使用 NetApp AI 进行情感分析

NetApp Solutions

NetApp
September 10, 2024

This PDF was generated from https://docs.netapp.com/zh-cn/netapp-solutions/ai/ai-sent-support-center-analytics.html on September 10, 2024. Always check docs.netapp.com for the latest.

目录

使用 NetApp AI 进行情感分析 1
TR-4910：《利用 NetApp AI 进行客户沟通时的情感分析》 1
用例 1
架构 3
设计注意事项 9
部署支持中心观点分析 10
验证结果 12
视频和演示 13
结论 15
从何处查找追加信息 16

使用 NetApp AI 进行情感分析

TR-4910：《利用 NetApp AI 进行客户沟通时的情感分析》

SML 科学公司 Rick Huang， Sathish Thyagarajan 和 NetApp 公司迭戈 Sosa-Coba 的 David Arnette

本技术报告为客户在企业级全球支持中心内使用 NetApp 数据管理技术和 NVIDIA 软件框架以及传输学习和对话 AI 执行情感分析提供了设计指导。此解决方案适用于希望从代表聊天日志，电子邮件和其他文本或音频通信的录制语音或文本文件中获得客户见解的任何行业。我们实施了一个端到端管道，用于在采用 NetApp 云连接全闪存存储的 GPU 加速计算集群上展示自动语音识别，实时情感分析和深度学习的自然语言处理模式 - 再培训功能。可以对大规模的一流语言模型进行培训和优化，以便与全球支持中心快速进行推理，从而打造卓越的客户体验，并对员工的长期业绩进行客观评估。

情绪分析是自然语言处理（ NLP ）中的一个研究领域，通过该领域，可以从文本中提取积极，负面或中立的情绪。随着越来越多的人与人工智能系统互动，对话式人工智能系统的集成程度已提升到接近全球水平。情感分析有多种使用情形，从确定支持中心员工在与来电者对话中的表现，提供适当的自动聊天机器人响应，到根据公司代表与受众在季度收益通话中的互动预测公司的股票价格。此外，情感分析可用于确定客户对品牌提供的产品，服务或支持的看法。

此端到端解决方案使用 NLP 模型执行高级别的情感分析，从而支持支持中心分析框架。录音会处理成书面文本，并从对话的每一句话中提取情感。可以将结果汇总到信息板中，以便对对话感受进行分析，包括历史和实时分析。此解决方案可以概括为具有类似数据模式和输出需求的其他解决方案。利用适当的数据，可以完成其他使用情形。例如，可以使用相同的端到端管道分析公司收益调用的情绪。由于管道的灵活性，还可以进行其他形式的 NLP 分析，例如主题建模和命名实体识别（ NER ）。

NVIDIA Riva， NVIDIA TAO 工具包和 NetApp DataOps 工具包共同实现了这些 AI 实施。NVIDIA 的工具可用于使用预构建的型号和管道快速部署高性能 AI 解决方案。NetApp DataOps 工具包可简化各种数据管理任务，以加快开发速度。

客户价值

企业可以通过员工评估和客户响应工具在文本，音频和视频对话中获得价值，以便进行情绪分析。经理可以从信息板中提供的信息中受益，从而可以根据对话的两个方面评估员工和客户满意度。

此外， NetApp DataOps 工具包还可管理客户基础架构中数据的版本控制和分配。这样就会频繁更新信息板中提供的分析，而不会产生庞大的数据存储成本。

用例

由于这些支持中心处理的呼叫数量众多，如果手动执行，则可能需要很长时间才能评估呼叫性能。传统方法，例如字包计数和其他方法，可以实现一定程度的自动化，但这些方法不能捕获动态语言的更细微的方面和语义上下文。AI 建模技术可用于自动执行其中一些更精细的分析。此外，随着 NVIDIA， AWS， Google 等公司发布的最新的经过预先培训的建模工具，现在可以相对轻松地建立一个包含复杂模型的端到端管道并进行自定义。

支持中心情感分析的端到端管道，可在员工与来电者交谈时实时载入音频文件。然后，这些音频文件将进行处理，以便在将其转换为文本格式的语音到文本组件中使用。对话中的每一句话都会收到一个标签，用于指示情绪（

积极，消极或中立）。

情绪分析可以为对话提供一个重要方面，用于评估通话性能。这些情感为员工和来电者之间的互动增加了一个深度。人工智能辅助的情绪信息板可为经理实时跟踪对话中的情绪，并对员工过去的通话进行回顾性分析。

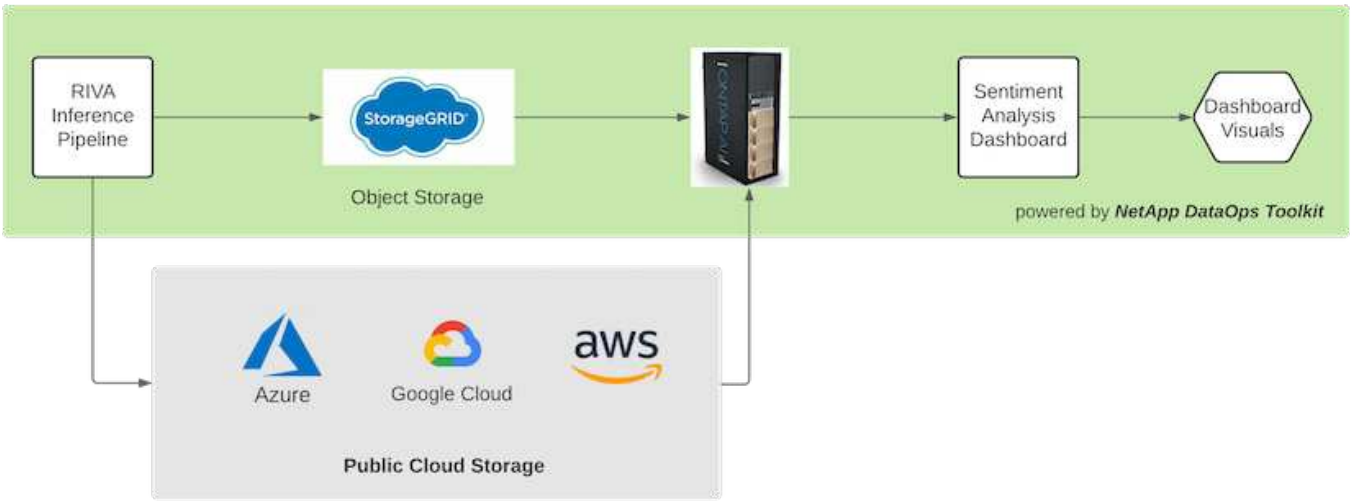
预构建的工具可以通过强大的方式组合在一起，快速创建端到端 AI 管道来解决此问题。在这种情况下，可以使用 NVIDIA Riva 库执行两项串联任务：音频抄写和情感分析。第一种是受监控学习信号处理算法，第二种是受监控学习 NLP 分类算法。这些即装即用的算法可以使用 NVIDIA TAO 工具包针对与业务相关的数据的任何相关使用情形进行微调。这样，构建的解决方案就会更加准确，功能更强大，只需成本和资源的一小部分。客户可以采用 "NVIDIA Maxine" 支持中心设计中 GPU 加速视频会议应用程序的框架。

