Table of Contents

Namespace Uralstech.UAI.LiteRTLM

Classes

AsyncDecodeCallback
: Callback for RunDecodeAsync(AsyncDecodeCallback?, CancellationToken).

Content
: Represents a content in the Message of the conversation.

ContentArray
: An array of Contents.

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.

InputData
: A sealed class representing the input data that can be provided to the LiteRT-LM.

InputDataArray
: An array of InputData.

JNIHelpers

JavaArrayList<T>
: An array of `T`s backed by a native `java.util.ArrayList`.

JavaObject
: A C# object backed by a native Java/Kotlin object.

Message
: Represents a message in the conversation. A message can contain multiple Contents.

MessageCallbacks
: Callbacks for receiving streaming message responses.

ResponseCallbacks
: Callbacks for receiving streaming responses.

SamplerConfig
: Configuration for the sampling process.

Session

Manages the lifecycle of a LiteRT-LM session, providing an interface for interacting with the native library.

Enums

Content.ContentType

The data type of the Content.

Engine.Backend

Backend for the LiteRT-LM engine.

Engine.LogSeverity

InputData.DataType

The data type of the InputData.

Class AsyncDecodeCallback

Namespace: Uralstech.UAI.LiteRTLM

Callback for RunDecodeAsync(AsyncDecodeCallback?, CancellationToken).

```
public sealed class AsyncDecodeCallback : AndroidJavaProxy
```

Inheritance

object ← AsyncDecodeCallback

Constructors

AsyncDecodeCallback()

```
public AsyncDecodeCallback()
```

Methods

Invoke(string, nint)

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

Parameters

`methodName` string

`javaArgs` nint

Returns

nint

Events

OnDone

Called when the decode operation is completed with the result.

```
public event Action<string>? OnDone
```

Event Type

Action<string>

Remarks

This event may be invoked from a background thread.

Class Content

Namespace: Uralstech.UAI.LiteRTLM

Represents a content in the Message of the conversation.

```
public sealed class Content : JavaObject
```

Inheritance
object ← JavaObject ← Content

Inherited Members
JavaObject.IsDisposed , JavaObject.Dispose()

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/JavaArrayList<T> 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>

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

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

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 sealed class ContentArray : JavaArrayList<Content>
```

Inheritance

object ← JavaObject ← JavaArrayList<Content> ← ContentArray

Inherited Members

JavaArrayList<Content>.Elements , JavaArrayList<Content>.HandleElementsDispose , JavaArrayList<Content>.Handle , JavaArrayList<Content>.Dispose() , JavaObject.IsDisposed

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> elements, bool handleChildDispose = true)
```

Parameters

`elements` IReadOnlyList<Content>

The elements contained in this array.

`handleChildDispose` bool

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

ContentArray(JavaArrayList<Content>)

Creates a new ContentArray from an existing one.

```
public ContentArray(JavaArrayList<Content> other)
```

Parameters

`other` JavaArrayList<Content>

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).

Methods

ElementFactory(AndroidJavaObject)

Creates a new Content from a native handle.

```
protected override Content ElementFactory(AndroidJavaObject native)
```

Parameters

`native` AndroidJavaObject

Returns

Content

ElementFactory(Content)

Creates a new Content from an existing instance.

```
protected override Content ElementFactory(Content other)
```

Parameters

`other` Content

Returns

Content

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 sealed class Conversation : JavaObject
```

Inheritance
object ← JavaObject ← Conversation

Inherited Members
JavaObject.IsDisposed

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.

Properties

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

IsAlive

Returns true if the session is alive and ready to be used; false otherwise.

```
public bool IsAlive { get; }
```

Property Value

bool

Methods

CancelProcess()

Cancels any ongoing inference process.

```
public bool CancelProcess()
```

Returns

bool

Dispose()

Disposes the handle and its associated resources.

```
public override 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 sealed class Engine : JavaObject
```

Inheritance

object ← JavaObject ← Engine

Inherited Members

JavaObject.IsDisposed

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

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

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.

Remarks

Cancellation of this operation **does not** cancel the initialization of the engine, but cancels the awaiting of it.

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.

CreateSession(SamplerConfig?)

Creates a new Session from the initialized engine.

```
public Session? CreateSession(SamplerConfig? samplerConfig = null)
```

Parameters

`samplerConfig` SamplerConfig

The optional configuration for the sampling process. If null, then uses the engine's default values.

Returns

Session

The session or null if the call failed.

Dispose()

Disposes the handle and its associated resources.

