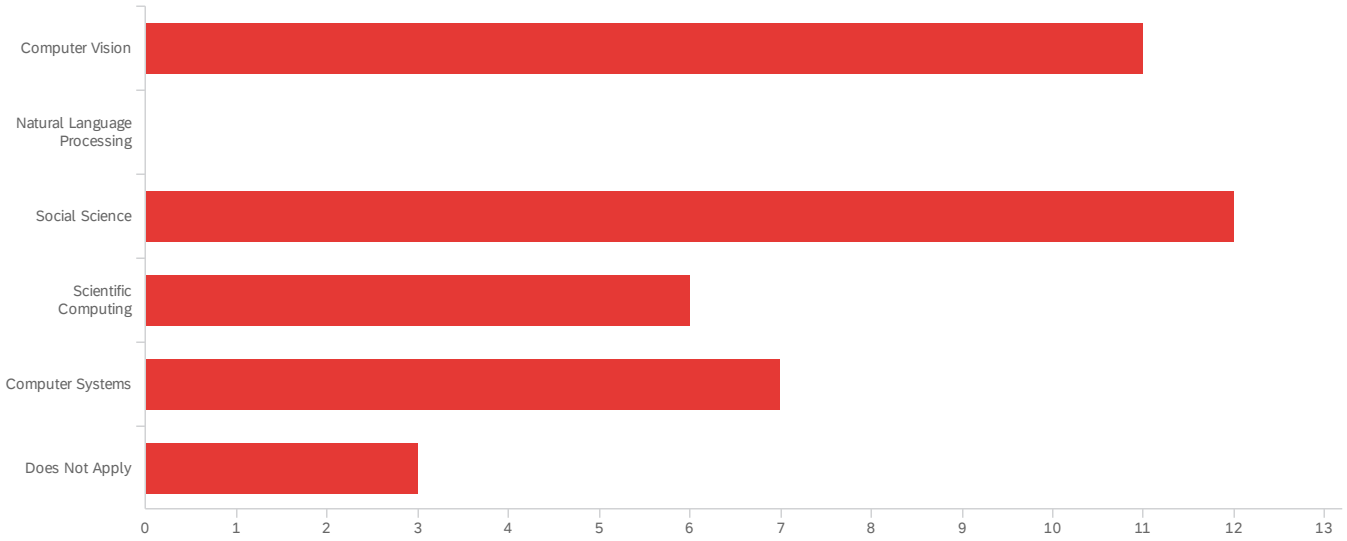
Default Report

Machine Learning Training Workload Survey

June 30, 2021 10:24 AM MDT

Q1 - Please select the most relevant one if none of them perfectly fit

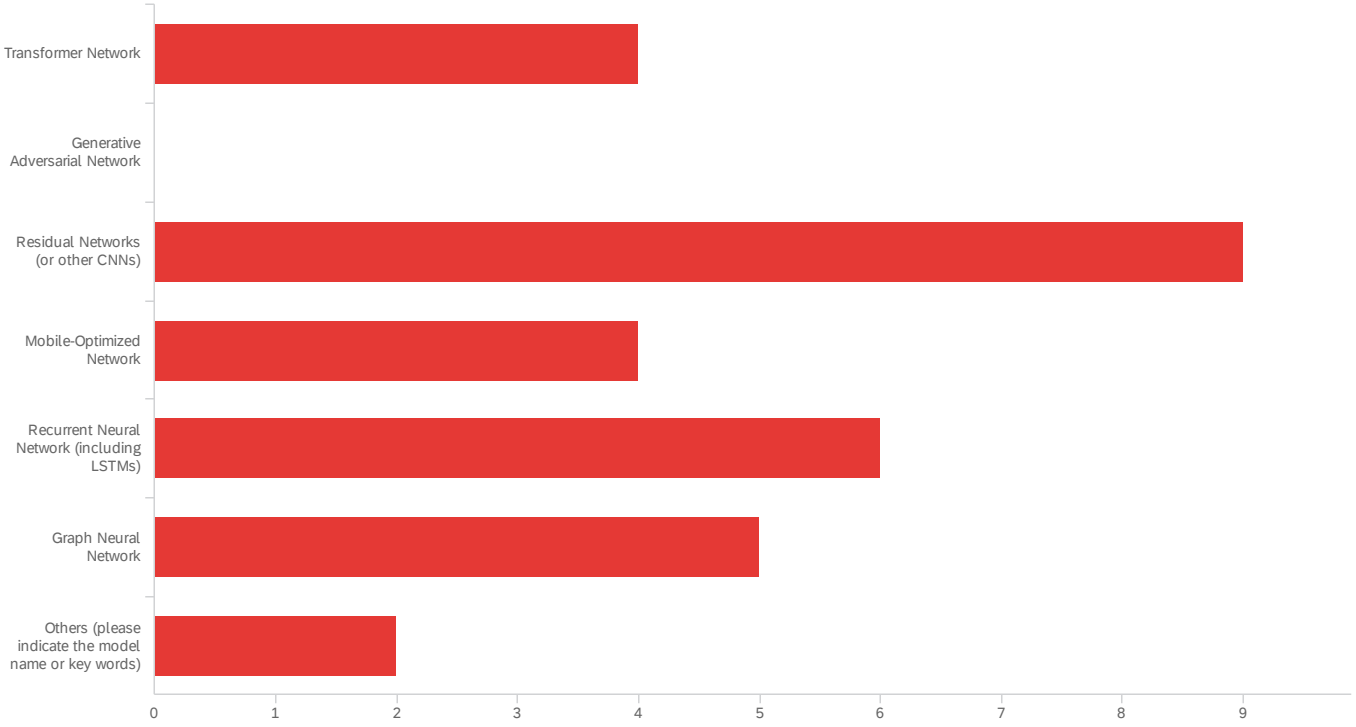


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Please select the most relevant one if none of them perfectly fit	1.00	6.00	3.18	1.63	2.66	39

#	Field		Choice Count
1	Computer Vision	28.21%	11
2	Natural Language Processing	0.00%	0
3	Social Science	30.77%	12
4	Scientific Computing	15.38%	6
5	Computer Systems	17.95%	7
6	Does Not Apply	7.69%	3
			39

Showing rows 1 - 7 of 7

