Modeling Strong and Human-Like Gameplay with KL-Regularized Search

International Conference on Machine Learning 2022

Athul Paul Jacob*, David J. Wu*, Gabriele Farina*, Adam Lerer, Hengyuan Hu, Anton Bakhtin, Jacob Andreas, Noam Brown

* Equal Contribution

Massachusetts Institute of Technology

Carnegie Mellon University

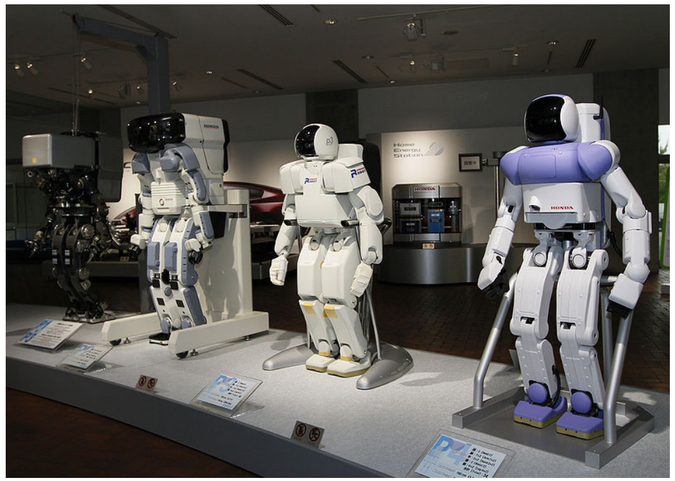
Meta AI

MOTIVATION

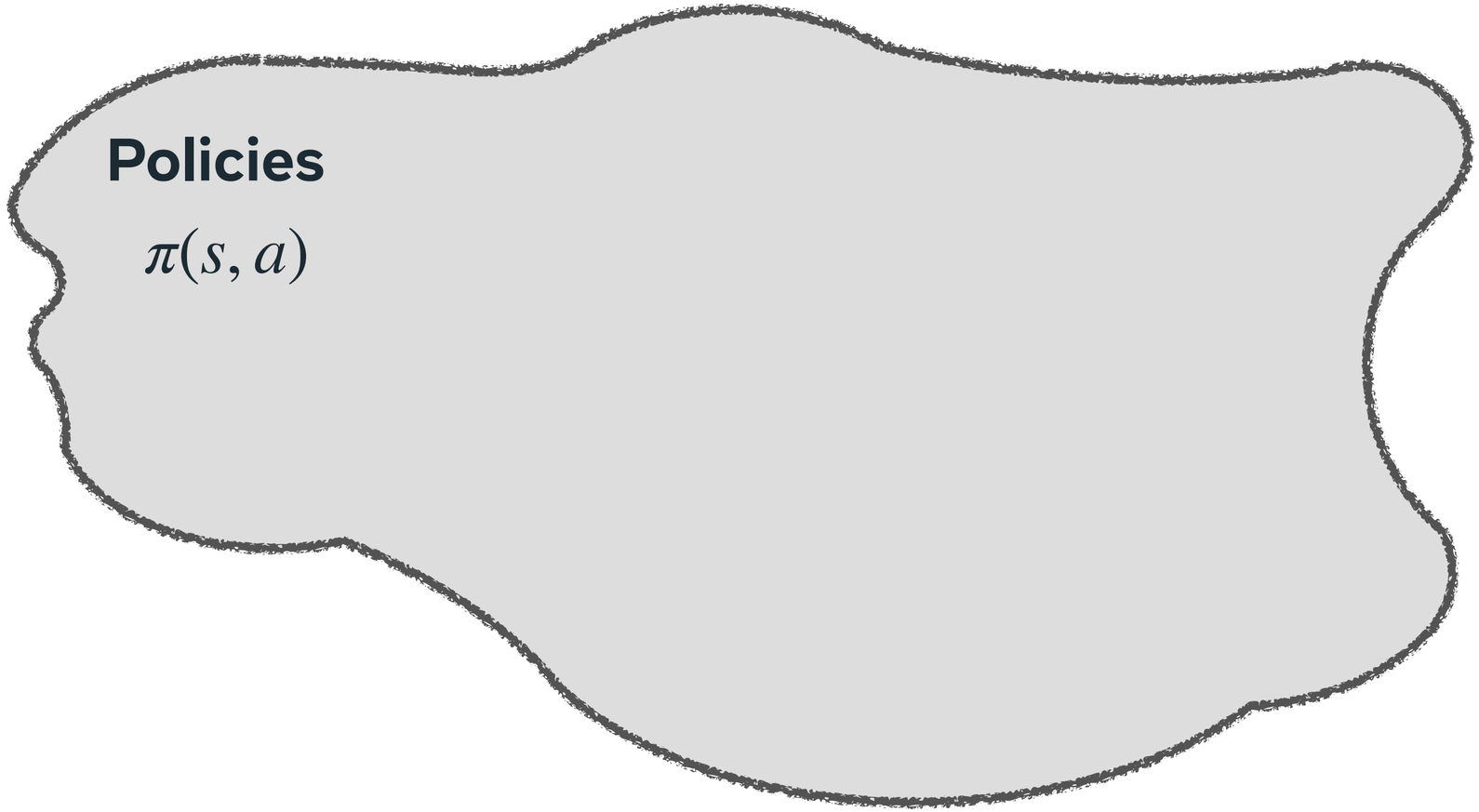
In more and more domains, AI is achieving superhuman performance

MOTIVATION

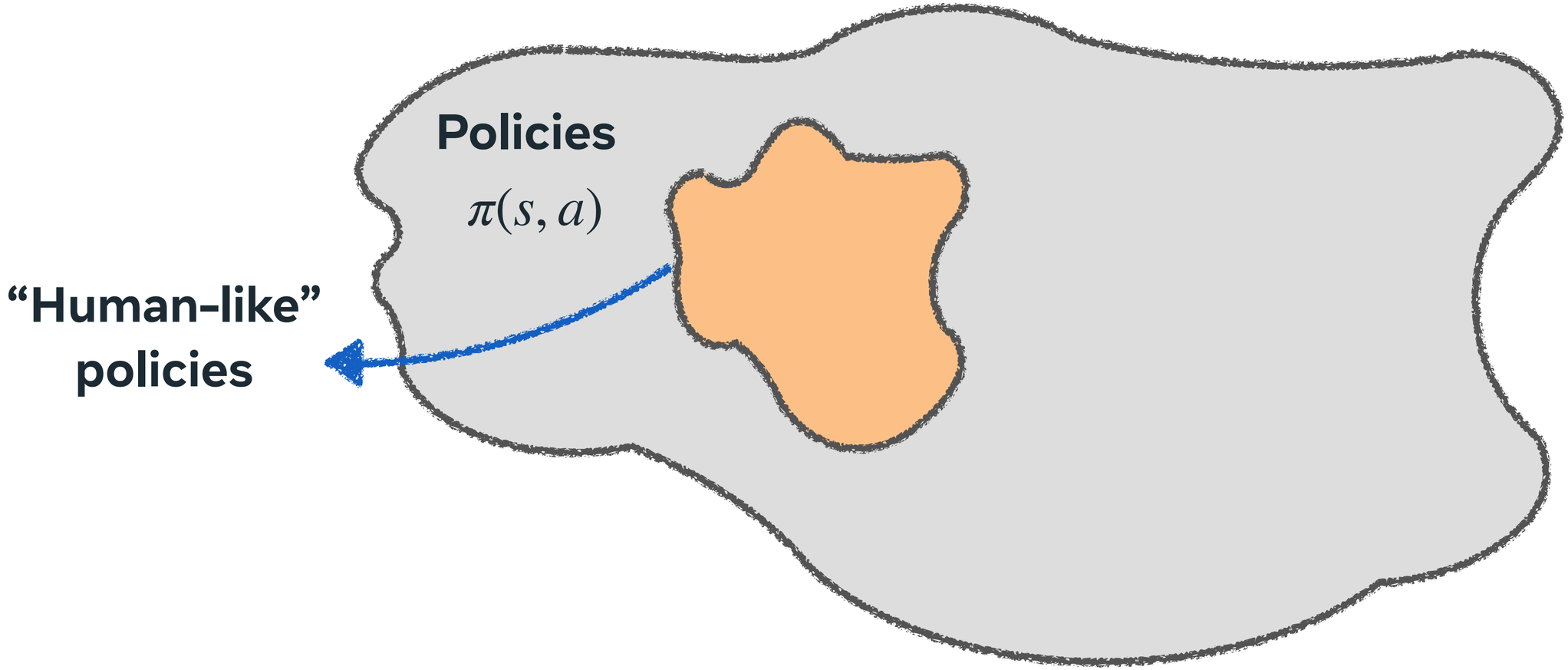
But we also want AI agents that can model humans well



