

Table of Contents

Uralstech.UAI.LiteRTLM	2
Content	3
Content.ContentType	9
ContentArray	10
Conversation	13
Engine	16
Engine.Backend	20
Engine.LogSeverity	21
Message	22
MessageCallbacks	25
SamplerConfig	28

Namespace Uralstech.UAI.LiteRTLM

Classes

[Content](#)

Represents a content in the [Message](#) of the conversation.

[ContentArray](#)

An array of [Content](#)s.

[Conversation](#)

Represents a conversation with the LiteRT-LM model.

[Engine](#)

Manages the lifecycle of a LiteRT-LM engine, providing an interface for interacting with the underlying native (C++) library.

[Message](#)

Represents a message in the conversation. A message can contain multiple [Contents](#).

[MessageCallbacks](#)

Callbacks for receiving streaming message responses.

[SamplerConfig](#)

Configuration for the sampling process.

Enums

[Content.ContentType](#)

The data type of the [Content](#).

[Engine.Backend](#)

Backend for the LiteRT-LM engine.

[Engine.LogSeverity](#)

Class Content

Namespace: [Uralstech.UAI.LiteRTLM](#)

Represents a content in the [Message](#) of the conversation.

```
public class Content
```

Inheritance

object ← Content

Remarks

This can store text or binary content, based on its [Type](#). This object manages a native `com.google.ai.edge.litertlm.Content` object and must be disposed after usage OR must be managed by a [ContentArray](#) to handle its disposal.

Constructors

Content(Content)

Creates a new [Content](#) from an existing one.

```
public Content(Content other)
```

Parameters

`other` [Content](#)

Remarks

This creates a shallow copy of `other`. A new AndroidJavaObject which refers to the same native Kotlin object as `other` is created, and the text and binary data of `other` is also copied by reference. The new object takes on the same [Type](#) as `other`.

Fields

StringContent

String content ([Text](#), [ImagePath](#), [AudioPath](#)).

```
public readonly string? StringContent
```

Field Value

string

Type

The type of the data contained in this object.

```
public readonly Content.ContentType Type
```

Field Value

[Content.ContentType](#)

Properties

BytesContent

Binary content ([ImageBytes](#), [AudioBytes](#)).

```
public ReadOnlySpan<byte> BytesContent { get; }
```

Property Value

ReadOnlySpan<byte>

Methods

AudioBytes(byte[])

Creates a [Content](#) for audio from bytes.

```
public static Content AudioBytes(byte[] data)
```

Parameters

data byte[]

Returns

[Content](#)

AudioFile(string)

Creates a [Content](#) for audio from a filepath.

```
public static Content AudioFile(string path)
```

Parameters

path string

Returns

[Content](#)

Dispose()

Performs application-defined tasks associated with freeing, releasing, or resetting unmanaged resources.

```
public void Dispose()
```

ImageBytes(byte[])

Creates a [Content](#) for an image from bytes.

```
public static Content ImageBytes(byte[] data)
```

Parameters

data byte[]

Returns

[Content](#)

ImageFile(string)

Creates a [Content](#) for an image from a filepath.

```
public static Content ImageFile(string path)
```

Parameters

path string

Returns

[Content](#)

Text(string)

Creates a [Content](#) for text.

```
public static Content Text(string content)
```

Parameters

content string

Returns

[Content](#)

ToString()

Returns a string that represents the current object.

```
public override string? ToString()
```

Returns

string

A string that represents the current object.

Operators

implicit operator Content(string)

```
public static implicit operator Content(string current)
```

Parameters

current string

Returns

[Content](#)

implicit operator ReadOnlySpan<byte>(Content)

```
public static implicit operator ReadOnlySpan<byte>(Content current)
```

Parameters

current [Content](#)

Returns

ReadOnlySpan<byte>

implicit operator string?(Content)

```
public static implicit operator string?(Content current)
```

Parameters

current [Content](#)

Returns

string

Enum Content.ContentType

Namespace: [Uralstech.UAI.LiteRTLM](#)

The data type of the [Content](#).

```
public enum Content.ContentType
```

Fields

AudioBytes = 3

Audio provided as raw bytes.

AudioPath = 4

Audio provided by a file.

ImageBytes = 1

Image provided as raw bytes.

ImagePath = 2

Image provided by a file.

Text = 0

Text.

