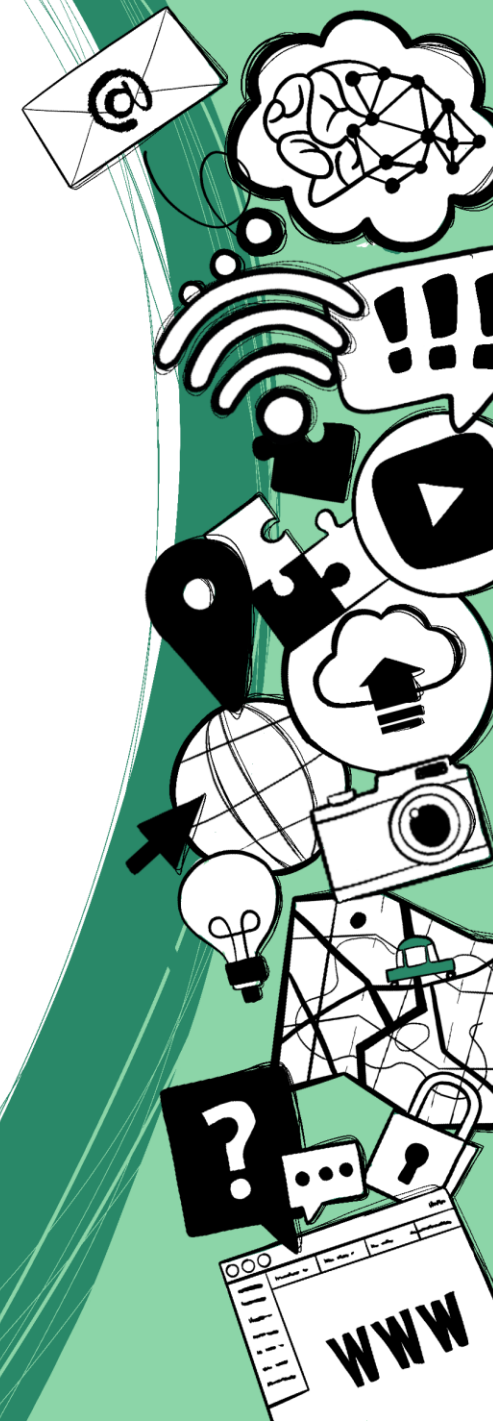


Megerősítőes tanulás



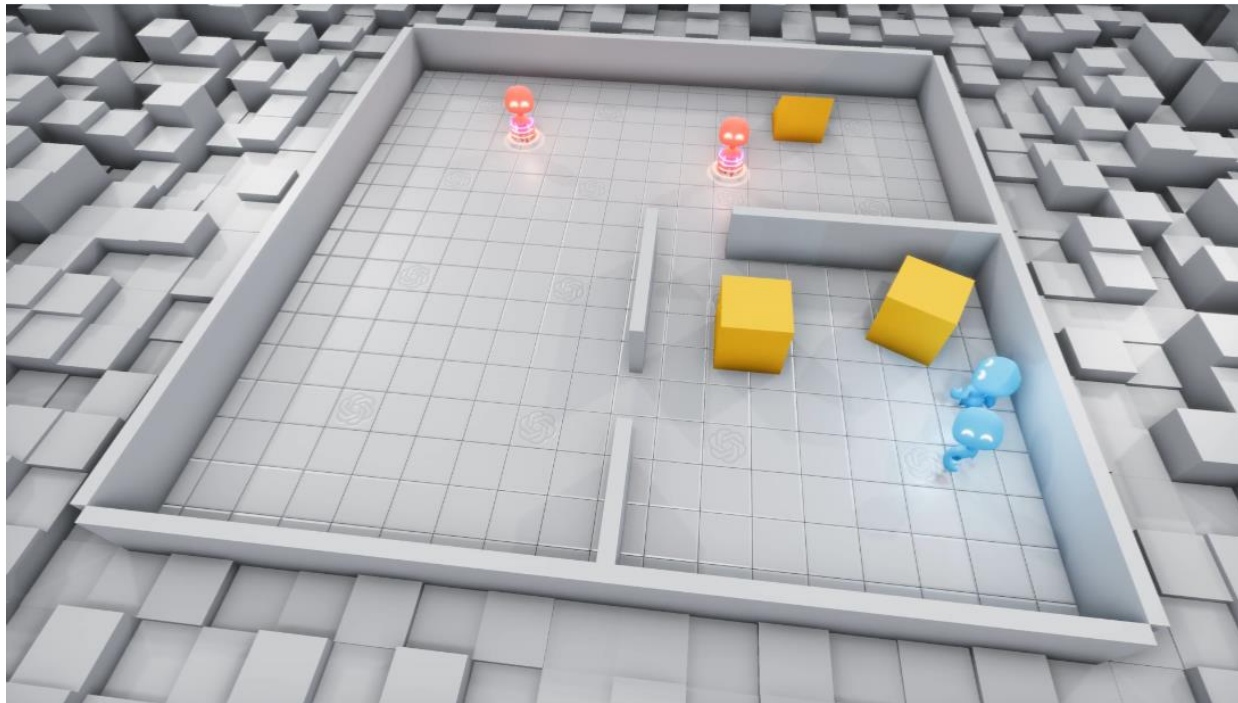
Példák





OpenAI Elrejt és megkeres (Hide and Seek)

<https://openai.com/blog/emergent-tool-use/>



Leela Sakk zéró (Chess Zero) (Lc0)