Q6 - Which of the following best describes the last model you trained?



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Which of the following best describes the last model you trained? - Selected Choice	1.00	7.00	4.03	1.72	2.97	30

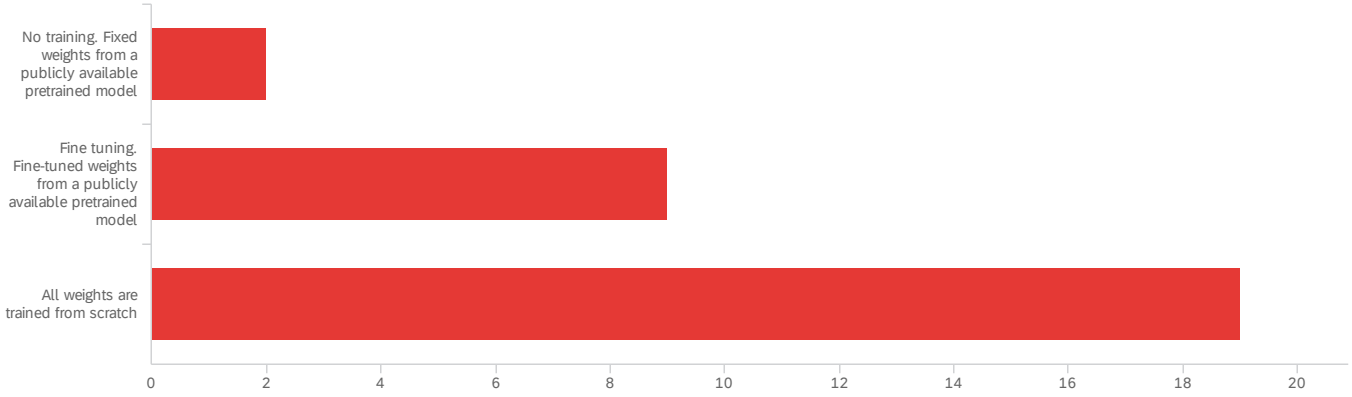
#	Field		Choice Count
1	Transformer Network	13.33%	4
2	Generative Adversarial Network	0.00%	0
3	Residual Networks (or other CNNs)	30.00%	9
4	Mobile-Optimized Network	13.33%	4
5	Recurrent Neural Network (including LSTMs)	20.00%	6
6	Graph Neural Network	16.67%	5
7	Others (please indicate the model name or key words)	6.67%	2
			30