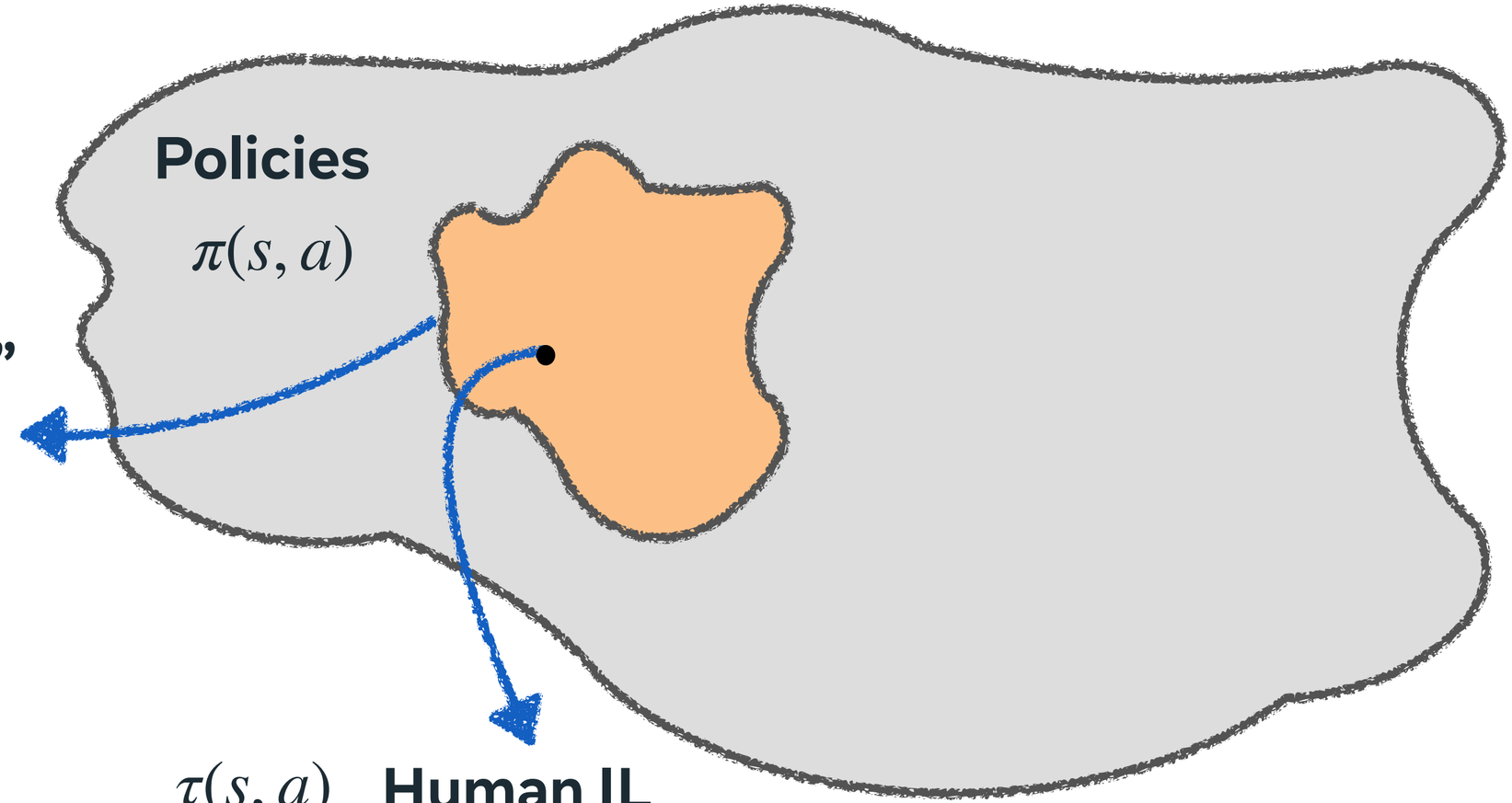
Building Strong, Human-like Policies



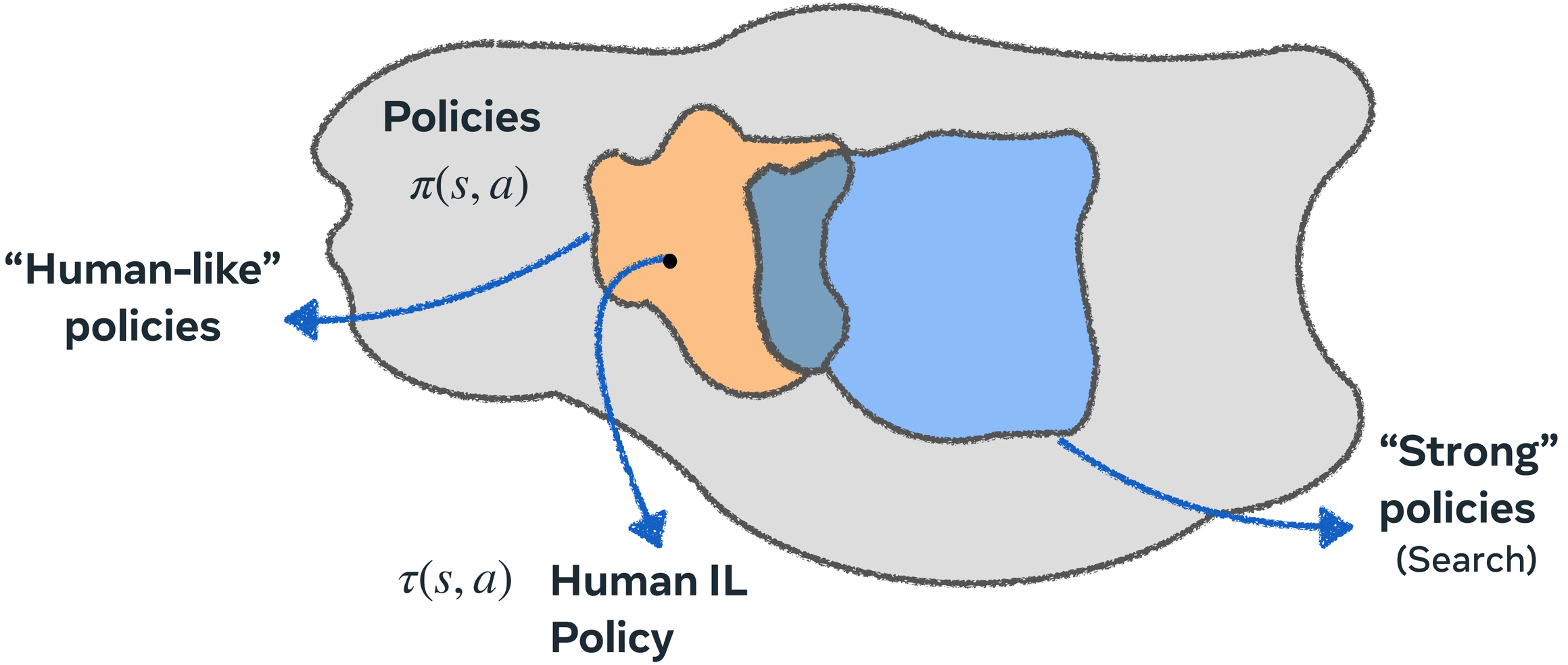
Building Strong, Human-like Policies



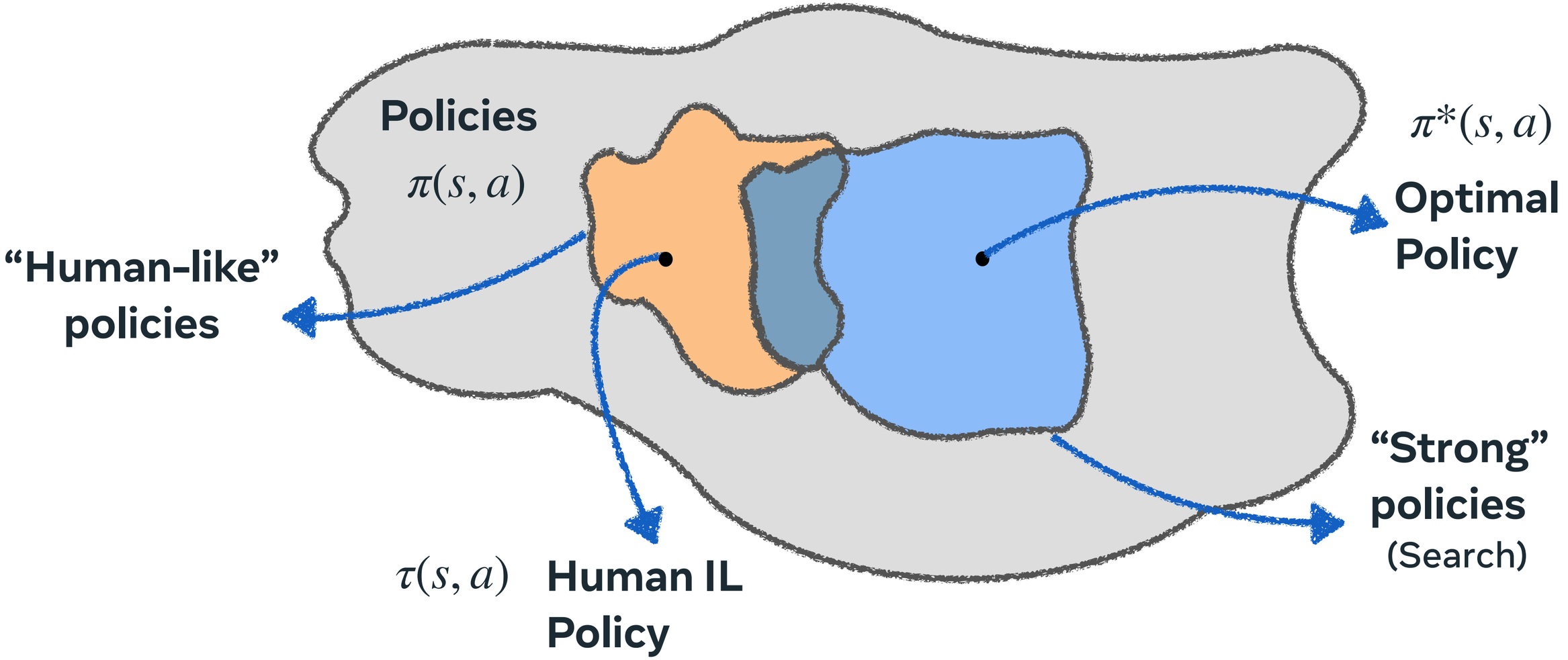
Building Strong, Human-like Policies



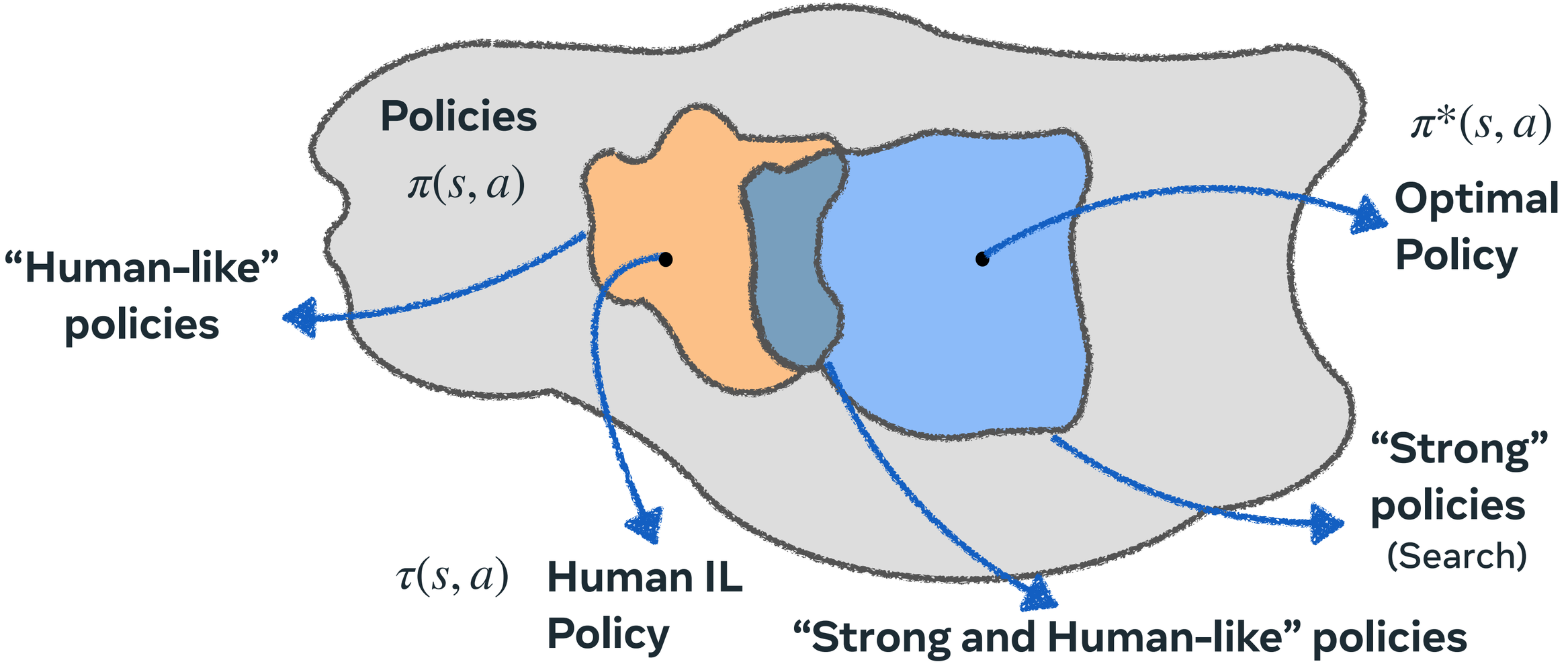
Building Strong, Human-like Policies



