

&lt; Blog



BLOG POST  
RESEARCH

24 JAN 2019

# AlphaStar: Mastering the Real-Time Strategy Game StarCraft II

Games have been used for decades as an important way to test and evaluate the performance of artificial intelligence systems. As capabilities have increased, the research community has sought games with increasing complexity that capture different elements of intelligence required to solve scientific and real-world problems. In recent years, StarCraft, considered to be one of the most challenging Real-Time Strategy (RTS) games and one of the longest-played esports of all time, has emerged by consensus as a “grand challenge” for AI research.

Now, we introduce our [StarCraft II](#) program AlphaStar, the first Artificial Intelligence to

strongest professional StarCraft players, 5-0, following a successful benchmark match against his team-mate Dario “TLO” Wünsch. The matches took place under professional match conditions on a competitive ladder map and without any game restrictions.

Although there have been significant successes in video games such as Atari, Mario, Quake III Arena Capture the Flag, and Dota 2, until now, AI techniques have struggled to cope with the complexity of StarCraft. The best results were made possible by hand-crafting major elements of the system, imposing significant restrictions on the game rules, giving systems superhuman capabilities, or by playing on simplified maps. Even with these modifications, no system has come anywhere close to rivalling the skill of professional players. In contrast, AlphaStar plays the full game of StarCraft II, using a deep neural network that is trained directly from raw game data by supervised learning and reinforcement learning.



## Demonstration game against MaNa

# The Challenge of StarCraft

StarCraft II, created by [Blizzard Entertainment](#), is set in a fictional sci-fi universe and features rich, multi-layered gameplay designed to challenge human intellect. Along with the original title, it is among the biggest and most successful games of all time, with players competing in esports tournaments for more than 20 years.



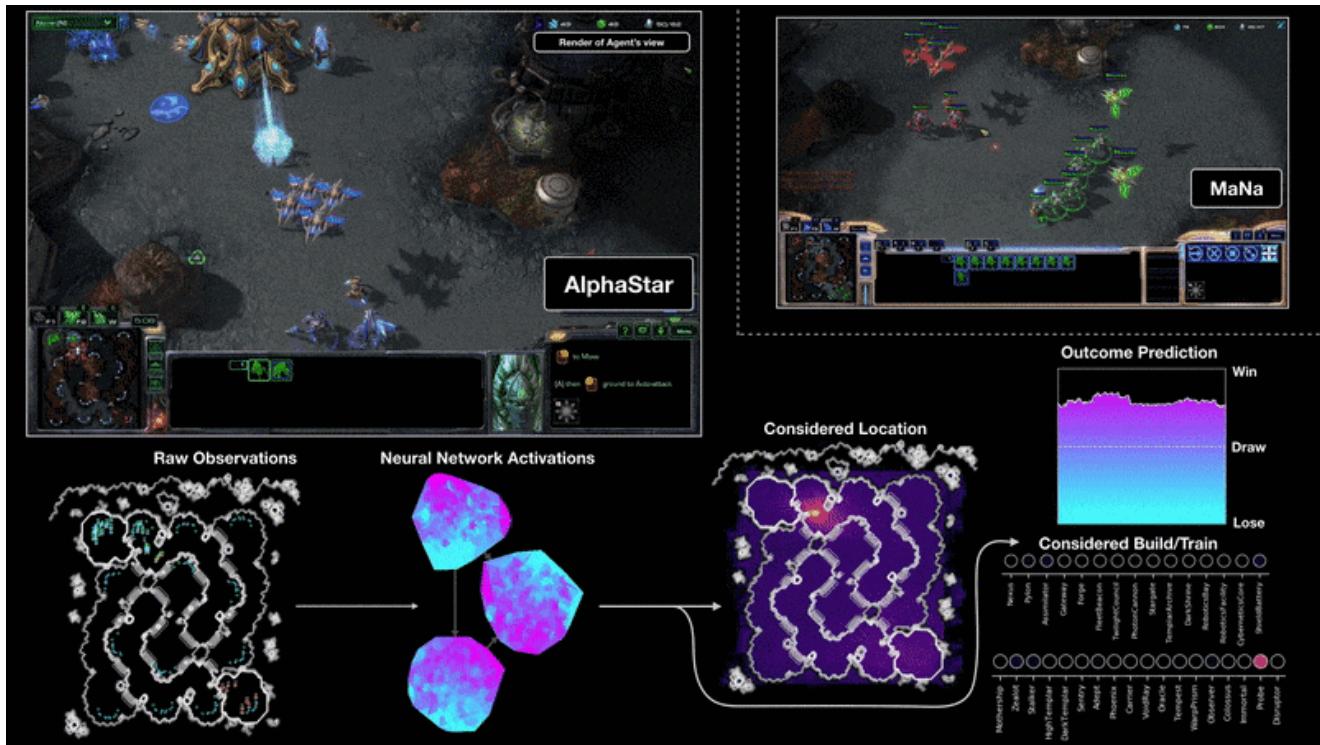
There are several different ways to play the game, but in esports the most common is a 1v1 tournament played over five games. To start, a player must choose to play one of three different alien “races” – Zerg, Protoss or Terran, all of which have distinctive characteristics and abilities (although professional players tend to specialise in one race). Each player starts with a number of worker units, which gather basic resources to build more units and structures and create new technologies. These in turn allow a player to harvest other resources, build more sophisticated bases and structures, and develop new capabilities that can be used to outwit the opponent. To win, a player must carefully balance big-picture management of their economy – known as macro – along with low-level control of their individual units – known as micro.

