



ISSN No. : 2321-9653

# IJRASET

**International Journal for Research in Applied  
Science & Engineering Technology**

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

## Certificate

*It is here by certified that the paper ID : IJRASET63985, entitled  
Open-AI model Efficient Memory Reduce Management for the Large Language Models (LLMs)  
Serving with Paged Attention of sharing the KV Cashes*

*by*

*Dr. K. Naveen Kumar*

*after review is found suitable and has been published in  
Volume 12, Issue VIII, August 2024*

*in*

*International Journal for Research in Applied Science &  
Engineering Technology*

*(International Peer Reviewed and Refereed Journal)*

*Good luck for your future endeavors*

*By [Signature]*

Editor in Chief, IJRASET



ISRA Journal Impact  
Factor: 7.429



45.98  
INDEX COPERNICUS



THOMSON REUTERS  
Researcher ID: N-9681-2016



TOGETHER WE REACH THE GOAL  
SJIF 7.429



ISSN No. : 2321-9653

# IJRASET

**International Journal for Research in Applied  
Science & Engineering Technology**

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

## Certificate

*It is here by certified that the paper ID : IJRASET63985, entitled  
Open-AI model Efficient Memory Reduce Management for the Large Language Models (LLMs)  
Serving with Paged Attention of sharing the KV Cashes*

*by*

*Mr. Sreedhar Ambala*

*after review is found suitable and has been published in  
Volume 12, Issue VIII, August 2024*

*in*

*International Journal for Research in Applied Science &  
Engineering Technology*

*(International Peer Reviewed and Refereed Journal)*

*Good luck for your future endeavors*

*By [Signature]*

Editor in Chief, IJRASET



ISRA Journal Impact  
Factor: 7.429



45.98  
INDEX COPERNICUS



THOMSON REUTERS  
Researcher ID: N-9681-2016



TOGETHER WE REACH THE GOAL  
SJIF 7.429



ISSN No. : 2321-9653

# IJRASET

**International Journal for Research in Applied  
Science & Engineering Technology**

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

## Certificate

*It is here by certified that the paper ID : IJRASET63985, entitled  
Open-AI model Efficient Memory Reduce Management for the Large Language Models (LLMs)  
Serving with Paged Attention of sharing the KV Cashes*

*by  
Dr. M. B Raju*

*after review is found suitable and has been published in  
Volume 12, Issue VIII, August 2024*

*in  
International Journal for Research in Applied Science &  
Engineering Technology*

*(International Peer Reviewed and Refereed Journal)*

*Good luck for your future endeavors*

*By [Signature]*

Editor in Chief, IJRASET



ISRA Journal Impact  
Factor: 7.429



45.98  
INDEX COPERNICUS



THOMSON REUTERS  
Researcher ID: N-9681-2016



TOGETHER WE REACH THE GOAL  
SJIF 7.429



ISSN No. : 2321-9653

# IJRASET

**International Journal for Research in Applied  
Science & Engineering Technology**

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : [www.ijraset.com](http://www.ijraset.com), E-mail : [ijraset@gmail.com](mailto:ijraset@gmail.com)

## Certificate

*It is here by certified that the paper ID : IJRASET63985, entitled  
Open-AI model Efficient Memory Reduce Management for the Large Language Models (LLMs)  
Serving with Paged Attention of sharing the KV Cashes*

*by*

*Dr. Sai Hareesh Anamandra*

*after review is found suitable and has been published in  
Volume 12, Issue VIII, August 2024*

*in*

*International Journal for Research in Applied Science &  
Engineering Technology*

*(International Peer Reviewed and Refereed Journal)*

*Good luck for your future endeavors*

*By [Signature]*

Editor in Chief, IJRASET



ISRA Journal Impact  
Factor: 7.429



45.98  
INDEX COPERNICUS



THOMSON REUTERS  
Researcher ID: N-9681-2016



TOGETHER WE REACH THE GOAL  
SJIF 7.429



ISSN No. : 2321-9653

# IJRASET

**International Journal for Research in Applied  
Science & Engineering Technology**

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

## Certificate

*It is here by certified that the paper ID : IJRASET63985, entitled  
Open-AI model Efficient Memory Reduce Management for the Large Language Models (LLMs)  
Serving with Paged Attention of sharing the KV Cashes*

*by*

*Mr. Dhanunjaya Rao Kodali*

*after review is found suitable and has been published in  
Volume 12, Issue VIII, August 2024*

*in*

*International Journal for Research in Applied Science &  
Engineering Technology*

*(International Peer Reviewed and Refereed Journal)*

*Good luck for your future endeavors*

*By [Signature]*

Editor in Chief, IJRASET



ISRA Journal Impact  
Factor: 7.429



45.98  
INDEX COPERNICUS



THOMSON REUTERS  
Researcher ID: N-9681-2016



TOGETHER WE REACH THE GOAL  
SJIF 7.429