Kasugai Municipal Hospital

On-premises AI infrastructure that securely protects medical information Developing generative AI services in-house to create an environment so staff can focus on medical care

Kasugai Municipal Hospital is pursuing business transformation through generative AI, aiming to create an environment where healthcare professionals can focus on providing medical care. To securely handle medical information, the hospital built an on-premises AI infrastructure by equipping Cisco UCS with NVIDIA GPUs. On this foundation, it operates generative AI services that it has developed in-house, promoting site-led hospital DX.

Kasugai Municipal Hospital

Location
1-1-1 Takaki-cho, Kasugai City

Opened
1951

Number of beds
558 beds (includingsixbeds for infectious diseases)

Kasugai City, with a population of approximately 310,000, is located in the northeasterrn part of Nagoya in Aichi Prefecture. At the heart of regional healthcare is Kasugai Municipal Hospital. As a core hospital in the region, it provides acute care, and in emergency medicine in particular, it upholds the motto of “never turning away emergency patients,” receiving around 10,000 ambulances each year.

Challenges

- Use generative Alto streamline administrative work that currently consumesmuch of healthcare professionals’ time
- An on-premisesAI infrastructure is needed to secure medical information
- Front line healthcare professionals want to take the lead in developing generativeAI services

Solutions

- Adopted Cisco UCS C240 M7 as the server and combined it with NVIDIA L40S GPUs to build a high-performance AI infrastructure
- Realized a smooth and stable infrastructure by leveraging Cisco’s pre-validation expertise
- Cisco supported not only the product deployment but also the resolution of technical challenges arising during the development of generative AI services

Results

- Developed services in-house that can generate summaries, minutes, and more, and made them available to all staff
- Achieved significant operational efficiency by automatically generating more than 1,000 discharge summaries per month for physicians and others using generative AI
- Accelerated site-led hospital DX as healthcare professionals take the lead in developing generative AI services

The Future

- Aim to realize AI agents that autonomously execute tasks
- Further expand AI utilization to achieve both greater efficiency and labor savings and higher quality of care

We want to use the power of generative AI to create an environment where we can focus on medicine.

Dr. Tomohiko Naruse

Kasugai Municipal Hospital
Director

Challenges

Aiming to reduce the burden on frontline staff and create an environment where they can focus on medical care

Kasugai Municipal Hospital actively adopts advanced technologies, including surgical support robots for urology and orthopedics. Generative AI is another technology the hospital is focusing on.

"To improve the quality of medical care, it is essential to create an environment where physicians and paramedical staff can concentrate on medicine itself. However, in reality, the frontline spends a great deal of time on administrative tasks such as recording information and creating documents. By streamlining these tasks with generative AI, we expect that healthcare professionals will be able to devote even more time to medical care," says Hospital Director Dr. Tomohiko Naruse.

Leading the use of generative AI is the Management Strategy Office, which also functions as the DX division. A key characteristic of the information team responsible for actual digital utilization is that its members are healthcare professionals who deeply understand the frontline. "Nurses, pharmacists, radiological technologists, and physical therapists are taking the lead in digital utilization. Our strength lies in the ability of frontline staff to identify issues themselves and solve them using digital tools," explains Hayato Baba, who is a certified radiological technologist.

To fully harness the power of generative AI, the hospital began by building a data infrastructure. "We integrated more than 50 internal systems so that not only medical information but also HR and management data are consolidated. Staff no longer need to search through multiple systems; aggregated data can be referenced across systems or combined for analysis," explains Masakazu Kogiso, a licensed nurse.

The hospital also began building an AI infrastructure to apply generative AI to operations. Since generative AI excels at natural language processing, the hospital believed it could streamline administrative tasks such as document summarization and minutes authoring. The most critical concern during this process was security. "Medical information is composed entirely of personal data and cannot simply be placed in the cloud. For that reason, we needed to build an on-premises AI infrastructure capable of safely handling data," says Hiroki Mizukusa, a licensed pharmacist.