The need to balance short and long-term goals and adapt to unexpected situations, poses a huge challenge for systems that have often tended to be brittle and inflexible. Mastering this problem requires breakthroughs in several AI research challenges including:

expand the frontiers of strategic knowledge.

- **Imperfect information:** Unlike games like chess or Go where players see everything, crucial information is hidden from a StarCraft player and must be actively discovered by “scouting”.
- **Long term planning:** Like many real-world problems cause-and-effect is not instantaneous. Games can also take anywhere up to one hour to complete, meaning actions taken early in the game may not pay off for a long time.
- **Real time:** Unlike traditional board games where players alternate turns between subsequent moves, StarCraft players must perform actions continually as the game clock progresses.
- **Large action space:** Hundreds of different units and buildings must be controlled at once, in real-time, resulting in a combinatorial space of possibilities. On top of this, actions are hierarchical and can be modified and augmented. Our parameterization of the game has an average of approximately 10 to the 26 legal actions at every time-step.

Due to these immense challenges, StarCraft has emerged as a “grand challenge” for AI research. Ongoing competitions in both StarCraft and StarCraft II have assessed progress since the launch of the BroodWar API in 2009, including the AIIDE StarCraft AI Competition, CIG StarCraft Competition, Student StarCraft AI Tournament, and the Starcraft II AI Ladder. To help the community explore these problems further, we worked with Blizzard in 2016 and 2017 to release an open-source set of tools known as PySC2, including the largest set of anonymised game replays ever released. We have now built on this work, combining engineering and algorithmic breakthroughs to produce AlphaStar.

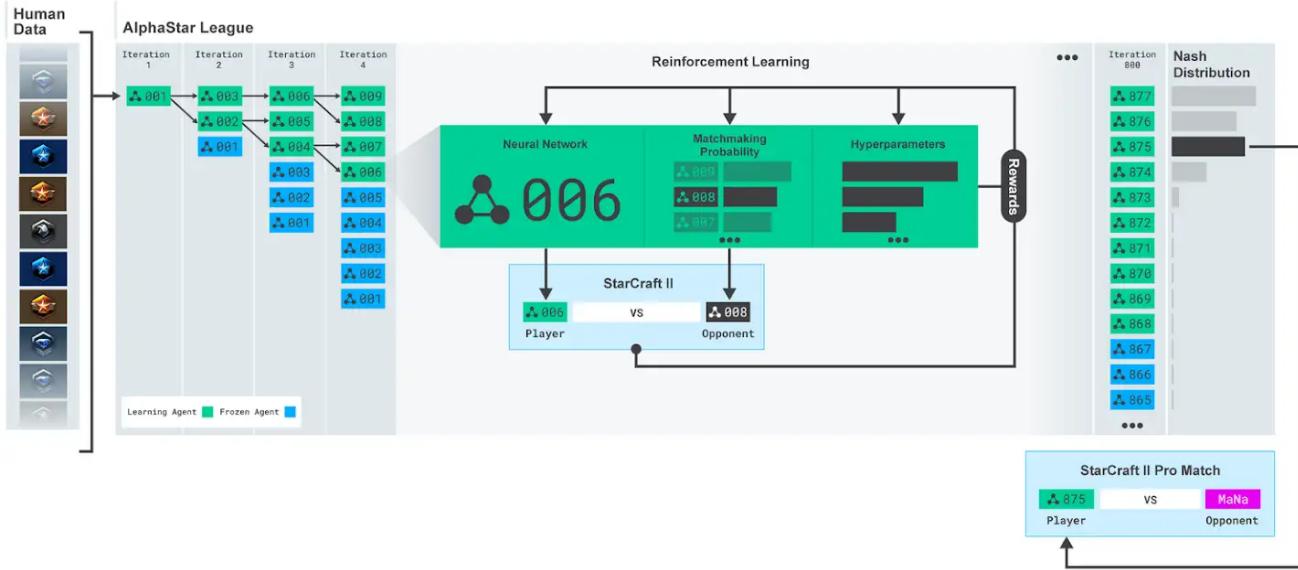


A VISUALISATION OF THE ALPHASTAR AGENT DURING GAME TWO OF THE MATCH AGAINST MANA. THIS SHOWS THE GAME FROM THE AGENT'S POINT OF VIEW: THE RAW OBSERVATION INPUT TO THE NEURAL NETWORK, THE NEURAL NETWORK'S INTERNAL ACTIVATIONS, SOME OF THE CONSIDERED ACTIONS THE AGENT CAN TAKE SUCH AS WHERE TO CLICK AND WHAT TO BUILD, AND THE PREDICTED OUTCOME. MANA'S VIEW OF THE GAME IS ALSO SHOWN, ALTHOUGH THIS IS NOT ACCESSIBLE TO THE AGENT.