以下用例是此解决方案的核心。这两种使用情形都使用 TAO 工具包进行模型微调，使用 Riva 进行模型部署。

- 语音到文本
- 情感分析

为了分析员工和客户之间的支持中心互动，可以通过管道进行每一次以音频通话形式进行的客户对话，以得出句子级别的感受。然后，人类可以验证这些情感，为这些情感提供合理的理由，或者根据需要进行调整。然后，标记的数据将传递到微调步骤，以改善情绪预测。如果已存在标记的情感数据，则可以加快模型微调。无论哪种情况，管道都可通用于需要输入音频和对句子进行分类的其他解决方案。

AI Sentiment 输出会上传到外部云数据库或公司管理的存储系统。情感输出将从这一更大的数据库传输到本地存储，以在为经理显示情感分析的信息板中使用。信息板的主要功能是与客户服务员工实时互动。经理可以通过实时更新每句话的感受，并对员工过去的表现或客户反应进行历史回顾，评估员工的通话并提供反馈。



。 "NetApp DataOps 工具包" 即使在 Riva 推理管道生成情感标签之后，也可以继续管理数据存储系统。这些 AI 结果可以上传到 NetApp DataOps 工具包管理的数据存储系统。数据存储系统必须能够管理数百个插件并每分钟进行一次选择。本地设备存储系统实时查询较大的数据存储以进行提取。此外，还可以查询较大的数据存储实例以查找历史数据，从而进一步增强信息板体验。NetApp DataOps 工具包可通过快速克隆数据并将其分发到使用该数据的所有信息板，从而为这两种用途提供便利。

目标受众

解决方案的目标受众包括以下组：

- 员工经理
- 数据工程师 / 数据科学家
- IT 管理员（内部，云或混合）

在整个对话过程中跟踪情绪是评估员工绩效的宝贵工具。通过使用 AI 信息板，经理可以实时了解员工和来电者如何改变自己的感受，从而可以开展实时评估和指导会议。此外，企业还可以从参与语音对话，文本聊天机器人和视频会议的客户中获得宝贵的客户见解。此类客户分析可利用大规模多模式处理功能以及现代化的一流 AI 模型和工作流。

在数据方面，支持中心每天都会处理大量音频文件。NetApp DataOps 工具包可帮助执行此数据处理任务，以便定期对模型和情感分析信息板进行微调。

IT 管理员还可以从 NetApp DataOps 工具包中受益，因为它允许他们在部署和生产环境之间快速移动数据。此外，还必须对 NVIDIA 环境和服务器进行管理和分布，以便进行实时推理。

架构

此支持中心解决方案的架构以 NVIDIA 的预构建工具和 NetApp DataOps 工具包为核心。NVIDIA 的工具可用于使用预构建的型号和管道快速部署高性能 AI 解决方案。NetApp DataOps 工具包可简化各种数据管理任务，以加快开发速度。

解决方案技术

"NVIDIA RIVA" 是一款 GPU 加速 SDK ，用于构建多模式对话 AI 应用程序，在 GPU 上提供实时性能。NVIDIA Train ， Aadapt ， and Optimize （ TAO ）工具包提供了一种更快，更简单的方法来加快训练速度，并快速创建高度准确且性能优异的特定域 AI 模型。

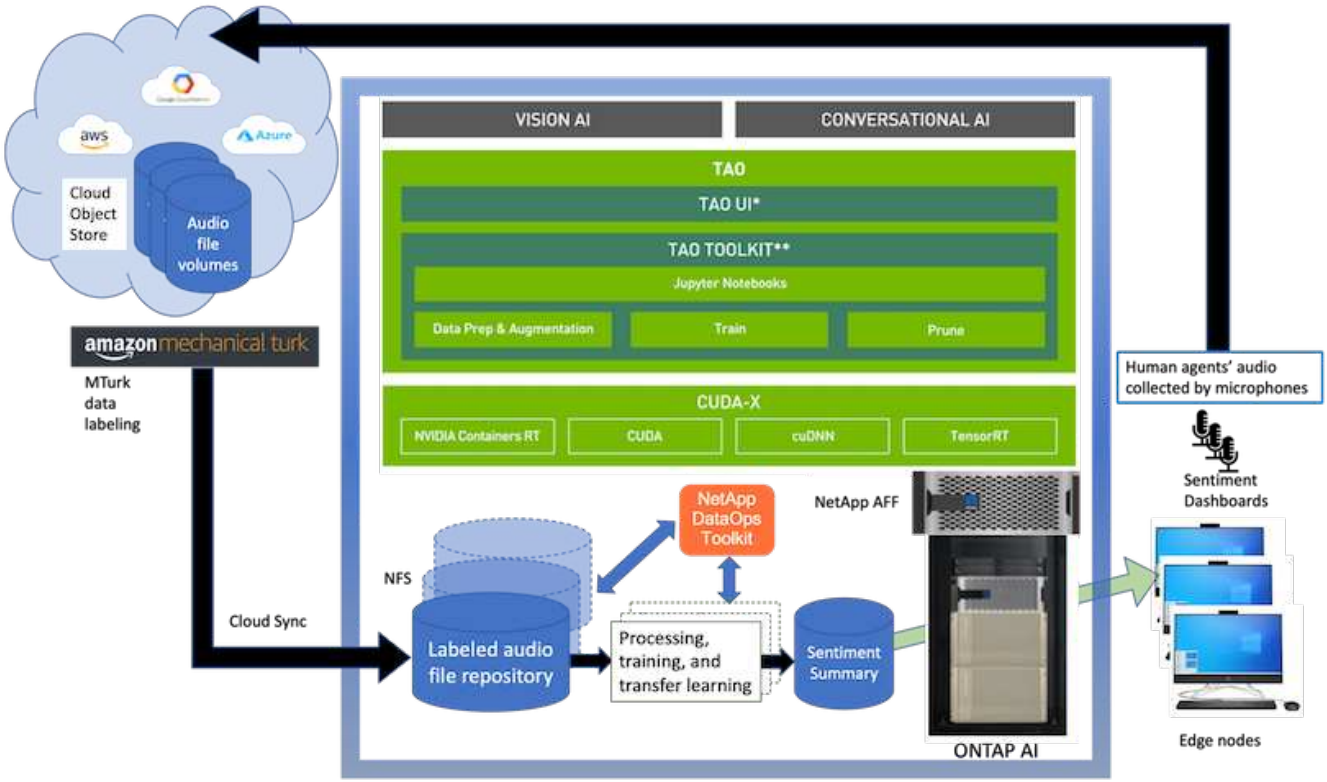
NetApp DataOps 工具包是一个 Python 库，可使开发人员，数据科学家，开发运营工程师和数据工程师轻松执行各种数据管理任务。其中包括近乎即时地配置新的数据卷或 JupyterLab 工作空间，近乎即时地克隆数据卷或 JupyterLab 工作空间，以及近乎即时地对数据卷或 JupyterLab 工作空间进行快照，以实现可追溯性和基线化。