Class ContentArray

Namespace: [Uralstech.UAI.LiteRTLM](#)

An array of [Contents](#).

```
public class ContentArray
```

Inheritance

object ← ContentArray

Remarks

This object manages a native `java.util.ArrayList` object and must be disposed after usage OR must be managed by a [Message](#) to handle its disposal.

Constructors

ContentArray(IReadOnlyList<Content>, bool)

Creates a new [ContentArray](#) object.

```
public ContentArray(IReadOnlyList<Content> contents, bool handleChildDispose = true)
```

Parameters

`contents` IReadOnlyList<[Content](#)>

The contents contained in this array.

`handleChildDispose` bool

Should disposal of `contents` be handled by this instance?

ContentArray(ContentArray)

Creates a new [ContentArray](#) from an existing one.

```
public ContentArray(ContentArray other)
```

Parameters

other [ContentArray](#)

Remarks

This creates a semi-deep copy of **other**. A new AndroidJavaObject which refers to the same native Kotlin object as **other** is created, and a shallow copy of each of **other**'s elements is added into a new array and stored as [Elements](#). The new instance's [HandleElementsDispose](#) is set to [true](#).

For more detail on how the elements are shallow copied, see [Content\(Content\)](#).

Fields

Elements

The contents contained in this array.

```
public readonly IReadOnlyList<Content> Elements
```

Field Value

[IReadOnlyList<Content>](#)

HandleElementsDispose

Is disposal of the elements of [Elements](#) handled by this instance?

```
public readonly bool HandleElementsDispose
```

Field Value

bool

Methods

Dispose()

Performs application-defined tasks associated with freeing, releasing, or resetting unmanaged resources.

```
public void Dispose()
```

Operators

implicit operator ContentArray(List<Content>)

```
public static implicit operator ContentArray(List<Content> current)
```

Parameters

`current` List<[Content](#)>

Returns

[ContentArray](#)

implicit operator ContentArray(Content[])

```
public static implicit operator ContentArray(Content[] current)
```

Parameters

`current` Content[]

Returns

[ContentArray](#)

Class Conversation

Namespace: [Uralstech.UAI.LiteRTLM](#)

Represents a conversation with the LiteRT-LM model.

```
public class Conversation
```

Inheritance

object ← Conversation

Remarks

This object manages a native wrapper for a `com.google.ai.edge.litertlm.Conversation` object and must be disposed after usage to close the `Conversation` object and to release the wrapper object.

Methods

CancelProcess()

Cancels any ongoing inference process.

```
public bool CancelProcess()
```

Returns

bool

Dispose()

Performs application-defined tasks associated with freeing, releasing, or resetting unmanaged resources.

```
public void Dispose()
```

SendMessage(Message)

Sends a message to the model and returns the response. This is a synchronous call.

```
public Message? SendMessage(Message message)
```

Parameters

message [Message](#)

The message to send to the model.

Returns

[Message](#)

The model's response message or [null](#) if the call failed

SendMessageAsync(Message, MessageCallbacks)

Sends a message to the model and returns the response aysnc with callbacks.

```
public bool SendMessageAsync(Message message, MessageCallbacks callbacks)
```

Parameters

message [Message](#)

The message to send to the model.

callbacks [MessageCallbacks](#)

The callback to receive the streaming responses.

Returns

bool

Returns [true](#) if the call succeeded; [false](#) otherwise.

StreamSendMessageAsync(Message, MessageCallbacks?, CancellationToken)

Sends a message to the model and returns the partial response messages as an System.Collections.Generic.IAsyncEnumerable<T>.

```
public IAsyncEnumerable<Message> StreamSendMessageAsync(Message message, MessageCallbacks?  
callbacks = null, CancellationToken token = default)
```

Parameters

message [Message](#)

The message to send to the model.

callbacks [MessageCallbacks](#)

Callback object to use in processing. Creates new if not provided.

token CancellationToken

Returns

IAsyncEnumerable<[Message](#)>

Returns the streamed [Message](#) objects. Their disposal is the responsibility of the consumer.

Remarks

[CancelProcess\(\)](#) is automatically called if this method is cancelled using **token**.

Class Engine

Namespace: [Uralstech.UAI.LiteRTLM](#)

Manages the lifecycle of a LiteRT-LM engine, providing an interface for interacting with the underlying native (C++) library.

```
public class Engine
```

Inheritance

object ← Engine

Remarks

This object manages a native wrapper for a [com.google.ai.edge.litertlm.Engine](#) object and must be disposed after usage to close the [Engine](#) object and to release the wrapper object.