# How AlphaStar is trained

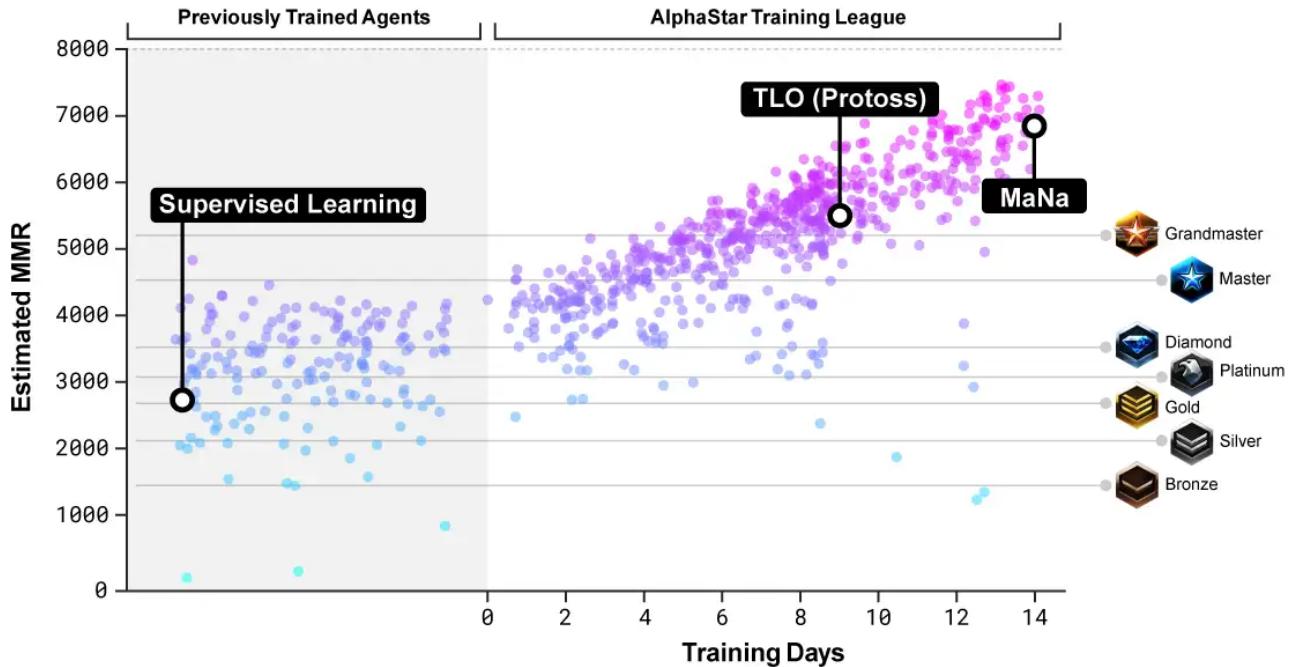
AlphaStar's behaviour is generated by a deep neural network that receives input data from the raw game interface (a list of units and their properties), and outputs a sequence of instructions that constitute an action within the game. More specifically, the neural network architecture applies a transformer torso to the units (similar to relational deep reinforcement learning), combined with a deep LSTM core, an auto-regressive policy head with a pointer network, and a centralised value baseline. We believe that this advanced model will help with many other challenges in machine learning research that involve long-term sequence modelling and large output spaces such as translation, language modelling and visual representations.

AlphaStar also uses a novel multi-agent learning algorithm. The neural network was initially trained by supervised learning from anonymised human games released by Blizzard. This allowed AlphaStar to learn, by imitation, the basic micro and macro-strategies used by players on the StarCraft ladder. This initial agent defeated the built-in



THE ALPHASTAR LEAGUE. AGENTS ARE INITIALLY TRAINED FROM HUMAN GAME REPLAYS, AND THEN TRAINED AGAINST OTHER COMPETITORS IN THE LEAGUE. AT EACH ITERATION, NEW COMPETITORS ARE BRANCHED, ORIGINAL COMPETITORS ARE FROZEN, AND THE MATCHMAKING PROBABILITIES AND HYPERPARAMETERS DETERMINING THE LEARNING OBJECTIVE FOR EACH AGENT MAY BE ADAPTED, INCREASING THE DIFFICULTY WHILE PRESERVING DIVERSITY. THE PARAMETERS OF THE AGENT ARE UPDATED BY REINFORCEMENT LEARNING FROM THE GAME OUTCOMES AGAINST COMPETITORS. THE FINAL AGENT IS SAMPLED (WITHOUT REPLACEMENT) FROM THE NASH DISTRIBUTION OF THE LEAGUE.

These were then used to seed a multi-agent reinforcement learning process. A continuous league was created, with the agents of the league – competitors – playing games against each other, akin to how humans experience the game of StarCraft by playing on the [StarCraft ladder](#). New competitors were dynamically added to the league, by branching from existing competitors; each agent then learns from games against other competitors. This new form of training takes the ideas of [population-based](#) and [multi-agent](#) reinforcement learning further, creating a process that continually explores the huge strategic space of StarCraft gameplay, while ensuring that each competitor performs well against the strongest strategies, and does not forget how to defeat earlier ones.