Engine PV

LCZero	Stockfish
0.30-dag-9a9c42d_784968	dev16_20221027
0.43	0.73
Evaluation: 44.5	Evaluation: 44.5
38/81	55/61
66.3knps	63.1Mnps
26.4M	2G
1.8k	893.8k
11.0 (3599)	15.0 (3625)
01:14:40	08:53:32
00:56	15:44

Eval | Mob | Time | D/SD | Speed | Nodes | TB | Kibitz

Agree

1 e4 a6 2 d4 b5 3 Nf3 Bb7 4 Bd3 e6 5 O-O c5 6 c3 Nf6 7 Re1 d6 8 a4 c4 9 Bc2 Nbd7 10 d5 exd5 11 exd5+ Be7 12 axb5 axb5 13 Rxa8 Bxa8 14 Qe2 Bxd5 15 Nd4 Nc5 16 Na3 Ne6 17 Ndx5 O-O 18 Nxc4 Bc6 19 Ne3 Qd7 20 Na3 d5 21 Ng4 Nxc4 22 Qxg4 Re8 23 Bd2 Rb8 24 Bf5 Rxb2 25 Bc1 Ra2 26 **Nc2***

0.73 Stockfish [6.2% W | 93.7% D | 0.1% B]

[10] 26 ... Bf6 27 Nd4 Qd6 28 Nxe6 fxe6 29 Bxe6+ Kh8 30 Bf5 g6 31 Bb1 @ Ra1 32 Qc8+ Kg7 33 Qh3 Bd7 34 Qh6+ Kg8 35 Bb2 Ra6 36 Qd2 Qc5 37 h3 Be6 38 Bd3 Rc6 39 Rc1 Rc8 40 Qf4 Bg7 41 Rc2 Qb6 42 h4 Bf7 43 g3 h5 44 Qc1 Qe6 45 Qd2 Qc6 46

0.43 LCZero [28.5% W | 71.5% D | 0.0% B]

[10] 26 **Nc2** Bf6 27 Nd4 Qd6 28 Nxe6 fxe6 29 Bxe6+ Kh8 30 Bf5 g6 31 Bb1 @ **Bd7** 32 Qf3 Ra1 33 Bf4 Qc5 34 Rc1 Qb6 35 Bg5 Kg7 36 Bxf6+ Qxf6 37 Qe3 Be6 38 Qd2 Bf7 39 Bd3 Ra5 40 g3 Qc5 41 h4 h5 42 Qb2 Rc6 43 Bf1 Rc4 44 Qb4 Rc6 45 Rc2 Rb6 46

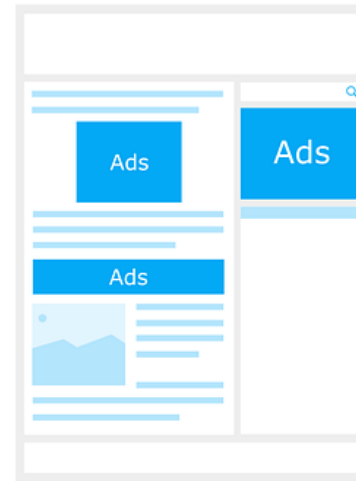
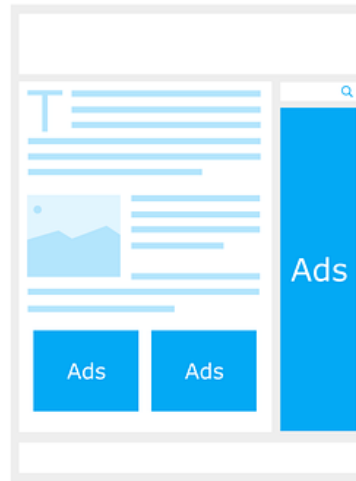
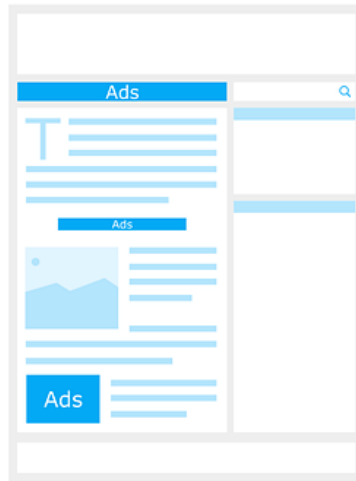
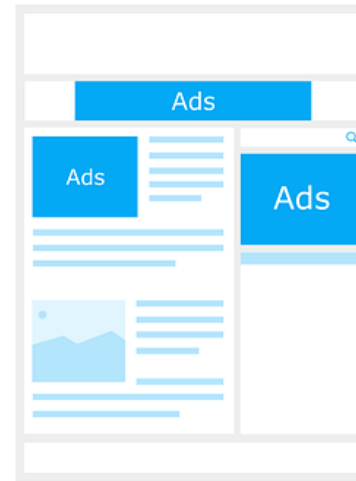
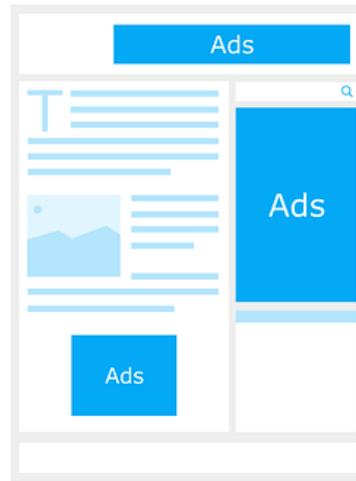
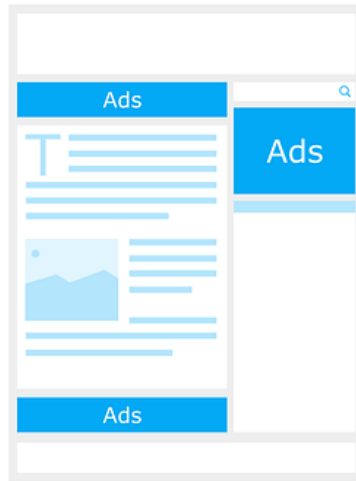
lc0



