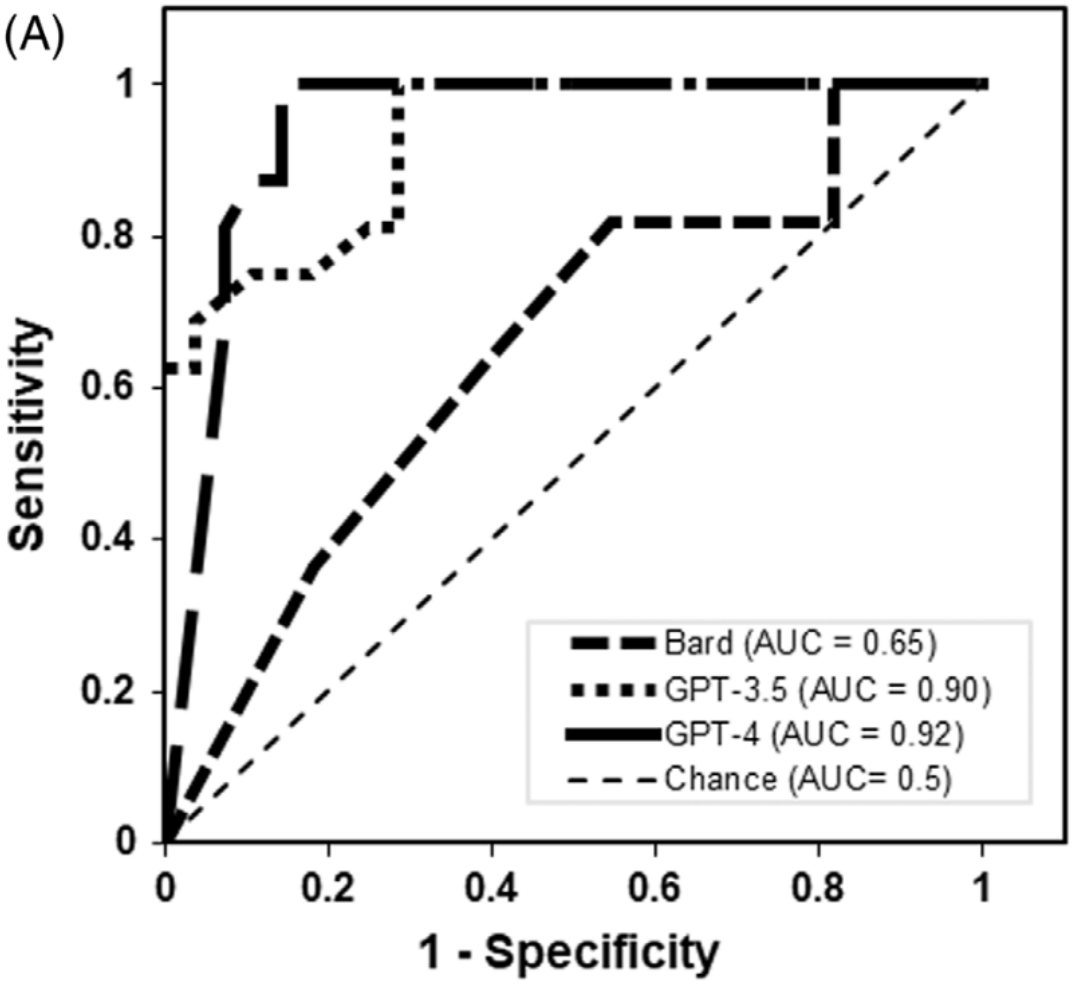


Hurley, Transfusion, 2024

LLMs „understand“ transfusion indications

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

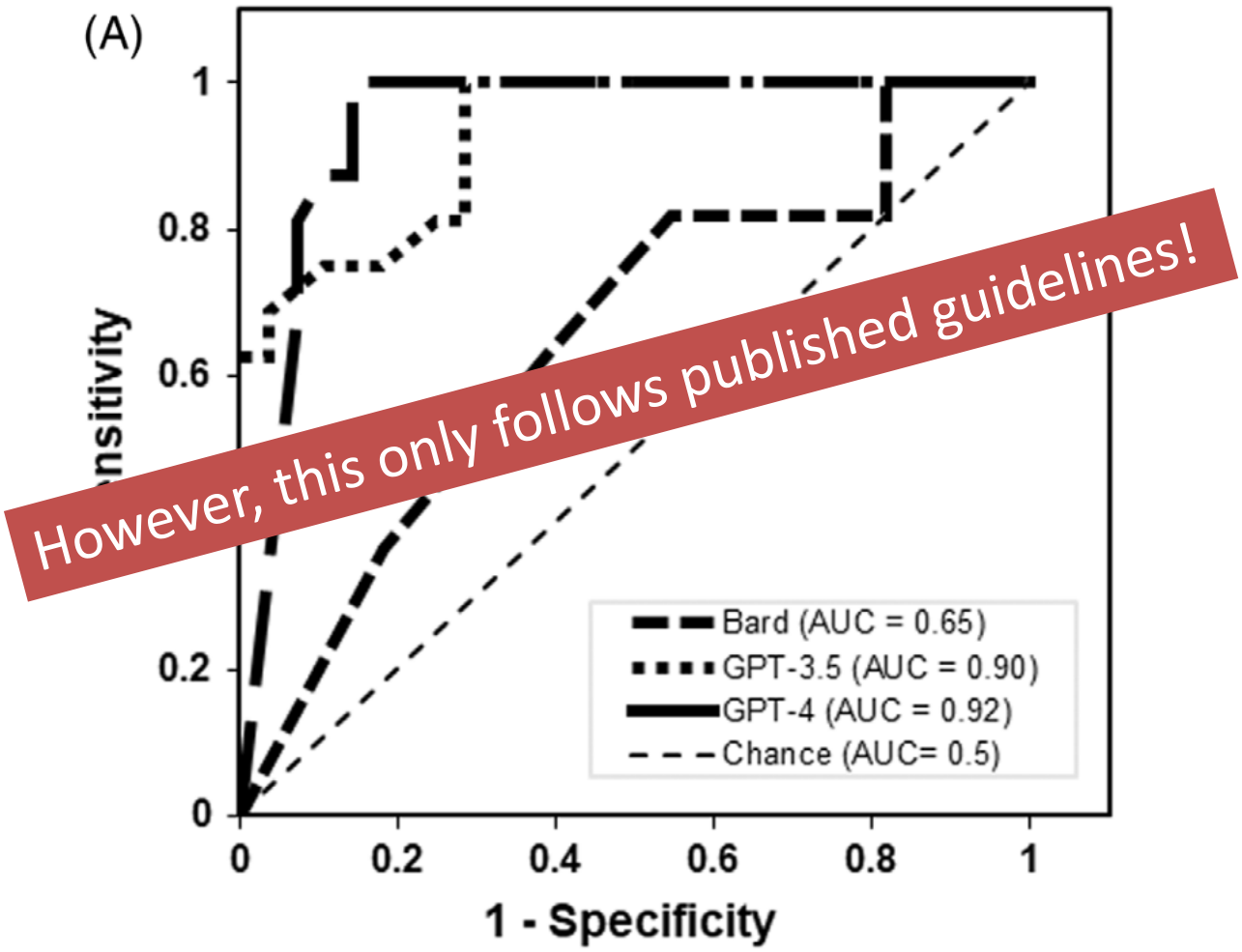


Hurley, Transfusion, 2024

LLMs „understand“ transfusion indications

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

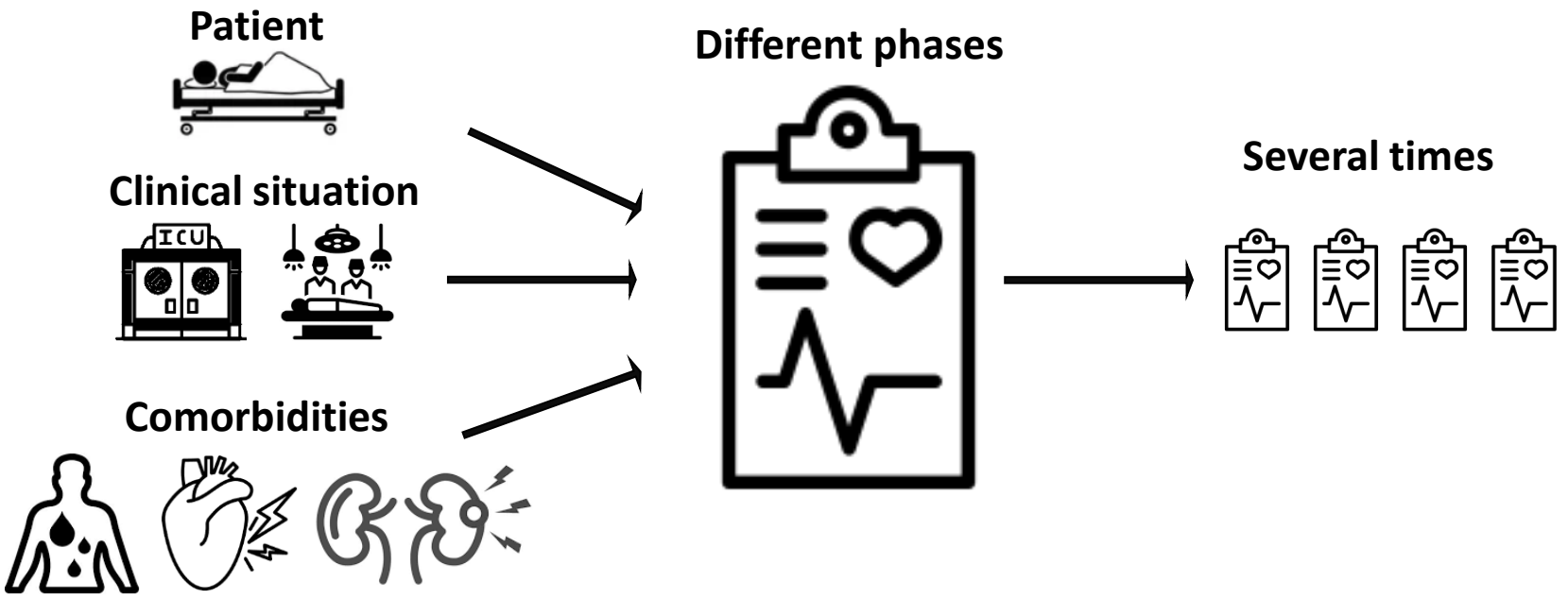


However, this only follows published guidelines!

Hurley, Transfusion, 2024

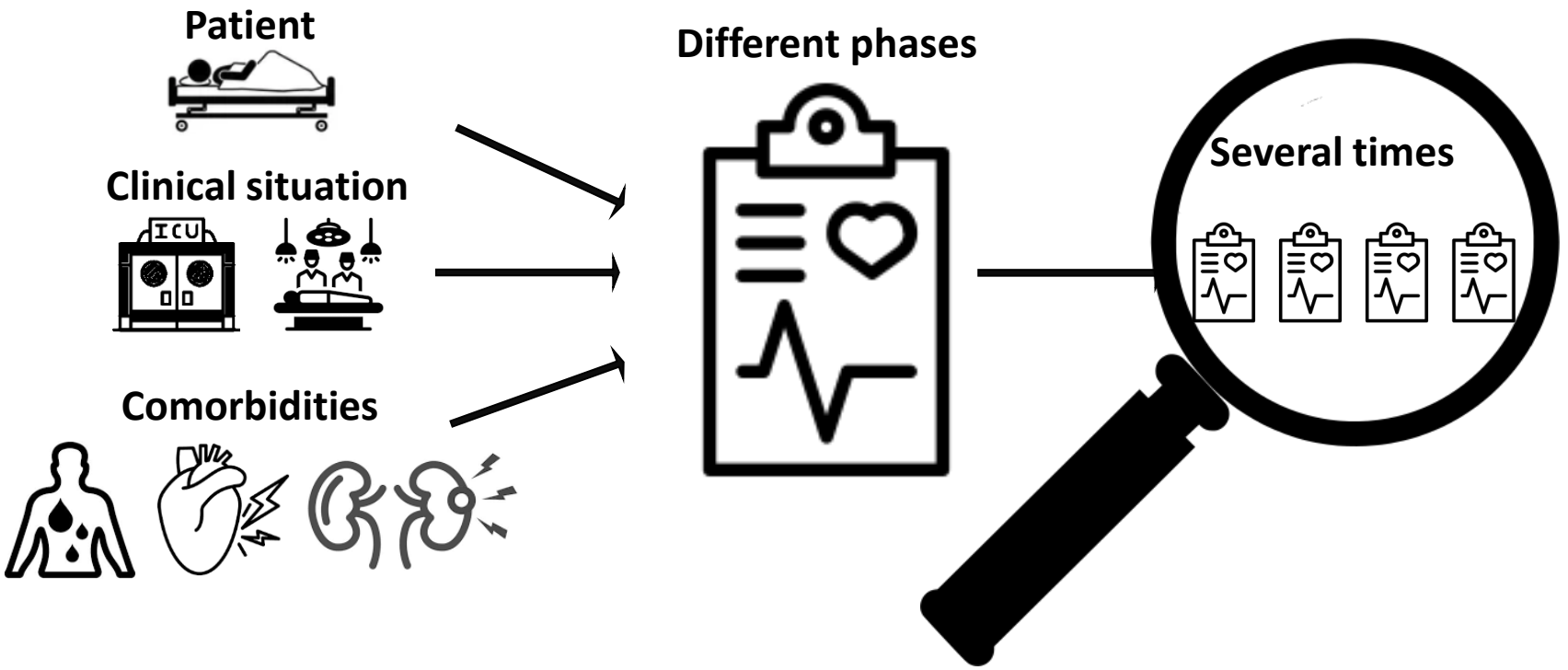
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum