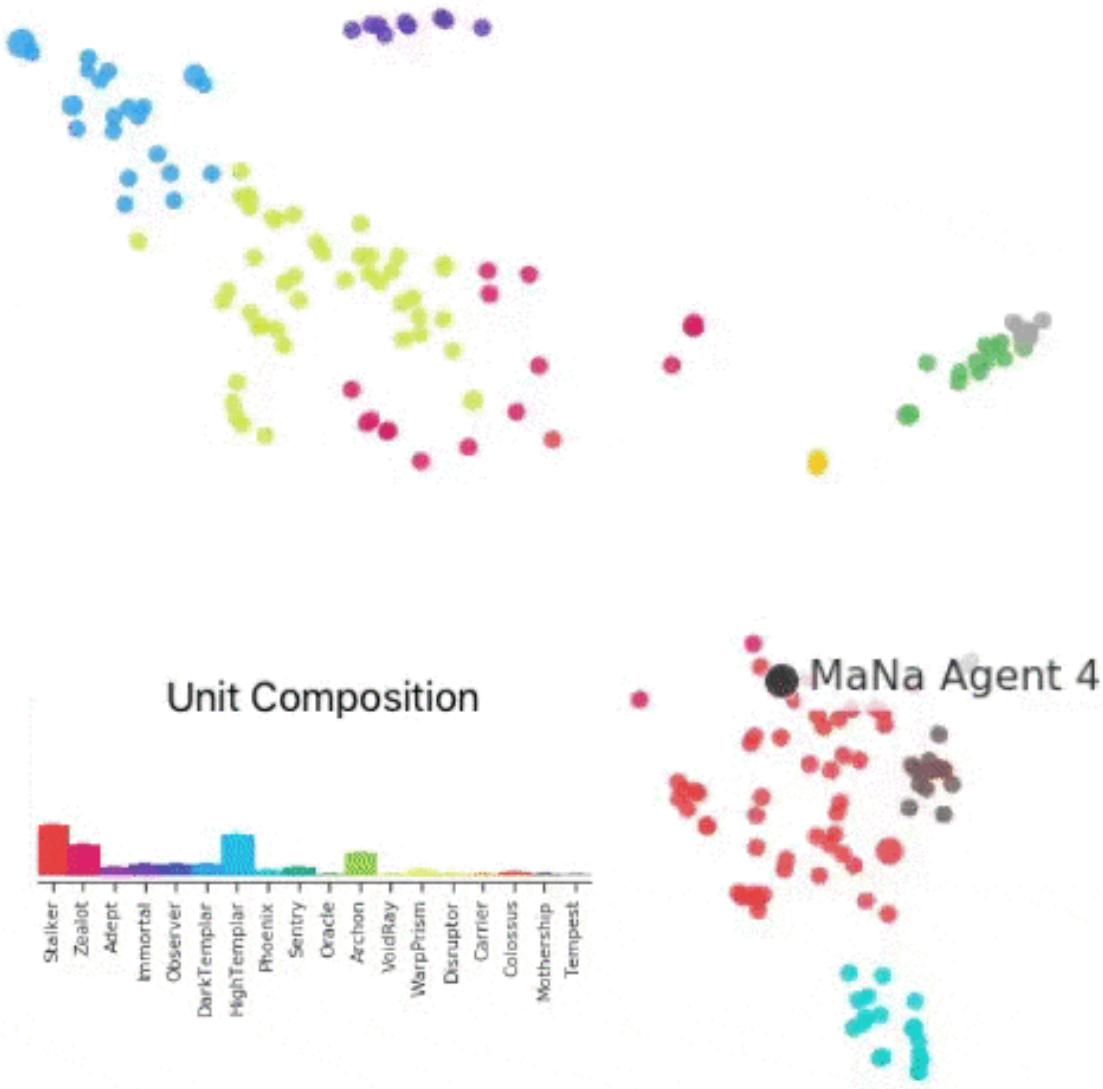


ESTIMATE OF THE MATCH MAKING RATING (MMR) – AN APPROXIMATE MEASURE OF A PLAYER'S SKILL – FOR COMPETITORS IN THE ALPHASTAR LEAGUE, THROUGHOUT TRAINING, IN COMPARISON TO BLIZZARD'S ONLINE LEAGUES.

As the league progresses and new competitors are created, new counter-strategies emerge that are able to defeat the earlier strategies. While some new competitors execute a strategy that is merely a refinement of a previous strategy, others discover drastically new strategies consisting of entirely new build orders, unit compositions, and micro-management plans. For example, early on in the AlphaStar league, "cheesy" strategies such as very quick rushes with Photon Cannons or Dark Templars were favoured. These risky strategies were discarded as training progressed, leading to other strategies: for example, gaining economic strength by over-extending a base with more workers, or sacrificing two Oracles to disrupt an opponent's workers and economy. This process is similar to the way in which players have discovered new strategies, and were able to defeat previously favoured approaches, over the years since StarCraft was released.