Alfa csillag (Alpha Star)



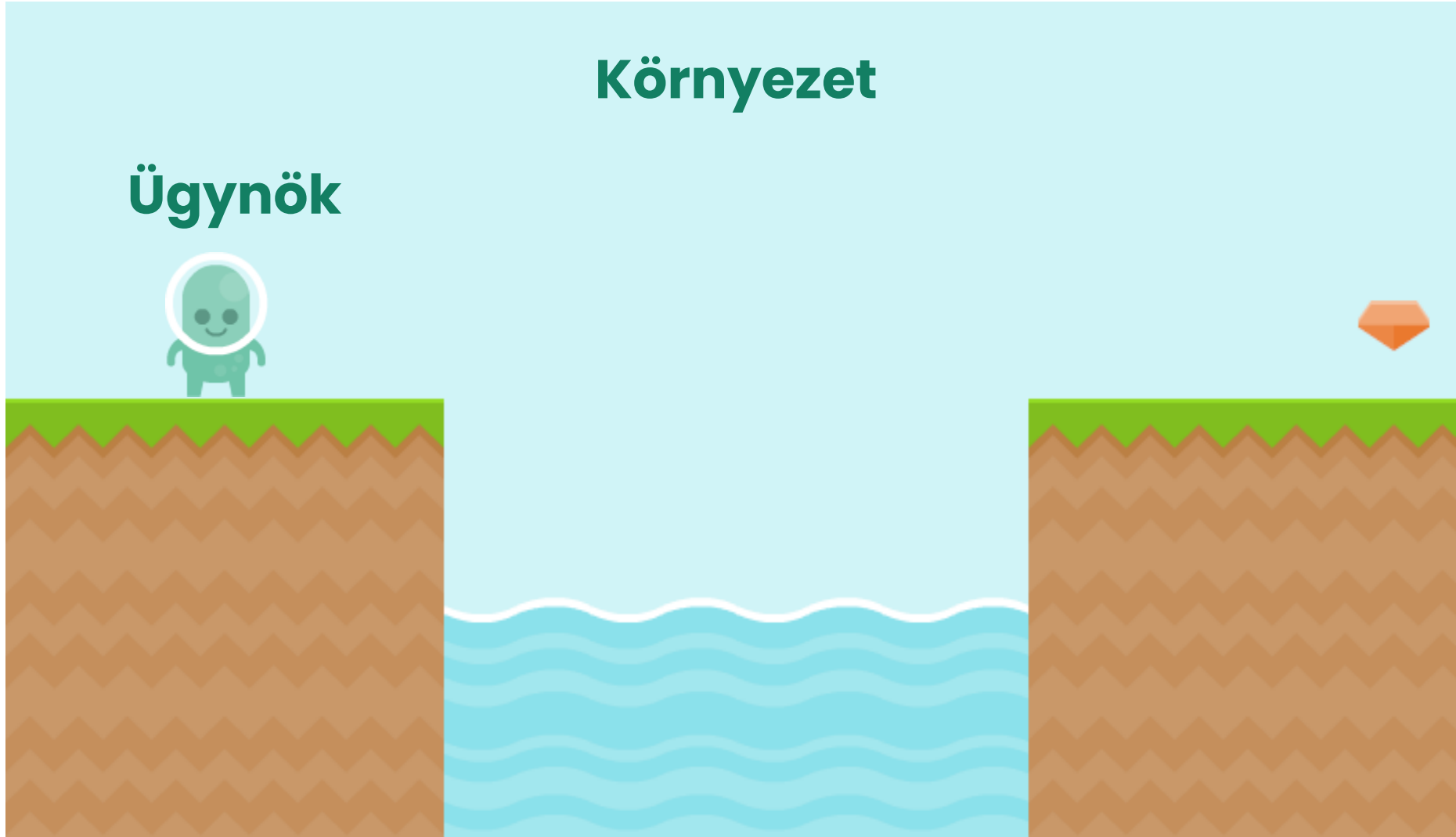
Egyedi hirdetés