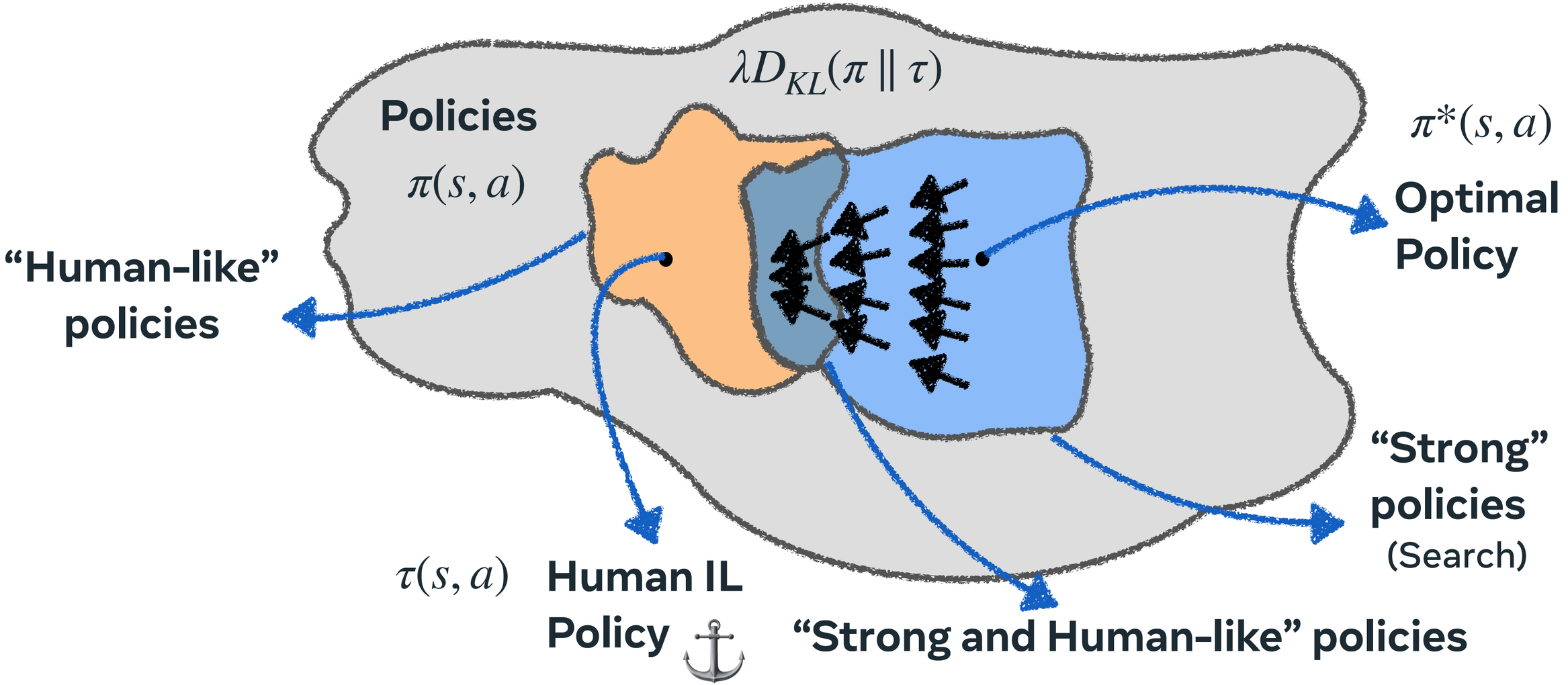
Building Strong, Human-like Policies



Building Strong, Human-like Policies



Building Strong, Human-like Policies

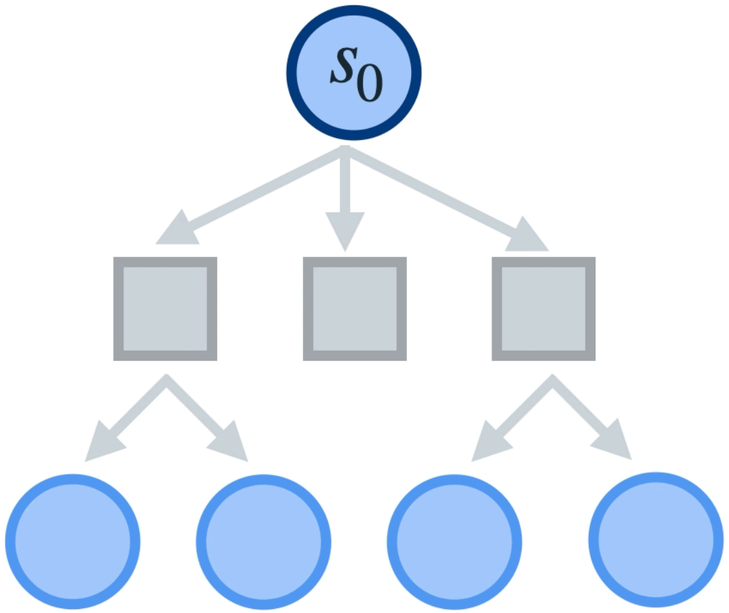


Sequential Games

Chess

Go

Monte Carlo Tree Search



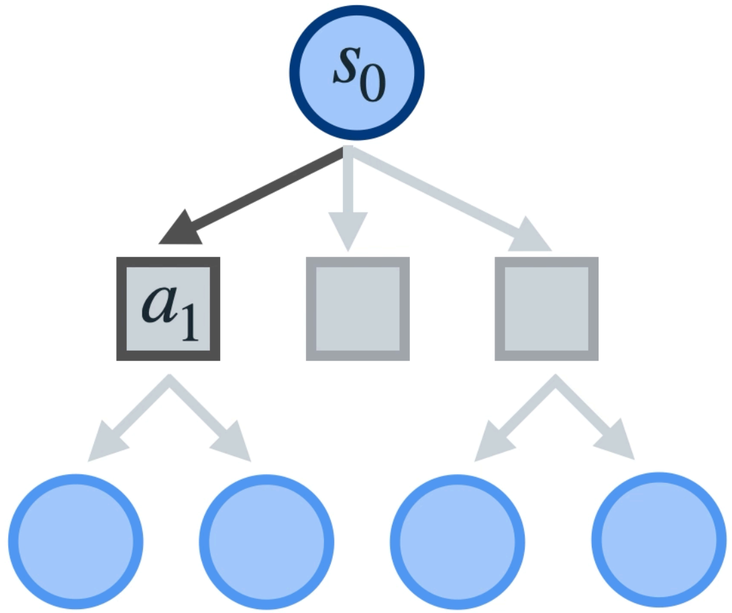
Sequential Games

Chess

Go

Monte Carlo Tree Search