## MaNa Agent 4 Training Progression

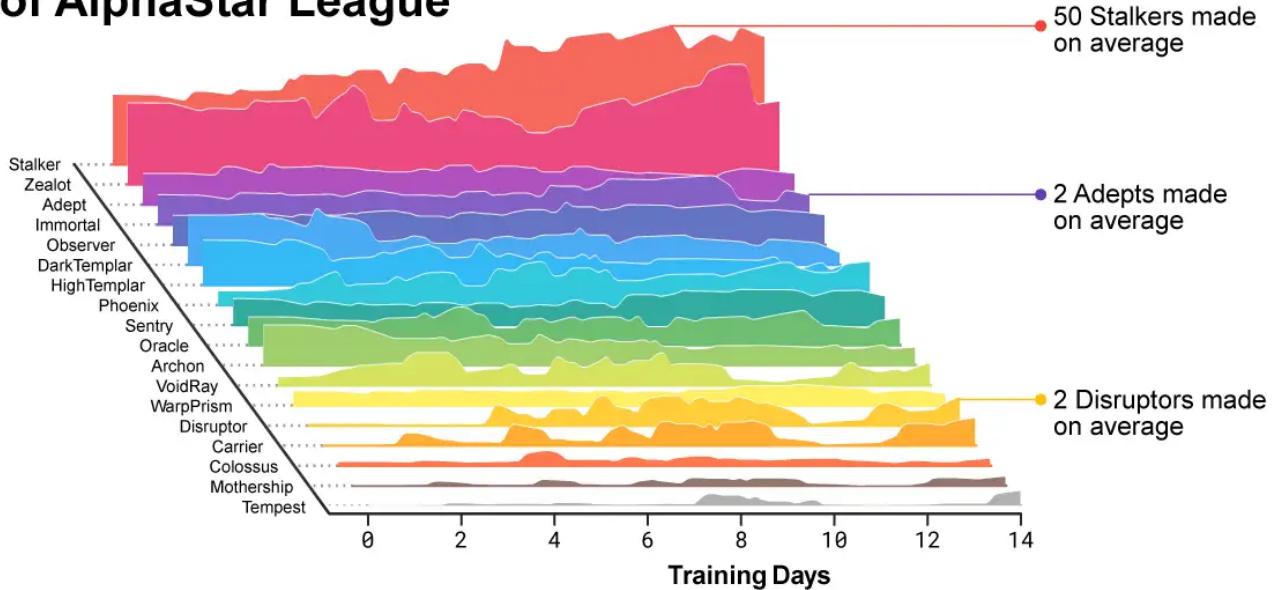
Size indicates Matchmaking Distribution



THE FIGURE SHOWS HOW ONE AGENT (BLACK DOT), WHICH WAS ULTIMATELY SELECTED TO PLAY AGAINST MANA, EVOLVED ITS STRATEGY AND COMPETITORS (COLOURED DOTS) DURING THE COURSE OF TRAINING. EACH DOT REPRESENTS A COMPETITOR IN THE ALPHASTAR LEAGUE. THE POSITION OF THE DOT REPRESENTS ITS STRATEGY (INSET), AND THE SIZE OF THE DOT REPRESENTS HOW FREQUENTLY IT IS SELECTED AS AN OPPONENT FOR THE MANA AGENT DURING TRAINING.

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 experienced 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.

## Units Counts of Nash of AlphaStar League

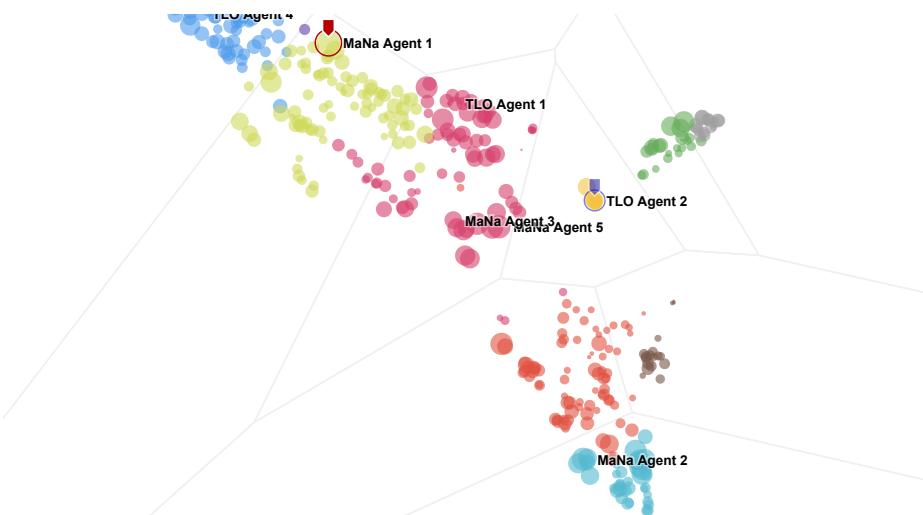


AS TRAINING PROGRESSED, THE LEAGUE CREATING ALPHASTAR CHANGED THE BLEND OF UNITS THAT IT BUILDS.

To encourage diversity in the league, each agent has its own learning objective: for example, which competitors should this agent aim to beat, and any additional internal motivations that bias how the agent plays. One agent may have an objective to beat one specific competitor, while another agent may have to beat a whole distribution of competitors, but do so by building more of a particular game unit. These learning objectives are adapted during training.