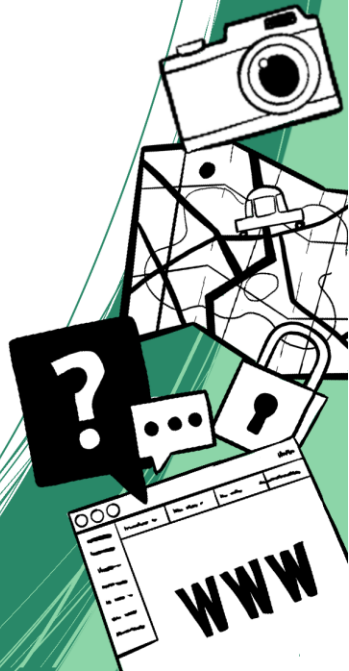
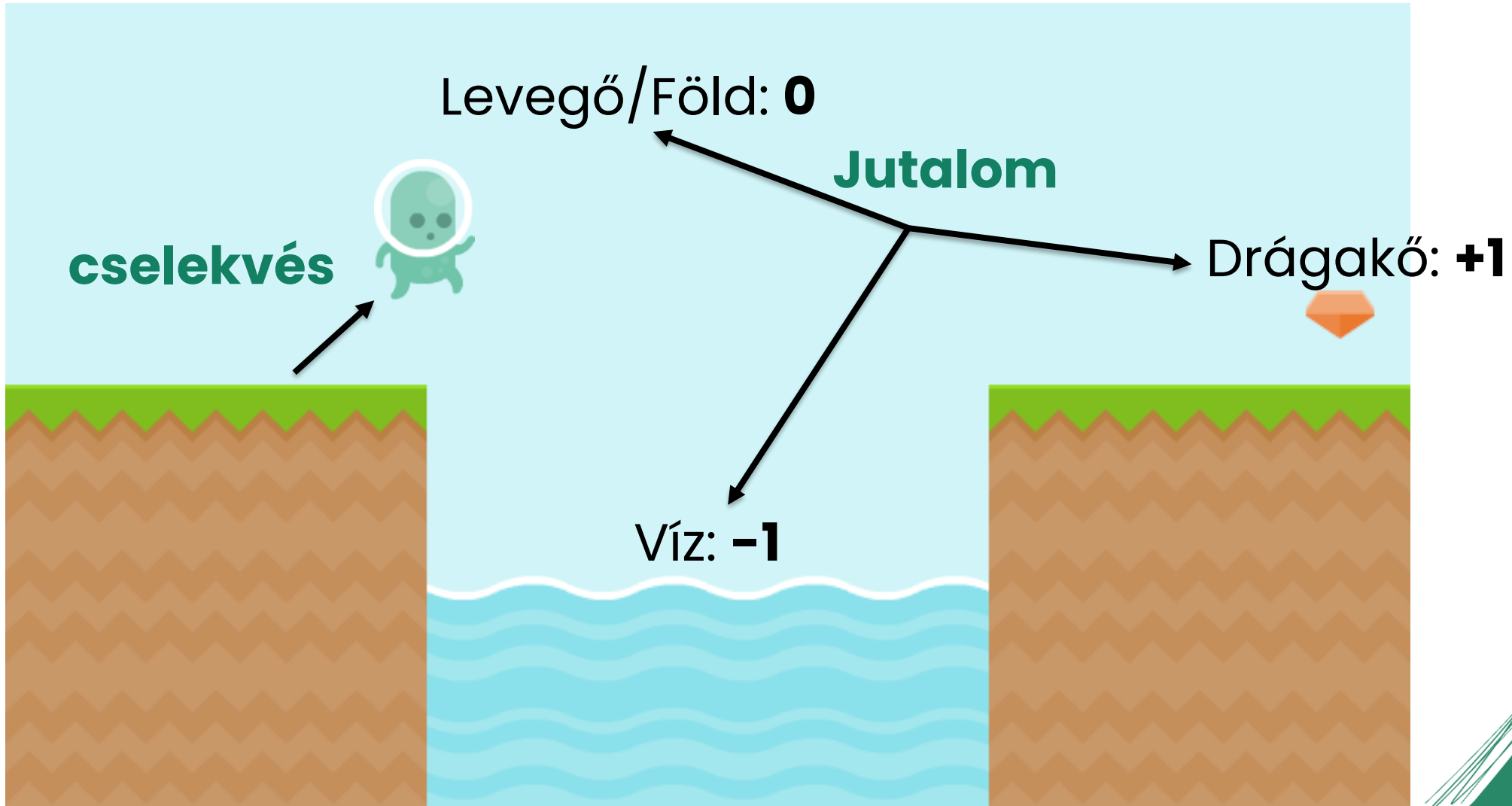
Megerősítőes tanulás (MT) alapok