JOHANNES KEPLER UNIVERSITÄT LINZ

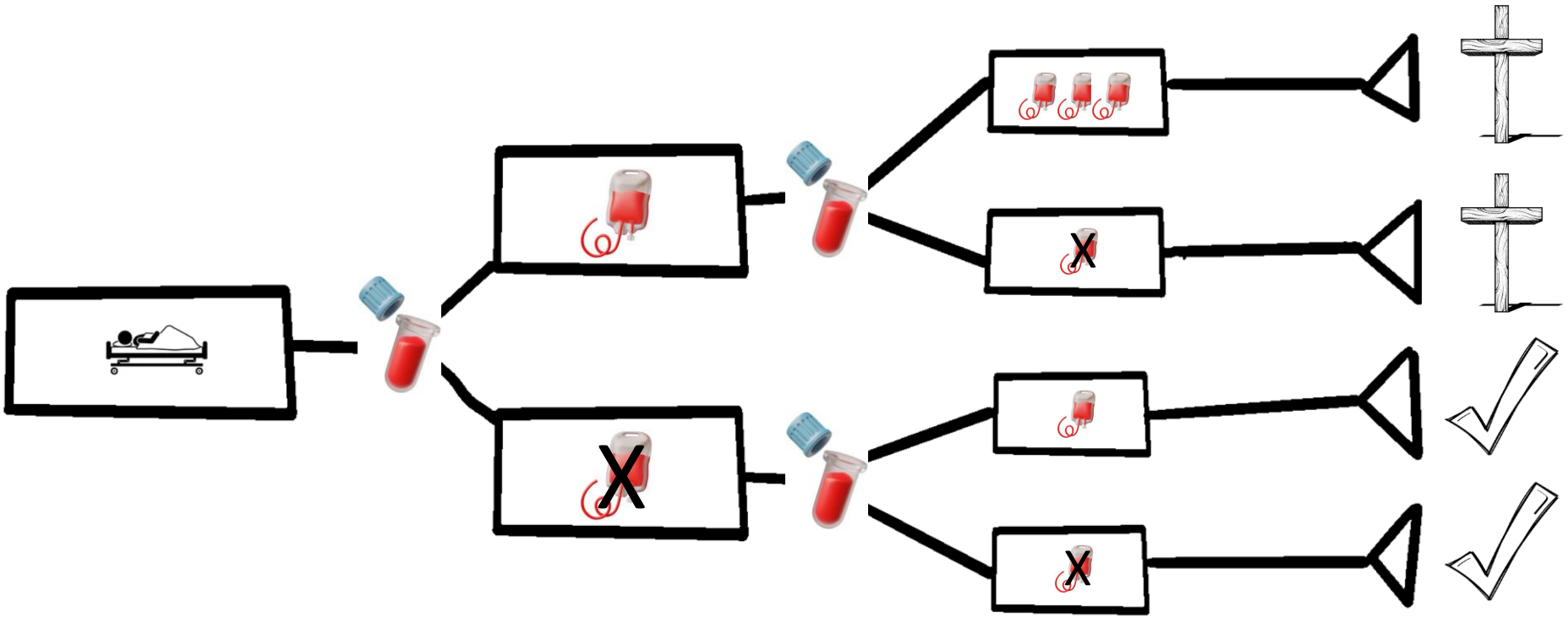
Kepler Universitäts Klinikum



In daily clinical practice: We have to decide often!

JOHANNES KEPLER UNIVERSITÄT LINZ

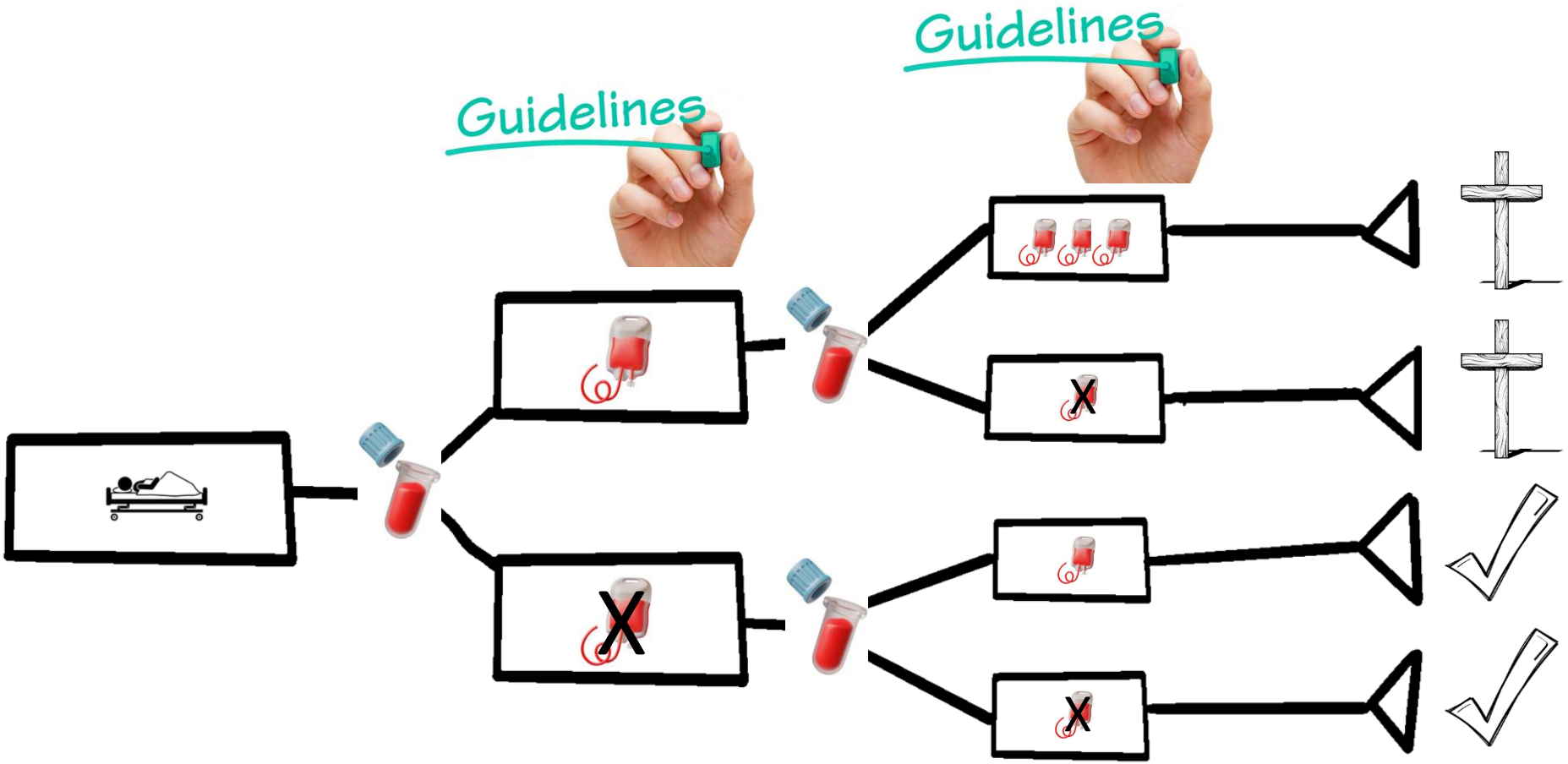
Kepler Universitäts Klinikum



In daily clinical practice: We have to decide often!

JOHANNES KEPLER UNIVERSITÄT LINZ

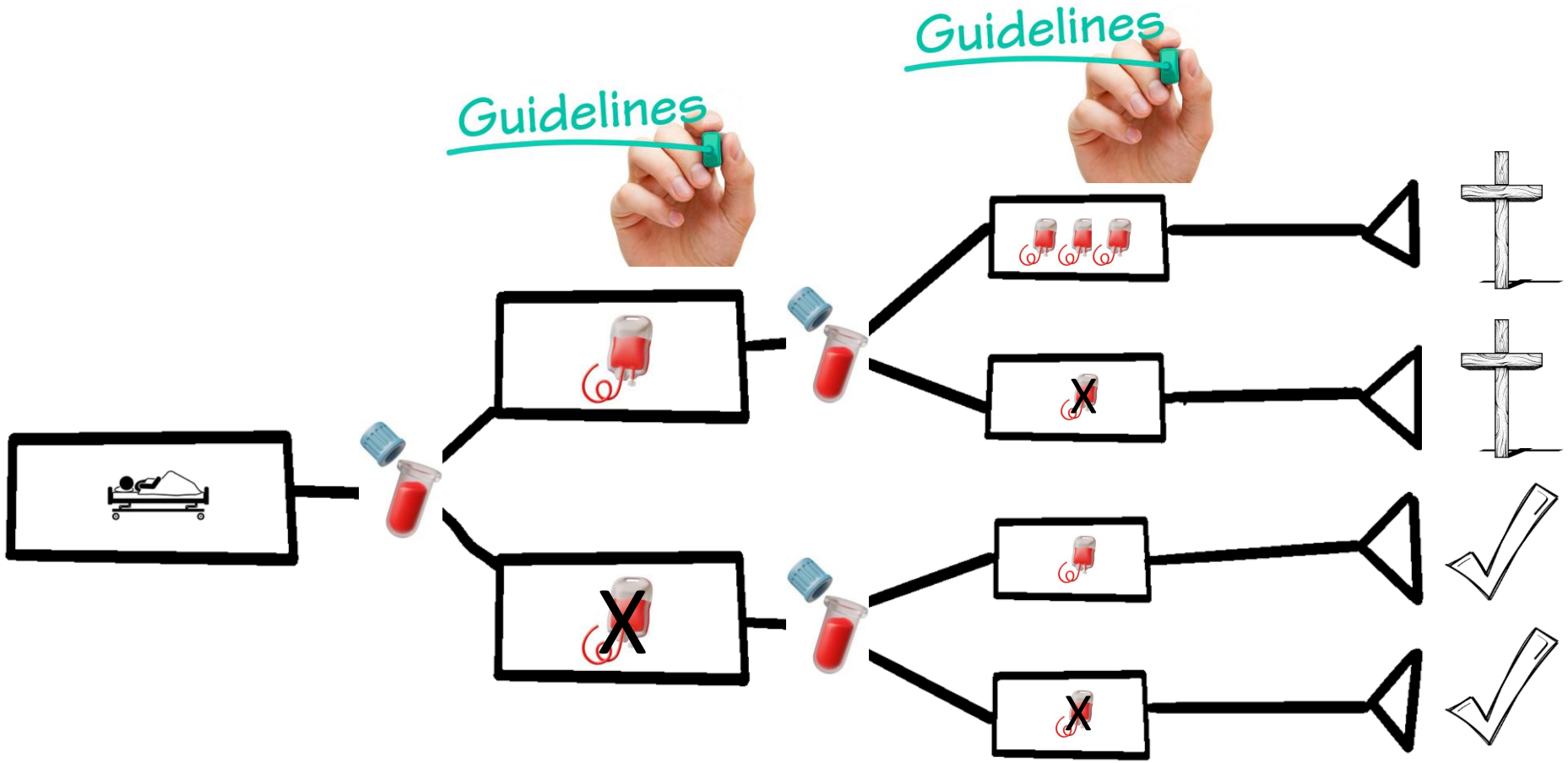
Kepler Universitäts Klinikum



In daily clinical practice: We have to decide often!