## AlphaStar League Strategy Map

Coloured by strategy clusters ▾




**MaNa Agent 1    TLO Agent 2**  
 MMR                    MMR  

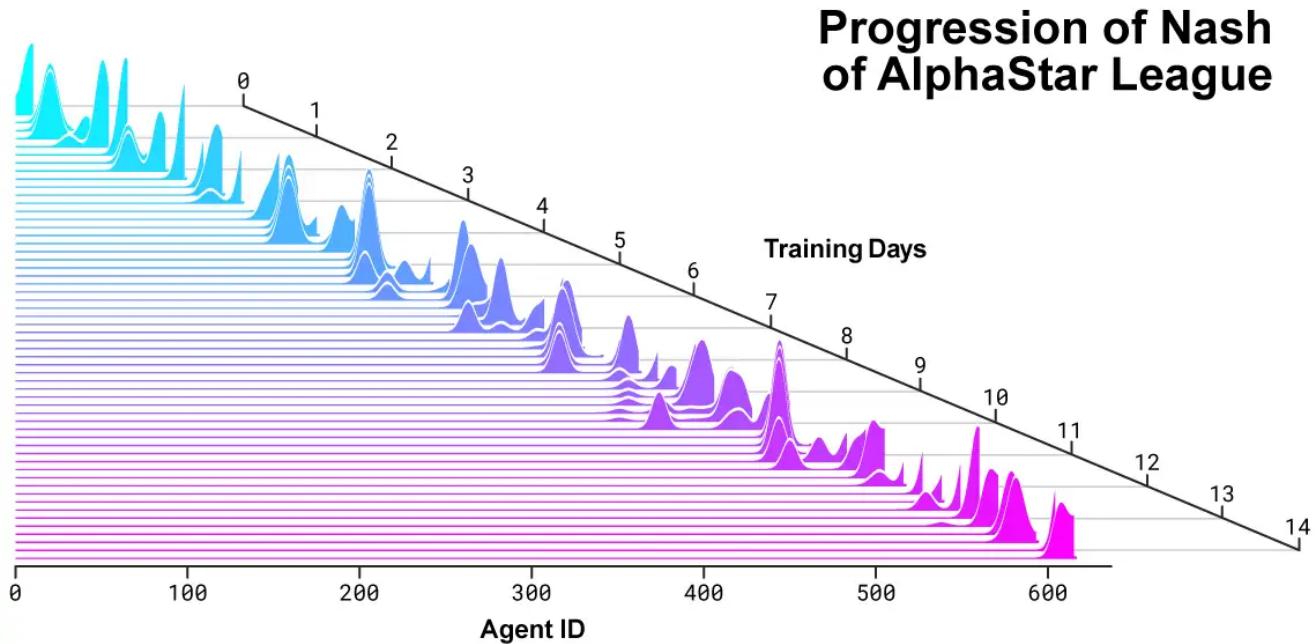
**Win rate of MaNa Agent 1 against TLO Agent 2**

Unit    average count    ▾ per game

	Stalker
Zealot	
Adept	
Immortal	
Observer	
DarkTemplar	
HighTemplar	
Phoenix	
Sentry	
Oracle	
Archon	
VoidRay	
WarpPrism	
Disruptor	

INTERACTIVE VISUALISATION WHICH SHOWS THE COMPETITORS OF THE ALPHASTAR LEAGUE.  
THE AGENTS THAT PLAYED AGAINST TLO AND MANA ARE SPECIFICALLY LABELLED.

The neural network weights of each agent are updated by reinforcement learning from its games against competitors, to optimise its personal learning objective. The weight update rule is an efficient and novel off-policy actor-critic reinforcement learning algorithm with experience replay, self-imitation learning and policy distillation.

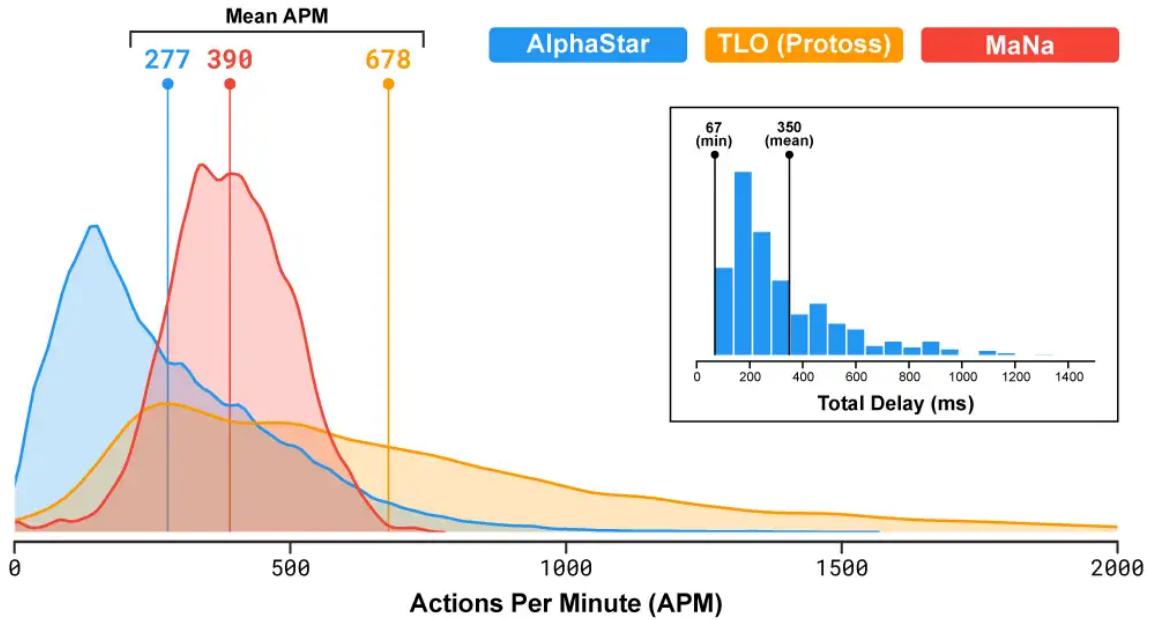


THE NASH DISTRIBUTION OVER COMPETITORS AS THE ALPHASTAR LEAGUE PROGRESSED AND NEW COMPETITORS WERE CREATED. THE NASH DISTRIBUTION, WHICH IS THE LEAST EXPLOITABLE SET OF COMPLEMENTARY COMPETITORS, WEIGHTS THE NEWEST COMPETITORS MOST HIGHLY, DEMONSTRATING CONTINUAL PROGRESS AGAINST ALL PREVIOUS COMPETITORS.