Showing rows 1 - 8 of 8

Q6_7_TEXT - Others (please indicate the model name or key words)

Others (please indicate the model name or key words)
Bayesian hierarchical model (variational inference)

Q22 - Which of the following best describes how you trained your last model?

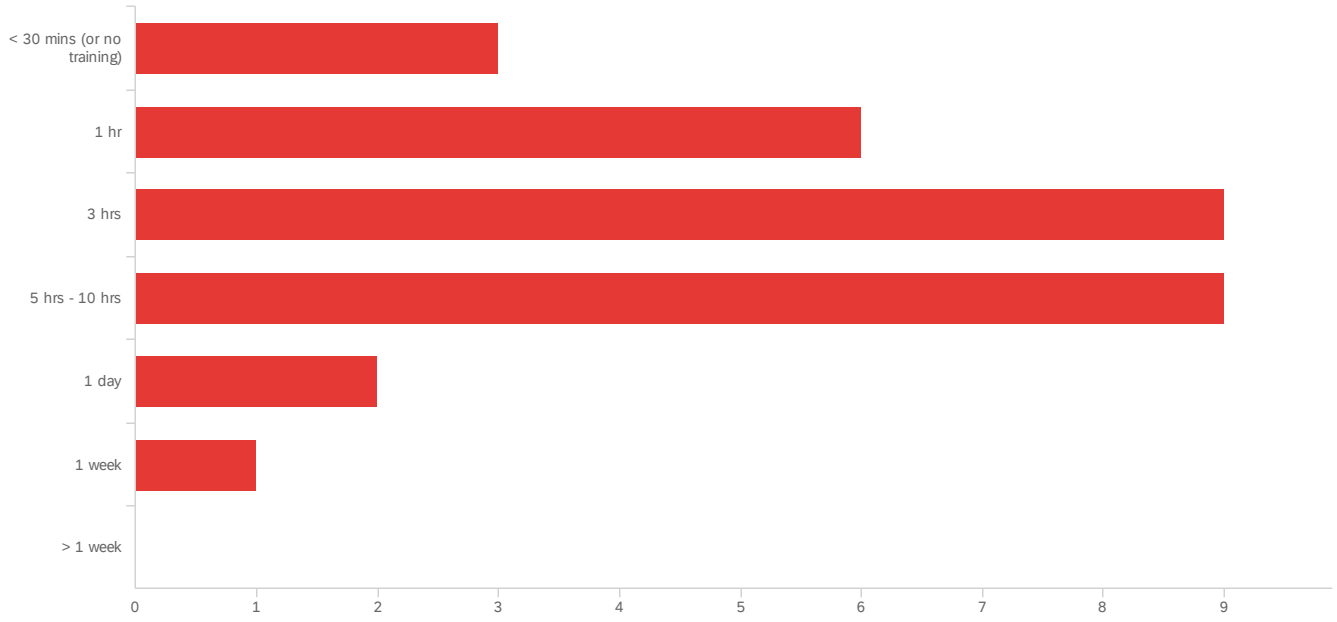


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Which of the following best describes how you trained your last model?	1.00	3.00	2.57	0.62	0.38	30

#	Field		Choice Count
1	No training. Fixed weights from a publicly available pretrained model	6.67%	2
2	Fine tuning. Fine-tuned weights from a publicly available pretrained model	30.00%	9
3	All weights are trained from scratch	63.33%	19
			30

Showing rows 1 - 4 of 4

Q17 - Which of the following best describes the amount of time needed to fully (i.e., further training will not improve the validation error) train your last model?