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum



1. Same recommendation ▶ various effects!

2. Guidelines are a coarse instrument

The role of these guidelines…

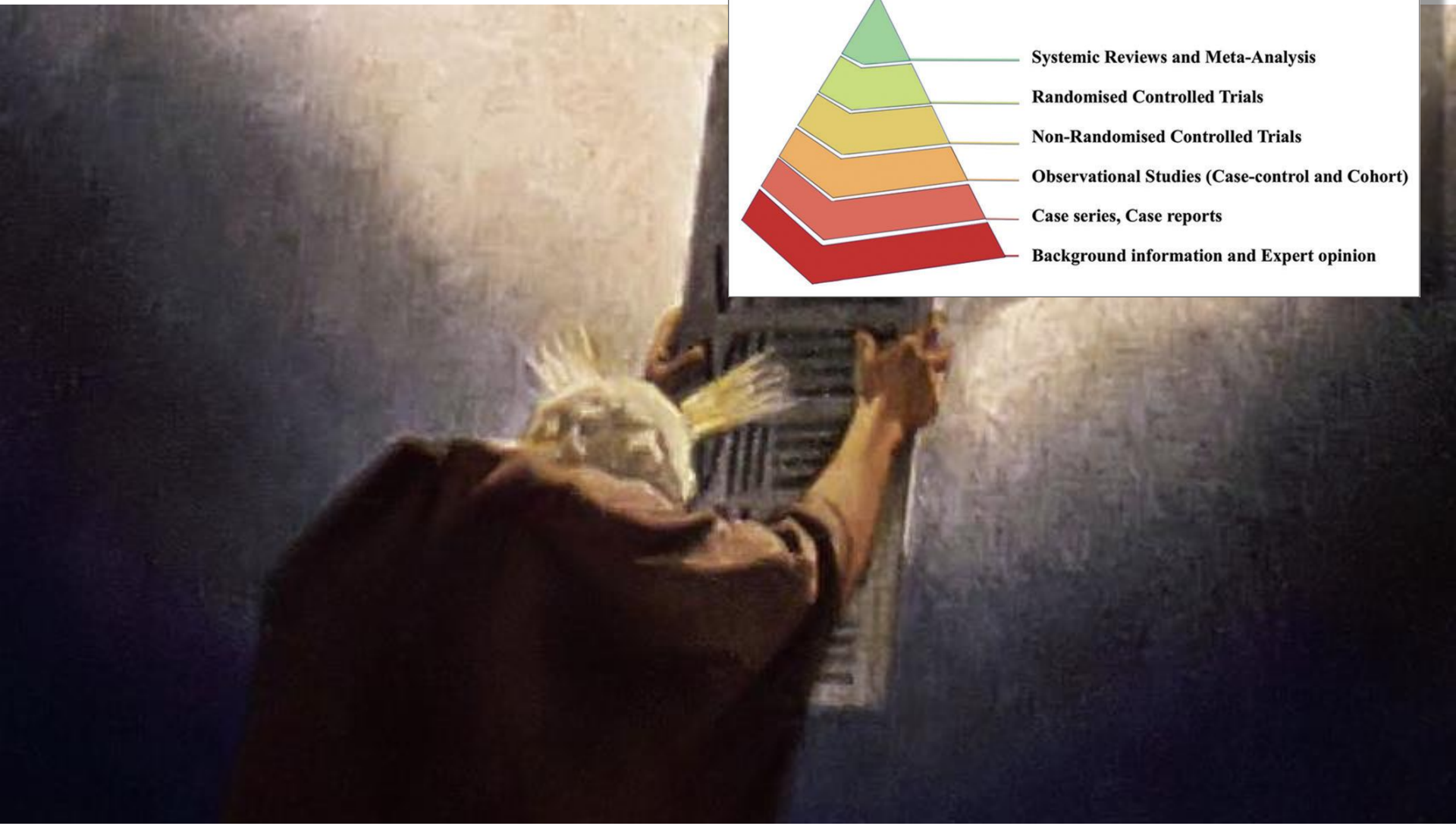
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

How do we get these guidelines?

JOHANNES KEPLER UNIVERSITÄT LINZ

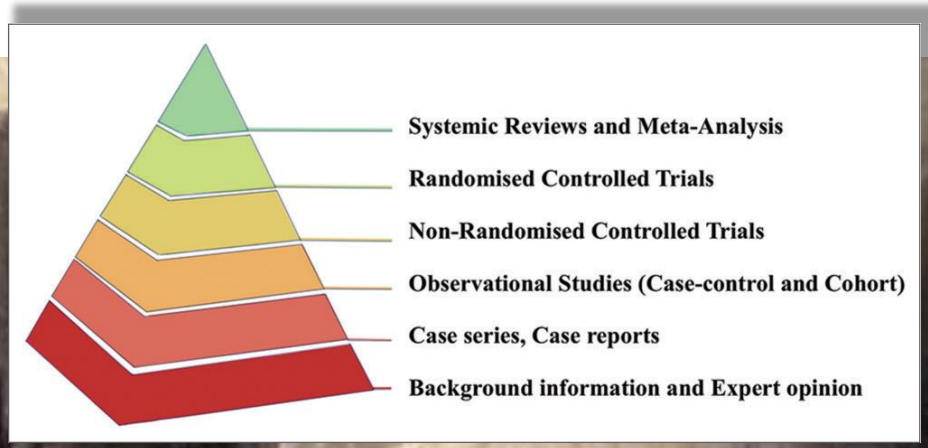
Kepler Universitäts Klinikum

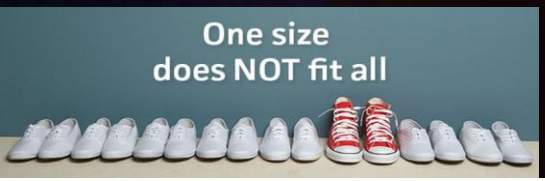


Shortcoming of this approach

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

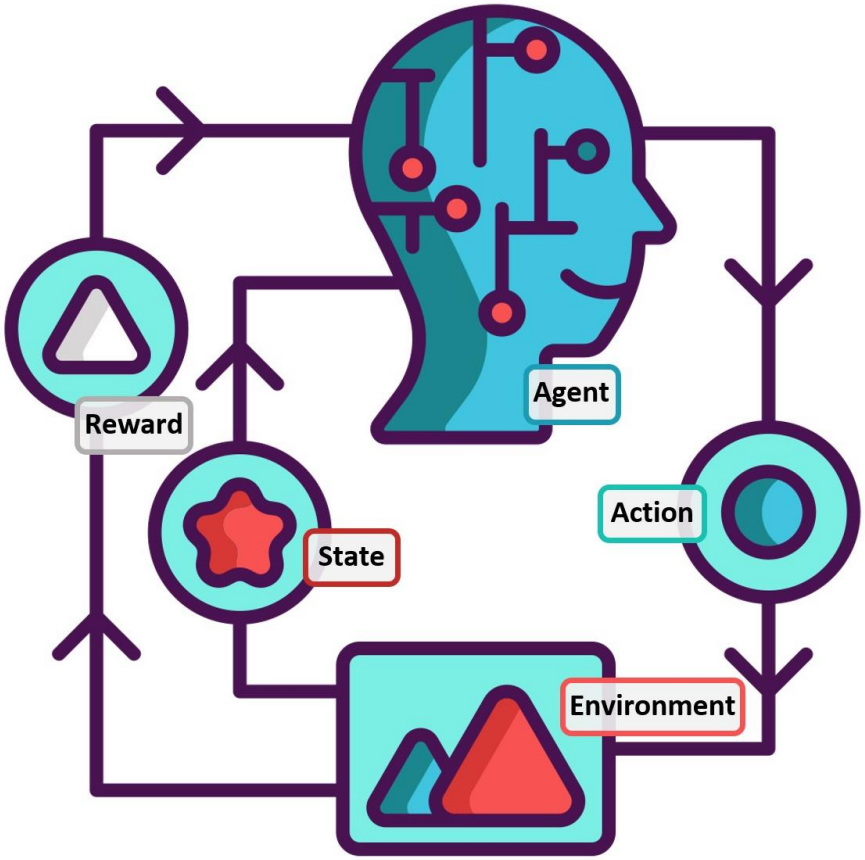


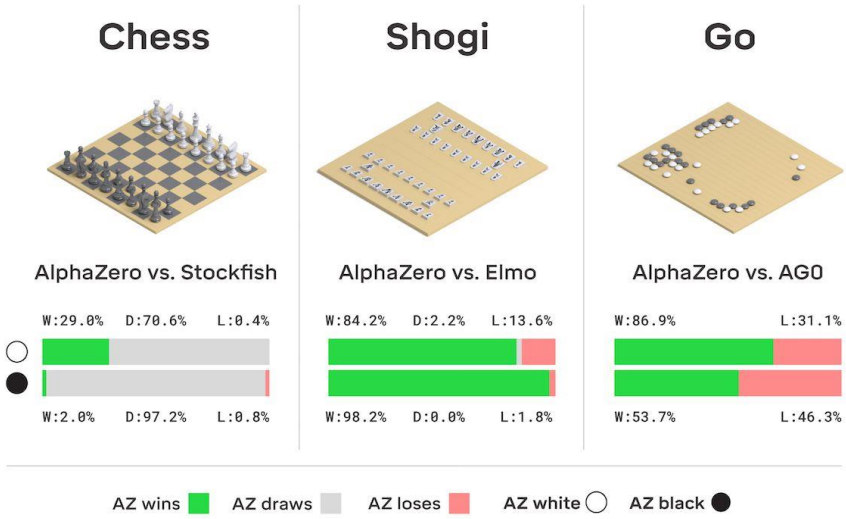


The new paradigm
Reinforcement Learning!

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum





Silver, Science, 2018

The algorithm: reinforcement learning

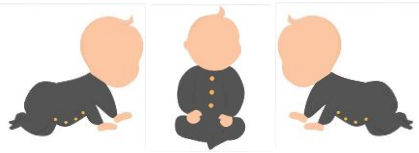
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder

The algorithm: reinforcement learning

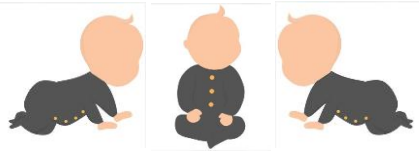
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder

The algorithm: reinforcement learning

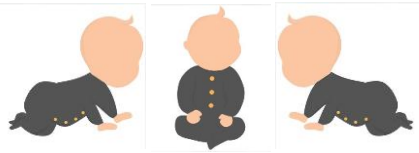
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

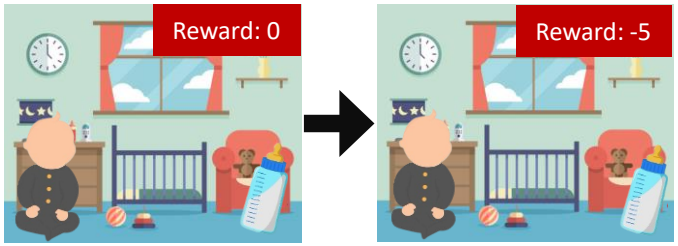
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

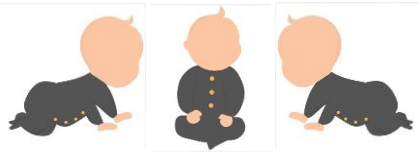
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder

The algorithm: reinforcement learning

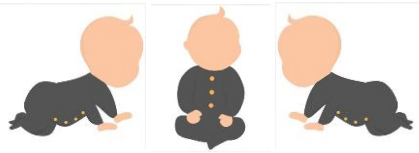
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

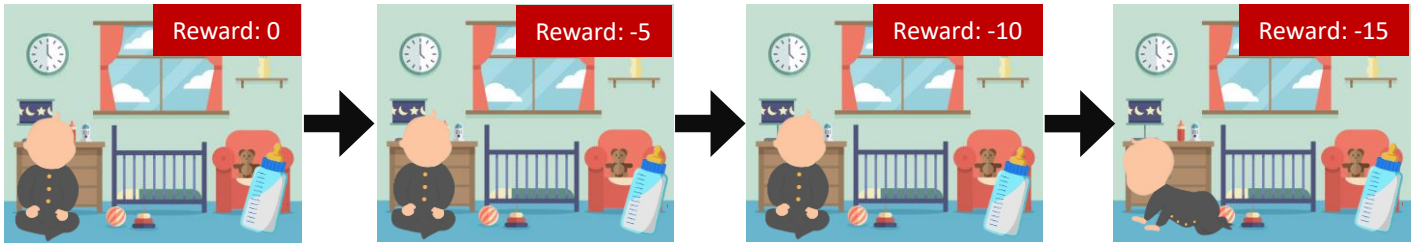
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

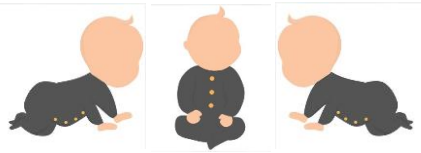
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

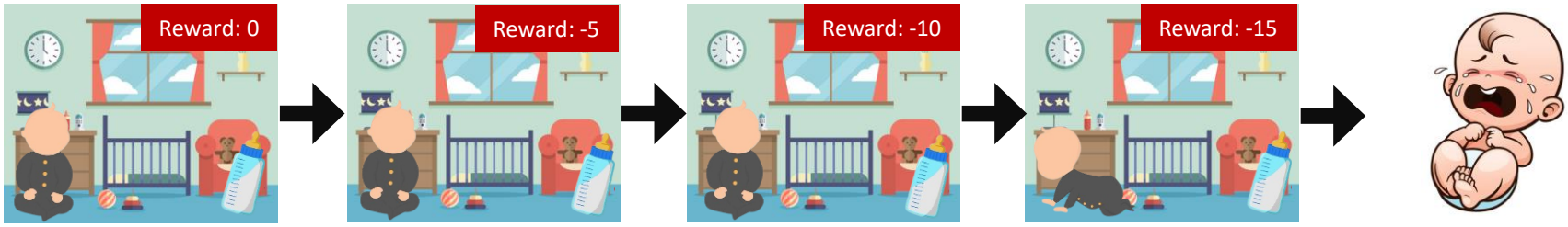
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

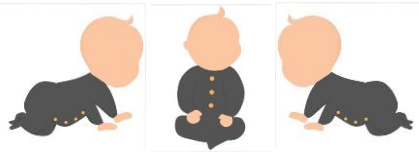
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