Properties

IsInitialized

Returns [true](#) if the engine is initialized and ready for use; [false](#) otherwise.

```
public bool IsInitialized { get; }
```

Property Value

bool

Methods

Create(string, Backend, Backend, Backend, int, bool)

Creates a new LiteRT LM engine.

```
public static Engine? Create(string modelPath, Engine.Backend backend = Backend.CPU,  
    Engine.Backend visionBackend = Backend.Undefined, Engine.Backend audioBackend =
```

```
Backend.Undefined, int maxTokens = 0, bool useExternalCacheDir = true)
```

Parameters

modelPath string

The absolute file path to the LiteRT-LM model.

backend [Engine.Backend](#)

The backend to use for the engine.

visionBackend [Engine.Backend](#)

The backend to use for the vision executor. If [Undefined](#), vision executor will not be initialized.

audioBackend [Engine.Backend](#)

The backend to use for the audio executor. If [Undefined](#), audio executor will not be initialized.

maxTokens int

The maximum number of the sum of input and output tokens. It is equivalent to the size of the kv-cache. When 0, use the default value from the model or the engine.

useExternalCacheDir bool

Should cache files be placed in the external or internal cache dir appointed to the app by Android?

Returns

[Engine](#)

The uninitialized engine or [null](#) if the call failed.

Remarks

The engine can take a long time to initialize. Check [IsInitialized](#) to see if it's done.

CreateAsync(string, Backend, Backend, Backend, int, bool, CancellationToken)

Creates a new LiteRT LM engine and waits for it to initialize.

```
public static Awaitable<Engine?> CreateAsync(string modelPath, Engine.Backend backend = Backend.CPU, Engine.Backend visionBackend = Backend.Undefined, Engine.Backend audioBackend = Backend.Undefined, int maxTokens = 0, bool useExternalCacheDir = true, CancellationToken token = default)
```

Parameters

modelPath string

The absolute file path to the LiteRT-LM model.

backend [Engine.Backend](#)

The backend to use for the engine.

visionBackend [Engine.Backend](#)

The backend to use for the vision executor. If [Undefined](#), vision executor will not be initialized.

audioBackend [Engine.Backend](#)

The backend to use for the audio executor. If [Undefined](#), audio executor will not be initialized.

maxTokens int

The maximum number of the sum of input and output tokens. It is equivalent to the size of the kv-cache. When 0, use the default value from the model or the engine.

useExternalCacheDir bool

Should cache files be placed in the external or internal cache dir appointed to the app by Android?

token CancellationToken

Returns

Awaitable<[Engine](#)>

The initialized engine or [null](#) if the call failed.

CreateConversation(Message?, SamplerConfig?)

Creates a new [Conversation](#) from the initialized engine.

```
public Conversation? CreateConversation(Message? systemMessage = null, SamplerConfig? samplerConfig = null)
```

Parameters

`systemMessage` [Message](#)

The optional system message to be used in the conversation.

`samplerConfig` [SamplerConfig](#)

The optional configuration for the sampling process. If [null](#), then uses the engine's default values.

Returns

[Conversation](#)

The conversation or [null](#) if the call failed.

Dispose()

Performs application-defined tasks associated with freeing, releasing, or resetting unmanaged resources.

```
public void Dispose()
```

SetNativeLogSeverity(LogSeverity)

Sets the minimum log severity for the native (C++) libraries. This affects global logging for all engine instances. If not set, it uses the native libraries' default.

```
public static void SetNativeLogSeverity(Engine.LogSeverity severity)
```

Parameters

`severity` [Engine.LogSeverity](#)

Enum Engine.Backend

Namespace: [Uralstech.UAI.LiteRTLM](#)

Backend for the LiteRT-LM engine.

```
public enum Engine.Backend
```

Fields

CPU = 0

CPU LiteRT backend.

GPU = 1

GPU LiteRT backend.

NPU = 2

NPU LiteRT backend.

Undefined = -1

Undefined value, equivalent to [null](#).

Enum Engine.LogSeverity

Namespace: [Uralstech.UAI.LiteRTLM](#)

```
public enum Engine.LogSeverity
```

Fields

Debug = 1

Error = 4

Fatal = 5

Infinity = 1000

Info = 2

Verbose = 0

Warning = 3

Class Message

Namespace: [Uralstech.UAI.LiteRTLM](#)

Represents a message in the conversation. A message can contain multiple [Contents](#).

```
public class Message
```

Inheritance

object ← Message

Remarks

This object manages a native `com.google.ai.edge.litertlm.Message` object and must be disposed after usage.