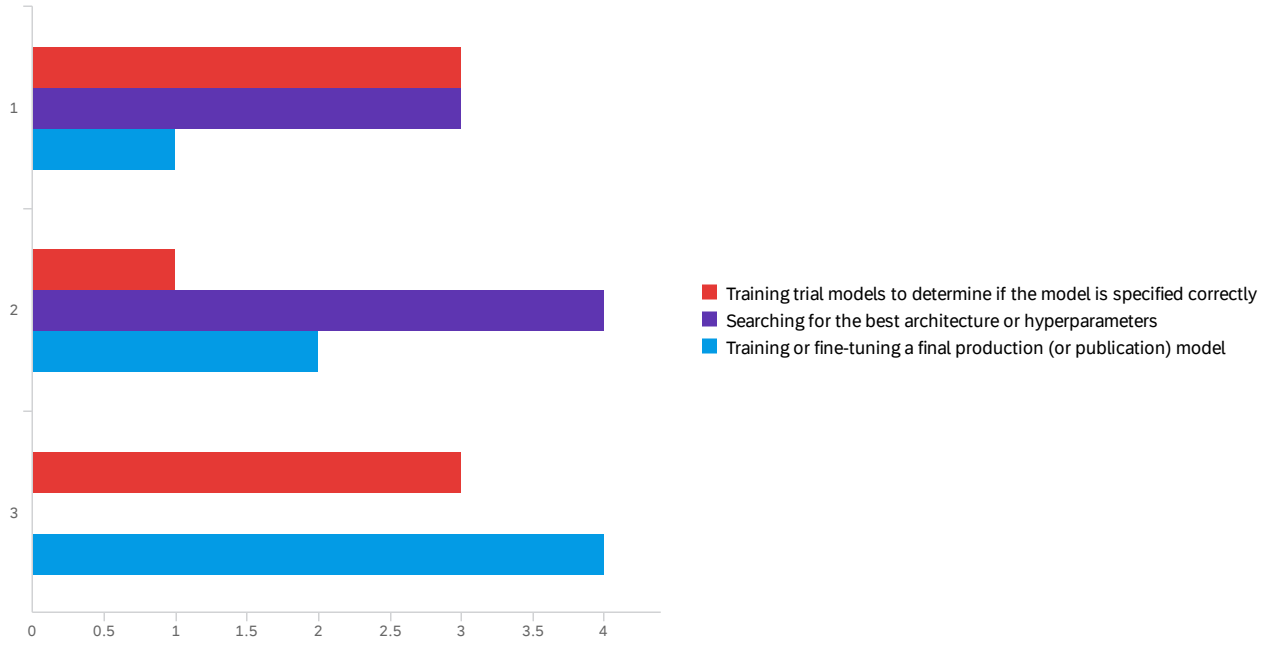


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Which of the following best describes the amount of time needed to fully (i.e., further training will not improve the validation error) train your last model?	1.00	6.00	3.13	1.20	1.45	30

#	Field		Choice Count
1	< 30 mins (or no training)	10.00%	3
2	1 hr	20.00%	6
3	3 hrs	30.00%	9
4	5 hrs - 10 hrs	30.00%	9
5	1 day	6.67%	2
6	1 week	3.33%	1
7	> 1 week	0.00%	0
			30

Showing rows 1 - 8 of 8

Q26 - Please rank the fraction of your time spent in each step of the development process