Selection



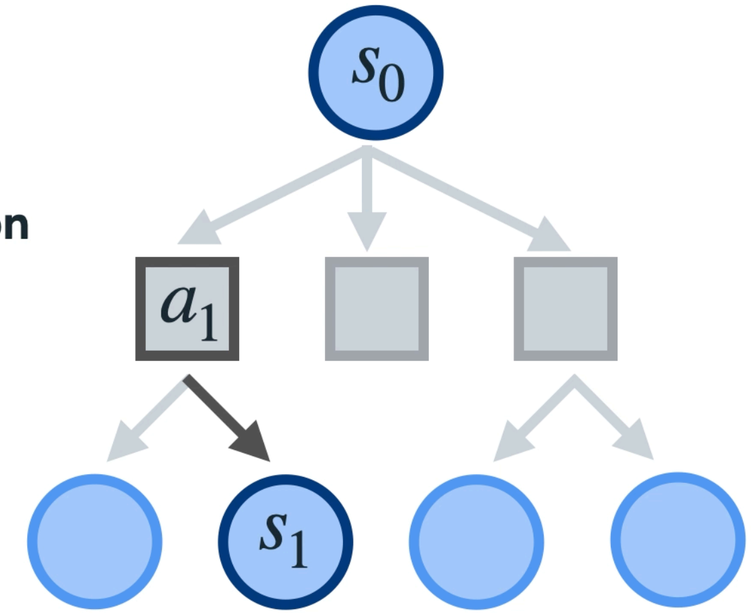
Sequential Games

Chess

Go

Monte Carlo Tree Search

Expansion



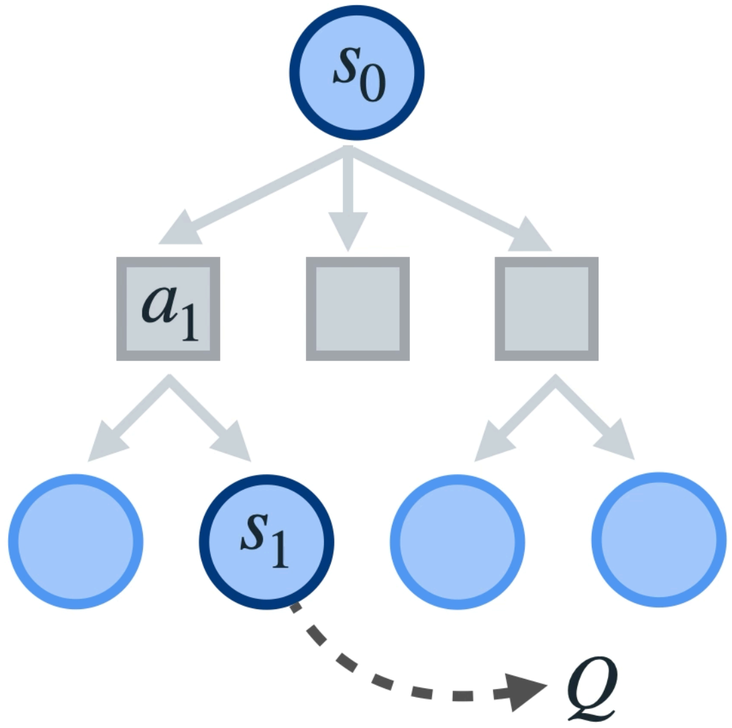
Sequential Games

Chess

Go

Monte Carlo Tree Search

Rollout



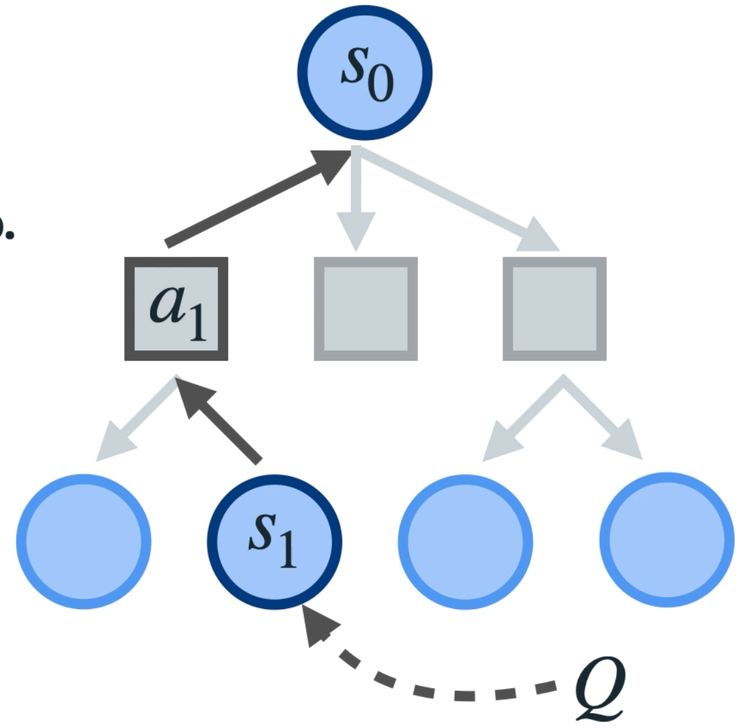
Sequential Games

Chess

Go

Monte Carlo Tree Search

Backprop.