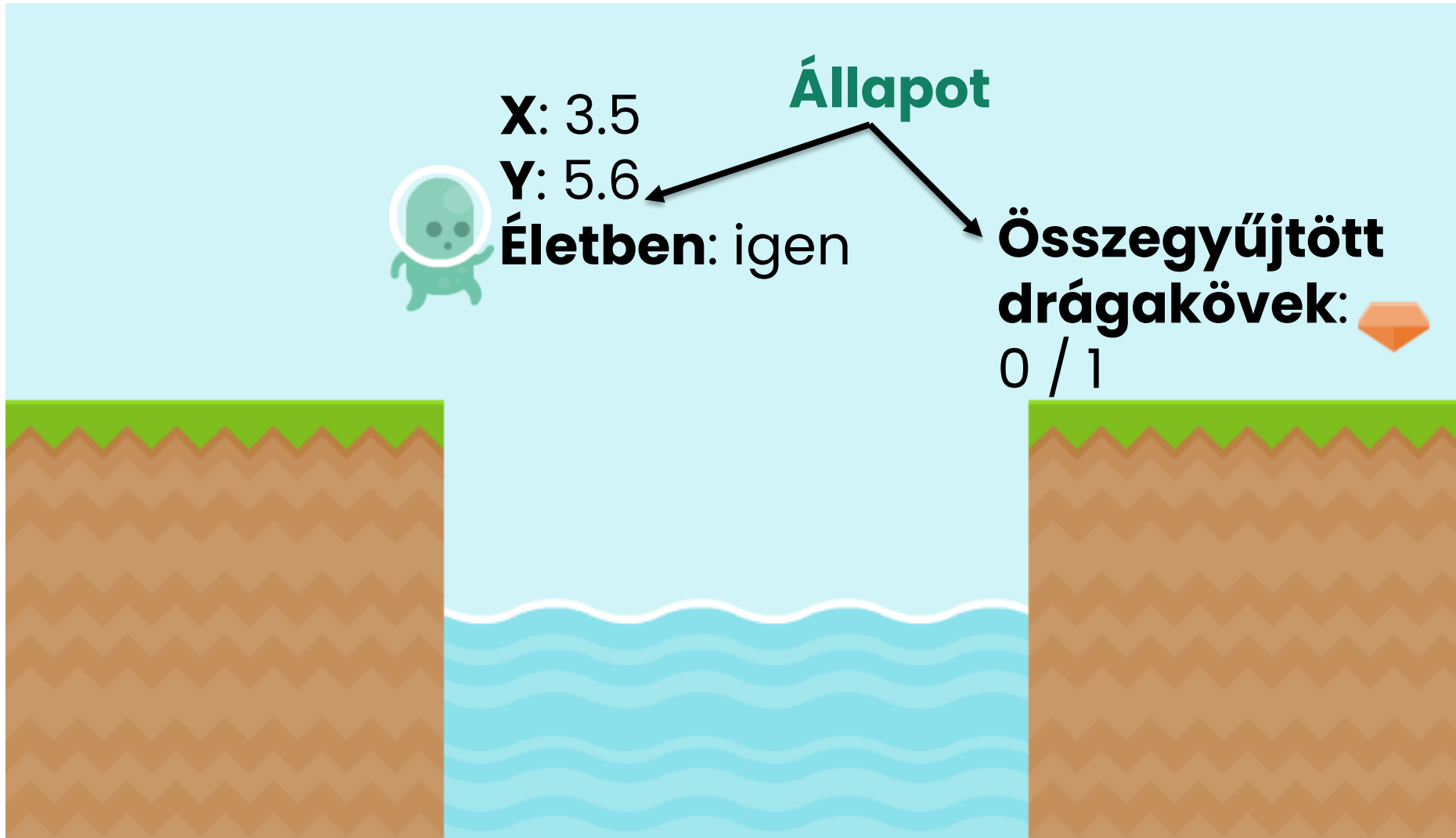
Amivel rendelkezünk



Amivel rendelkezünk

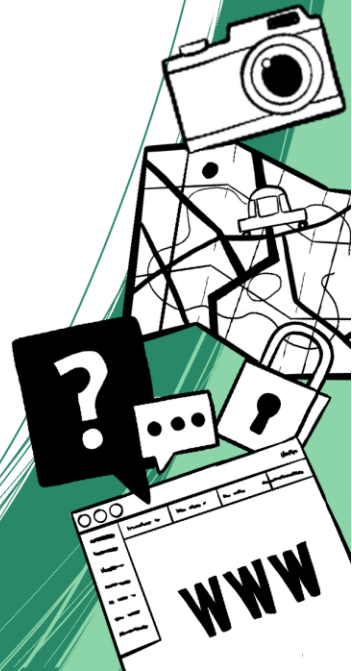


Amivel rendelkezünk





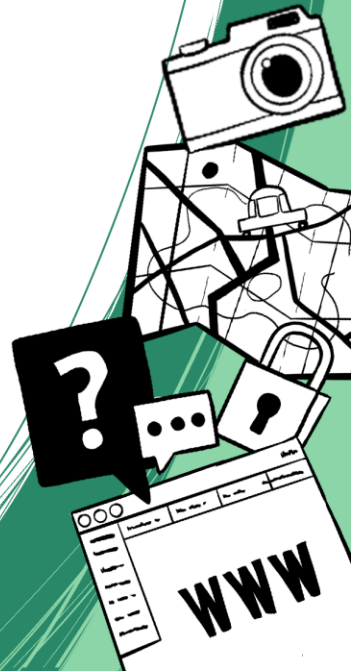
Példa – Amőba játék





Példa - Amőba játék

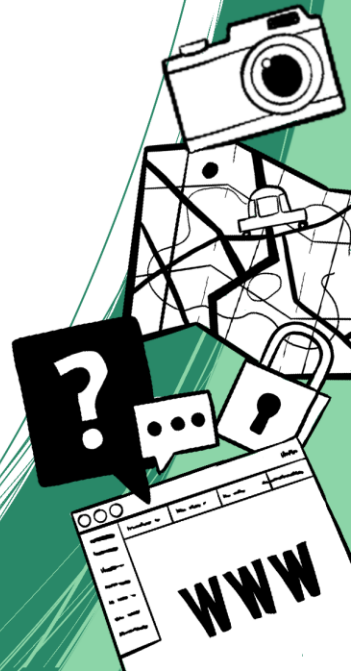
Ügynök



Példa - Amőba játék

Ügynök

- Játékosok (X, O)



Példa - Amőba játék

Ügynök

- Játékosok (X, O)

Állapot

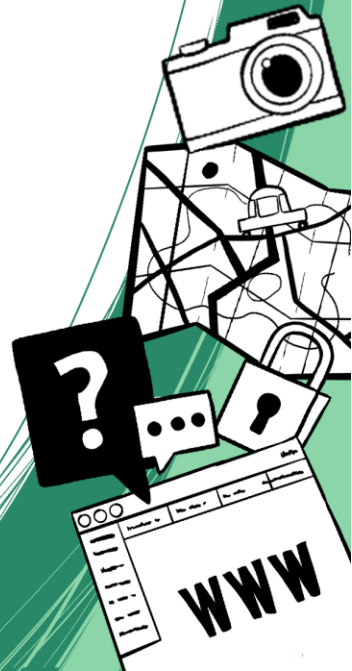
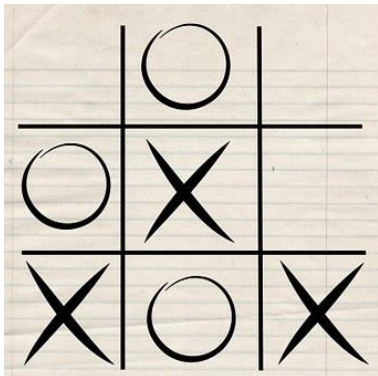