架构图

下图显示了解决方案架构。环境分为三大类：云，核心和边缘。每个类别都可能在不同的地理位置上分散。例如、云包含的对象存储在不同区域的分段中包含音频文件、而核心可能包含通过高速网络或NetApp BlueXP复制和同步链接的数据中心。边缘节点表示各个人工代理的日常工作平台，在这些平台中，交互式信息板工具和麦克风可用于直观显示情绪并从与客户的对话中收集音频数据。

在 GPU 加速数据中心中，企业可以使用 NVIDIA "Riva" 用于构建对话式 AI 应用程序的框架 "TAO 工具包" 使用传输 L 学习技术连接模型修剪和重新培训。这些计算应用程序和工作流由提供支持 "NetApp DataOps 工具包"，实现 ONTAP 提供的最佳数据管理功能。通过该工具包，企业数据团队可以利用相关的结构化和非结构化数据，通过快照和克隆快速构建模型的原型，以实现可追溯性，版本控制， A/B 测试，从而提供安全性，监管， 和合规性。请参见一节 "存储设计" 有关详细信息：

此解决方案演示了音频文件处理， NLP 模型培训，传输学习和数据管理详细步骤。生成的端到端管道生成一个情感摘要，该摘要会实时显示在人力支持代理的信息板上。



硬件要求

下表列出了实施解决方案所需的硬件组件。在任何特定解决方案实施中使用的硬件组件可能会因客户要求而异。

响应延迟测试	时间（毫秒）
数据处理	10
推理	10

这些响应时间测试针对 560 个对话中 50 ， 000 多个音频文件运行。每个音频文件的大小为 ~100KB ，转换为 wav 时大小为 ~1 MB 。数据处理步骤会将 MP3 转换为 wav 文件。推理步骤会将音频文件转换为文本，并从文本中提取感受。这些步骤彼此独立，可以并行处理以加快此过程。

考虑到在商店之间传输数据的延迟，经理应该能够在句子结尾的一秒内看到实时情感分析的更新。

NVIDIA Riva 硬件

硬件	要求
os	Linux x86_64
GPU 内存（As1）	流模型： ~5600 MB 非流模型： ~3100 MB
GPU 内存（NLP）	~每个 Bert 型号 500 MB

NVIDIA TAO 工具包硬件

硬件	要求
系统 RAM	32 GB
GPU RAM	32 GB
CPU	8 个核心
GPU	NVIDIA （A100 ， V100 和 RTX 30x0）
SSD	100 GB

闪存存储系统

NetApp ONTAP 9.

ONTAP 9.9 是 NetApp 推出的最新一代存储管理软件，可帮助企业打造现代化的基础架构并过渡到云就绪数据中心。借助行业领先的数据管理功能，无论数据位于何处， ONTAP 都可以通过一组工具来管理和保护数据。您还可以将数据自由移动到需要的任何位置：边缘，核心或云。ONTAP 9.9 包含许多功能，可简化数据管理，加快和保护关键数据，并在混合云架构中实现下一代基础架构功能。

NetApp BlueXP复制和同步

"BlueXP复制和同步" 是一项 NetApp 服务，用于快速安全地同步数据，使您可以在内部 NFS 或 SMB 文件共享之间将文件传输到以下任一目标：

- NetApp StorageGRID
- NetApp ONTAP S3
- NetApp Cloud Volumes Service
- Azure NetApp Files

- Amazon Simple Storage Service （Amazon S3）
- Amazon Elastic File System （Amazon EFS）
- Azure Blob
- Google Cloud 存储
- IBM 云对象存储

BlueXP复制和同步功能可将文件快速安全地移动到您需要的位置。数据传输完成后，即可在源和目标上完全使用。BlueXP复制和同步功能会根据预定义的计划持续同步数据、仅移动增量、从而最大限度地减少数据复制所需的时间和资金。BlueXP Copy and Sync是一款软件即服务(SaaS)工具、易于设置和使用。由BlueXP复制和同步触发的数据传输由数据代理执行。您可以在AWS、Azure、Google Cloud Platform或内部部署BlueXP Copy and Sync数据代理。

NetApp StorageGRID

StorageGRID 软件定义的对象存储套件可无缝支持公有，私有云和混合多云环境中的各种用例。借助行业领先的创新技术， NetApp StorageGRID 可存储，保护和保留非结构化数据，以供多用途使用，包括长期的自动化生命周期管理。有关详细信息，请参见 "NetApp StorageGRID" 站点

软件要求

下表列出了实施此解决方案所需的软件组件。在任何特定解决方案实施中使用的软件组件可能会因客户要求而异。

主机	要求
Riva （原 JarVis）	1.4.0
TAO 工具包（以前称为传输学习工具包）	3.0
ONTAP	9.9.1
DGX 操作系统	5.1
DOTK	2.0.0