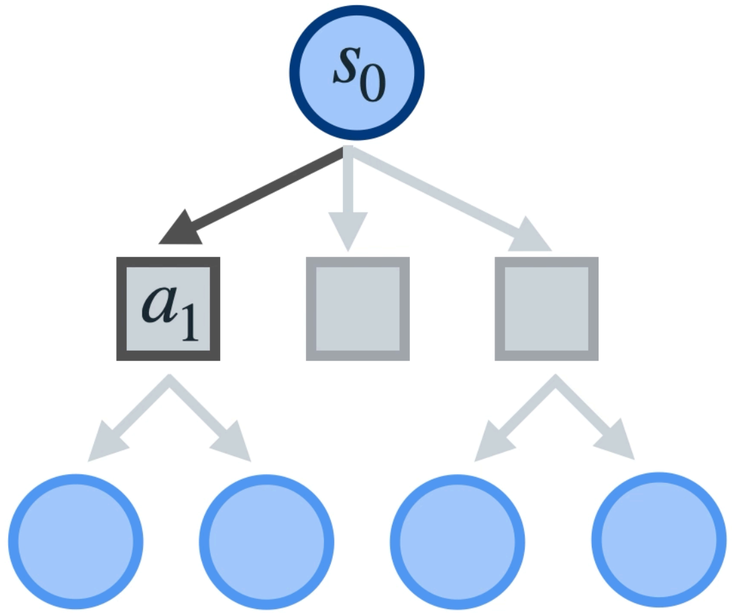
Sequential Games

Chess

Go

Monte Carlo Tree Search

Selection

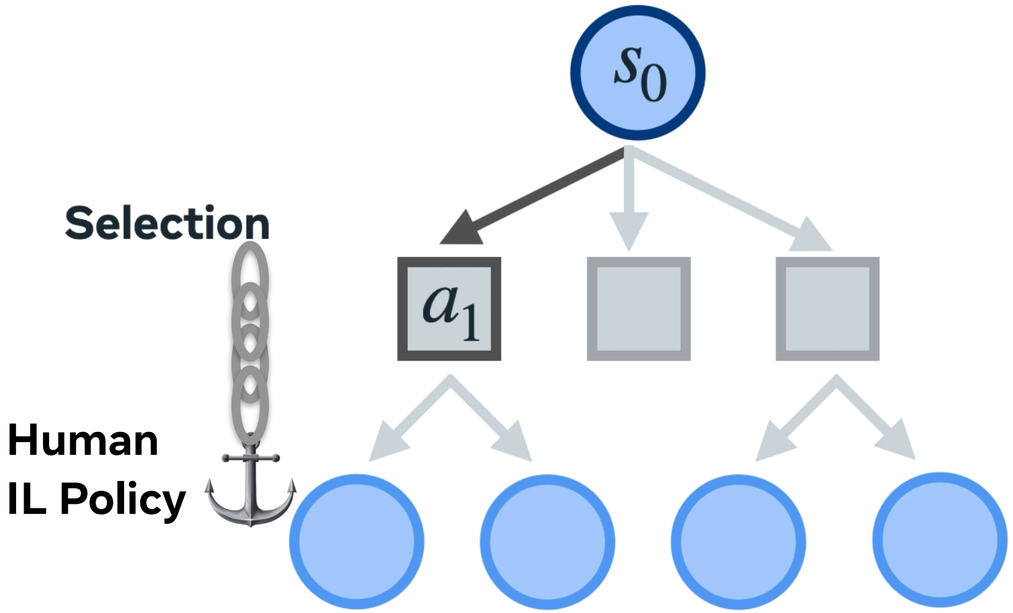


Sequential Games

Chess

Go

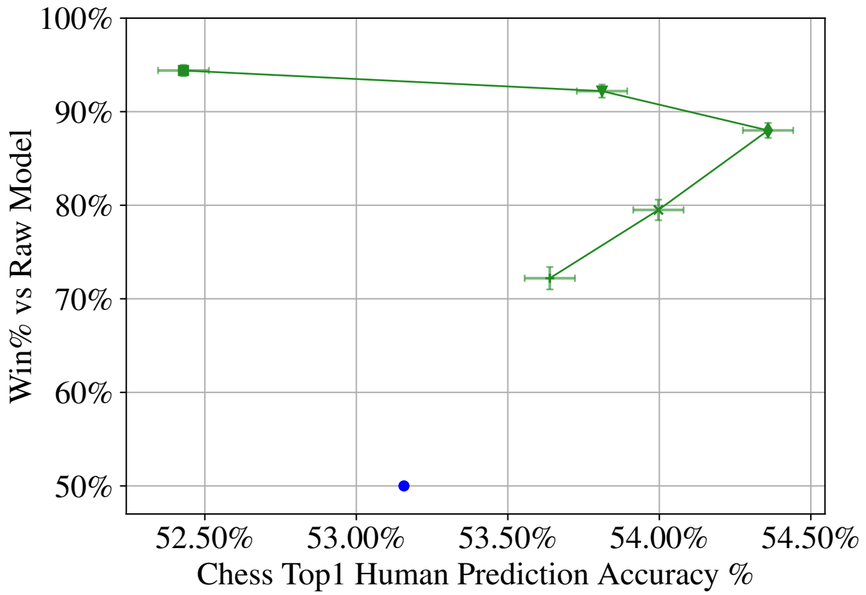
Human Policy Regularized Monte Carlo Tree Search