Példa - Amőba játék

Ügynök

- Játékosok (X, O)

Állapot



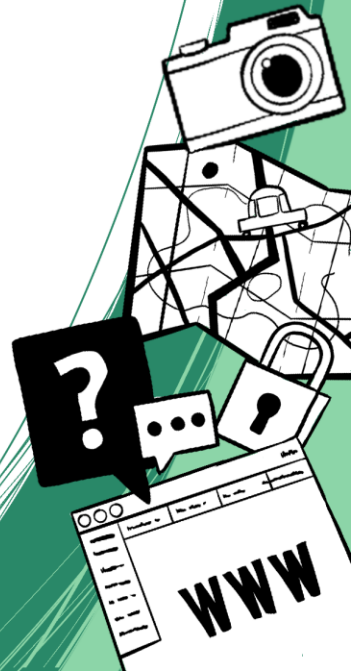
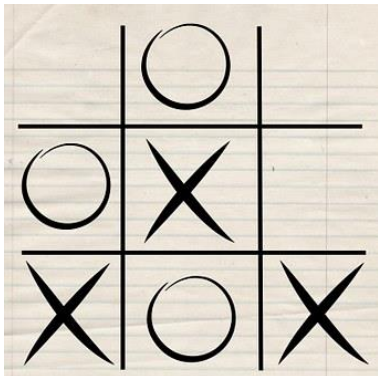
Példa - Amőba játék

Ügynök

- Játékosok (X, O)

Cselekvés

Állapot

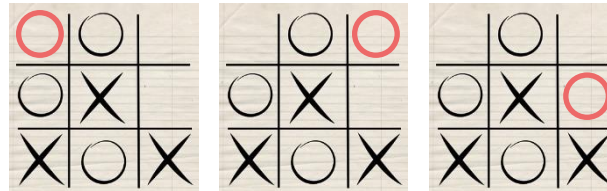


Példa - Amőba játék

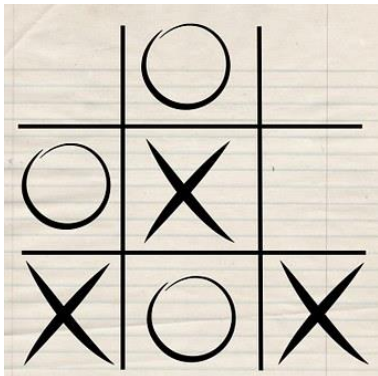
Ügynök

- Játékosok (X, O)

Cselekvések



Állapot

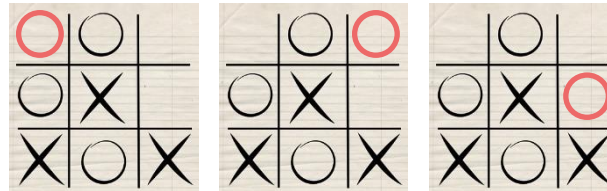


Példa - Amőba játék

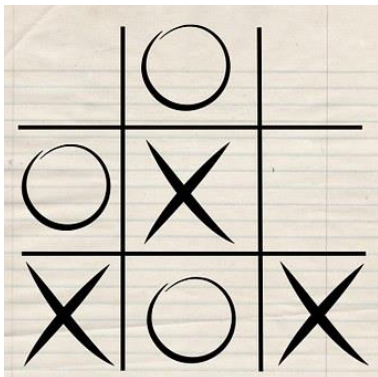
Ügynök

- Játékosok (X, O)

Cselekvések



Állapot



Jutalmak

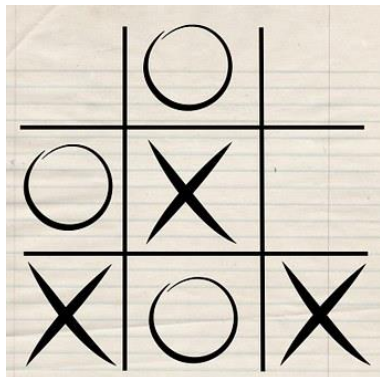


Példa - Amőba játék

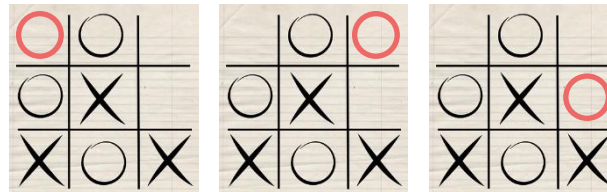
Ügynök

- Players (X, O)