Valuing a support structure that helps immediately when needed

Solutions

Assessing the validated architecture and flexible support structure

To build an on-premises AI infrastructure, the hospital chose Cisco's AI infrastructure. Specifically, it deployed Cisco's server product, Cisco Unified Computing System (UCS) C240 M7, equipped with NVIDIA L40S GPUs. With high computational performance and scalability, the environment supports learning and inference using large language models (LLMs) and various other AI processes in a stable manner.

Cisco is currently placing strong emphasis on providing AI infrastructure and continuously performs pre-validation to establish optimal configurations for servers, networks, storage, and GPUs. In partnership with key vendors, Cisco executes various AI workloads, such as training, inference, and analytics, in real environments. By validating and optimizing them, Cisco can deliver AI infrastructure quickly and enable stable performance from the moment of deployment. Cisco and NVIDIA, who supplies the indispensable GPUs for AI infrastructure, have also formed a strategic partnership to jointly advance the sophistication of AI platforms.

In the hospital's on-premises AI infrastructure project, Cisco applied all of the expertise developed through these initiatives. While addressing the hospital's specific requirements, Cisco achieved the rapid deployment of a high-performance, stable AI infrastructure.

In addition to technical capabilities and proven results, the hospital also highly values Cisco's support system.

"We have used Cisco products for many years, starting with network equipment. Not only do we recognize the high reliability of their products, but we have always appreciated their deep understanding of medical environments and their supportive, collaborative approach. For our AI infrastructure project, we believed Cisco would not only provide the products but also walk alongside us and offer thorough support throughout the process. That's why we chose Cisco," says Baba.

Results and the Future

Major achievements in tasks such as authoring discharge summaries

The AI infrastructure combining Cisco UCS with NVIDIA GPUs is already in operation. The hospital selected Ubuntu as the server OS, achieving a flexible environment while managing operating costs. The hospital uses multiple open-source LLMs, selecting optimal models according to use cases—LLMs adept at long-form processing for minutes and physician summaries, and lightweight, fast-responding LLMs for chats and tasks requiring immediacy.

Using this infrastructure, the hospital has been rapidly rolling out its own generative AI services. These can be easily launched via menu in the electronic medical record system, enabling all staff to use them freely in their daily operations.

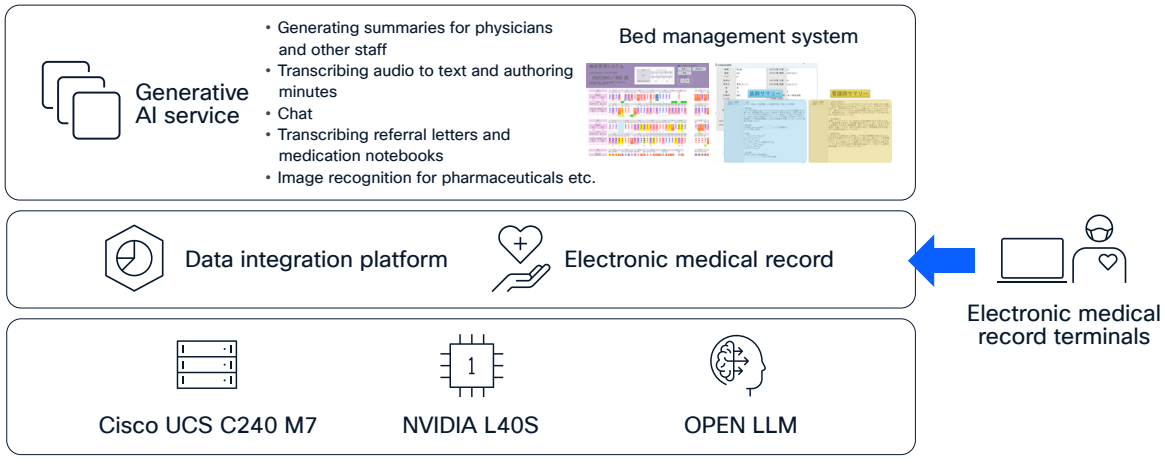
The main services currently in operation include:
- Generating summaries for physicians and other staff
- Transcribing audio to text and authoring minutes
- Chat
- Transcribing referral letters and medication notebooks
- Image recognition for pharmaceuticals

