THE GREAT EXHIBITION ROAD FESTIVAL 2026 | IMPERIAL COLLEGE LONDON

Can You Train AI to Stay Safe?

Choose Pac-Man or MetaDrive, tune risk and reward, then climb the leaderboard.

Formal Methods in AI | IMPERIAL | Safe & Trusted AI UKRI Centre for Doctoral Training | THE GREAT EXHIBITION ROAD FESTIVAL

reward | safety cost | human choices | leaderboard

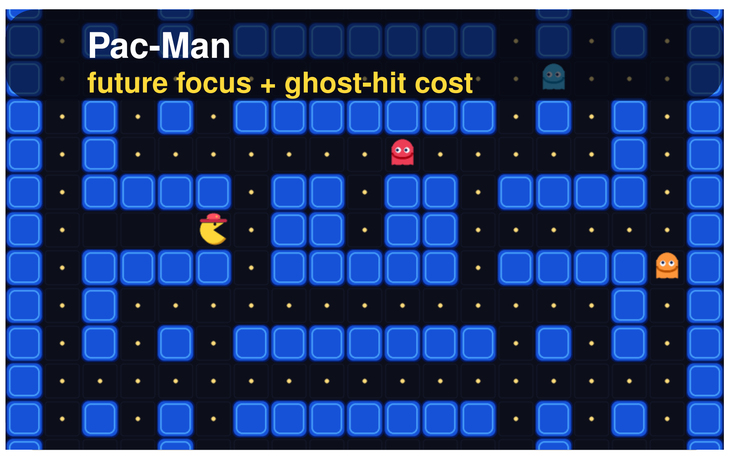
The Challenge

1. Pick a world
Choose Pac-Man or MetaDrive, then train.

2. Select choices
Set safety focus, traffic patience, and how much mistakes cost.

3. Watch training
See the results of your choices - your agent will be trained live!

4. Compare scores
Try again and climb the Pac-Man or MetaDrive leaderboard.



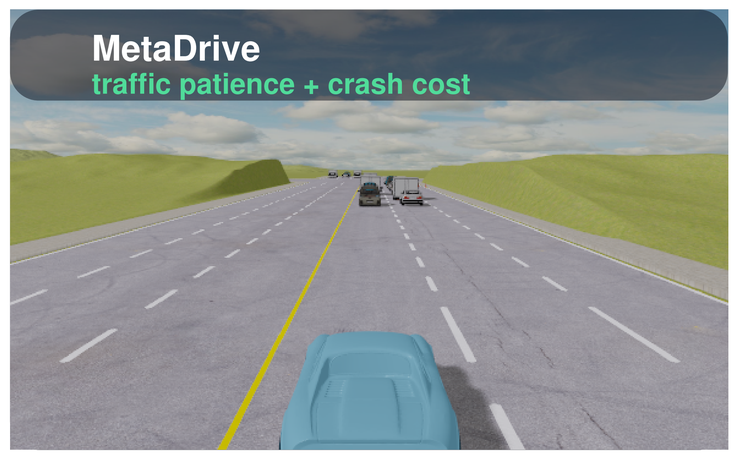
Choose your world

Pick Pac-Man or MetaDrive, then train safety-aware performance.

collect coins

drive safely

avoid costly mistakes



Learning from Experience

Reinforcement Learning is trial and error with feedback. An agent (for example a robot) tries actions, receives feedback (a 'reward'), then changes what it does next. The feedback is something given by a human. What should the agent try the hardest at, what behaviours should it avoid doing, how fast should it try to learn?

Tune the trade-off

Future focus — 0.9
next coin vs long-term plan

Ghost-hit cost — 5/10
cheap bump to serious mistake

Traffic patience — steady
take the gap to wait safely

Learning speed — steady
gentle, steady, or wild updates

The score we optimise

safe score = reward - safety cost

Pac-Man safety cost comes from ghost hits. MetaDrive safety cost comes from crashes or going off-road. Higher reward (like Pac-Man collecting more coins) helps only when the behaviour stays following the safety rule.

Policies to compare

Reward chaser / Speed chaser
Grabs points or progress quickly, even when risk spikes.
High reward pressure with a permissive cost penalty.

Balanced explorer / Careful driver
Still moves forward, but leaves room for the safety signal.
Reward and constraint terms are kept in tension.

Safety shield / Safety cruiser
Gives up speed or coins to stay clear of risky states.
Acts as if mistakes have a high multiplier.

Leaderboards

Each world has its own leaderboard. Scores are submitted after the final training result, so you can compare your results against other festival visitors!

Talk to us

We are happy to answer any questions.

Question for visitors

How much freedom should an AI get when mistakes have real costs?

play it safe | take some risks | ask a human first

For curious visitors

This is Constrained Reinforcement Learning: optimise reward while accounting for costs that should not be ignored.

Beyond the demo

fairness | reliability | physical risk

Safety signals can stand in for privacy, permission, fairness, reliability, or physical risk.

Why compare?

Good performance is only impressive when it is still safety compliant.

Try the demo: can you train a high-scoring agent that still follows the rule?