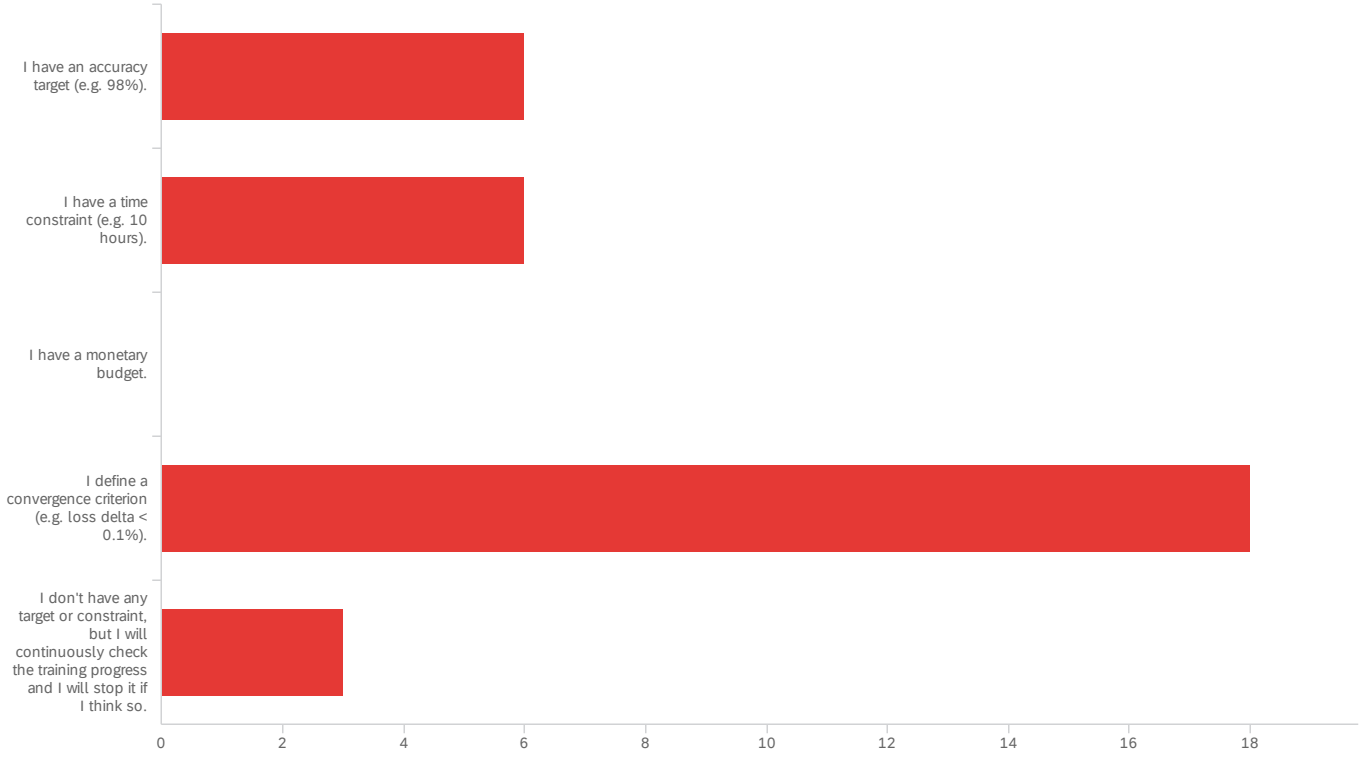


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Training trial models to determine if the model is specified correctly	1.00	3.00	2.00	0.93	0.86	7
2	Searching for the best architecture or hyperparameters	1.00	2.00	1.57	0.49	0.24	7
3	Training or fine-tuning a final production (or publication) model	1.00	3.00	2.43	0.73	0.53	7