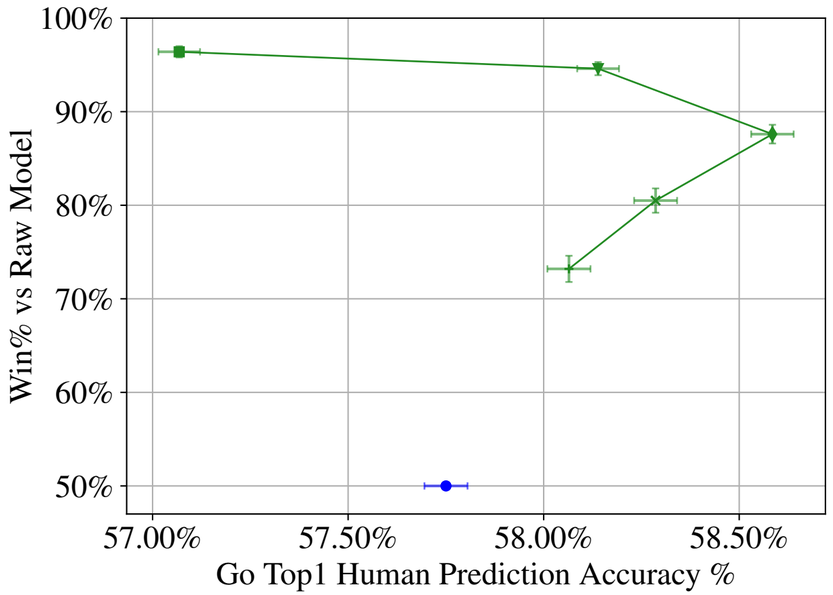


Chess

Go

Strength →



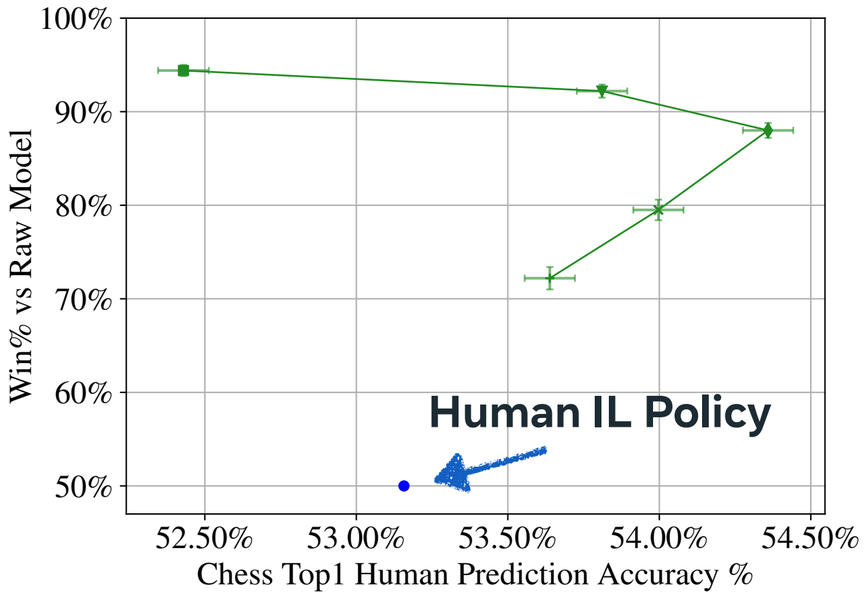


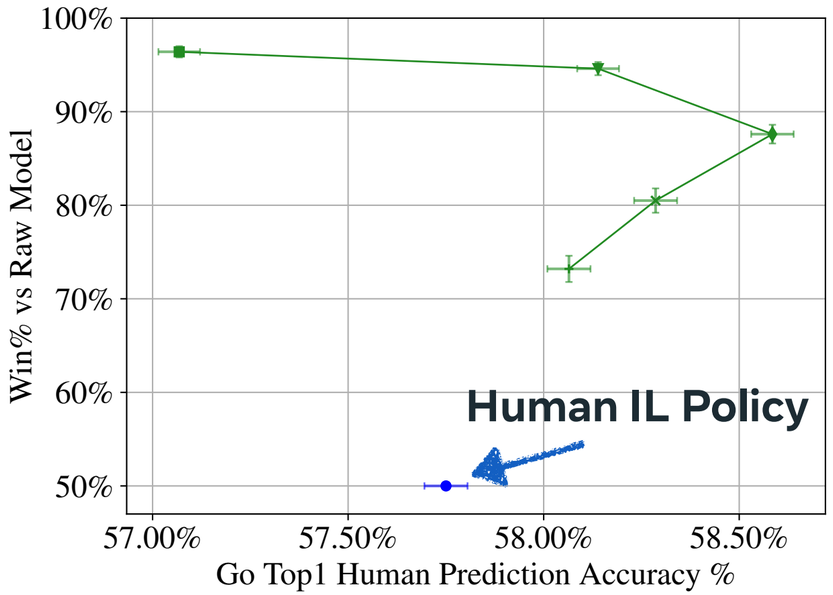
Human Action Prediction →

Chess

Go

Strength →



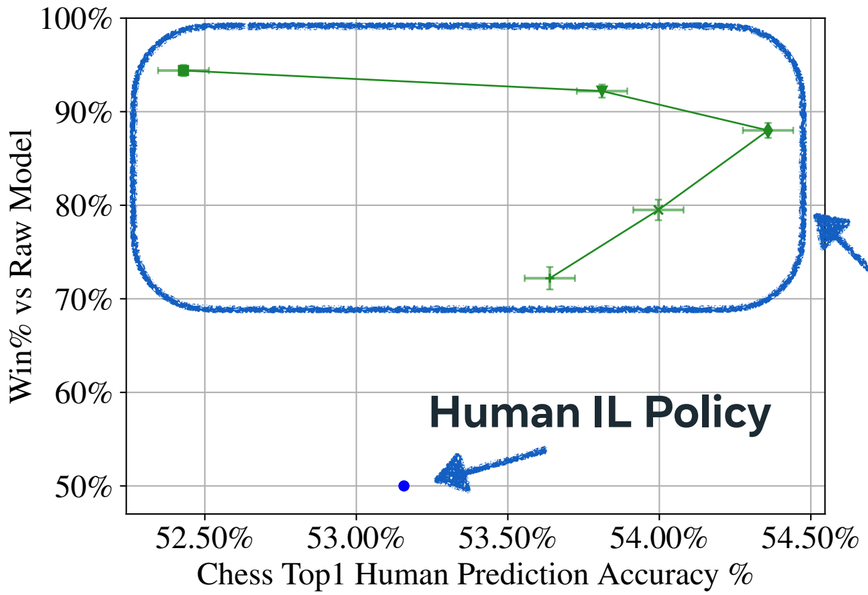


Human Action Prediction →

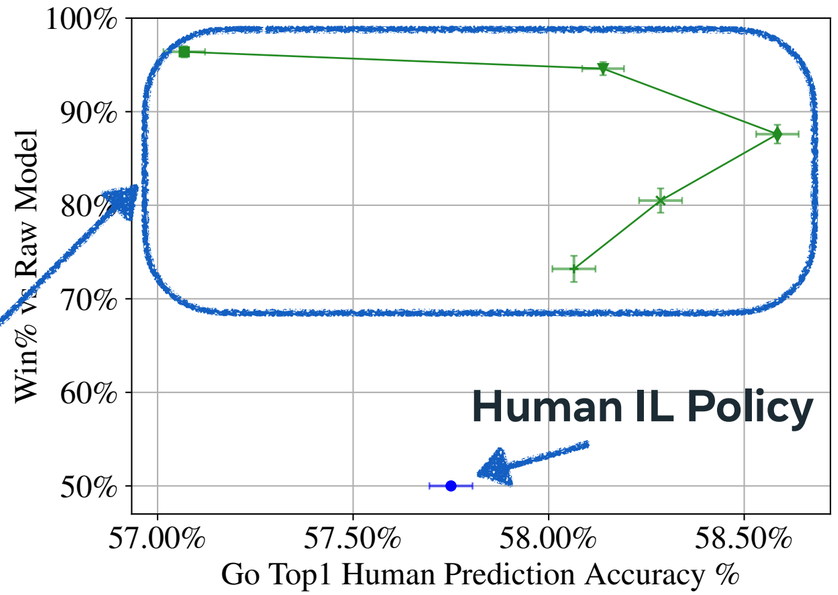
Chess

Go

Strength →



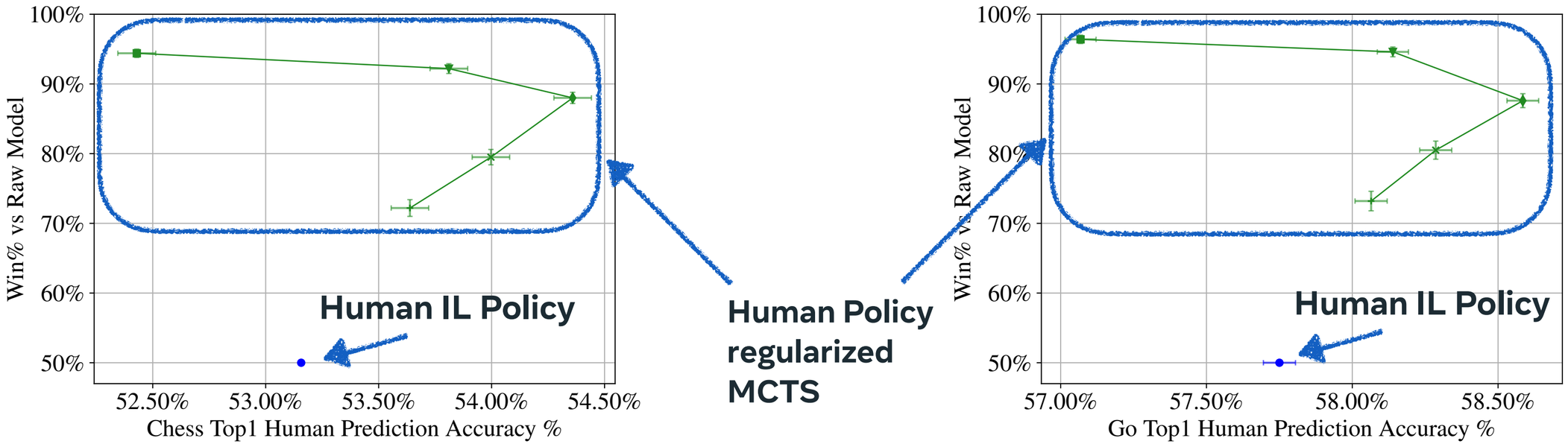
Human Policy regularized MCTS



Human Action Prediction →

KL-regularized Search produces stronger and more human-like agents in both Chess and go!