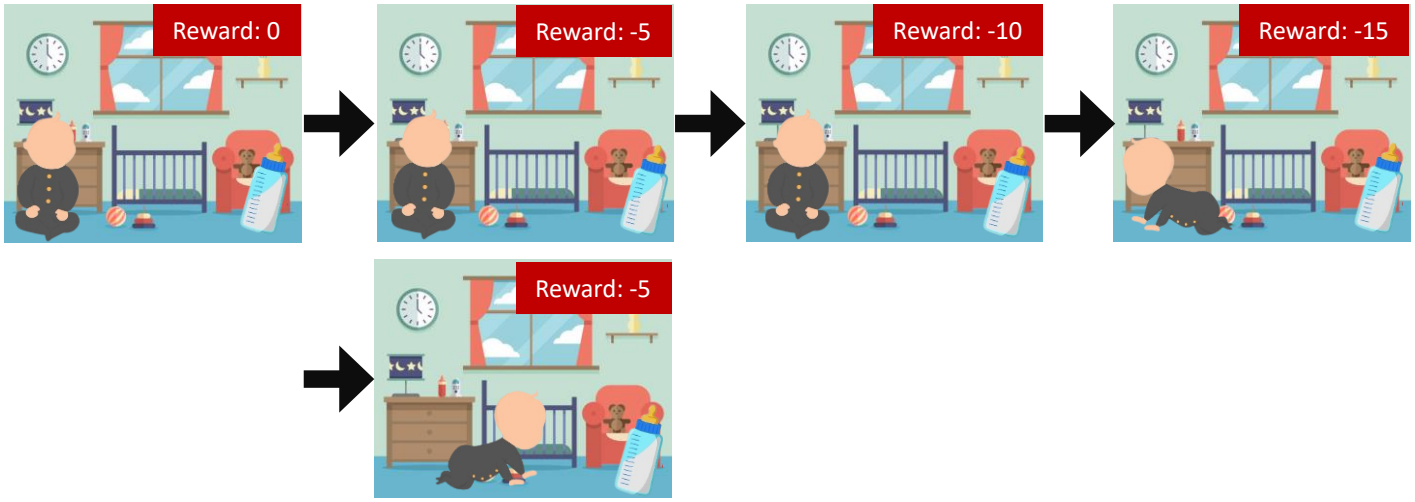
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

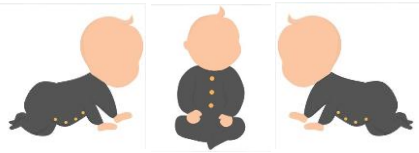
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

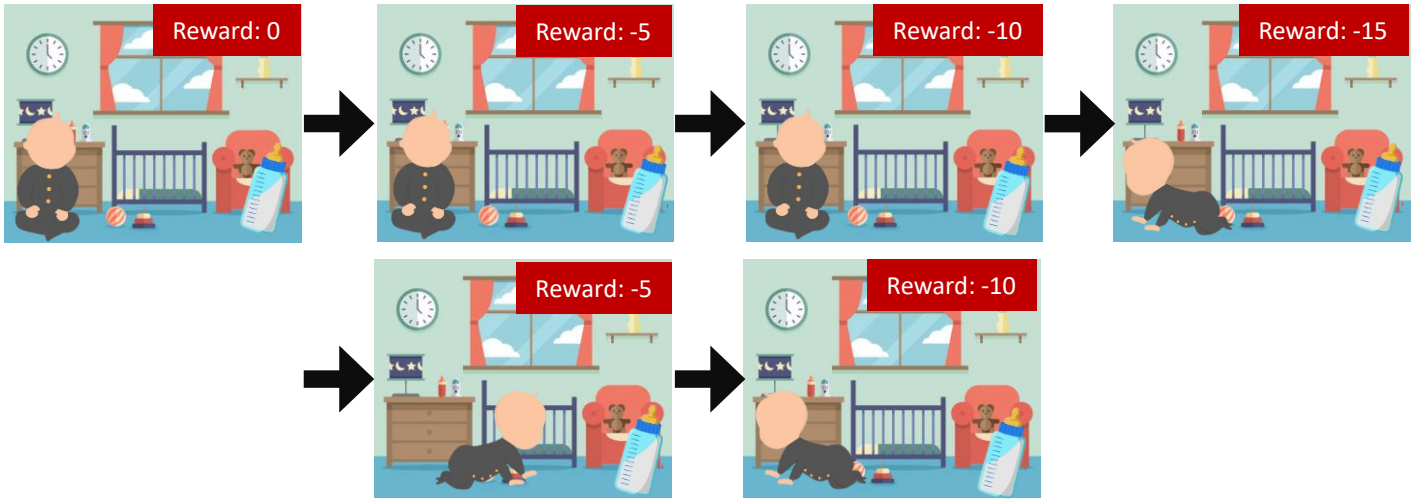
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

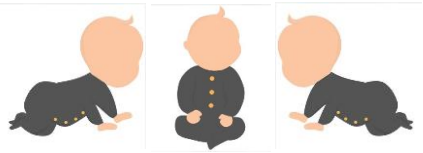
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

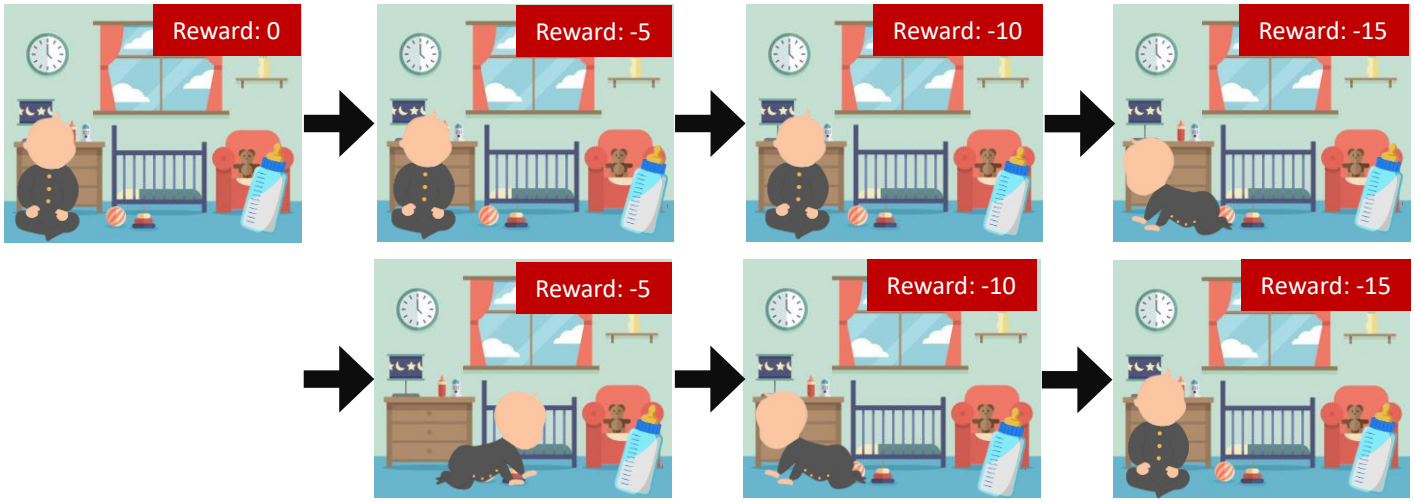
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

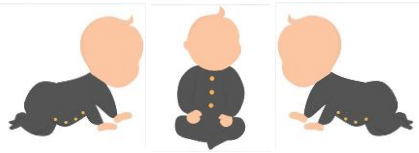
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

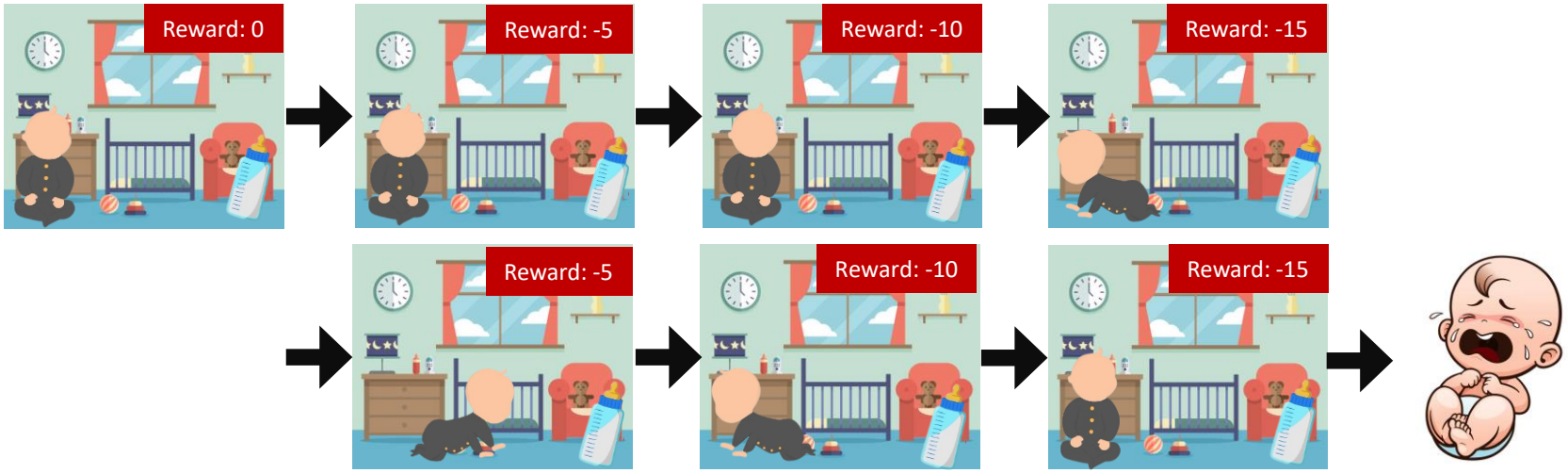
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

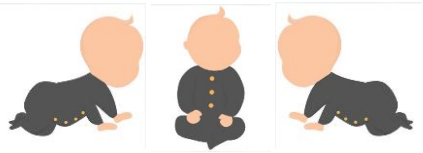
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

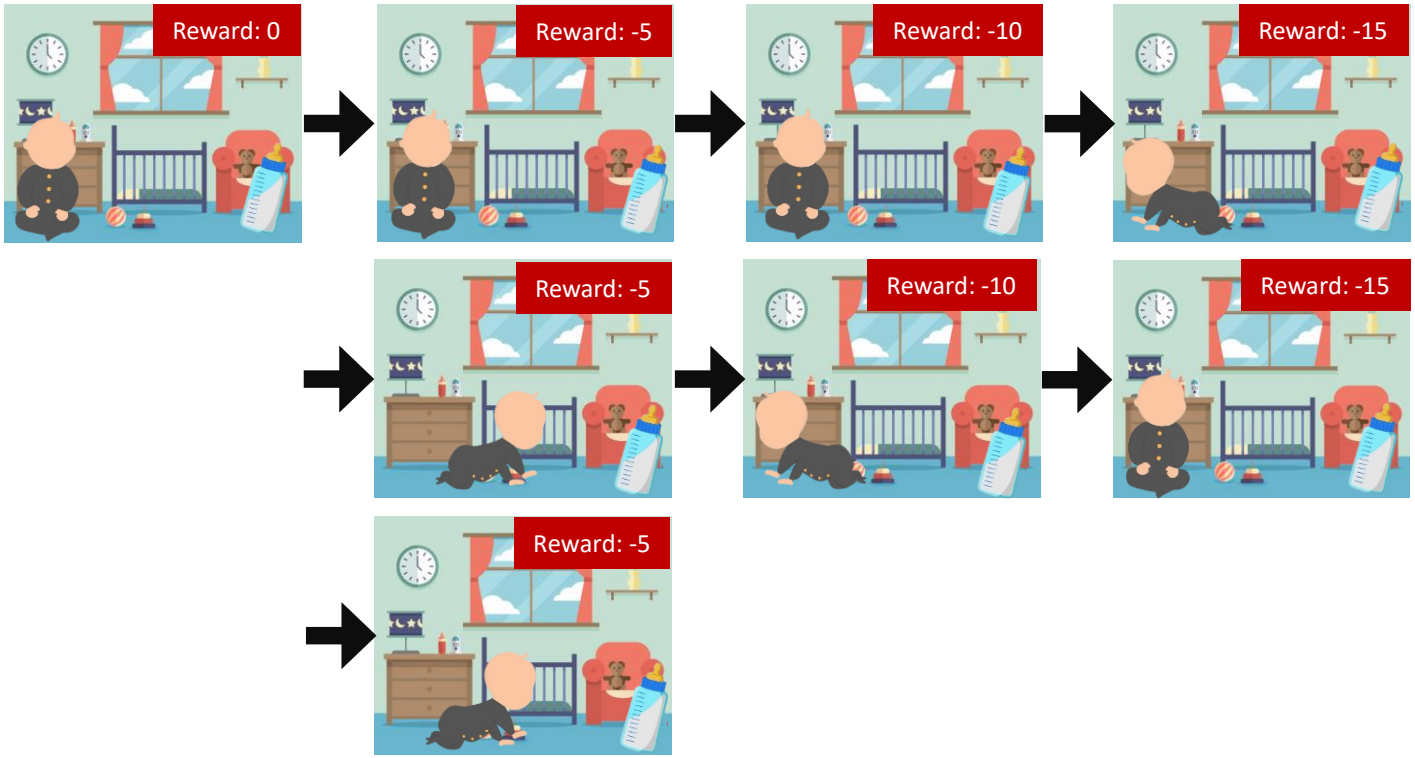
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