NVIDIA Riva 软件

软件	要求
Docker	如果不使用 DGX，则大于 19.02 （安装了 NVIDI-Docker） >=19.03
NVIDIA 驱动程序	对于数据中心 GPU，则为 465.19.01+ 418.40+，440.33+， 450.51+， 460.27+
容器操作系统	Ubuntu 20.04
CUDA	11.3.0
cucBLAS	11.5.1.101
cuDNN	8.2.0.41
NCCL	2.9.6
TensorRT	7.2.3.4

软件	要求
Triton 推理服务器	2.9.0

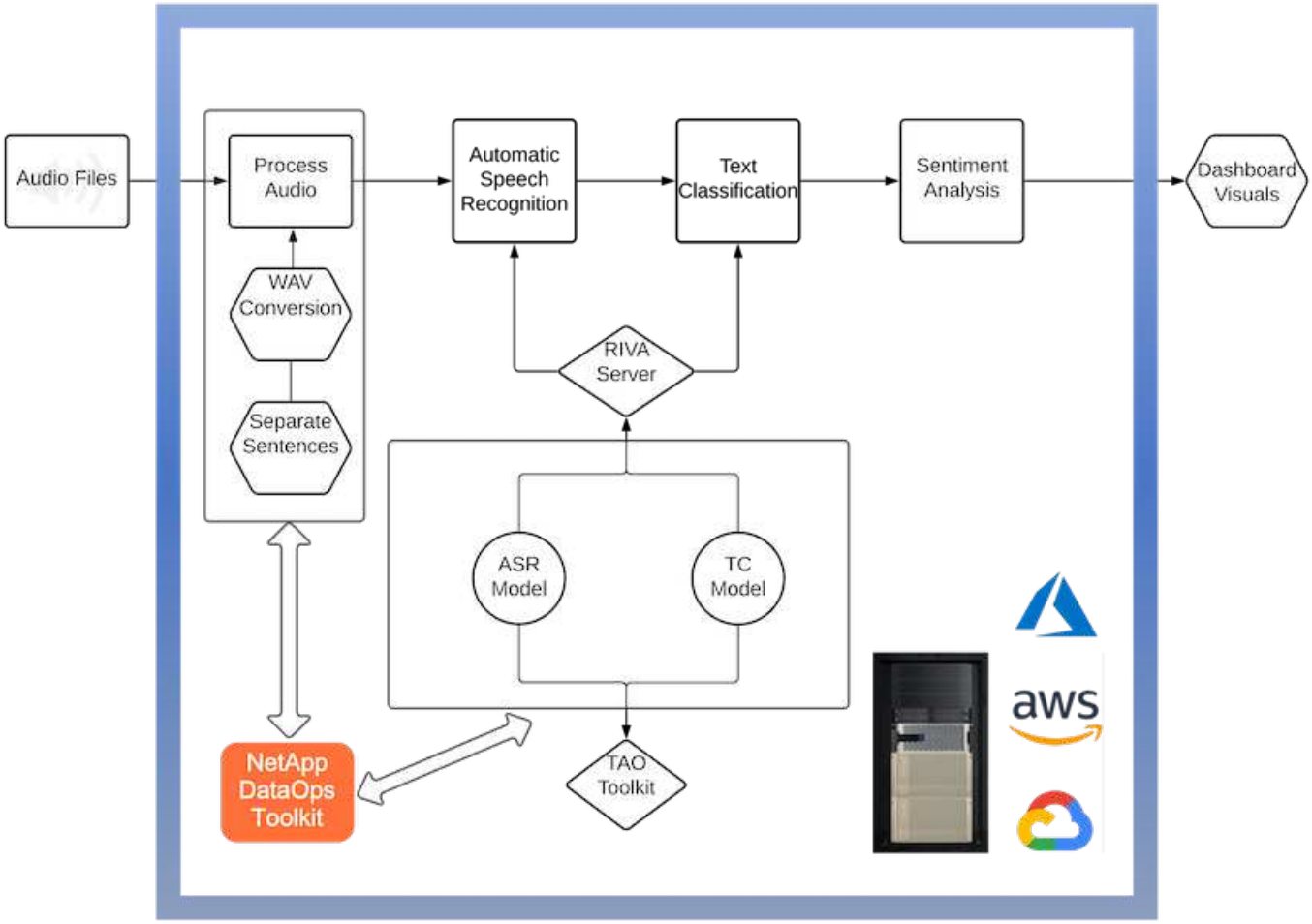
NVIDIA TAO 工具包软件

软件	要求
Ubuntu 18.04 LTS	18.04
Python	>=3.6.9
Docker — ce	19.03.5
Docker API	1.40
nvidia-container-toolkit	>1.3.0-1
nvidia-container-runtime	3.4.0-1.
nvidia-docker2.	2.5.0-1.
NVIDIA 驱动程序	>455
Python-pip	>21.06
nvidia-pyindex	最新版本

