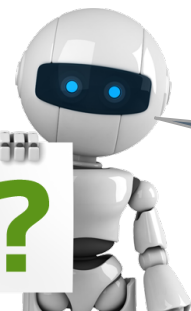
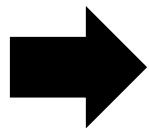


Next Steps

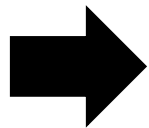


Anomaly Detection



Tennis Ball

Today



I do not know what this is.

Anomaly Detection

Explainable A.I.



Shoe

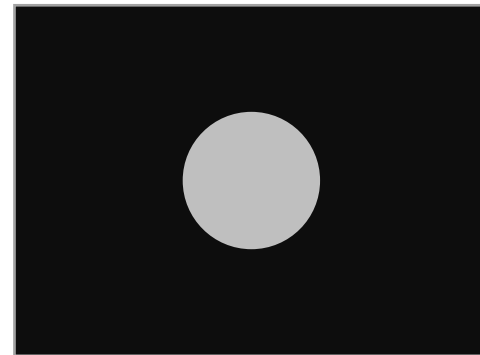
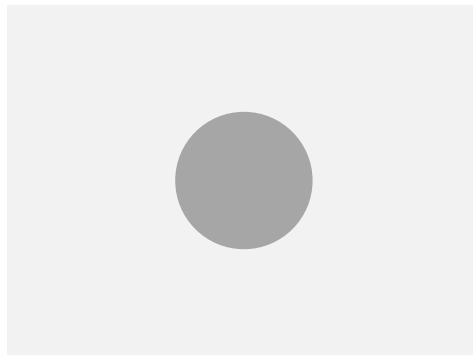
Puma

?

How to Prevent Adversarial Attack?

Adversarial Attack

Human



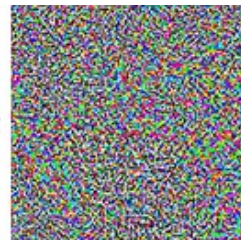
Machine



"panda"

57.7% confidence

+ ϵ



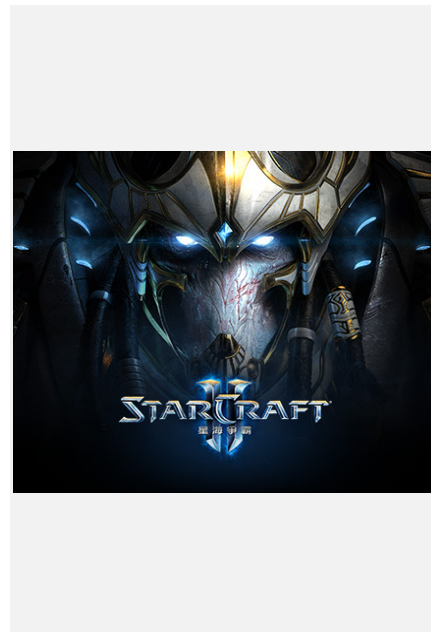
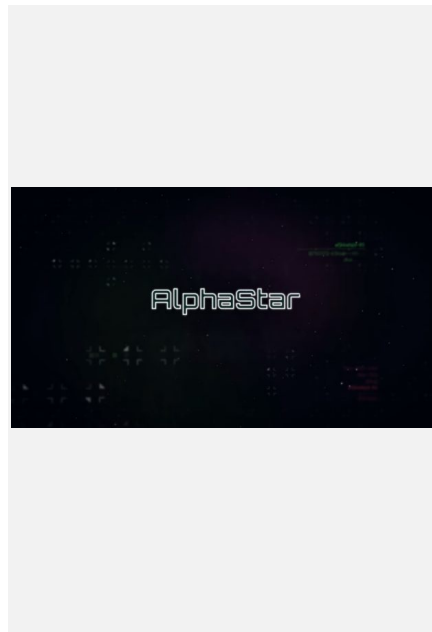
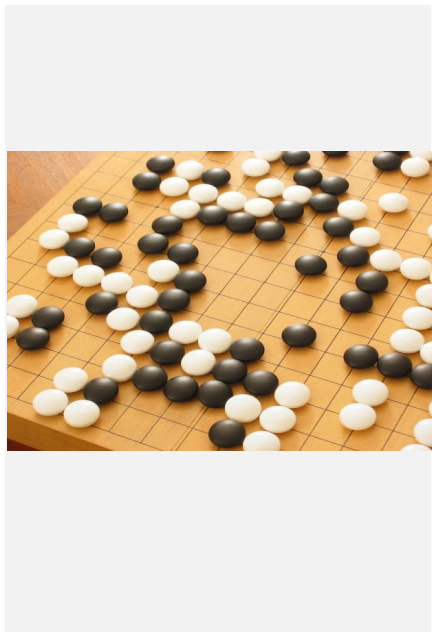
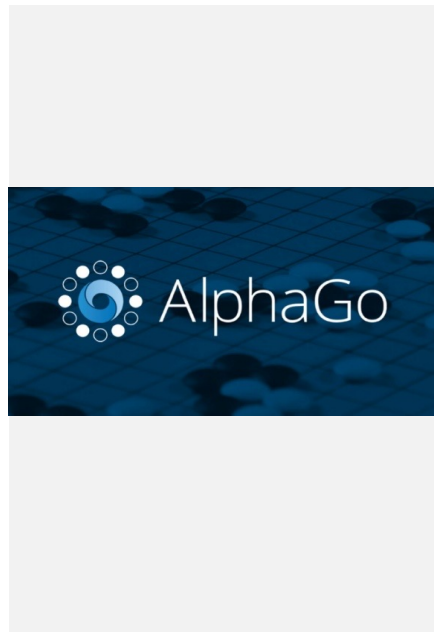
=