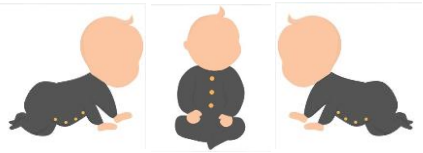
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

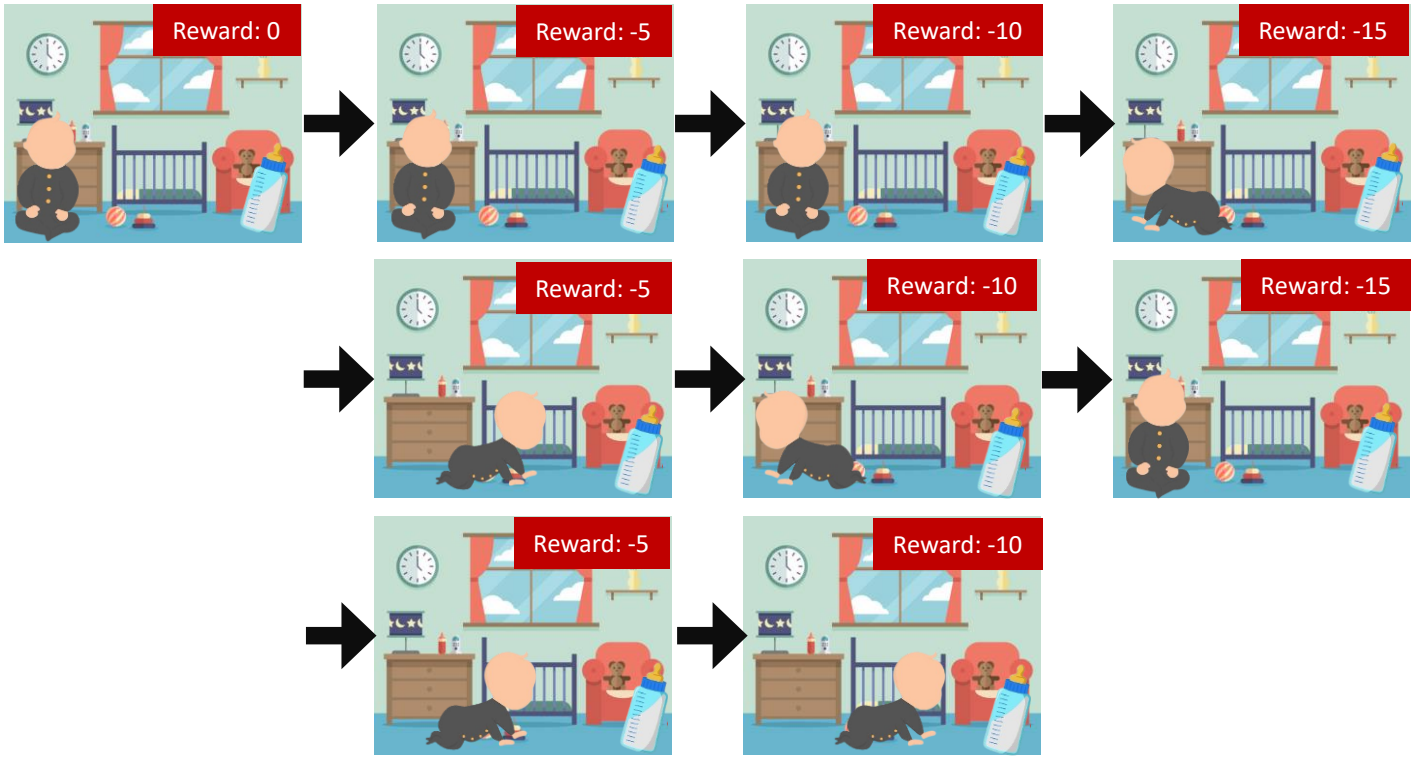
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

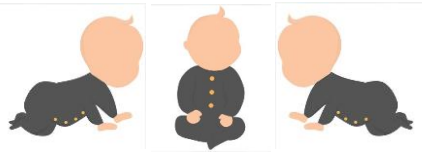
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

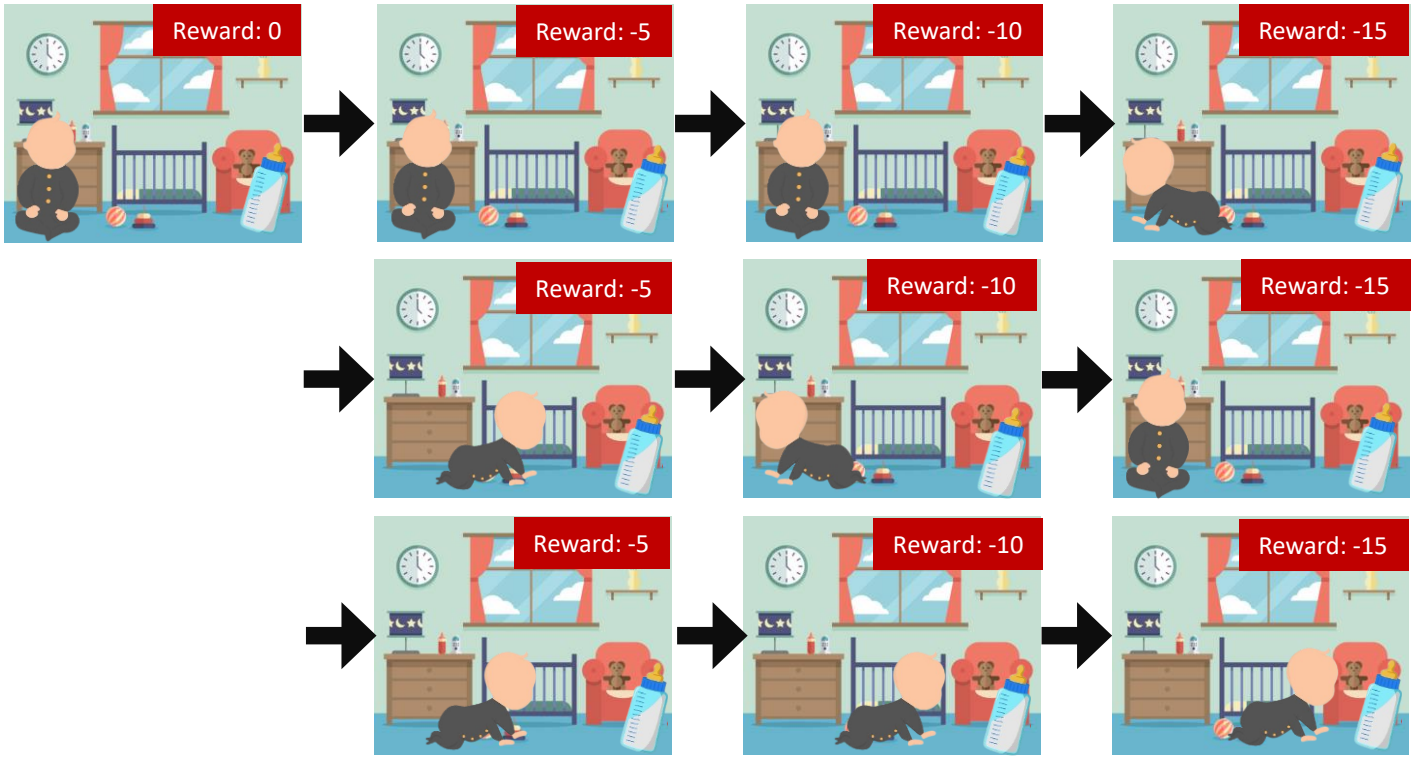
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

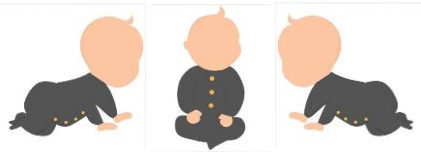
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

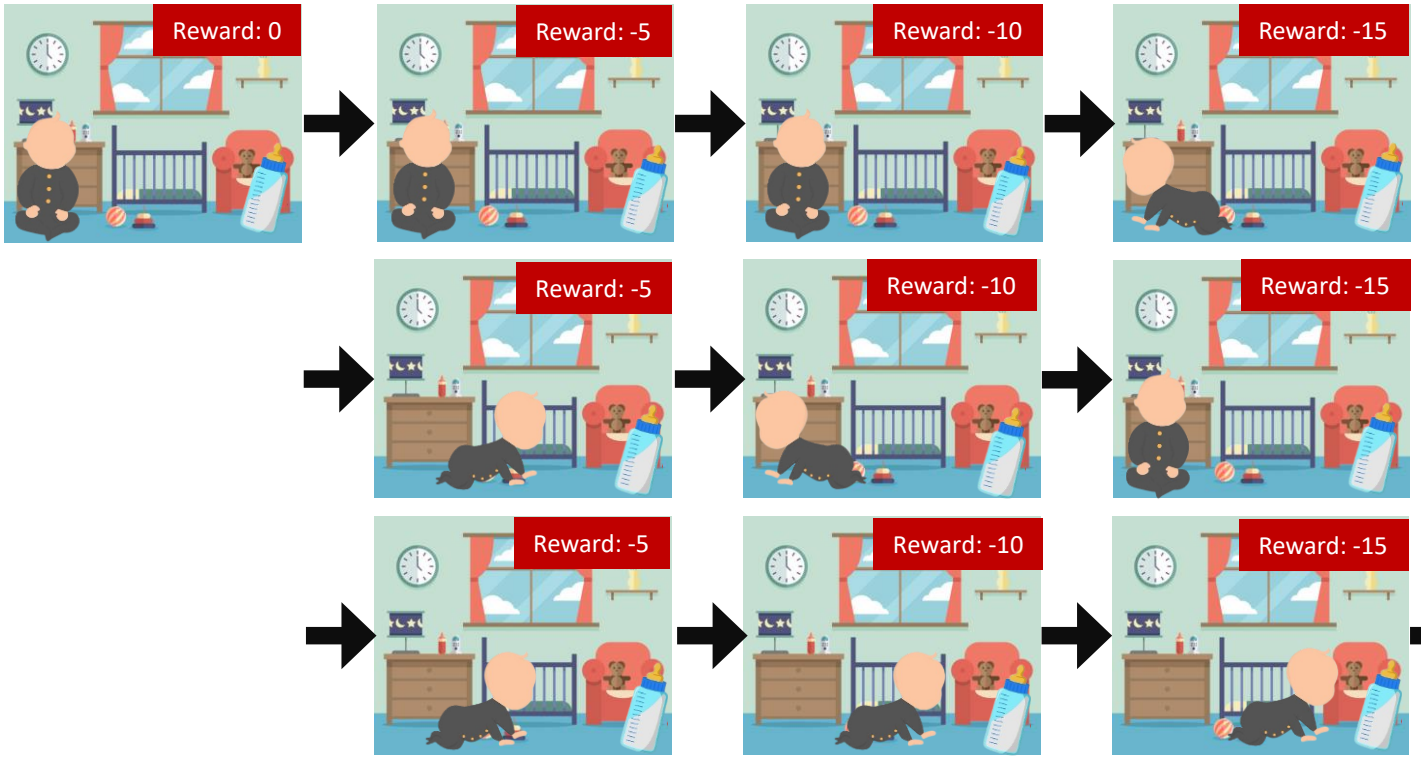
Environment
baby‘s room

State
hungry, somewhere



Action
Crawling, sitting

Reward
Feeder



The algorithm: reinforcement learning

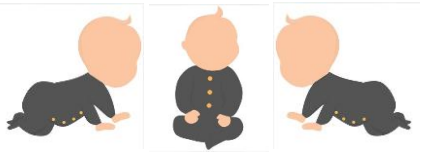
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

