

NEURAL MAGIC SUPPORT SERVICES POLICY

Updated: July 1, 2024

This document describes Neural Magic's technical support policies about Neural Magic Products ("NM Products"), DeepSparse and nm-vllm.

Free, open-source versions and cloud marketplace distributions of DeepSparse come with Community Support that can be accessed by:

- Opening an issue in the <u>Neural Magic GitHub</u> repository under the appropriate repository.
- Joining our <u>Neural Magic Community Slack</u>.

With the purchase of a Neural Magic Enterprise License or Neural Magic consulting agreement, Customers are entitled to Support Services as follows. Available technical support levels and Neural Magic's terms and conditions for support are also described below. Capitalized terms not defined herein have the meaning set forth in the agreement that applies to Customer's use of NM Products.

The following does not include Enterprise Support unless Customer enters into a separate support agreement with Neural Magic:

- Cloud Distribution of DeepSparse Enterprise offered through our cloud marketplaces
- vLLM as offered through GitHub and distributed through PyPI

1. Support Services.

- 1.1 Neural Magic offers one contracted Support Services plan as Enterprise Support for the NM Products and consulting services. During onboarding, the Neural Magic team will review Support Services to ensure customer success with purchased offerings.
- 1.2 Customer Support Channels: Neural Magic shall provide Support Services through
 - Private Slack channel
 - Self-ticketing
 - Audio/Video Conferencing: Support discussions will be scheduled in advance at a time
 mutually agreed by the parties and for durations and at a frequency that is commercially
 reasonable for Neural Magic. Support Services will be provided in English.
- 1.3 Business Days and Hours of Operation: Customer support is available 8:00 AM to 8:00 PM ET (GMT -5) Monday through Friday Americas region (excluding US Federal Holidays). Support Requests received outside of office hours will be collected, processed the following business day; however no action can be guaranteed until the next working day.
- 1.4 Support Request Prioritization & Neural Magic Actions: Support Requests will be categorized by priority level in accordance with the following definitions with the following actions:

Support Request Priority Definitions & Neural Magic Actions

M Neural Magic

| Priority Level | Definition | Neural Magic Actions |
|----------------|--|--|
| Critical | Critical means that, due to an Issue, NM Products used by Customer in production is severely impacted or completely shut down. | Neural Magic will: (i) assign engineers to work continuously to correct the Issue; (ii) provide ongoing communication on the status of the Update or Issue resolution; and (iii) simultaneously begin work to provide a temporary Workaround or Fix. |
| High | High means (i) due to an Issue, NM Products are functioning with limited capabilities, or is unstable with periodic interruptions, or (ii) there is an Issue in an application in development that is in final testing, facing a critical time frame of going into production use. | Neural Magic will: (i) assign engineers to correct the Issue; (ii) provide ongoing communication on the status of the Update or Issue resolution; and (iii) simultaneously begin work to provide a temporary Workaround or Fix. |
| Medium | Medium means there (i) are Issues with Workaround solutions in production, (ii) there are Issues in non-critical functions, or (iii) there is a time-sensitive Issue affecting performance or deliverables. | Neural Magic will use resources during Business Hours until the Issue is resolved or a Workaround is in place. |
| Low | Low means there (i) is a need to clarify procedures or information in Documentation, (ii) there is a request for a product enhancement or new feature, (iii) cosmetic or non-functional Issues; or (iv) issues in Documentation. | Neural Magic will triage the request, provide clarification when possible, and may include a resolution in a future Update. |

During the submission process, Customer may assign a priority level to a Support Request. Neural Magic will review Customer's priority designation and respond in accordance with the applicable Target Initial Response Time. However, Neural Magic may re-assign the priority level if it believes Customer's designation to be incorrect based on the definitions specified in this Support Policy. Neural Magic will notify Customer of such a change in its response to the Support Request.

Submitting Critical Support Requests

For issues that meet the criteria of a Critical, Customers may report the issues using an online ticket (preferred method) or Slack. 24 x 7 x 365 support is available only to Customers with a Neural Magic Enterprise License.

Always submit a new ticket when experiencing a Critical issue. This helps keep communication and troubleshooting efforts focused on the current issue which may not necessarily be related to a previous issue. The urgency of existing tickets can't be increased to Critical.

1.5 Responses. A "Response" is an initial reply to the Support Request. The "Target First Response

M Neural Magic

Times" shall be measured by the elapsed time between Neural Magic's receipt of a Support Request and the time when Neural Magic begins to address it, by responding and initiating communication with Customer about the Support Request. The actual time required to fully resolve an Issue or Support Request, if such full resolution occurs, may be longer than the Target First Response Time. Customer understands and agrees that resolution of an Issue or Request is not guaranteed and may not occur.

| Target First Response Times | | |
|-----------------------------|---------------------------|--|
| Priority Level | Enterprise Support | |
| Critical | Within 30 minutes, 24 x 7 | |
| High | Within 2 Business hours | |
| Medium | Within 2 Business days | |
| Low | Within 5 Business days | |

- 1.6 Customer Responsibilities: Neural Magic's obligation to provide Support Services is conditioned upon Customer satisfying the following responsibilities with respect to each Issue:
 - A. Customer making reasonable efforts to resolve the Issue before reporting the Issue to Neural Magic, including having the Issue reviewed by the representative of the Customer that submits the Support Request;
 - B. Customer providing Neural Magic with sufficient information, including any reproducible test cases requested by Neural Magic;
 - C. (For Critical and High Requests only) Customer designating personnel resources to provide necessary diagnostic information until a Fix or Workaround is made available.

2.0 Exclusions.

Notwithstanding anything to the contrary in this Support Services policy or the Agreement, Neural Magic is not obligated to continue work on a Support Request when Neural Magic determines that:

- A. The reported issue has been caused by Customer's negligence, hardware malfunction, network latency or causes beyond the reasonable control of Neural Magic;
- B. The reported issue has been caused by third-party software not managed by Neural Magic as part of NM Products, unless the Documentation requires the software for proper use of NM Products; For nm-vllm, Neural Magic will only offer support on the supported versions of nm-vllm tethered to the same vLLM version at https://docs.vllm.ai/en/latest/.
- C. The reported issue has been caused by Customer's use of NM Products other than in accordance with the configuration and operation guidelines described in the Documentation.
- D. Customer has not paid the fees for the Neural Magic Enterprise License when due.

3.0 <u>Definitions</u>.

- "Business Day" means Monday through Friday in the Americas Eastern time zone.
- "Business Hours" means 8:00 AM to 8:00 PM on Business Days.
- "Customer Representative" means the individual employee of Customer that submits a Support Request via support channels such as Slack or self-ticketing

M Neural Magic

- "Documentation" means the published documentation describing the functionality of NM Products, which may be modified from time to time. Documentation is located at https://docs.neuralmagic.com-
- "Issue" means a failure of NM Products to conform to the specifications set forth in the Documentation.
- "Support Request" means a support request or Issue submitted by Customer as described in this Support Services Policy.
- "Support Services" means the support services purchased by Customer and described in this Support Services Policy.
- 4.0 <u>Changes to Support Services Policy</u>. This Support Services Policy may be updated from time to time at Neural Magic's sole discretion, provided that any such updates will not materially reduce the level of Support Services during the period for which Customer has purchased Support Services.