These services have significantly improved efficiency across various tasks. One area where the effect has been especially notable is the creation of discharge summaries, which summarize treatment details and progress during hospitalization. It is completed upon discharge and serves an important role in handing over information to the next medical institution during transfers or referrals. “Physicians write the treatment progress, nurses write the nursing records, and physical therapists write rehabilitation details. More than 1,000 discharge summaries are created each month, and until now, staff have written them all manually. By automatically generating a first draft summary using generative AI, allowing staff to review and edit it, the time required for summary creation has been greatly reduced,” says Baba.

The summary creation service is also linked with the hospital’s in-house bed management system, which visualizes the status of all inpatients. For each patient, generative AI summarizes the past week’s information, which is used in admission/discharge management and conference preparation.

Generative AI is also being used for transcription of scanned referral letters and for authoring meeting minutes, and the hospital continues developing new services. For example, in the pharmacy domain, a system is being planned to evaluate whether newly adopted medications resemble existing ones in appearance, since similarities in shape or color can increase the risk of mix-ups during dispensing or administration.

Concept for the generative AI environment at Kasugai Municipal Hospital



Although members of the Management Strategy Office design and develop these generative AI services, Cisco's extensive support has been a major source of strength throughout development. "Because we are not IT specialists, we encountered many challenges during the build process. Each time, Cisco's engineers responded quickly and provided support," says Mizukusa.

Taking on the challenge of realizing AI agents

The hospital is now looking toward AI agents as the next stage of generative AI utilization. Whereas the current focus is on automating single tasks such as summarization or transcription, the next step is to link multiple operational processes so that AI can autonomously execute tasks. "For example, upon receiving an inquiry, AI would automatically search the data, create graphs, and generate a report. Instead of merely helping humans work more efficiently, the goal is to enable AI to take over tasks, enabling healthcare professionals to focus even more deeply on medical care. We want to proactively advance efforts while determining which tasks should be handled by humans and which are suited to AI," says Kogiso.

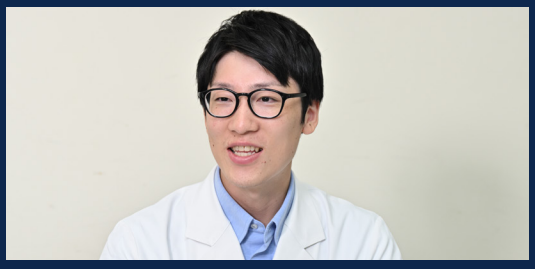
Strengthening operations is another major theme. Cisco has proposed Cisco Intersight, which enables visualization of server and AI workload status via the cloud and supports efficient remote management.

"There are many approaches to improving the quality of healthcare, and generative AI utilization is one of the initiatives in which we place great expectations. Beyond operational efficiency, we foresee applications in disaster response and many other areas. We will continue actively advancing these efforts," concludes Dr. Naruse.

Kasugai Municipal Hospital
Director
Dr. Tomohiko Naruse

Kasugai Municipal Hospital
Management Strategy Office and Administrative Services Division
Information Team, Radiological Technologist / Healthcare Information Technologist / Certified Medical Information Manager
Hayato Baba

Kasugai Municipal Hospital
Management Strategy Office
Nurse
Masakazu Kogiso



Kasugai Municipal Hospital
Management Strategy Office and Pharmacy Department,
Pharmacist / Certified Medical Information Manager
Hiroki Mizukusa