用例详细信息

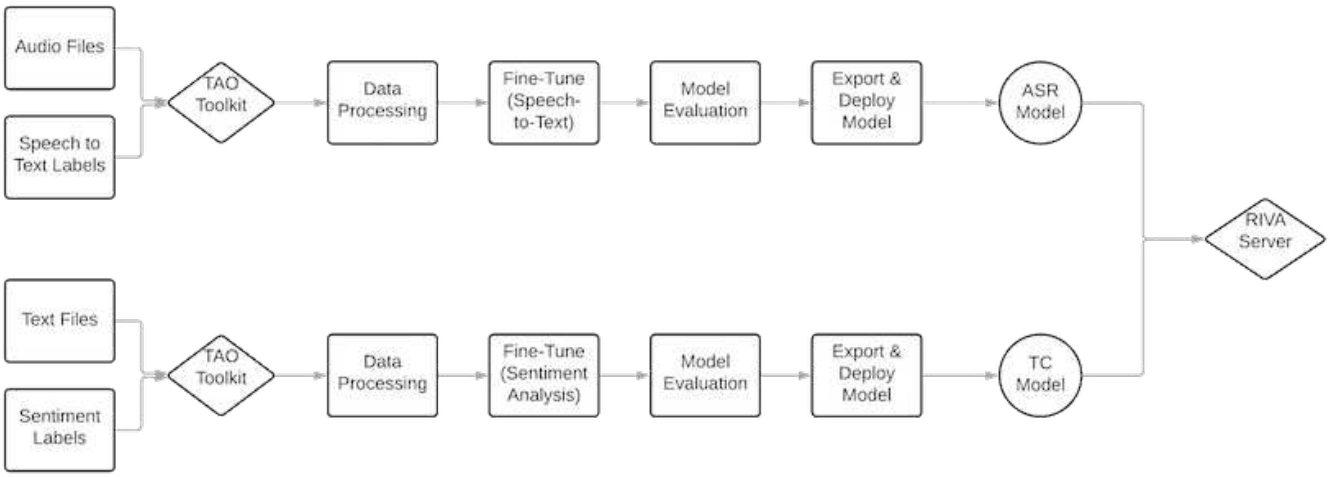
此解决方案适用场景的使用情形如下：

- 语音到文本
- 情感分析



语音到文本使用情形首先会为支持中心载入音频文件。然后，根据 Riva 所需的结构对该音频进行处理。如果尚未将音频文件拆分为其分析单元，则必须先执行此操作，然后再将音频传递给 Riva 。处理完此音频文件后，它将作为 API 调用传递到 Riva 服务器。该服务器采用其托管的众多型号之一，并返回响应。此语音到文本（自动语音识别的一部分）将返回音频的文本表示形式。此时，管道将切换到情感分析部分。

对于情感分析，自动语音识别输出的文本将作为文本分类的输入。文本分类是用于将文本分类为任意数量的类别的 NVIDIA 组件。对于支持中心对话，情绪类别从正面到负面不等。可以使用一个组套来评估模型的性能，以确定微调步骤是否成功。



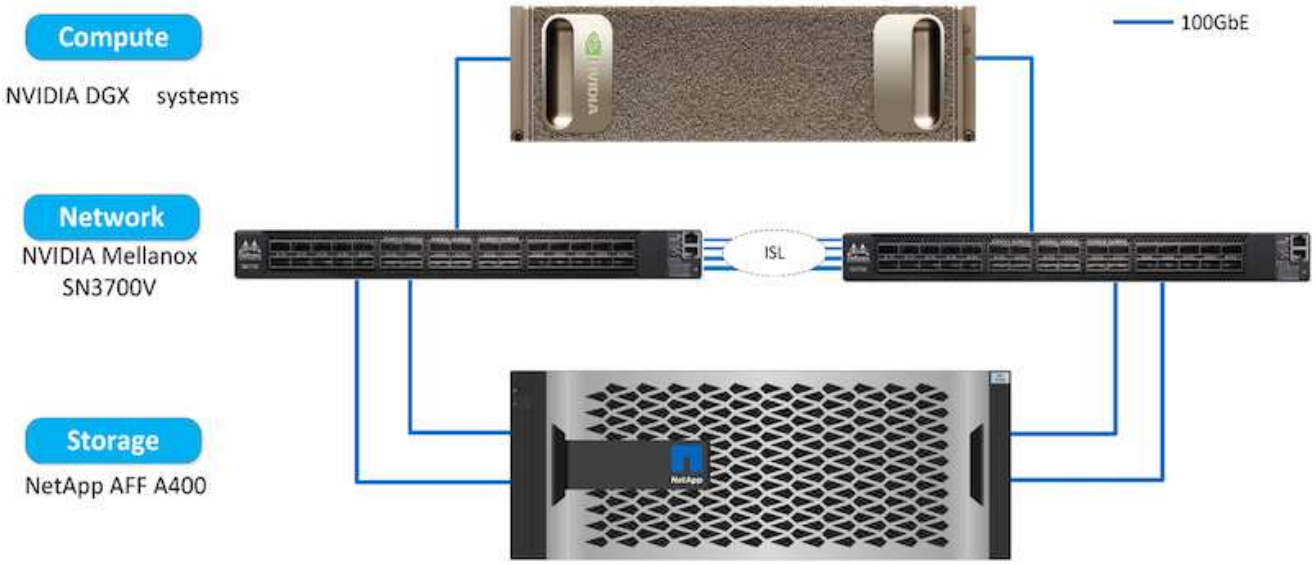
在 TAO 工具包中，语音到文本和情感分析也使用类似的管道。主要区别在于使用了微调模型所需的标签。TAO 工具包管道从处理数据文件开始。然后是经过预先培训的型号（来自 "NVIDIA NGC 目录"）。经过微调的模型会根据其相应的性能指标进行评估，如果性能比预先训练的模型更高，则会部署到 Riva 服务器。

设计注意事项

本节介绍此解决方案的不同组件的设计注意事项。

网络和计算设计

根据数据安全性的限制，所有数据都必须保留在客户的基础架构或安全环境中。



存储设计

NetApp DataOps 工具包是用于管理存储系统的主要服务。DataOps 工具包是一个 Python 库，可使开发人员，数据科学家，开发运营工程师和数据工程师轻松执行各种数据管理任务，例如近乎即时地配置新的数据卷或 JupyterLab 工作空间，近乎即时地克隆数据卷或 JupyterLab 工作空间， 以及接近瞬时的数据卷快照或 JupyterLab 工作空间快照，以实现可追溯性或基线化。此 Python 库可以用作命令行实用程序或函数库，可以导入到任何 Python 程序或 Jupyter Notebook 中。