## How AlphaStar plays and observes the game

Professional StarCraft players such as TLO and MaNa are able to issue hundreds of actions per minute (APM) on average. This is far fewer than the majority of existing bots, which control each unit independently and consistently maintain thousands or even tens of thousands of APMs.

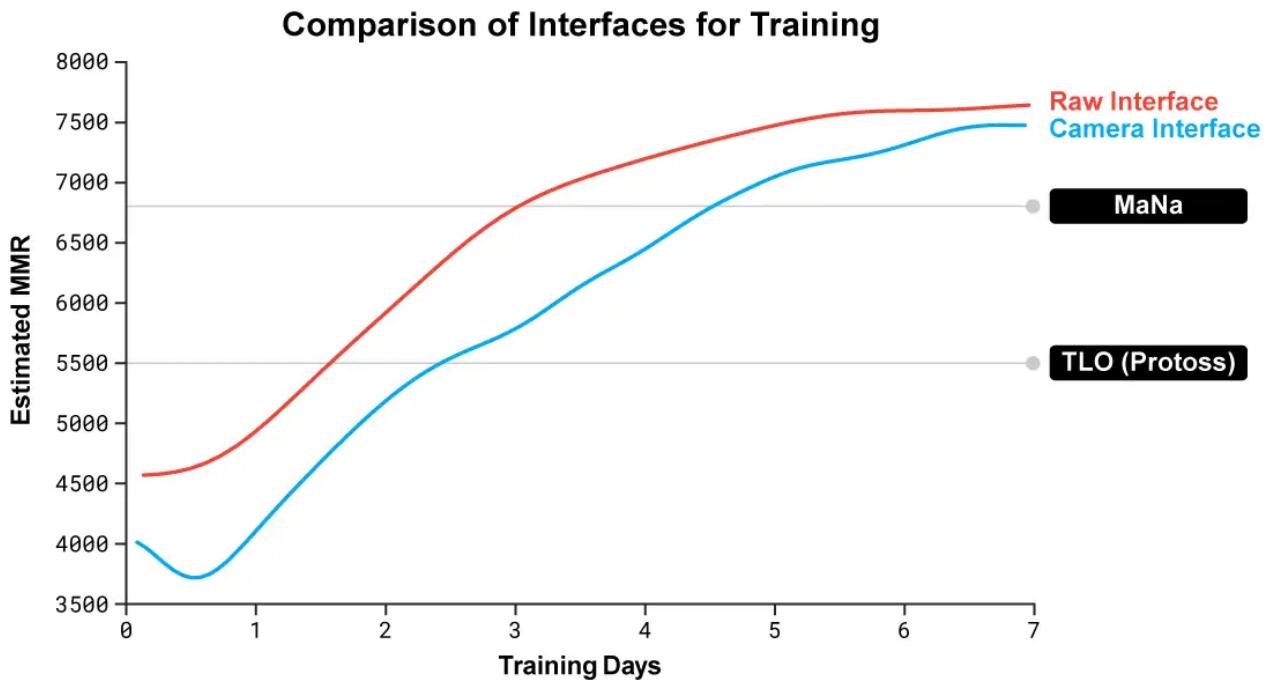
In its games against TLO and MaNa, AlphaStar had an average APM of around 280, significantly lower than the professional players, although its actions may be more precise. This lower APM is, in part, because AlphaStar starts its training using replays and thus mimics the way humans play the game. Additionally, AlphaStar reacts with a delay between observation and action of 350ms on average.



THE DISTRIBUTION OF ALPHASTAR'S APMS IN ITS MATCHES AGAINST MANA AND TLO AND THE TOTAL DELAY BETWEEN OBSERVATIONS AND ACTIONS. CLARIFICATION (29/01/19): TLO'S APM APPEARS HIGHER THAN BOTH ALPHASTAR AND MANA BECAUSE OF HIS USE OF RAPID-FIRE HOT-KEYS AND USE OF THE "REMOVE AND ADD TO CONTROL GROUP" KEY BINDINGS. ALSO NOTE THAT ALPHASTAR'S EFFECTIVE APM BURSTS ARE SOMETIMES HIGHER THAN BOTH PLAYERS.

During the matches against TLO and MaNa, AlphaStar interacted with the StarCraft game engine directly via its raw interface, meaning that it could observe the attributes of its own and its opponent's visible units on the map directly, without having to move the camera – effectively playing with a zoomed out view of the game. In contrast, human players must explicitly manage an "economy of attention" to decide where to focus the camera. However, analysis of AlphaStar's games suggests that it manages an implicit focus of attention. On average, agents "switched context" about 30 times per minute, similar to MaNa or TLO.

Additionally, and subsequent to the matches, we developed a second version of AlphaStar. Like human players, this version of AlphaStar chooses when and where to move the camera, its perception is restricted to on-screen information, and action locations are restricted to its viewable region.



PERFORMANCE OF ALPHASTAR USING THE RAW INTERFACE AND THE CAMERA INTERFACE, SHOWING THE NEWLY TRAINED CAMERA AGENT RAPIDLY CATCHING UP WITH AND ALMOST EQUALLING THE PERFORMANCE OF THE AGENT USING THE RAW INTERFACE.

We trained two new agents, one using the raw interface and one that must learn to control the camera, against the AlphaStar league. Each agent was initially trained by supervised learning from human data followed by the reinforcement learning procedure outlined above. The version of AlphaStar using the camera interface was almost as strong as the raw interface, exceeding 7000 MMR on our internal leaderboard. In an exhibition match, MaNa defeated a prototype version of AlphaStar using the camera interface, that was trained for just 7 days. We hope to evaluate a fully trained instance of the camera interface in the near future.