Constructors

Message(Message)

Creates a new [Message](#) from an existing one.

```
public Message(Message other)
```

Parameters

`other` [Message](#)

Remarks

This creates a semi-deep copy of `other`. A new `AndroidJavaObject` which refers to the same native Kotlin object as `other` is created, and a semi-deep copy of `other`'s [Contents](#) is created. The new instance's [HandleContentsDispose](#) is set to `true`.

For more detail on how [Contents](#) is semi-deep copied, see [ContentArray\(ContentArray\)](#).

Fields

Contents

The managed content array stored by this object.

```
public readonly ContentArray Contents
```

Field Value

[ContentArray](#)

HandleContentsDispose

Is disposal of [Contents](#) handled by this instance?

```
public readonly bool HandleContentsDispose
```

Field Value

bool

Methods

Dispose()

Performs application-defined tasks associated with freeing, releasing, or resetting unmanaged resources.

```
public void Dispose()
```

Of(string)

Creates a [Message](#) from a text string.

```
public static Message Of(string textMessage)
```

Parameters

`textMessage` string

Returns

[Message](#)

Of(ContentArray, bool)

Creates a [Message](#) from the [ContentArray](#).

```
public static Message Of(ContentArray contents, bool handleContentsDispose = true)
```

Parameters

`contents` [ContentArray](#)

`handleContentsDispose` bool

Should the message object handle the disposing of the array?

Returns

[Message](#)

ToString()

Returns a string that represents the current object.

```
public override string ToString()
```

Returns

string

A string that represents the current object.

Class MessageCallbacks

Namespace: [Uralstech.UAI.LiteRTLM](#)

Callbacks for receiving streaming message responses.

```
public class MessageCallbacks : AndroidJavaProxy
```

Inheritance

object ← MessageCallbacks

Constructors

MessageCallbacks()

```
public MessageCallbacks()
```

Methods

Invoke(string, nint)

```
public override nint Invoke(string methodName, nint javaArgs)
```

Parameters

methodName string

javaArgs nint

Returns

nint

Events

OnDone

Called when all message chunks are sent for a given SendMessageAsync call.

```
public event Action? OnDone
```

Event Type

Action

OnError

Called when an error occurs during the response streaming process, with the Kotlin `Throwable` and any error message.

```
public event Action<AndroidJavaObject, string?>? OnError
```

Event Type

Action<AndroidJavaObject, string>

Remarks

The `AndroidJavaObject` is disposed of immediately after the event's `Invoke` is completed.

OnMessage

Called when a new message chunk is available from the model, along with the message.

```
public event Action<Message>? OnMessage
```

Event Type

Action<[Message](#)>

Remarks

This method may be called multiple times for a single SendMessageAsync call as the model streams its response. The [Message](#) object is disposed of immediately after the event's Invoke is completed.

Class SamplerConfig

Namespace: [Uralstech.UAI.LiteRTLM](#)

Configuration for the sampling process.

```
public class SamplerConfig
```

Inheritance

object ← SamplerConfig

Remarks

This object manages a native `com.google.ai.edge.litertlm.SamplerConfig` object and must be disposed after usage.

Constructors

SamplerConfig(double, double, int, int)

Creates a new [SamplerConfig](#) object.

```
public SamplerConfig(double temperature = 1, double topP = 0.949999988079071, int topK = 64,  
int seed = 0)
```

Parameters

`temperature` double

The temperature to use for sampling.

`topP` double

The cumulative probability threshold for nucleus sampling.

`topK` int

The number of top logits used during sampling.

`seed` int

The seed to use for randomization. Default to 0 (same default as engine code).

Fields

Seed

The seed to use for randomization. Default to 0 (same default as engine code).

```
public readonly int Seed
```

Field Value

int

Temperature

The temperature to use for sampling.

```
public readonly double Temperature
```

Field Value

double

TopK

The number of top logits used during sampling.

```
public readonly int TopK
```

Field Value

int

TopP

The cumulative probability threshold for nucleus sampling.

```
public readonly double TopP
```

Field Value

double

Methods

Dispose()

Performs application-defined tasks associated with freeing, releasing, or resetting unmanaged resources.

```
public void Dispose()
```