

ПРИЛОЖЕНИЕ №7.EVO.11.1.
 к Договору

**Тарифы на Evolution
 Evolution ML Inference**

№	Наименование работ, услуг	Единица тарификации	Период тарификации	Цена, без НДС, руб.	НДС* (20%), руб.	Цена с НДС*, руб.
1.	Инференс моделей машинного обучения Инференс ML модели на V100 NVLink 1Gb GPU 0,125 vCPU 2G RAM	шт.	час	6,25	1,25	7,50
2.	Инференс моделей машинного обучения Инференс ML модели на H100 NVLink 1Gb GPU 0,5 vCPU 2G RAM	шт.	час	8,75	1,75	10,50
3.	Инференс моделей машинного обучения Инференс ML модели на H100 PCIe 1Gb GPU 0,13 vCPU 1,5G RAM	шт.	час	5,625	1,125	6,75
4.	Инференс моделей машинного обучения Инференс ML модели на A100 PCIe 1Gb GPU 0,13 vCPU 1,5G RAM	шт.	час	3,25	0,65	3,90
5.	Инференс моделей машинного обучения Запросы 1 миллион	шт.	час	12,8	2,56	15,36
6.	Инференс моделей машинного обучения Инференс ML на H100 NVLink 1xH100 20 vCPU 190G RAM	шт.	час	700,00	140,00	840,00
7.	Инференс моделей машинного обучения Инференс ML на H100 NVLink 2xH100 40 vCPU 380G RAM	шт.	час	1400,00	280,00	1680,00
8.	Инференс моделей машинного обучения Инференс ML на H100 NVLink 4xH100 80 vCPU 760G RAM	шт.	час	2800,00	560,00	3360,00
9.	Инференс моделей машинного обучения Инференс ML на H100 NVLink 6xH100 120 vCPU 1140G RAM	шт.	час	4200,00	840,00	5040,00
10.	Инференс моделей машинного обучения Инференс ML на H100 NVLink 8xH100 160 vCPU 1520G RAM	шт.	час	5600,00	1120,00	6720,00
11.	Инференс моделей машинного обучения Инференс ML на V100 NVLink 1xV100 4 vCPU 64G RAM	шт.	час	200,00	40,00	240,00
12.	Инференс моделей машинного обучения Инференс ML на V100 NVLink 2xV100 8 vCPU 128G RAM	шт.	час	400,00	80,00	480,00
13.	Инференс моделей машинного обучения Инференс ML на V100 NVLink 4xV100 16 vCPU 256G RAM	шт.	час	800,00	160,00	960,00
14.	Инференс моделей машинного обучения Инференс ML на V100 NVLink 8xV100 32 vCPU 512G RAM	шт.	час	1600,00	320,00	1920,00
15.	Инференс моделей машинного обучения Инференс ML на V100 NVLink 16xV100 64 vCPU 1024G RAM	шт.	час	3200,00	640,00	3840,00
16.	Инференс моделей машинного обучения Инференс ML на A100 PCIe 1xA100 20 vCPU 125G RAM	шт.	час	260,00	52,00	312,00
17.	Инференс моделей машинного обучения Инференс ML на A100 PCIe 2xA100 40 vCPU 250G RAM	шт.	час	520,00	104,00	624,00
18.	Инференс моделей машинного обучения Инференс ML на A100 PCIe 4xA100 80 vCPU 500G RAM	шт.	час	1040,00	208,00	1248,00

19.	Инференс моделей машинного обучения Инференс ML на A100 PCIe 6xA100 120 vCPU 750G RAM	шт.	час	1560,00	312,00	1872,00
20.	Инференс моделей машинного обучения Инференс ML на A100 PCIe 8xA100 160 vCPU 1000G RAM	шт.	час	2080,00	416,00	2496,00
21.	Инференс моделей машинного обучения Инференс ML на H100 PCIe 1xH100 20 vCPU 125G RAM	шт.	час	450,00	90,00	540,00
22.	Инференс моделей машинного обучения Инференс ML на H100 PCIe 2xH100 40 vCPU 250G RAM	шт.	час	900,00	180,00	1080,00
23.	Инференс моделей машинного обучения Инференс ML на H100 PCIe 4xH100 80 vCPU 500G RAM	шт.	час	1800,00	360,00	2160,00
24.	Инференс моделей машинного обучения Инференс ML на H100 PCIe 6xH100 120 vCPU 700G RAM	шт.	час	2700,00	540,00	3240,00
25.	Инференс моделей машинного обучения Инференс ML на H100 PCIe 8xH100 160 vCPU 1000G RAM	шт.	час	3600,00	720,00	4320,00
26.	Инференс моделей машинного обучения Кэш моделей 1Гб	Гб	час	0,013	0,0026	0,0156

Примечание: «*» - если применимо.