These results suggest that AlphaStar's success against MaNa and TLO was in fact due to superior macro and micro-strategic decision-making, rather than superior click-rate, faster reaction times, or the raw interface.

## Evaluating AlphaStar against professional players

The game of StarCraft allows players to select one of three alien races: Terran, Zerg or Protoss. We elected for AlphaStar to specialise in playing a single race for now – Protoss – to reduce training time and variance when reporting results from our internal league.

map. To evaluate AlphaStar's performance, we initially tested our agents against TLO: a top professional Zerg player and a GrandMaster level Protoss player. AlphaStar won the match 5-0, using a wide variety of units and build orders. "I was surprised by how strong the agent was," he said. "AlphaStar takes well-known strategies and turns them on their head. The agent demonstrated strategies I hadn't thought of before, which means there may still be new ways of playing the game that we haven't fully explored yet."

After training our agents for an additional week, we played against MaNa, one of the world's strongest StarCraft II players, and among the 10 strongest Protoss players. AlphaStar again won by 5 games to 0, demonstrating strong micro and macro-strategic skills. "I was impressed to see AlphaStar pull off advanced moves and different strategies across almost every game, using a very human style of gameplay I wouldn't have expected," he said. "I've realised how much my gameplay relies on forcing mistakes and being able to exploit human reactions, so this has put the game in a whole new light for me. We're all excited to see what comes next."

While StarCraft is just a game, albeit a complex one, we think that the techniques behind AlphaStar could be useful in solving other problems. For example, its neural network architecture is capable of modelling very long sequences of likely actions – with games often lasting up to an hour with tens of thousands of moves – based on imperfect information. Each frame of StarCraft is used as one step of input, with the neural network predicting the expected sequence of actions for the rest of the game after every frame. The fundamental problem of making complex predictions over very long sequences of data appears in many real world challenges, such as weather prediction, climate modelling, language understanding and more. We're very excited about the potential to make significant advances in these domains using learnings and developments from the AlphaStar project.

We also think some of our training methods may prove useful in the study of safe and robust AI. One of the great challenges in AI is the number of ways in which systems could go wrong, and StarCraft pros have previously found it easy to beat AI systems by finding inventive ways to provoke these mistakes. AlphaStar's innovative league-based training process finds the approaches that are most reliable and least likely to go wrong. We're excited by the potential for this kind of approach to help improve the safety and robustness of AI systems in general, particularly in safety-critical domains like energy, where it's essential to address complex edge cases.

Achieving the highest levels of StarCraft play represents a major breakthrough in one of the most complex video games ever created. We believe that these advances, alongside other recent progress in projects such as [AlphaZero](#) and [AlphaFold](#), represent a step forward in our mission to create intelligent systems that will one day help us unlock novel solutions to some of the world's most important and fundamental scientific problems.

***We are thankful for the support and immense skill of Team Liquid's TLO and MaNa.  
We are also grateful for the continued support of Blizzard and the StarCraft  
community for making this work possible.***

### **AlphaStar Team:**

*Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg,  
Wojtek Czarnecki, Andrew Dudzik, Aja Huang, Petko Georgiev, Richard Powell, Timo  
Ewalds, Dan Horgan, Manuel Kroiss, Ivo Danihelka, John Agapiou, Junhyuk Oh, Valentin*

Pohlen, Yuhuai Wu, Dani Yogatama, Julia Cohen, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Chris Apps, Koray Kavukcuoglu, Demis Hassabis, David Silver

### With thanks to:

Ali Razavi, Daniel Toyama, David Balduzzi, Doug Fritz, Eser Aygün, Florian Strub, Guillaume Alain, Haoran Tang, Jaume Sanchez, Jonathan Fildes, Julian Schrittwieser, Justin Novosad, Karen Simonyan, Karol Kurach, Philippe Hamel, Remi Leblond, Ricardo Barreira, Scott Reed, Sergey Bartunov, Shibli Mourad, Steve Gaffney, Thomas Hubert, the team that created PySC2 and the whole DeepMind Team, with special thanks to the research platform team, comms and events teams.



[DOWNLOAD ALL 11 REPLAYS OF THE GAMES AGAINST TLO AND MANA](#)



[WATCH THE EXHIBITION GAME AGAINST MANA](#)



[WATCH A VISUALISATION OF ALPHASTAR'S ENTIRE SECOND GAME AGAINST MANA](#)



[WATCH GAME HIGHLIGHTS OF ALPHASTAR VERSUS TLO AND MANA](#)



[WATCH GAME HIGHLIGHTS OF ALPHASTAR VERSUS TLO AND MANA](#)



[TO CITE THIS BLOG PLEASE USE THIS BIBTEX](#)

### Further reading





---

## AUTHORS

At [The AlphaStar team](#)

### About

Our story

Recent progress

Our global community

Leadership

Teams

Access to science

### Research

Blog

Open source

Publications

### Impact

Real-world impact

Significant breakthroughs

### Blog

Research

News

### Safety & Ethics

Overview

Technical safety

Ethics & Society

### Careers

A unique mission

Teams

Diversity matters

Locations

PRESS

TERMS & CONDITIONS

PRIVACY POLICY

MODERN SLAVERY STATEMENT

ALPHABET INC