Állapot

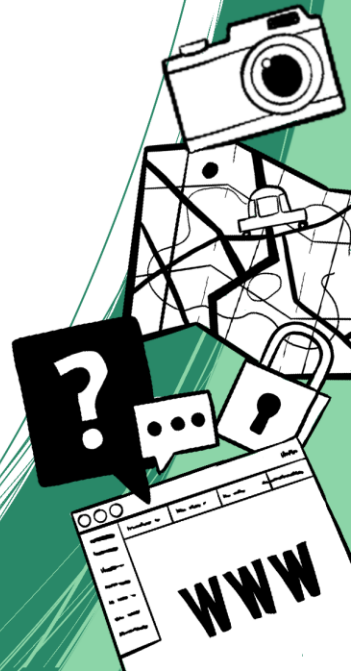


Cselekvések



Jutalmak

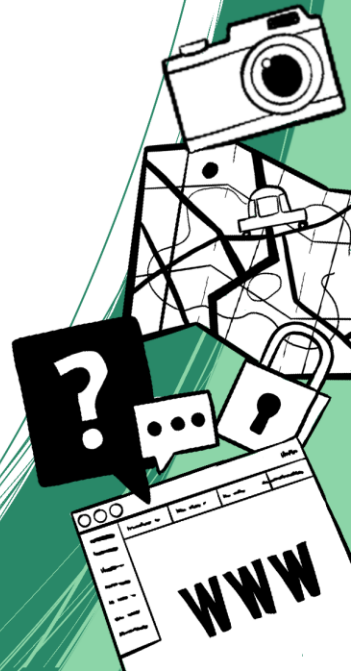
- Győzelem: **+1**
- Vereség: **-1**
- Egyéb: **0**



További példák

Mik az **ügynökök**, **állapot**, **cselekvések** és a lehetséges **jutalmak** a...

- ...Leela Chess Zero játékban
- ...OpenAI Hide and Seek játékban
- ...egyedi hirdetésekben



Q-Tanulás

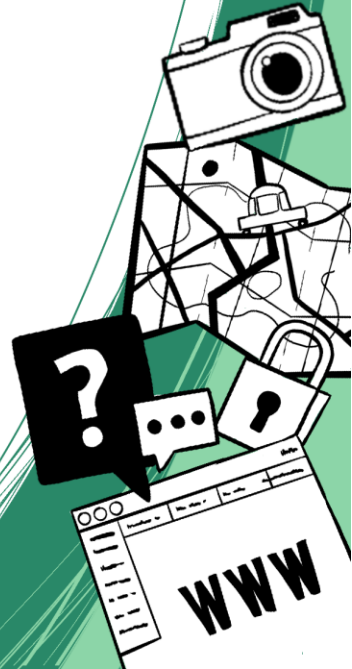
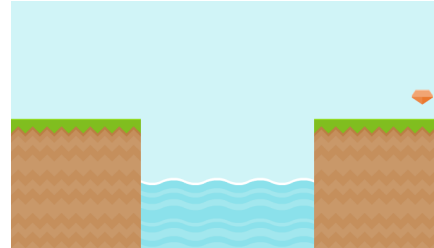


Megerősítőes tanulás hurok

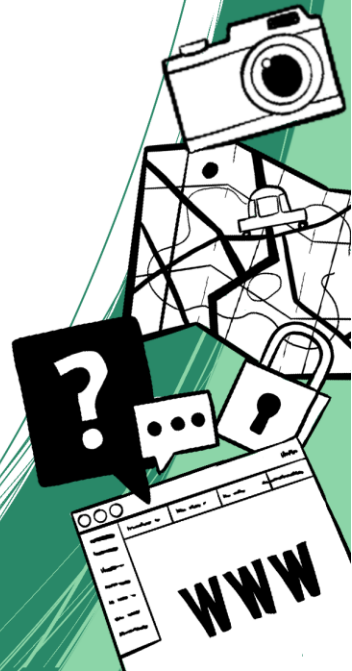
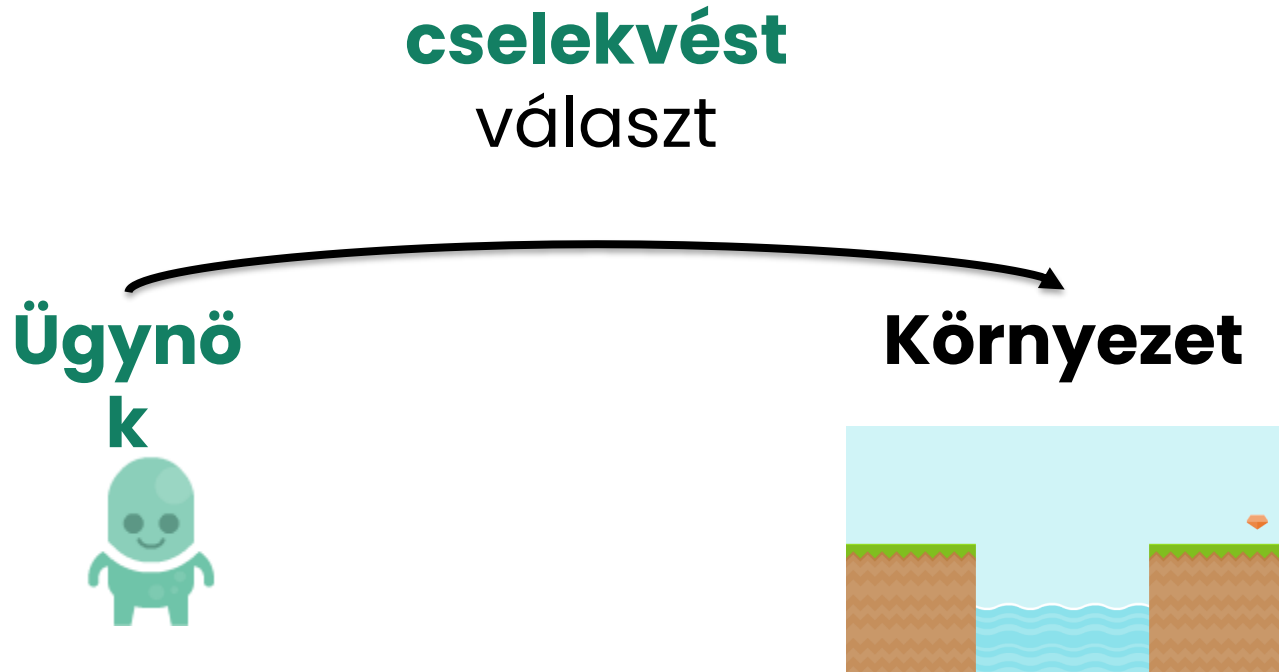
Ügynök