#	Field	1		2		3		Total
1	Training trial models to determine if the model is specified correctly	42.86%	3	14.29%	1	42.86%	3	7
2	Searching for the best architecture or hyperparameters	42.86%	3	57.14%	4	0.00%	0	7
3	Training or fine-tuning a final production (or publication) model	14.29%	1	28.57%	2	57.14%	4	7

Showing rows 1 - 3 of 3

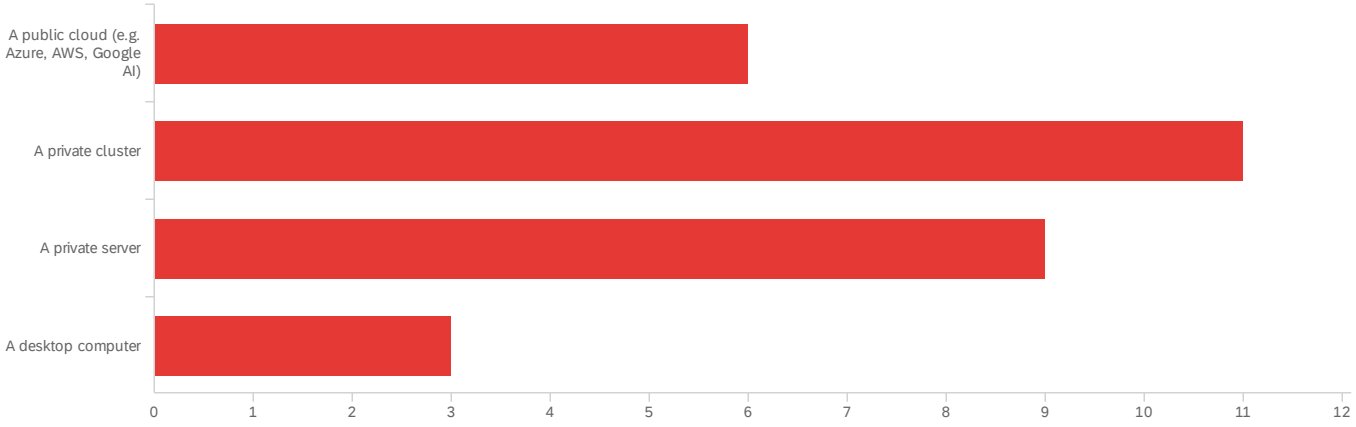
Q21 - How do you know when the training jobs should be stopped (Multiple Choice)



#	Field		Choice Count
1	I have an accuracy target (e.g. 98%).	18.18%	6
2	I have a time constraint (e.g. 10 hours).	18.18%	6
3	I have a monetary budget.	0.00%	0
4	I define a convergence criterion (e.g. loss delta < 0.1%).	54.55%	18
5	I don't have any target or constraint, but I will continuously check the training progress and I will stop it if I think so.	9.09%	3
			33

Showing rows 1 - 6 of 6

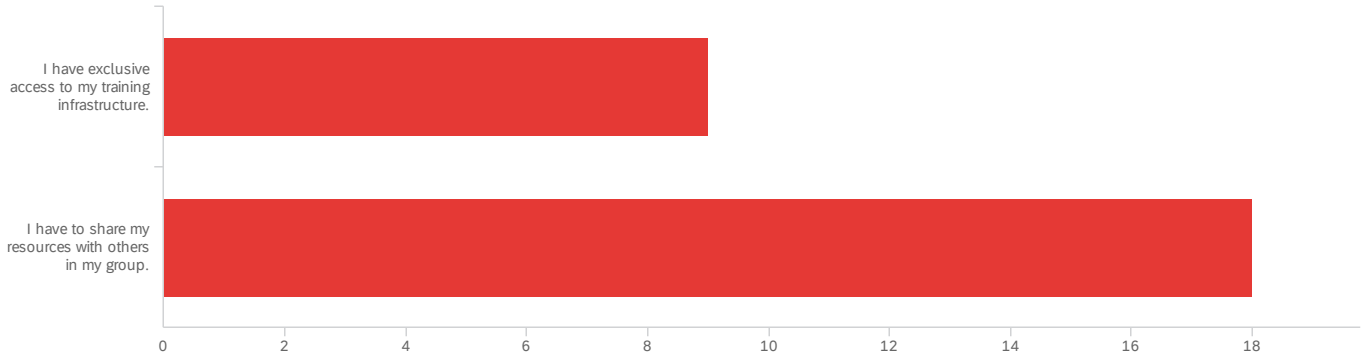
Q16 - Which of the following describes your infrastructure (Multiple Choice)?