Agent
baby

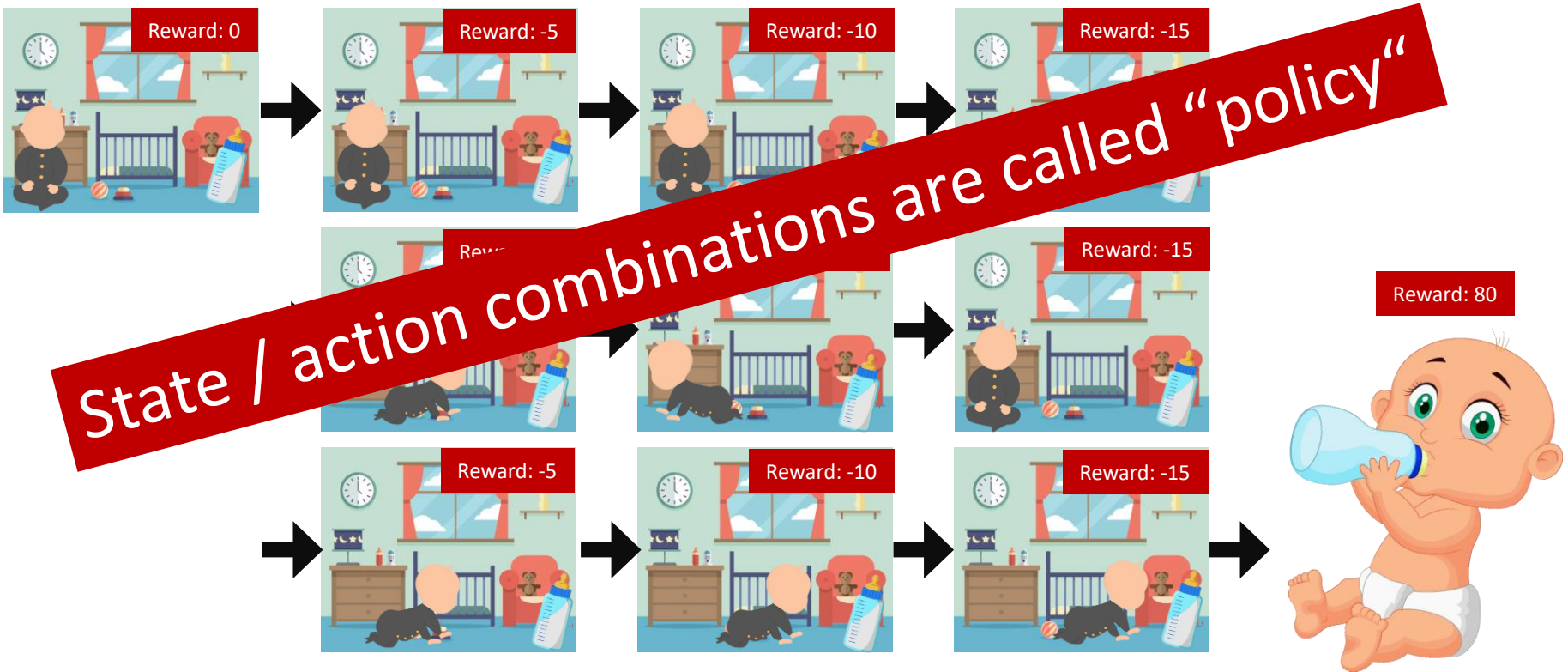
Environment
baby's room

State
hungry, somewhere



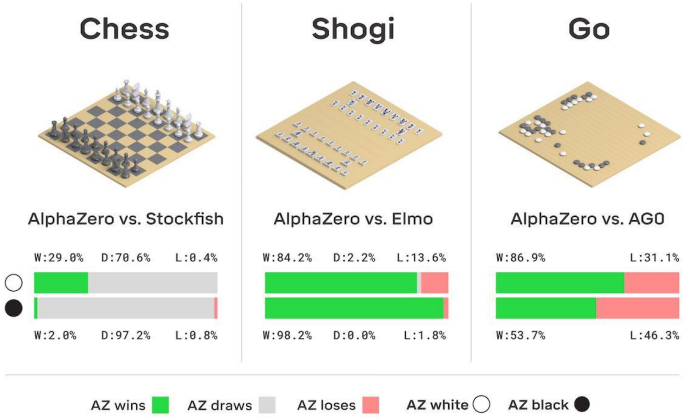
Action
Crawling, sitting

Reward
Feeder

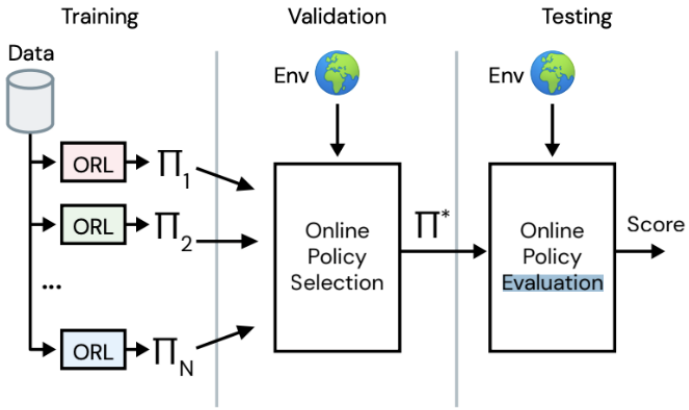


State / action combinations are called "policy"

The difference between online and offline reinforcement learning



Online reinforcement learning

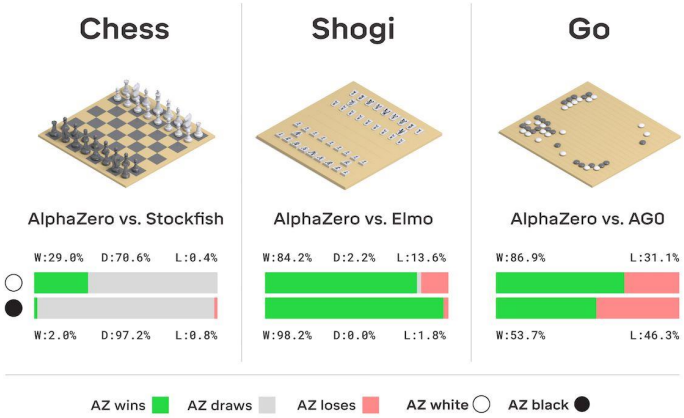


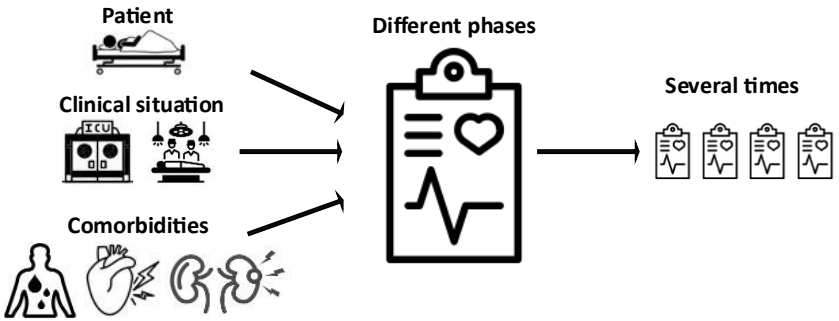
Gnanasekaran, Stanford, Project Posters and Reports, Fall 2017
Ghasemi, npj Digital Medicine, 2025

The difference between online and offline reinforcement learning

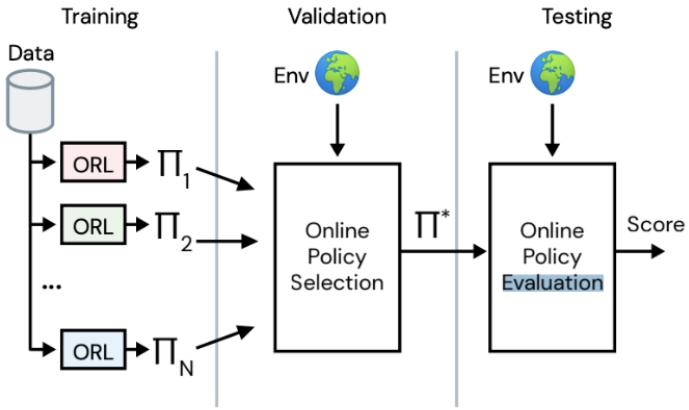
JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

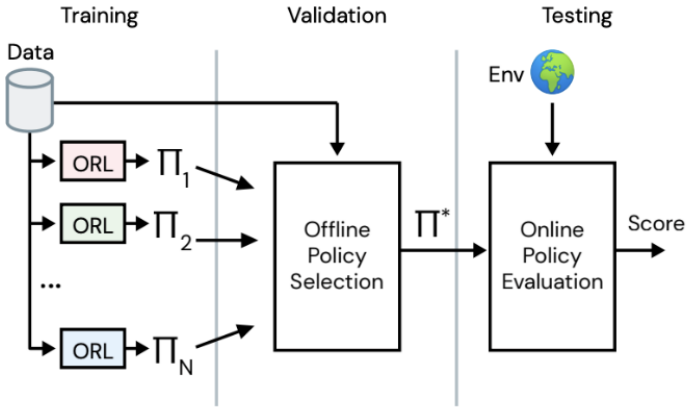




Online reinforcement learning



Offline reinforcement learning



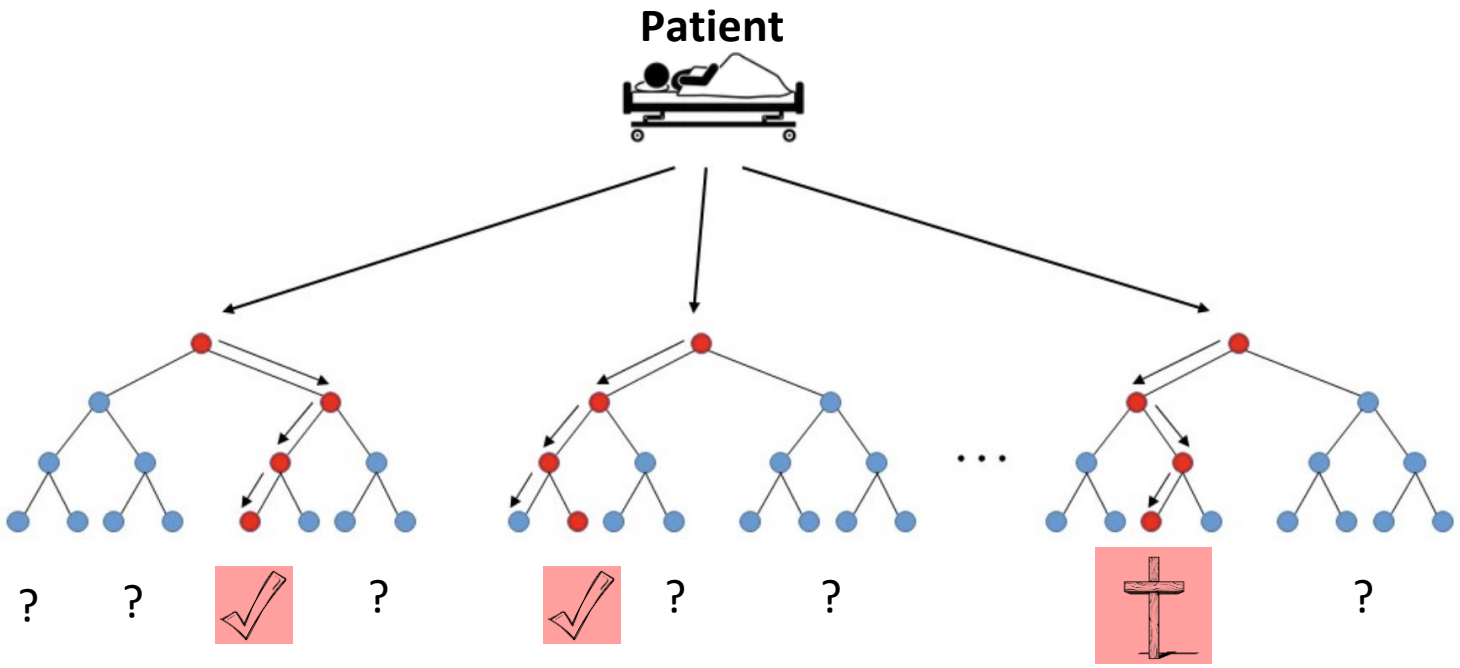
Gnanasekaran, Stanford, Project Posters and Reports, Fall 2017

Ghasemi, npj Digital Medicine, 2025

Can you evaluate theoretical policies on historical data?

JOHANNES KEPLER UNIVERSITÄT LINZ

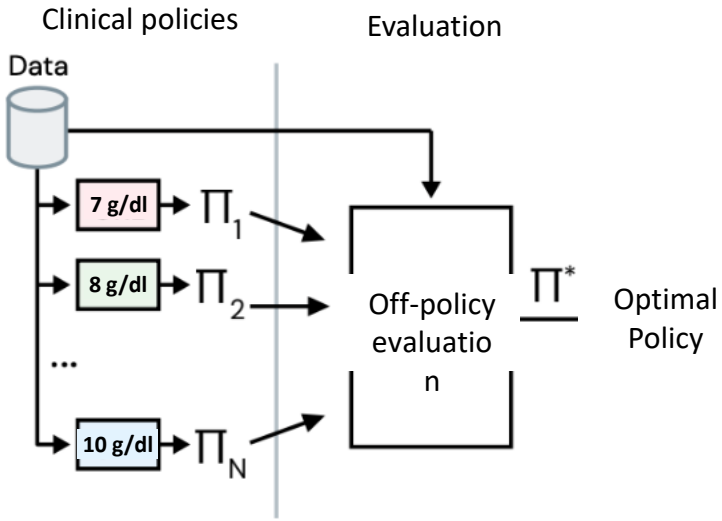
Kepler Universitäts Klinikum



Trajectory exists

Trajectory does not exist

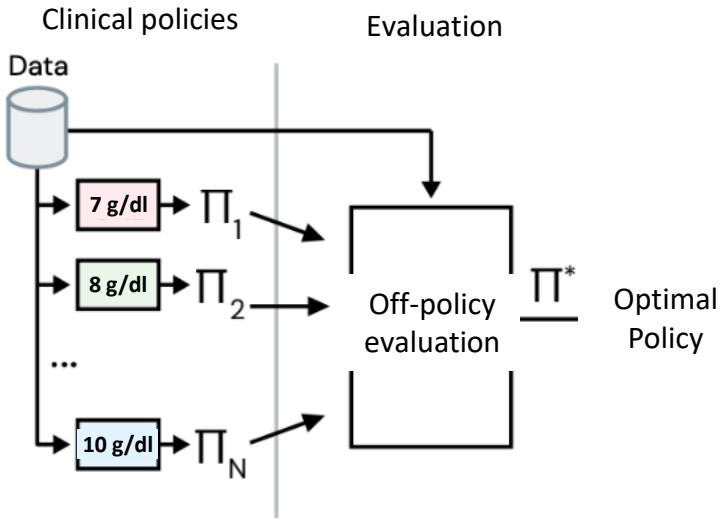
Can you evaluate theoretical policies on historical data?