Strength →

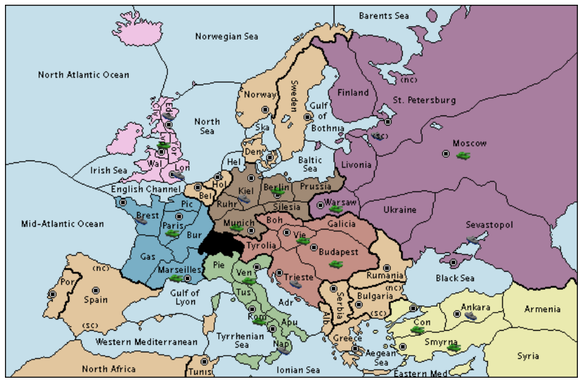


Human Action Prediction →

Imperfect-Information Games

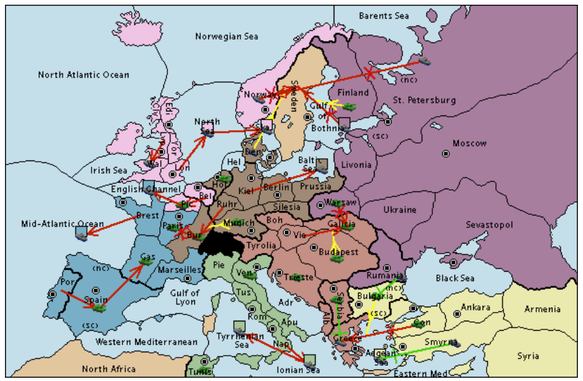
Imperfect-Information Games

No-press Diplomacy



Imperfect-Information Games

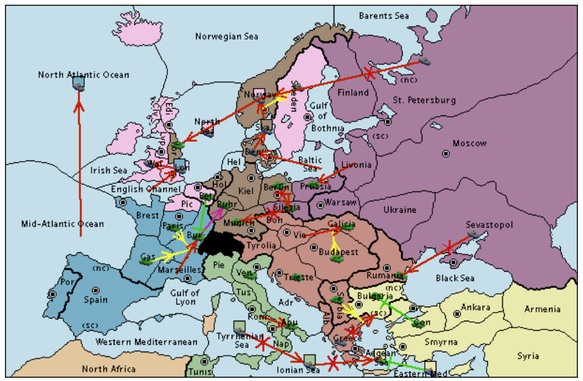
No-press Diplomacy



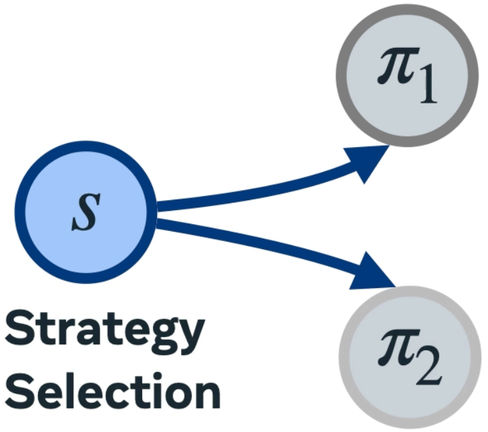
Hedge

Imperfect-Information Games

No-press Diplomacy

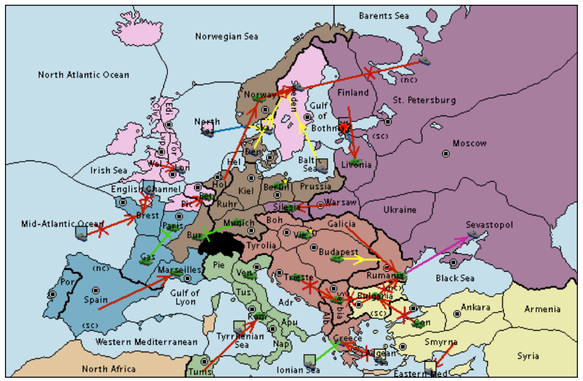


Hedge

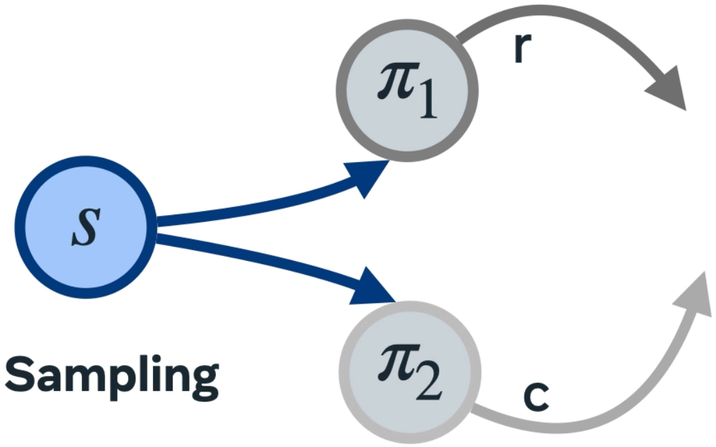


Imperfect-Information Games

No-press Diplomacy



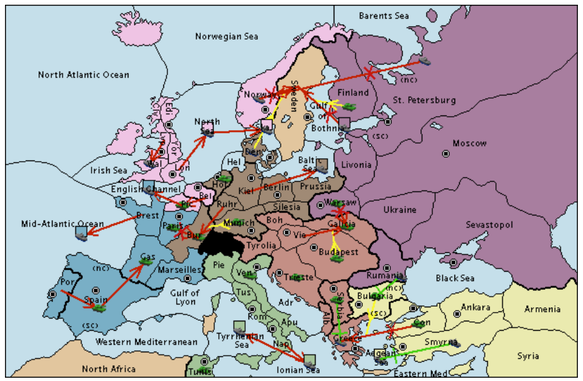
Hedge



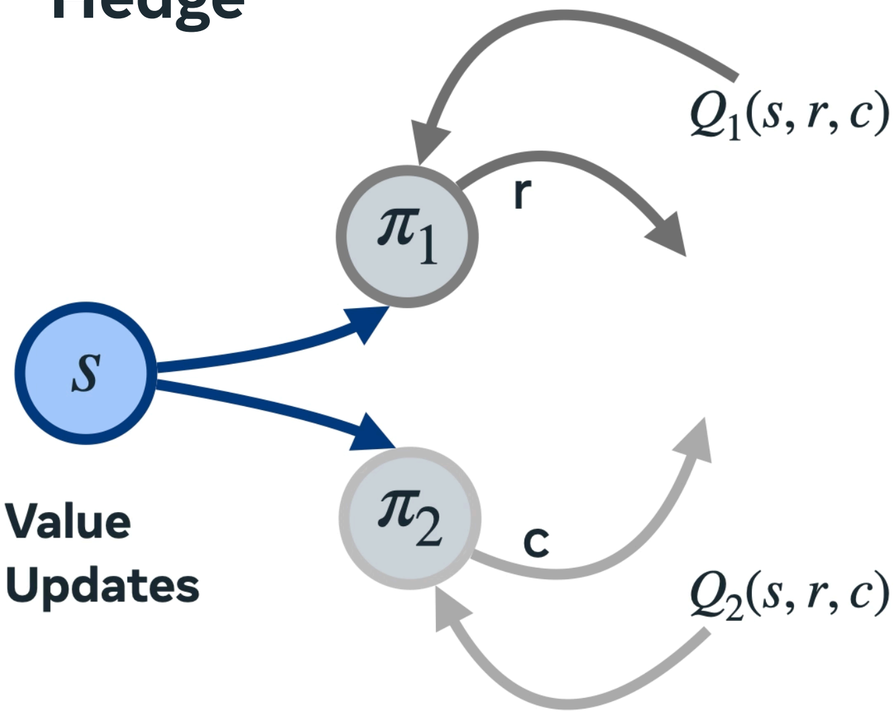
Sampling

Imperfect-Information Games

No-press Diplomacy

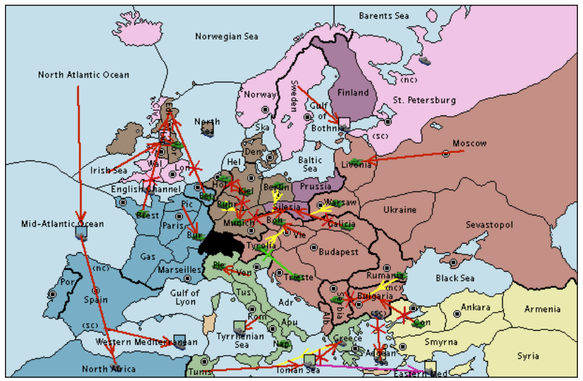


Hedge

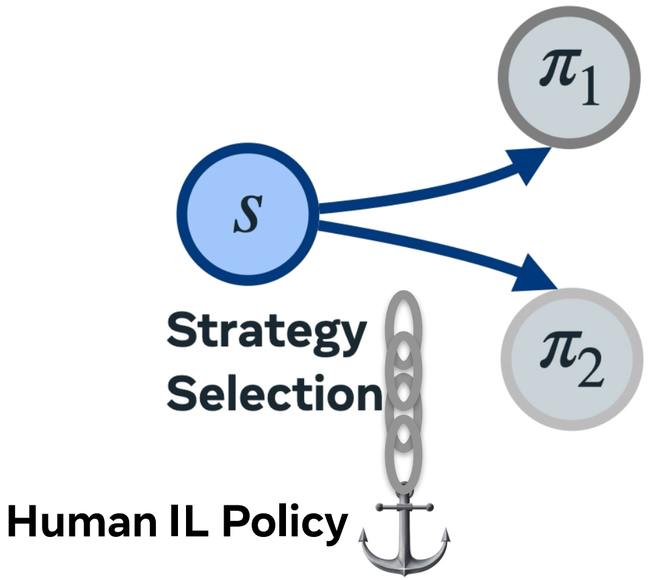


Imperfect-Information Games

No-press Diplomacy

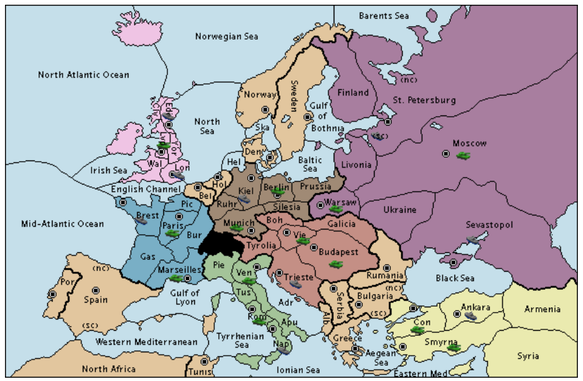


Human Policy Regularized Hedge (piKL-Hedge)



Imperfect-Information Games

No-press Diplomacy

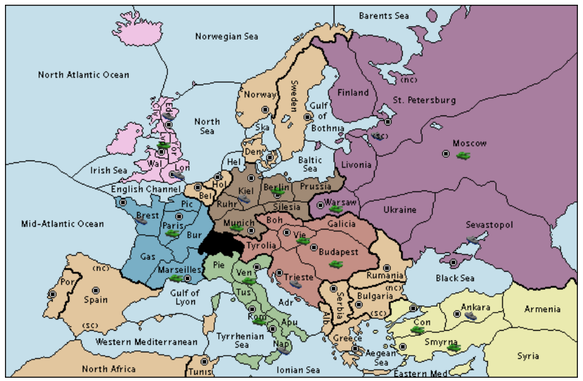


Human Policy Regularized Hedge (piKL-Hedge)

Theorem 1: In two player zero sum games, the average policy produced by piKL-Hedge stays close to the anchor policy.