春日井市民病院
Kasugai Municipal Hospital

Kasugai Municipal Hospital defines its mission as responding sincerely and continuously to the community's healthcare needs. In addition to operating a 24-bed intensive care unit and a dedicated stroke ward,it has six emergency outpatient beds as an Emergency and Critical Care Center, accepting around 10,000 emergency transports annually. It also serves as a coreregional hospital handlinga wide range of conditions, with designations including Aichi Prefectural Designated Cancer CareHospitals, Regional Medical Care Support Hospitals, Core Regional DisasterKey Hospitals, and DMAT-designated medical institutions.
URL https://www.hospital.kasugai.aichi.jp/

Products and services

- Cisco UCS C240 M7
- NVIDIA L40S
- Cisco Intersight

2510-0-xx

Kasugai City, with a population of approximately 310,000, is located in the northeasterrn part of Nagoya in Aichi Prefecture. At the heart of regional healthcare is Kasugai Municipal Hospital. As a core hospital in the region, it provides acute care, and in emergency medicine in particular, it upholds the motto of “never turning away emergency patients,” receiving around 10,000 ambulances each year.

Challenges

- Use generative AIto streamline administrative work that currently consumesmuch of healthcare professionals' time
- An on-premisesAI infrastructure is needed to secure medical information
- Front line healthcare professionals want to take the lead in developing generativeAI services

Solutions

- Adopted Cisco UCS C240 M7 as the server and combined it with NVIDIA L40S GPUs to build a high-performance AI infrastructure
- Realized a smooth and stable infrastructure by leveraging Cisco's pre-validation expertise
- Cisco supported not only the product deployment but also the resolution of technical challenges arising during the development of generative AI services

Results

- Developed services in-house that can generate summaries, minutes, and more, and made them available to all staff
- Achieved significant operational efficiency by automatically generating more than 1,000 discharge summaries per month for physicians and others using generative AI
- Accelerated site-led hospital DX as healthcare professionals take the lead in developing generative AI services

The Future

- Aim to realize AI agents that autonomously execute tasks
- Further expand AI utilization to achieve both greater efficiency and labor savings and higher quality of care

We want to use the power of generative AI to create an environment where we can focus on medicine.

Dr. Tomohiko Naruse
Kasugai Municipal Hospital
Director

Case study
Cisco public

春日井市民病院
Kasugai Municipal Hospital

Kasugai Municipal Hospital defines its mission as responding sincerely and continuously to the community's healthcare needs. In addition to operating a 24-bed intensive care unit and a dedicated stroke ward,it has six emergency outpatient beds as an Emergency and Critical Care Center, accepting around 10,000 emergency transports annually. It also serves as a coreregional hospital handlinga wide range of conditions, with designations including Aichi Prefectural Designated Cancer CareHospitals, Regional Medical Care Support Hospitals, Core Regional DisasterKey Hospitals, and DMAT-designated medical institutions.
URL https://www.hospital.kasugai.aichi.jp/

2510-0-xx

Products and services

- Cisco UCS C240 M7
- NVIDIA L40S
- Cisco Intersight

Case study
Cisco public

Kasugai Municipal Hospital

On-premises AI infrastructure that securely protects medical information Developing generative AI services in-house to create an environment so staff can focus on medical care

Kasugai Municipal Hospital is pursuing business transformation through generative AI, aiming to create an environment where healthcare professionals can focus on providing medical care. To securely handle medical information, the hospital built an on-premises AI infrastructure by equipping Cisco UCS with NVIDIA GPUs. On this foundation, it operates generative AI services that it has developed in-house, promoting site-led hospital DX.

Kasugai Municipal Hospital

Location
1-1-1 Takaki-cho, Kasugai City

Opened
1951

Number of beds
558 beds (includingsixbeds for infectious diseases)

Challenges

Aiming to reduce the burden on frontline staff and create an environment where they can focus on medical care

Kasugai Municipal Hospital actively adopts advanced technologies, including surgical support robots for urology and orthopedics. Generative AI is another technology the hospital is focusing on.