Can you evaluate theoretical policies on historical data?

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum



Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

Importance Sampling for Fair Policy Selection*

Shayan Doroudi[1,3], **Philip S. Thomas**[2] and **Emma Brunskill**[3]
[1] Carnegie Mellon University
[2] University of Massachusetts Amherst
[3] Stanford University
shayand@cs.cmu.edu, pthomas@cs.umass.edu, ebrun@cs.stanford.edu

$$\hat{V}_{\text{IS}}^{\pi_e} \triangleq \frac{1}{n}\sum_{i=1}^{n} w_i \sum_{t=1}^{T_i} R_{i,t}$$

$$w_i \triangleq \prod_{t=1}^{T_i} \frac{\pi_e(a_{i,t}|\tau_{i,1:t-1})}{\pi_b(a_{i,t}|\tau_{i,1:t-1})}$$

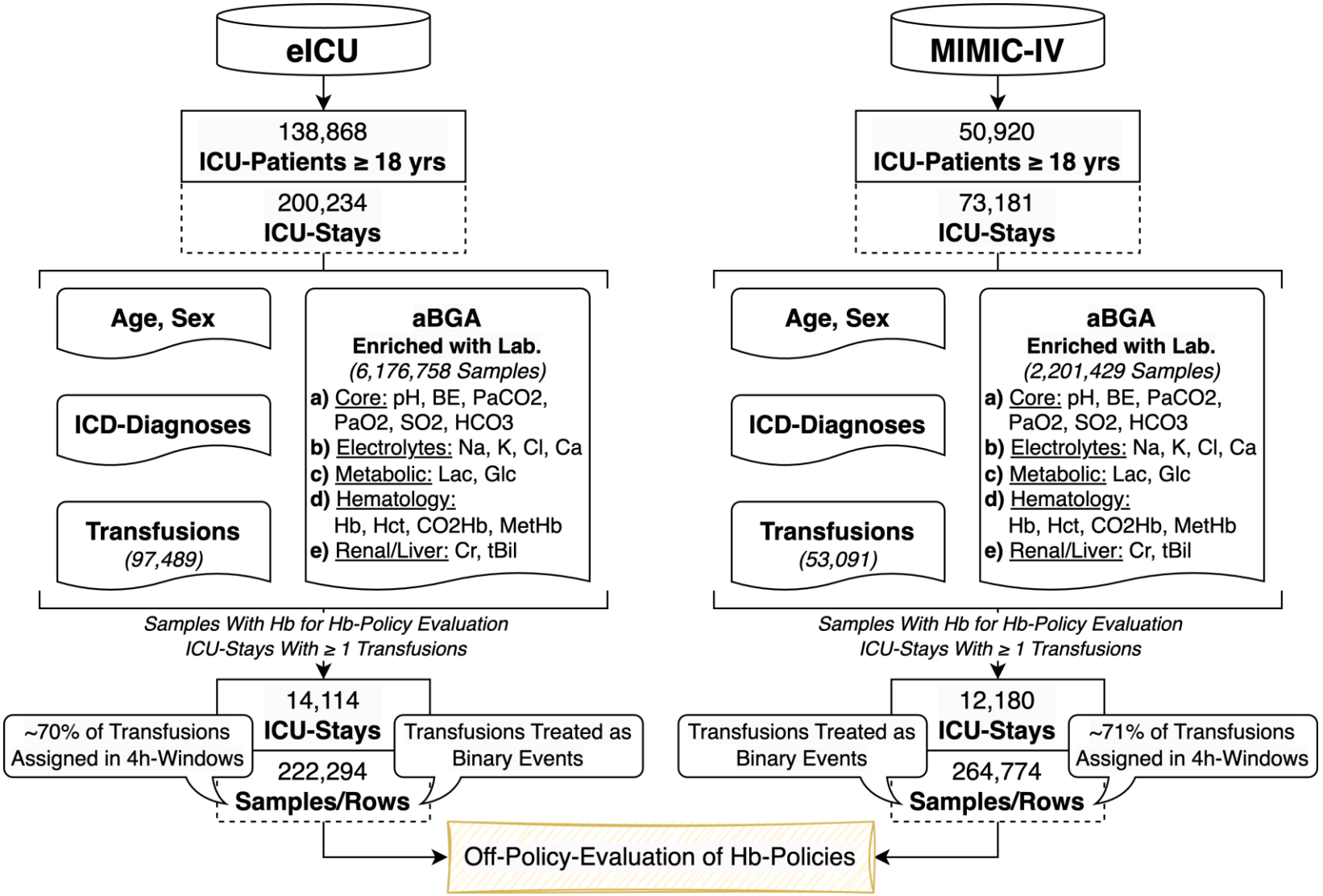
$$\hat{V}_{\text{PHWIS}} = \sum_{l \in L} W_l \underbrace{\frac{1}{\sum_{\{i|T_i=l\}} w_i} \sum_{\{i|T_i=l\}} w_i \sum_{t=1}^{T_i} R_{i,t}}_{\text{WIS estimate on } l\text{-length trajectories}}$$

per weighted horizon importance sampling: PWHIS

Off-policy evaluation of Hb policies

JOHANNES KEPLER UNIVERSITÄT LINZ

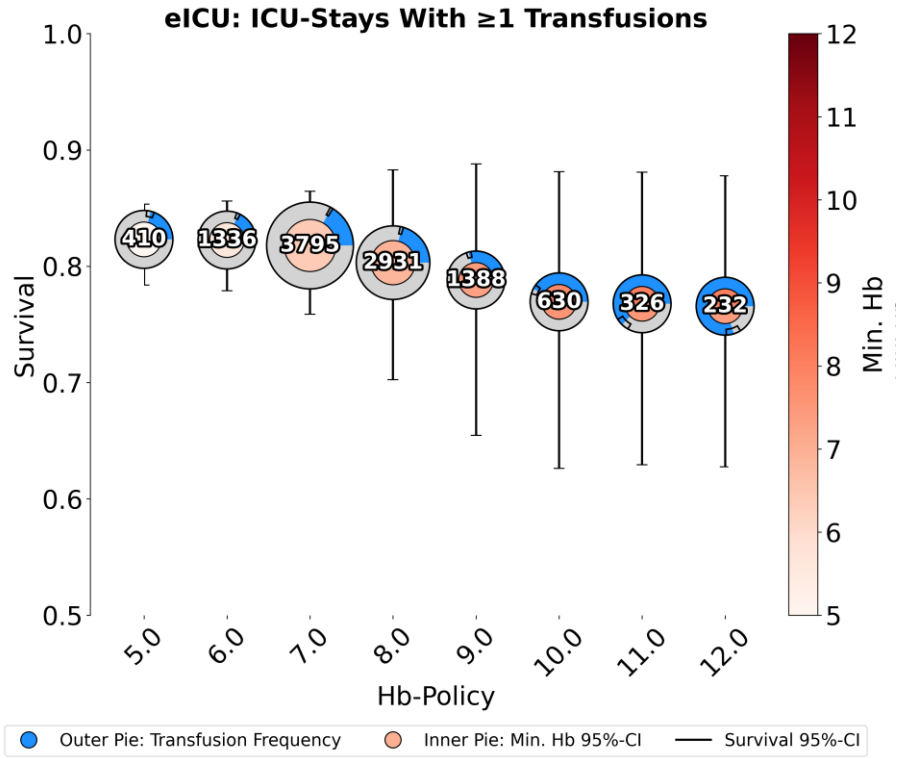
Kepler Universitäts Klinikum

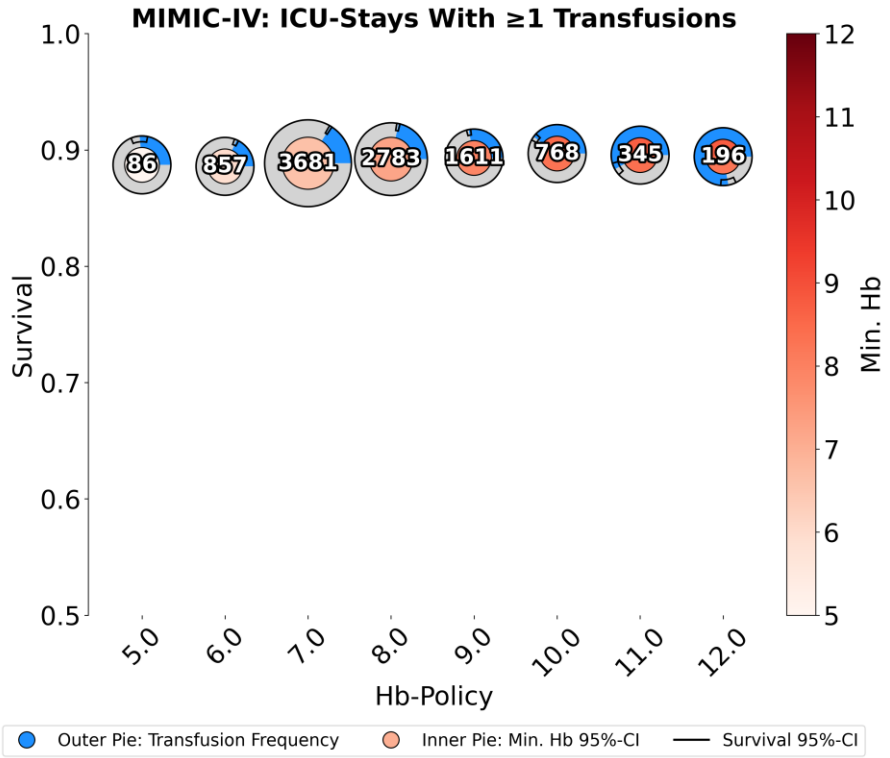


Hb policies evaluated for all patients

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

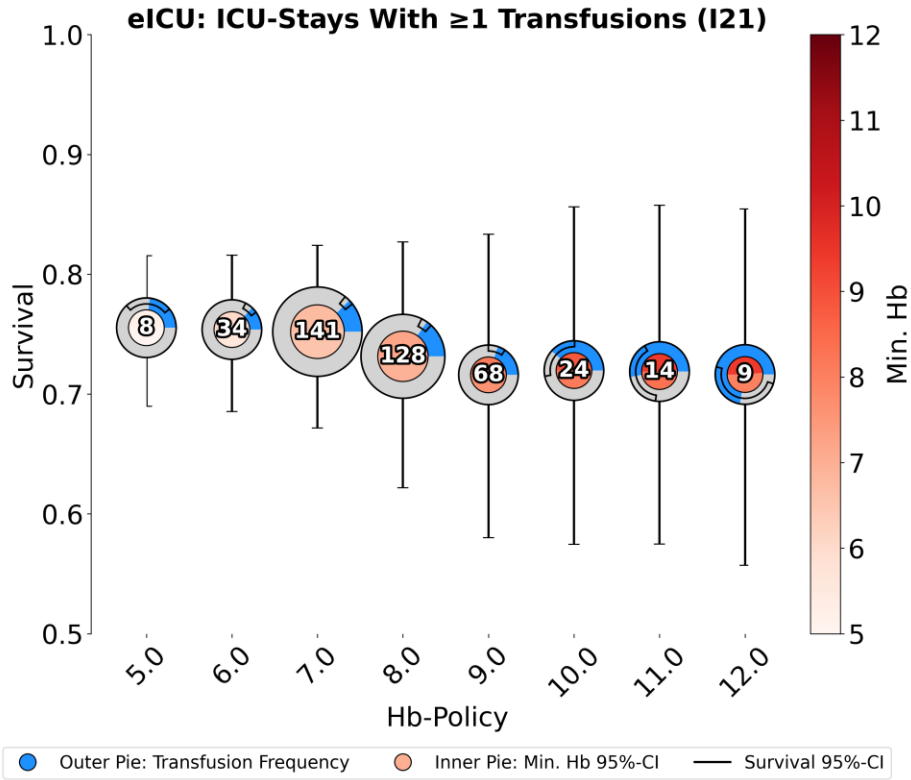


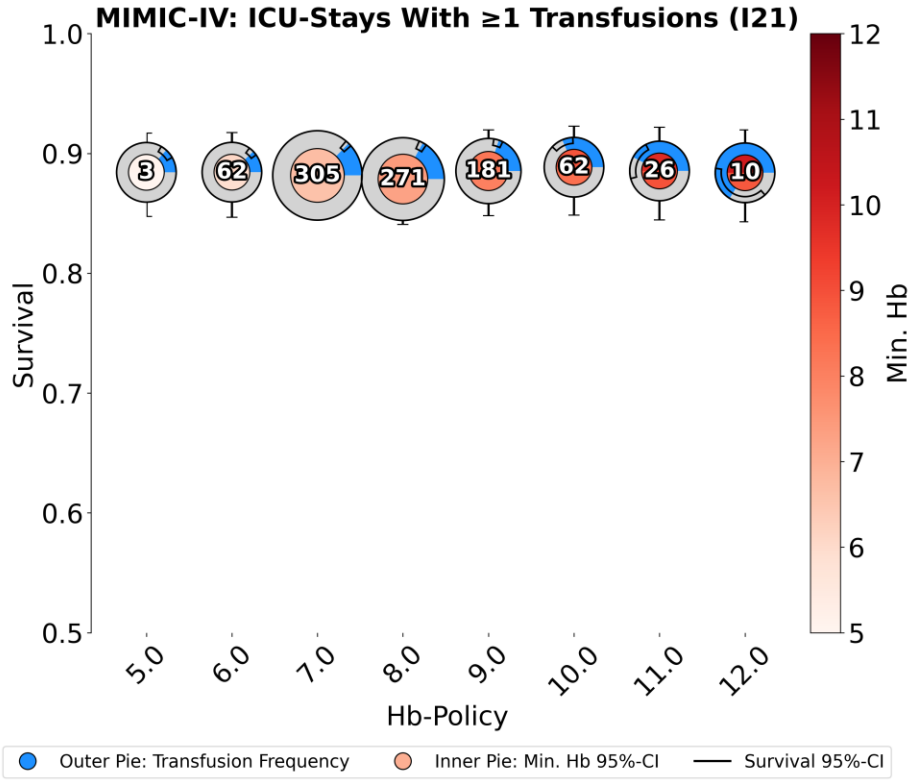


Hb policies evaluated for myocardial infarction patients

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum





Can you improve this with reinforcement learning?