RIVA 最佳实践

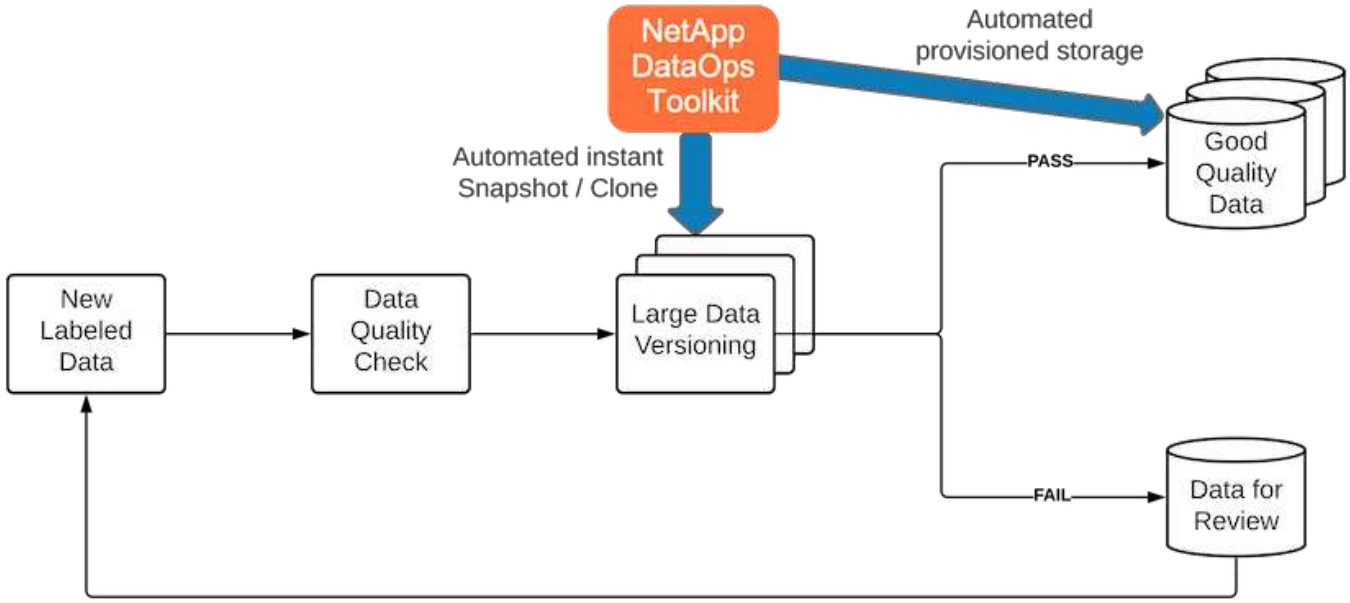
NVIDIA 提供了多种常规功能 "最佳数据实践" 要使用 Riva ，请执行以下操作：

- * 如果可能，请使用无损音频格式。 * 使用像 mp3 这样的有损编解码器可能会降低质量。
- * 扩充训练数据。 * 在音频训练数据中增加背景噪声最初会降低准确性，但同时提高稳定性。
- * 如果使用的是擦文本，请限制词汇大小。 * 许多在线源都包含拼写错误或辅助发音以及不常见的词。删除这些内容可以改进语言模式。
- * 如果可能，请使用最小采样速率 16kH* 。但是，请尝试不要重新采样，因为这样做会降低音频质量。

除了这些最佳实践之外，客户还必须优先收集具有代表性的样本数据集，并为管道的每个步骤提供准确的标签。

换言之，样本数据集应按比例反映目标数据集所示的指定特征。同样，数据集标注器也有责任平衡标记的准确性和速度，以便最大限度地提高数据的质量和数量。例如，此支持中心解决方案需要音频文件，带标签的文本和情感标签。此解决方案的顺序性意味着从管道开始的错误会一直传播到管道的末尾如果音频文件质量较差，则文本抄录和情感标签也会是。

此错误传播方式与此类似，适用场景是对接受过此数据培训的模型进行的。如果情感预测 100% 准确，但语音到文本模式表现不佳，则最终管道将受初始音频到文本抄录的限制。开发人员必须单独考虑每个模型的性能，并将其作为更大管道的组成部分。在这种特定情况下，最终目标是开发一个能够准确预测情绪的渠道。因此，评估管道的总体指标是感受的准确性，而语音到文本的文字记录直接影响到这一点。



NetApp DataOps 工具包利用其近乎瞬时的数据克隆技术补充了数据质量检查管道。必须对每个标记的文件进行评估，并将其与现有标记的文件进行比较。在各种数据存储系统之间分布这些质量检查可确保快速高效地执行这些检查。

部署支持中心观点分析

部署解决方案涉及以下组件：

1. NetApp DataOps 工具包
2. NGC 配置
3. NVIDIA Riva 服务器
4. NVIDIA TAO 工具包
5. 将 TAO 模型导出到 Riva

要执行部署，请完成以下步骤：

NetApp DataOps 工具包：支持中心情感分析

以使用 "NetApp DataOps 工具包"，完成以下步骤：

1. 通过 Pip 安装工具包。

```
python3 -m pip install netapp-dataops-traditional
```

2. 配置数据管理

```
netapp_dataops_cli.py config
```

NGC 配置：支持中心情感分析

进行设置 "NVIDIA NGC"，完成以下步骤：

1. 下载 NGC 。

```
wget -O ngccli_linux.zip
https://ngc.nvidia.com/downloads/ngccli_linux.zip && unzip -o
ngccli_linux.zip && chmod u+x ngc
```

2. 将当前目录添加到路径。

```
echo "export PATH=\"\$PATH:$(pwd)\"" >> ~/.bash_profile && source
~/.bash_profile
```

3. 您必须配置 NGC 命令行界面以供使用，才能运行命令。出现提示时，输入以下命令，包括 API 密钥。

```
ngc config set
```

对于不基于 Linux 的操作系统，请访问 "此处"。

NVIDIA Riva 服务器：支持中心情感分析

进行设置 "NVIDIA RIVA"，完成以下步骤：

1. 从 NGC 下载 Riva 文件。

```
ngc registry resource download-version
nvidia/riva/riva_quickstart:1.4.0-beta
```

2. 初始化 Riva 设置（`Riva_init.sh`）。
3. 启动 Riva 服务器（`Riva_start.sh`）。

4. 启动 Riva 客户端（`Riva_start_client.sh`）。
5. 在 Riva 客户端中，安装音频处理库（"FFmpeg"）

```
apt-get install ffmpeg
```

6. 启动 "Jupyter" 服务器
7. 运行 Riva 推理管道笔记本电脑。

NVIDIA TAO 工具包：支持中心情感分析

要设置 NVIDIA TAO 工具包，请完成以下步骤：

1. 准备并激活 "虚拟环境" 适用于 TAO 工具包。
2. 安装 "所需的软件包"。
3. 手动拉取训练和微调期间使用的图像。

```
docker pull nvcr.io/nvidia/tao/tao-toolkit-pyt:v3.21.08-py3
```