Imperfect-Information Games

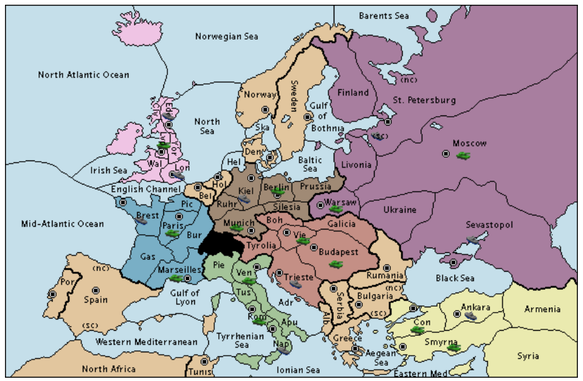
No-press Diplomacy



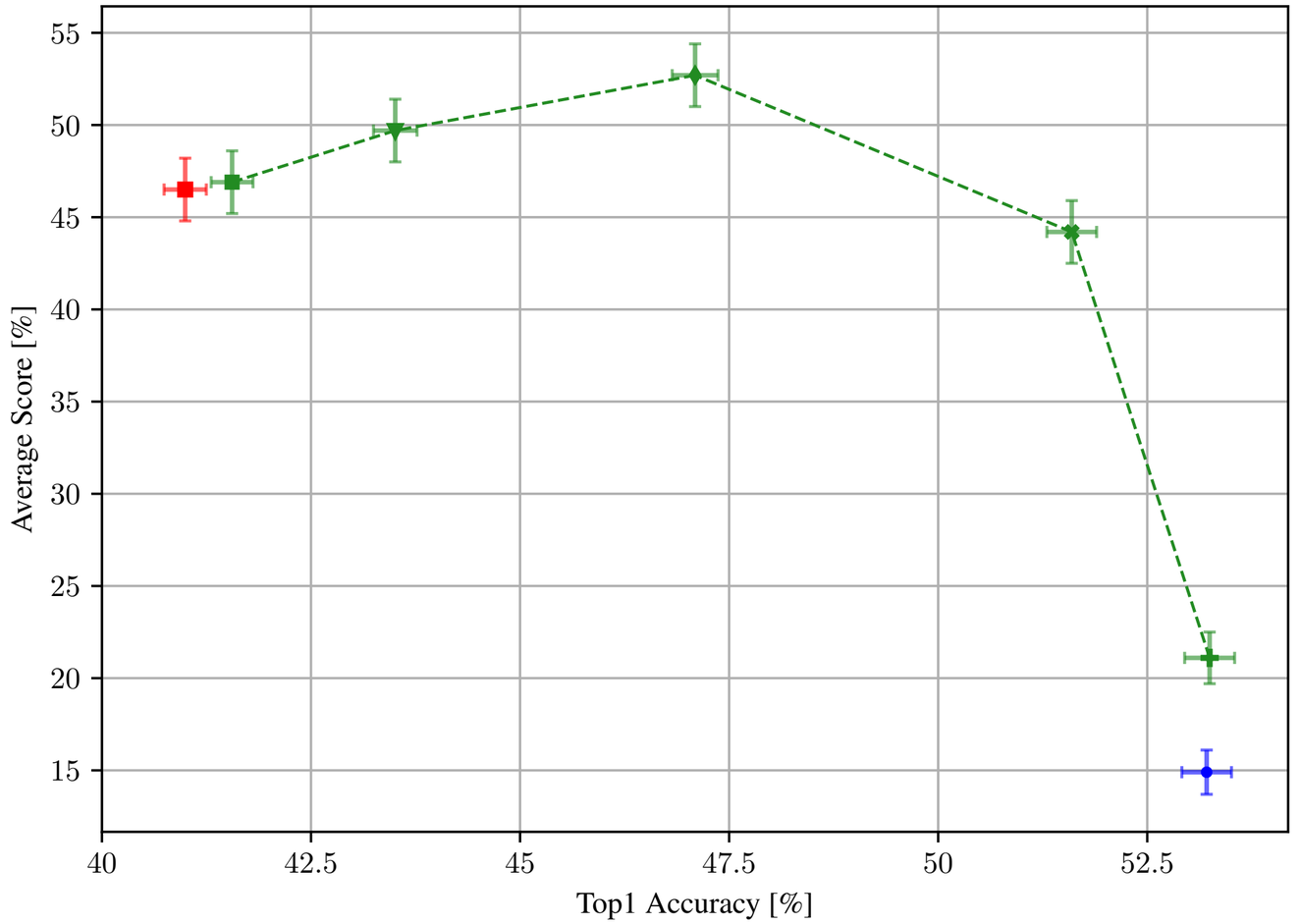
Human Policy Regularized Hedge (piKL-Hedge)

Theorem 2: In two player zero sum games, the average policies of the players produced by piKL-Hedge converges to an approx. Nash equilibrium under the original non-regularized utilities.

No-press Diplomacy

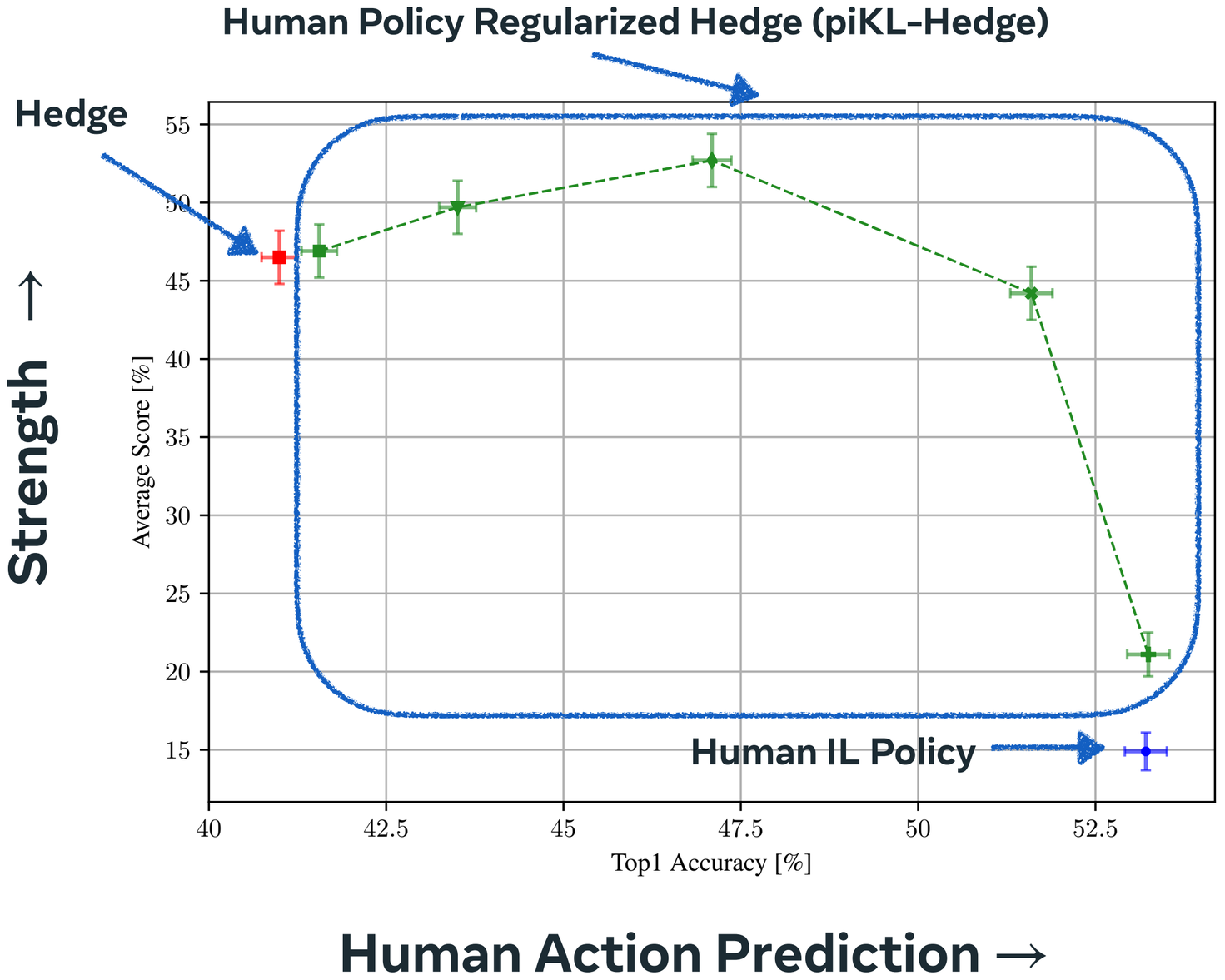


Strength →



Human Action Prediction →

piKL-Hedge produces strong, human-like gameplay in no-press Diplomacy



Conclusion

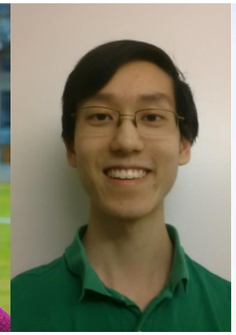
Can we build gameplay agents that are not only strong but also one that better models human actions?

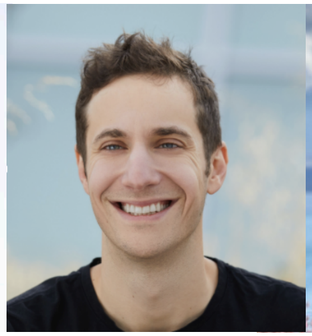
Yes! By using KL-Regularization Search towards a human imitation learned policy!

Image Credits

https://commons.wikimedia.org/wiki/File:Two_poker_cards_and_poker_chips_20170611.jpg - License: https://www.wikidata.org/wiki/Q18199165
https://commons.wikimedia.org/wiki/File:Chess-king.JPG – License: https://www.wikidata.org/wiki/Q19068220
https://commons.wikimedia.org/wiki/File:Go_game.jpg – License: https://www.wikidata.org/wiki/Q19125117
https://commons.wikimedia.org/wiki/File:Waymo_self-driving_car_side_view.gk.jpg – License: https://www.wikidata.org/wiki/Q18199165
https://commons.wikimedia.org/wiki/File:Honda_prototype_robots_Honda_Collection_Hall.jpg – License: https://www.wikidata.org/wiki/Q14946043
https://commons.wikimedia.org/wiki/File:IRobot_Roomba_870_(15860914940).jpg – License: https://www.wikidata.org/wiki/Q19125117
https://commons.wikimedia.org/wiki/File:Chess-american-krikor-2021-final.gif – License: https://www.wikidata.org/wiki/Q18199165
https://commons.wikimedia.org/wiki/File:Go_--_2021_--_6757.jpg – License: https://www.wikidata.org/wiki/Q18199165
https://en.wikipedia.org/wiki/Alpha%E2%80%93beta_pruning#/media/File:AB_pruning.svg - License: CC BY-SA 3.0
https://en.wikipedia.org/wiki/Monte_Carlo_tree_search#/media/File:MCTS_Algorithm.png - License: CC BY-SA 4.0

Modeling Strong and Human-Like Gameplay with KL-Regularized Search





Athul Paul Jacob* MIT
David J. Wu* FAIR
Gabriele Farina* CMU
Adam Lerer FAIR
Hengyuan Hu FAIR
Anton Bakhtin FAIR
Jacob Andreas MIT
Noam Brown FAIR

* Equal Contribution

Thank you!

Paper:

Poster details:
HALL E #816
6:30PM - 8:30PM