JKU JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

- $Q(S_t, A_t)$ (left side): New Q-value estimation
- $Q(S_t, A_t)$: Former Q-value estimation
- α: Learning Rate
- R_{t+1}: Immediate Reward
- $\gamma max_a Q(S_{t+1}, a)$: Discounted Estimate optimal Q-value of next state
- $Q(S_t, A_t)$ (last term): Former Q-value estimation
- $R_{t+1} + \gamma max_a Q(S_{t+1}, a)$: TD Target
- $R_{t+1} + \gamma max_a Q(S_{t+1}, a) - Q(S_t, A_t)$: TD Error

$$\pi^*(s) \leftarrow \underset{a}{\operatorname{argmax}}\, Q^{\pi^*}(s,a) \quad \forall s$$

UP

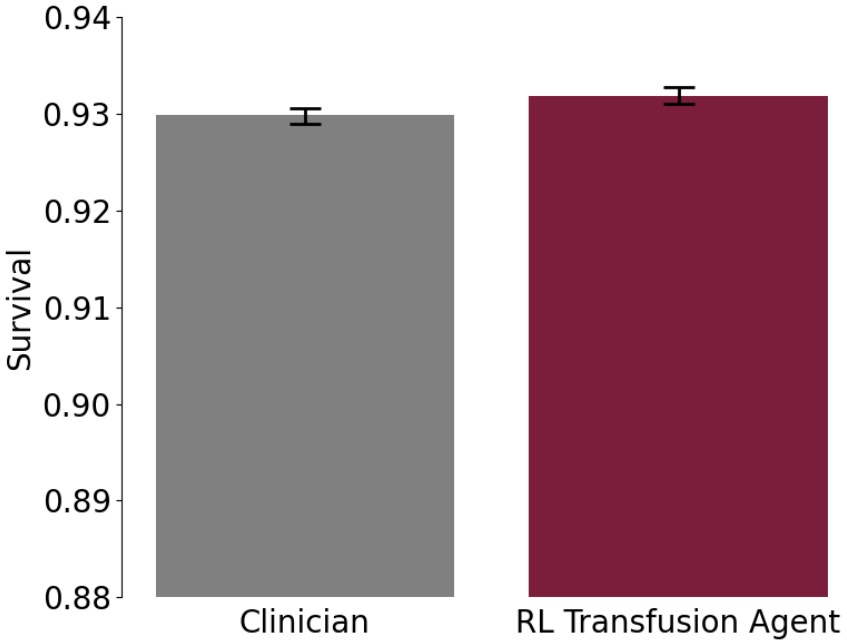
	0	1	2	3	4	5	6	7	8	9	10	11
0	U: -6.76 D: -6.73 R: -6.75 L: -6.71	U: -6.70 D: -6.74 R: -6.60 L: -6.62	U: -6.42 D: -6.53 R: -6.34 L: -6.34	U: -6.14 D: -6.09 R: -6.06 L: -6.12	U: -5.82 D: -5.77 R: -5.74 L: -5.78	U: -5.51 D: -5.37 R: -5.36 L: -5.53	U: -5.12 D: -4.97 R: -4.94 L: -5.32	U: -4.58 D: -4.49 R: -4.49 L: -4.69	U: -4.01 D: -4.02 R: -3.94 L: -4.28	U: -3.57 D: -3.37 R: -3.36 L: -3.70	U: -2.65 D: -2.69 R: -2.65 L: -3.01	U: -2.26 D: -1.90 R: -2.07 L: -2.15
1	U: -6.81 D: -6.96 R: -6.89 L: -6.89	U: -6.80 D: -6.71 R: -6.70 L: -6.75	U: -6.51 D: -6.43 R: -6.45 L: -6.67	U: -6.17 D: -6.08 R: -6.09 L: -6.39	U: -5.76 D: -5.68 R: -5.68 L: -5.98	U: -5.55 D: -5.21 R: -5.21 L: -5.63	U: -4.73 D: -4.68 R: -4.68 L: -4.92	U: -4.37 D: -4.09 R: -4.09 L: -4.23	U: -3.92 D: -3.44 R: -3.44 L: -3.98	U: -3.80 D: -2.71 R: -2.71 L: -2.93	U: -2.84 D: -1.90 R: -1.90 L: -3.24	U: -1.72 D: -1.00 R: -1.43 L: -2.27
2	U: -7.06 D: -7.40 R: -6.86 L: -7.14	U: -6.96 D: -99.95 R: -6.51 L: -7.15	U: -6.71 D: -93.75 R: -6.13 L: -6.86	U: -6.37 D: -96.88 R: -5.70 L: -6.16	U: -6.09 D: -99.61 R: -5.22 L: -6.12	U: -5.60 D: -99.22 R: -4.69 L: -5.68	U: -5.07 D: -99.22 R: -4.10 L: -5.18	U: -4.60 D: -99.22 R: -3.44 L: -4.07	U: -4.07 D: -96.88 R: -2.71 L: -3.98	U: -3.34 D: -98.44 R: -1.90 L: -3.35	U: -2.64 D: -98.44 R: -1.00 L: -2.63	U: -1.72 D: 0.00 R: -1.00 L: -1.81
3	U: -7.18 D: -7.46 R: -99.22 L: -7.45	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00	U: 0.00 D: 0.00 R: 0.00 L: 0.00

Evaluating Policy after Q-learning of the MIMIC IV data

JOHANNES KEPLER UNIVERSITÄT LINZ

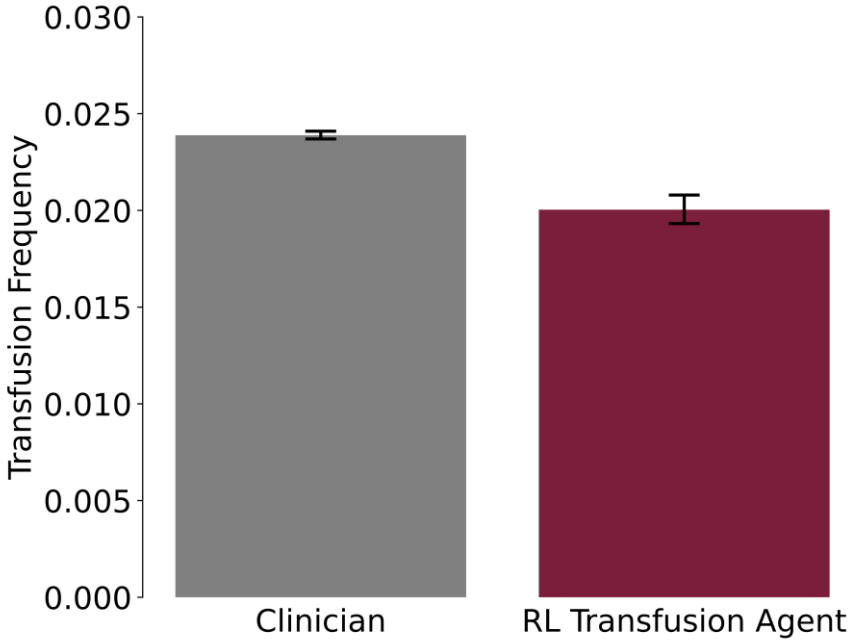
Kepler Universitäts Klinikum

Survival



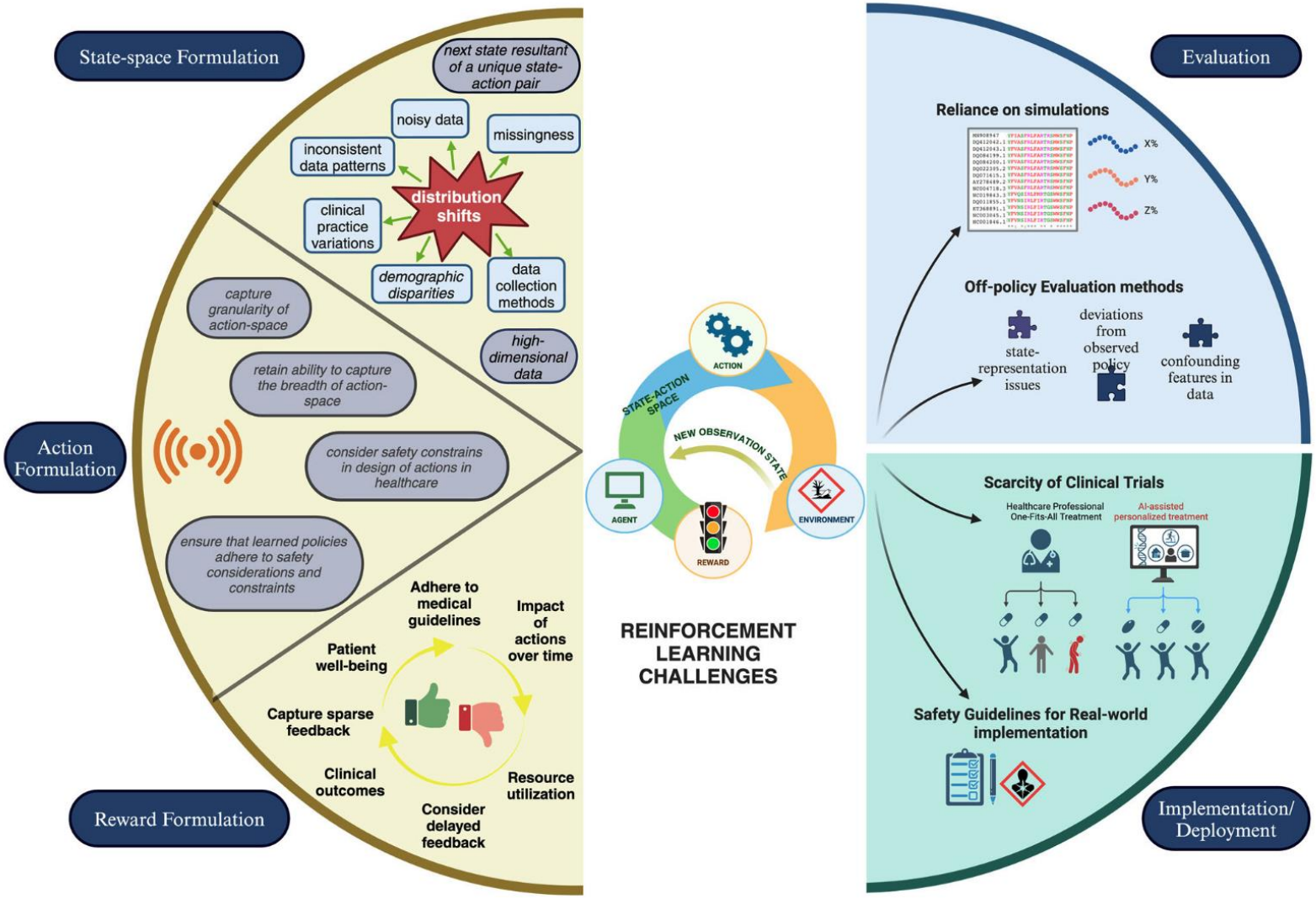
p<0.05

Transfusion Frequency



p<0.05

The problems of reinforcement learning



Jayaraman, Digital Medicine, 2024

Conclusion

JOHANNES KEPLER UNIVERSITÄT LINZ

Kepler Universitäts Klinikum

- ML models are well established for transfusion prediction
- LLMs might be able to guide transfusion in the future
- Reinforcement learning is the way to go!
- Hb based transfusion policies have (probably) no effect on survival
- one can build a reinforcement learning model, that
  - transfuses less
  - has no influence on survival