4. 启动 "Jupyter" 服务器
5. 运行 TAO 微调笔记本电脑。

将 TAO 模型导出到 Riva ： 支持中心情感分析

以使用 "Riva 中的 TAO 工具包模型"，完成以下步骤：

1. 将模型保存在 TAO 微调笔记本中。
2. 将经过 TAO 训练的型号复制到 Riva 型号目录。
3. 启动 Riva 服务器（`Riva_start.sh`）。

部署障碍

在开发您自己的解决方案时，请谨记以下几点：

- 首先安装 NetApp DataOps 工具包，以确保数据存储系统以最佳状态运行。
- 必须先安装 NVIDIA NGC ，再安装其他任何版本，因为它会对映像和型号的下载进行身份验证。
- 必须先安装 Riva ，然后再安装 TAO 工具包。Riva 安装会根据需要配置 Docker 守护进程以提取映像。
- DGX 和 Docker 必须能够访问 Internet 才能下载这些型号。

验证结果

如上一节所述，只要有两个或更多机器学习模型按顺序运行，错误就会传播到整个管道

中。对于此解决方案，句子的感受是衡量公司股票风险水平的最重要因素。语音到文本模式虽然对管道至关重要，但在预测情感之前，它是预处理单元。真正重要的是基本事实句子与预测的句子之间的情感差异。此选项用作错误率（WER）一词的代理。语音到文本的准确性非常重要，但 WER 不会直接用于最终管道指标。

```
PIPELINE_SENTIMENT_METRIC = MEAN(DIFF(GT_sentiment, ASR_sentiment))
```

可以根据每句话的 F1 分数，回顾和精确度计算这些情感指标。然后，可以汇总结果并将其显示在混乱列表中，同时还可以显示每个指标的置信区间。

使用传输学习的优势在于，模型性能提高，只需满足极少的数据要求，培训时间和成本。此外，还应将经过微调的模型与其基线版本进行比较，以确保传输学习可以提高性能，而不会影响性能。换言之，经过微调的模型在支持中心数据上的性能应比预先训练的模型更好。

渠道评估

测试用例	详细信息
测试编号	渠道感受指标
测试前提条件	经过微调的语音到文本和情感分析模型模型
预期结果	经过微调的模型的情感指标比原始的预先训练模型的性能更好。

渠道感受指标

1. 计算基线模型的情感指标。
2. 计算经过微调的模型的情感指标。
3. 计算这些指标之间的差异。
4. 计算所有句子之间的差异平均值。

视频和演示

有两台笔记本电脑包含情感分析管道： ""Support-Center-Mode-Transfer-Learning-and-fine-Tuned.ipynb "" 和 ""Support-Center-sentiment-Analysis-Pipeline.ipynb ""。这些笔记本电脑共同演示了如何开发管道，以便使用根据用户数据微调的最先进深度学习模型，载入支持中心数据并从每句话中提取情感。

支持中心—情感分析管道 .ipynb

此笔记本电脑包含推理 Riva 管道，用于载入音频，将其转换为文本以及提取情感以供外部信息板使用。如果尚未下载和处理数据集，则会自动下载和处理该数据集。该笔记本电脑的第一部分是语音到文本，用于将音频文件转换为文本。接下来是 " 情感分析 " 部分，该部分将提取每个文本句子的感受，并以类似于建议的信息板的格式显示这些结果。

此笔记本电脑必须在模型训练和微调之前运行，因为必须下载并转换为正确格式的 mp3 数据集。

Call Center - Sentiment Analysis Pipeline

This notebook demonstrates how to build a pipeline for sentiment analysis of call center conversations. The goal of this pipeline is to develop sentiment analysis for use within an external dashboard.

This tutorial will guide you through the use of NVIDIA's RIVA for automatic speech recognition and text classification. This tutorial uses NetApp cloud storage for data storage and a pre-trained RIVA model.

Channels

These are the channels on which RIVA is hosting models.

- speech: `51051`
- voice: `61051`

These channels **must** be aligned with `riva_speech_api_port` and `riva_vision_api_port` within `config.sh`

In [4]:
```
speech_channel = "localhost:51051"
voice_channel = "localhost:61051"
```

Speech-To-Text

Automatic Speech Recognition (ASR) takes as input an audio stream or audio buffer and returns one or more text transcripts, along with additional optional metadata. ASR represents a full speech recognition pipeline that is GPU accelerated with optimized performance and accuracy. ASR supports synchronous and streaming recognition modes.

For more information on NVIDIA RIVA's Automatic Speech Recognition, visit here.

Constants

Use these constants to affect different aspects of this pipeline:

- `DATA_DIR` : base folder where data is stored
- `DATASET_NAME` : name of the call center dataset
- `COMPANY_DATE` : folder name identifying the particular call center conversation