"gibbon"

99.3% confidence

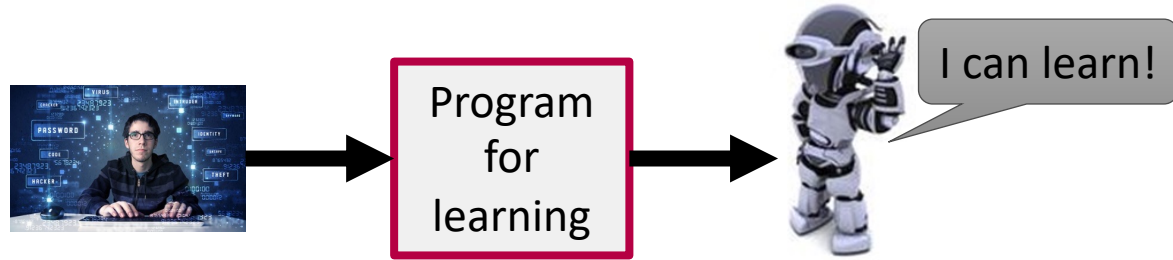
Lifelong Learning



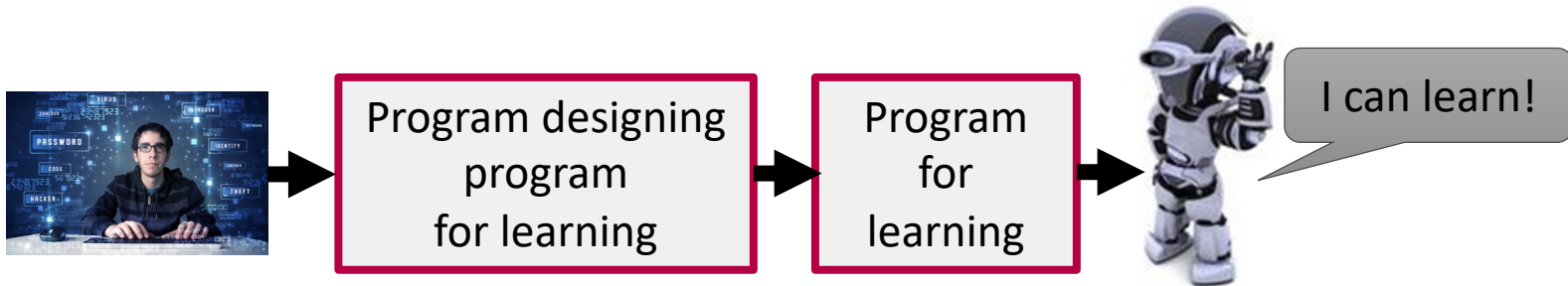
Meta-Learning/Learn to Learn

Learn to Learn

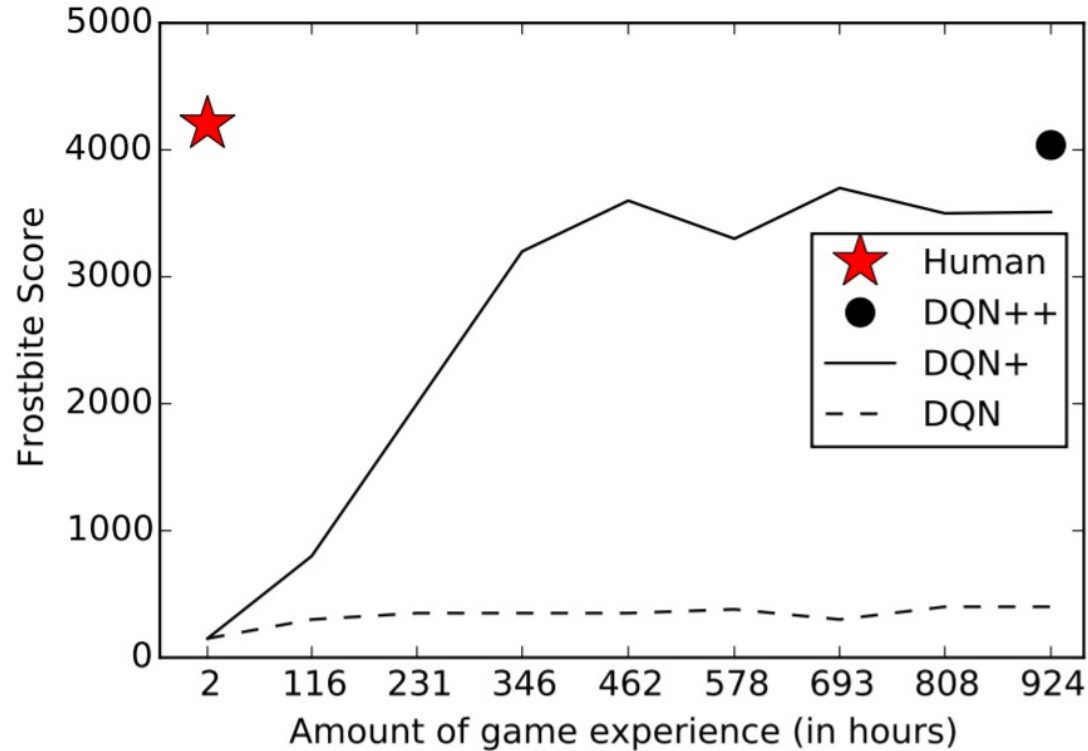
- Now we design the learning algorithm.



- Can the machine learn the learning algorithm?



Reinforcement Learning



Reinforcement Learning, Continued

In order to train AlphaStar, we built a highly scalable distributed training setup using **Google's v3 TPUs that** supports a population of agents learning from many thousands of parallel instances of StarCraft II. The AlphaStar league was run for 14 days, using 16 TPUs for each agent. During training, each agent experience up to **200 years** of real-time StarCraft play. The final AlphaStar agent consists of the components of the **Nash distribution of the league**—in other words, the most effective mixture of strategies that have been discovered—that run on a single desktop GPU.