“To improve the quality of medical care, it is essential to create an environment where physicians and paramedical staff can concentrate on medicine itself. However, in reality, the frontline spends a great deal of time on administrative tasks such as recording information and creating documents. By streamlining these tasks with generative AI, we expect that healthcare professionals will be able to devote even more time to medical care,” says Hospital Director Dr. Tomohiko Naruse.

Leading the use of generative AI is the Management Strategy Office, which also functions as the DX division. A key characteristic of the information team responsible for actual digital utilization is that its members are healthcare professionals who deeply understand the frontline. “Nurses, pharmacists, radiological technologists, and physical therapists are taking the lead in digital utilization. Our strength lies in the ability of frontline staff to identify issues themselves and solve them using digital tools,” explains Hayato Baba, who is a certified radiological technologist.

To fully harness the power of generative AI, the hospital began by building a data infrastructure. “We integrated more than 50 internal systems so that not only medical information but also HR and management data are consolidated. Staff no longer need to search through multiple systems; aggregated data can be referenced across systems or combined for analysis,” explains Masakazu Kogiso, a licensed nurse.

The hospital also began building an AI infrastructure to apply generative AI to operations. Since generative AI excels at natural language processing, the hospital believed it could streamline administrative tasks such as document summarization and minutes authoring. The most critical concern during this process was security. “Medical information is composed entirely of personal data and cannot simply be placed in the cloud. For that reason, we needed to build an on-premises AI infrastructure capable of safely handling data,” says Hiroki Mizukusa, a licensed pharmacist.

Valuing a support structure that helps immediately when needed

Solutions

Assessing the validated architecture and flexible support structure

To build an on-premises AI infrastructure, the hospital chose Cisco’s AI infrastructure. Specifically, it deployed Cisco’s server product, Cisco Unified Computing System (UCS) C240 M7, equipped with NVIDIA L40S GPUs. With high computational performance and scalability, the environment supports learning and inference using large language models (LLMs) and various other AI processes in a stable manner.

Cisco is currently placing strong emphasis on providing AI infrastructure and continuously performs pre-validation to establish optimal configurations for servers, networks, storage, and GPUs. In partnership with key vendors, Cisco executes various AI workloads, such as training, inference, and analytics, in real environments. By validating and optimizing them, Cisco can deliver AI infrastructure quickly and enable stable performance from the moment of deployment. Cisco and NVIDIA, who supplies the indispensable GPUs for AI infrastructure, have also formed a strategic partnership to jointly advance the sophistication of AI platforms.

In the hospital’s on-premises AI infrastructure project, Cisco applied all of the expertise developed through these initiatives. While addressing the hospital’s specific requirements, Cisco achieved the rapid deployment of a high-performance, stable AI infrastructure.

In addition to technical capabilities and proven results, the hospital also highly values Cisco’s support system.

“We have used Cisco products for many years, starting with network equipment. Not only do we recognize the high reliability of their products, but we have always appreciated their deep understanding of medical environments and their supportive, collaborative approach. For our AI infrastructure project, we believed Cisco would not only provide the products but also walk alongside us and offer thorough support throughout the process. That’s why we chose Cisco,” says Baba.