支持中心—模型训练和微调。 ipynb

在执行此笔记本电脑之前，必须先设置 TAO 工具包虚拟环境（有关安装说明，请参见 " 命令概述 " 中的 "TAO 工具包 " 一节）。

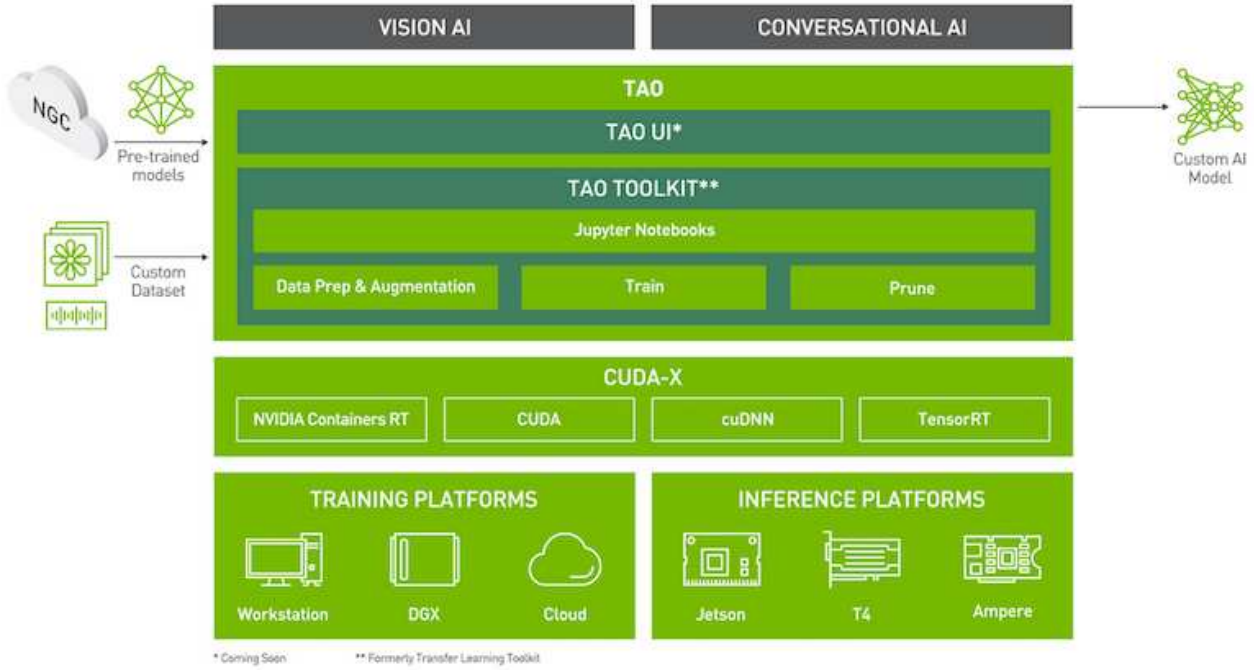
此笔记本电脑依靠 TAO 工具包根据客户数据微调深度学习模型。与上一个笔记本电脑一样，这一款笔记本电脑分为两个部分，分别用于语音到文本和情感分析组件。每个部分都将介绍数据处理，模型培训和微调，结果评估和模型导出。最后，我们将在最后一节中介绍如何部署这两种经过微调的型号，以便在 Riva 中使用。

Call Center - Model Transfer Learning and Fine-Tuning

TAO Toolkit is a python based AI toolkit for taking purpose-built pre-trained AI models and customizing them with your own data. Transfer learning extracts learned features from an existing neural network to a new one. Transfer learning is often used when creating a large training dataset is not feasible in order to enhance the base performance of state-of-the-art models.

For this call center solution, the speech-to-text and sentiment analysis models are fine-tuned on call center data to augment the model performance on business specific terminology.

For more information on the TAO Toolkit, please visit here.



Installing necessary dependencies

For ease of use, please install TAO Toolkit inside a python virtual environment. We recommend performing this step first and then launching the notebook from the virtual environment. Please refer to the README for these instructions.

结论

随着客户体验日益被视为关键的竞争战场， AI 增强型全球支持中心成为几乎每个行业的公司都无法忽视的关键组成部分。本技术报告中建议的解决方案已经证明能够为交付如此卓越的客户体验提供支持，现在的挑战是确保企业采取行动来实现 AI 基础架构和工作流的现代化。

AI 在客户服务中的最佳实施方式不是取代人工代理。相反， AI 可以通过实时情感分析，争议上报和多模式情感计算，让他们能够创造卓越的客户体验，从而检测语言，非语言和面部线索，通过这些线索，全面的 AI 模型可以大规模地提出建议，并补充个人代理可能缺少的内容。AI 还可以更好地匹配具有当前可用代理的特定客户。利用 AI ，企业可以从客户对提供商产品，服务和品牌形象的想法和印象中汲取宝贵的意见。

解决方案还可用于为支持代理构建时间序列数据，以用作客观的性能评估指标。传统的客户满意度调查通常缺乏足够的响应。通过收集员工和客户的长期情绪，雇主可以就支持代理的表现做出明智的决策。

NetApp ， SFL 科学，开源业务流程框架和 NVIDIA 相结合，将最新技术作为托管服务集于一身，并具有极大的灵活性，可加快技术采用速度，加快新 AI/ML 应用程序的上市速度。这些高级服务在内部环境中提供，可轻松移植到云原生环境以及混合部署架构中。

从何处查找追加信息

要了解有关本文档中所述信息的更多信息，请查看以下文档和 / 或网站：

- 3D 交互式演示

  "www.netapp.com/ai"

- 直接与 NetApp AI 专家联系

  "https://www.netapp.com/artificial-intelligence/"

- 《采用 NetApp 解决方案的 NVDIA 基本命令平台》简介

  https://www.netapp.com/pdf.html?item=/media/32792-DS-4145-NVIDIA-Base-Command-Platform-with-NetApp.pdf

- NetApp 的 AI 10 大理由信息图

  "https://www.netapp.com/us/media/netapp-ai-10-good-reasons.pdf"

- 医疗保健领域的人工智能：深入学习如何在 lung-CT 扫描中识别 COVID-19 病变白皮书

  https://www.netapp.com/pdf.html?item=/media/31240-WP-7342.pdf

- 医疗保健领域的人工智能：监控医疗保健环境中的面部掩模使用情况白皮书

  https://www.netapp.com/pdf.html?item=/media/37490-NA-611-Monitoring-face-mask-usage-in-healthcare-settings.pdf

- 医疗保健领域的人工智能：诊断成像技术报告

  https://www.netapp.com/pdf.html?item=/media/7395-tr4811.pdf

- AI for Retail ：使用 NVIDIA Riva 的 NetApp 对话 AI

  "内容提要"

- NetApp ONTAP AI 解决方案简介

  https://www.netapp.com/pdf.html?item=/media/6736-sb-3939.pdf

- NetApp DataOps 工具包解决方案简介

  https://www.netapp.com/pdf.html?item=/media/21480-SB-4111-1220-NA-Data-Science-Toolkit.pdf

- NetApp AI 控制平面解决方案简介

  https://www.netapp.com/pdf.html?item=/media/6737-sb-4055.pdf

- 《利用 Data Drive AI 转变行业》电子书

  "https://www.netapp.com/us/media/na-337.pdf"

- NetApp EF 系列 AI 解决方案简介

  https://www.netapp.com/pdf.html?item=/media/26708-SB-4136-NetApp-AI-E-Series.pdf

- NetApp AI 和联想 ThinkSystem for AI 推理解决方案简介

  https://www.netapp.com/pdf.html?item=/media/25316-SB-4129.pdf

- NetApp AI 和联想 ThinkSystem for Enterprise AI 和 ML 解决方案简介

  https://www.netapp.com/pdf.html?item=/media/25317-SB-4128.pdf

- NetApp 和 NVIDIA —重新定义人工智能视频的可能

  https://www.youtube.com/watch?v=38xw65SteUc

版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。 美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 http://www.netapp.com/TM 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。