#	Field		Choice Count
1	A public cloud (e.g. Azure, AWS, Google AI)	20.69%	6
2	A private cluster	37.93%	11
3	A private server	31.03%	9
4	A desktop computer	10.34%	3
			29

Showing rows 1 - 5 of 5

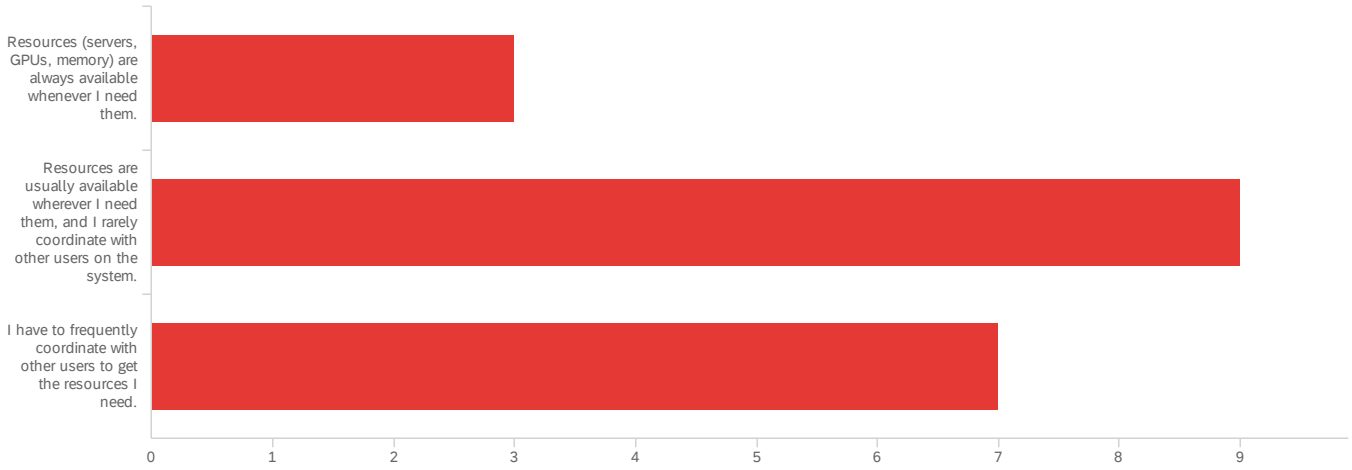
Q19 - Which of the following describes how you use your training infrastructure (Multiple Choice)?



#	Field		Choice Count
1	I have exclusive access to my training infrastructure.	33.33%	9
2	I have to share my resources with others in my group.	66.67%	18
			27

Showing rows 1 - 3 of 3

Q20 - If you do use shared infrastructure, select the response that describes your experiences (Multiple Choice):



#	Field		Choice Count
1	Resources (servers, GPUs, memory) are always available whenever I need them.	15.79%	3
2	Resources are usually available wherever I need them, and I rarely coordinate with other users on the system.	47.37%	9
3	I have to frequently coordinate with other users to get the resources I need.	36.84%	7
			19

Showing rows 1 - 4 of 4

Q12 - If your code of training jobs/models are open source, and you are comfortable with sharing them, would you like to allow us to access them via Github, Bitbucket, or other platform? (this is only for academic purpose)

If your code of training jobs/models are open source, and you are comfortab...
No
yes
yes
OpenPAI and Ray (ray.io)

End of Report