Wrong Assumption

Training data and testing data have the same distribution.

Training data



Testing data

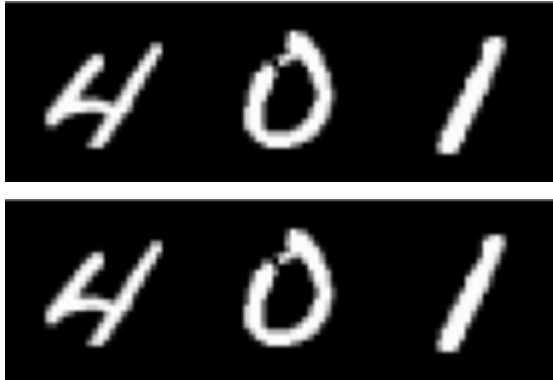


This is a lie.

Wrong Assumption, Continued

How to do unsupervised domain adaption?

Training data - 99.5%



Testing data – 57.5%



Ethical Considerations

Two Petty Theft Arrests

VERNON PRATER

Prior Offenses
2 armed robberies, 1
attempted armed
robbery

Subsequent Offenses
1 grand theft

LOW RISK

3

BRISHA BORDEN

Prior Offenses
4 juvenile
misdemeanors

Subsequent Offenses
None

HIGH RISK

8

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.



References

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *Pro Publica*.
- Ganin, Y., & Lempitsky, V. (n.d.). Unsupervised domain adaption by backpropagation.
- Lake, B.M., Ullman, T.D., Tenenbaum, J.B., Gershman, S.J. (n.d.). Building machines that learn and think like people.