Környezet



Megerősítőes tanulás hurok

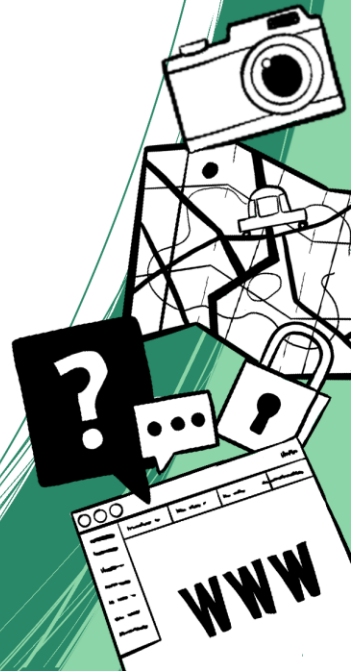


Megerősítőes tanulás hurok



Q-Tanulás

- Minden egyes **állapothoz** tárolunk egy **minőségi értéket (Q-érték)** cselekvésenként (mennyire jó a cselekvés az adott állapotot tekintve)



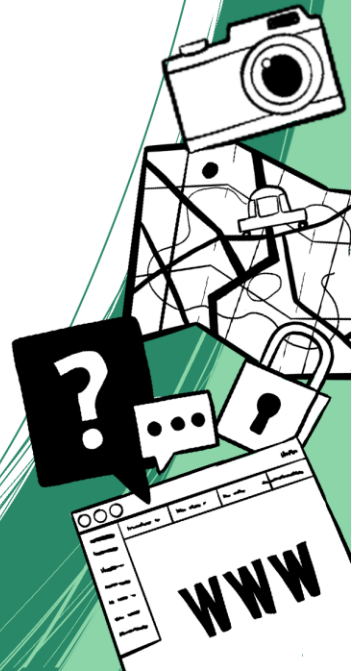
Q-Tanulás

- Minden egyes **állapothoz** tárolunk egy **minőségi értéket (Q-érték)** cselekvésenként (mennyire jó a cselekvés az adott állapotot tekintve)
- Ezután, amikor elérünk egy **állapotot**, kiválasztjuk a legmagasabb **Q-értékkel** rendelkező **cselekvést**



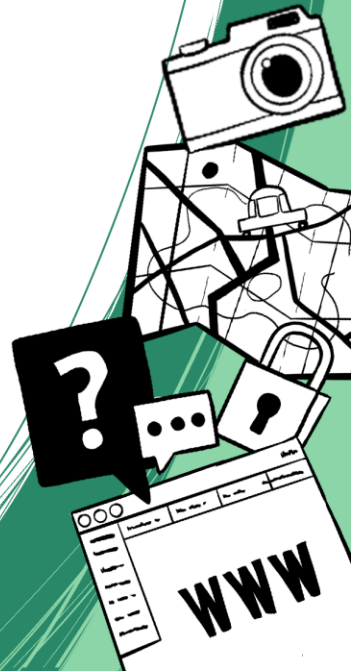
Q-Tanulás

- Minden egyes **állapothoz** tárolunk egy **minőségi értéket (Q-érték)** cselekvésenként (mennyire jó a cselekvés az adott állapotot tekintve)
- Ezután, amikor elérünk egy **állapotot**, kiválasztjuk a legmagasabb **Q-értékkel** rendelkező **cselekvést**
- Végül **növeljük/csökkentjük** a **Q-értéket** a **juttalom** tekintetében a **cselekvés** végrehajtása után



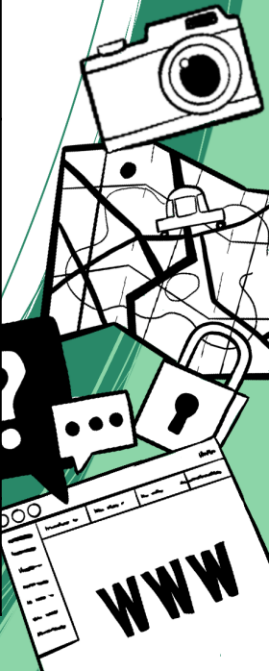
Q-Tanulás

- Minden egyes **állapothoz** tárolunk egy **minőségi értéket (Q-érték)** **cselekvésenként** (mennyire jó a cselekvés az adott állapotot tekintve)
- Ezután, amikor elérünk egy **állapotot**, kiválasztjuk a legmagasabb **Q-értékkel** rendelkező **cselekvést**
- Végül **növeljük/csökkentjük** a **Q-értéket** a **juttalom** tekintetében a **cselekvés** végrehajtása után
- Kisebb forgatókönyvek esetén ez egy táblázatban (**Q-táblázat**) tárolható



Q-Táblázat példa

Állapot	Lépés balra	Lépés jobbra	Ugrás balra	Ugrás jobbra
	0	1	-1	0
	-1	0	-1	1
...				



Q-Táblázat példa

Állapot	Lépés balra	Lépés jobbra	Ugrás balra	Ugrás jobbra
	-0.2	1	-1	0
	-1	0.5	-1	1
...				



Érme játék

Próbáld ki egy **valós példán!**