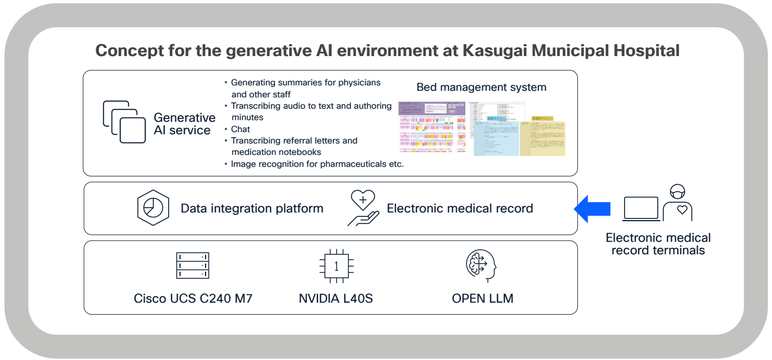
Results and the Future

Major achievements in tasks such as authoring discharge summaries

The AI infrastructure combining Cisco UCS with NVIDIA GPUs is already in operation. The hospital selected Ubuntu as the server OS, achieving a flexible environment while managing operating costs. The hospital uses multiple open-source LLMs, selecting optimal models according to use cases—LLMs adept at long-form processing for minutes and physician summaries, and lightweight, fast-responding LLMs for chats and tasks requiring immediacy.

Using this infrastructure, the hospital has been rapidly rolling out its own generative AI services. These can be easily launched via menu in the electronic medical record system, enabling all staff to use them freely in their daily operations.

The main services currently in operation include:
- Generating summaries for physicians and other staff
- Transcribing audio to text and authoring minutes
- Chat
- Transcribing referral letters and medication notebooks
- Image recognition for pharmaceuticals



These services have significantly improved efficiency across various tasks. One area where the effect has been especially notable is the creation of discharge summaries, which summarize treatment details and progress during hospitalization. It is completed upon discharge and serves an important role in handing over information to the next medical institution during transfers or referrals. “Physicians write the treatment progress, nurses write the nursing records, and physical therapists write rehabilitation details. More than 1,000 discharge summaries are created each month, and until now, staff have written them all manually. By automatically generating a first draft summary using generative AI, allowing staff to review and edit it, the time required for summary creation has been greatly reduced,” says Baba.

The summary creation service is also linked with the hospital’s in-house bed management system, which visualizes the status of all inpatients. For each patient, generative AI summarizes the past week’s information, which is used in admission/discharge management and conference preparation.

Generative AI is also being used for transcription of scanned referral letters and for authoring meeting minutes, and the hospital continues developing new services. For example, in the pharmacy domain, a system is being planned to evaluate whether newly adopted medications resemble existing ones in appearance, since similarities in shape or color can increase the risk of mix-ups during dispensing or administration.

Although members of the Management Strategy Office design and develop these generative AI services, Cisco’s extensive support has been a major source of strength throughout development. “Because we are not IT specialists, we encountered many challenges during the build process. Each time, Cisco’s engineers responded quickly and provided support,” says Mizukusa.

Taking on the challenge of realizing AI agents

The hospital is now looking toward AI agents as the next stage of generative AI utilization. Whereas the current focus is on automating single tasks such as summarization or transcription, the next step is to link multiple operational processes so that AI can autonomously execute tasks. “For example, upon receiving an inquiry, AI would automatically search the data, create graphs, and generate a report. Instead of merely helping humans work more efficiently, the goal is to enable AI to take over tasks, enabling healthcare professionals to focus even more deeply on medical care. We want to proactively advance efforts while determining which tasks should be handled by humans and which are suited to AI,” says Kogiso.

Strengthening operations is another major theme. Cisco has proposed Cisco Intersight, which enables visualization of server and AI workload status via the cloud and supports efficient remote management.

“There are many approaches to improving the quality of healthcare, and generative AI utilization is one of the initiatives in which we place great expectations. Beyond operational efficiency, we foresee applications in disaster response and many other areas. We will continue actively advancing these efforts,” concludes Dr. Naruse.

Kasugai Municipal Hospital
Director
Dr. Tomohiko Naruse

Kasugai Municipal Hospital
Management Strategy Office and Administrative Services Division
Information Team, Radiological Technologist / Healthcare Information Technologist / Certified Medical Information Manager
Hayato Baba

Kasugai Municipal Hospital
Management Strategy Office
Nurse
Masakazu Kogiso

Kasugai Municipal Hospital
Management Strategy Office and Pharmacy Department,
Pharmacist / Certified Medical Information Manager
Hiroki Mizukusa