```
public override 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 InputData

Namespace: Uralstech.UAI.LiteRTLM

A sealed class representing the input data that can be provided to the LiteRT-LM.

```
public sealed class InputData : JavaObject
```

Inheritance

object ← JavaObject ← InputData

Inherited Members

JavaObject.IsDisposed , JavaObject.Dispose()

Remarks

This can store text or binary content, based on its Type. This object manages a native `com.google.ai.edge.litertlm.InputData` object and must be disposed after usage OR must be managed by a InputDataArray/JavaArrayList<T> to handle its disposal.

Fields

BytesContent

Binary content (Image, Audio).

```
public readonly byte[]? BytesContent
```

Field Value

byte[]

StringContent

String content (Text).

```
public readonly string? StringContent
```

Field Value

string

Type

The type of the data contained in this object.

```
public readonly InputData.DataType Type
```

Field Value

InputData.DataType

Properties

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

Methods

Audio(byte[])

Creates InputData representing audio.

```
public static InputData Audio(byte[] data)
```

Parameters

`data` byte[]

Returns

InputData

Image(byte[])

Creates InputData representing an image.

```
public static InputData Image(byte[] data)
```

Parameters

`data` byte[]

Returns

InputData

Text(string)

Creates InputData representing text.

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

Parameters

`content` string

Returns

InputData

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 InputData(string)

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

Parameters

`current` string

Returns

InputData

Enum InputData.DataType

Namespace: Uralstech.UAI.LiteRTLM

The data type of the InputData.

```
public enum InputData.DataType
```

Fields

`Audio = 2`

Represents audio input.

Supported formats: WAV.

`Image = 1`

Represents image input.

Supported formats: PNG and JPG.

`Text = 0`

Represents text input.

Class InputDataArray

Namespace: Uralstech.UAI.LiteRTLM

An array of InputData.

```
public sealed class InputDataArray : JavaArrayList<InputData>
```

Inheritance

object ← JavaObject ← JavaArrayList<InputData> ← InputDataArray

Inherited Members

JavaArrayList<InputData>.Elements , JavaArrayList<InputData>.HandleElementsDispose , JavaArrayList<InputData>.Handle , JavaArrayList<InputData>.Dispose() , JavaObject.IsDisposed

Remarks

This object manages a native `java.util.ArrayList` object and must be disposed after usage.

Constructors

InputDataArray(IReadOnlyList<InputData>, bool)

Creates a new InputDataArray object.

```
public InputDataArray(IReadOnlyList<InputData> elements, bool handleChildDispose = true)
```

Parameters

`elements` IReadOnlyList<InputData>

The elements contained in this array.

`handleChildDispose` bool

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

Methods

ElementFactory(AndroidJavaObject)

This method is not supported for this type.

```
protected override InputData ElementFactory(AndroidJavaObject native)
```

Parameters

`native` AndroidJavaObject

Returns

InputData

ElementFactory(InputData)

This method is not supported for this type.

```
protected override InputData ElementFactory(InputData other)
```

Parameters

`other` InputData

Returns

InputData

Class JNIHelpers

Namespace: Uralstech.UAI.LiteRTLM

```
public static class JNIHelpers
```

Inheritance

object ← JNIHelpers

Methods

UnwrapObjectFromArray(nint, int)

```
public static AndroidJavaObject UnwrapObjectFromArray(nint array, int index)
```

Parameters

`array` nint

`index` int

Returns

AndroidJavaObject

UnwrapStringFromArray(nint, int)

```
public static string? UnwrapStringFromArray(nint array, int index)
```

Parameters

`array` nint

`index` int

Returns

string

Class JavaArrayList<T>

Namespace: Uralstech.UAI.LiteRTLM

An array of `T`s backed by a native `java.util.ArrayList`.

```
public abstract class JavaArrayList<T> : JavaObject where T : JavaObject
```

Type Parameters

`T`

Inheritance

object ← JavaObject ← JavaArrayList<T>

Derived

ContentArray, InputDataArray

Inherited Members

JavaObject.IsDisposed , JavaObject.ThrowIfDisposed()

Remarks

This object manages a native `java.util.ArrayList` object and must be disposed after usage.

Constructors

JavaArrayList(AndroidJavaObject, int)

Creates a new JavaArrayList<T> from a native handle.

```
protected JavaArrayList(AndroidJavaObject native, int size)
```

Parameters

`native` AndroidJavaObject

The native handle.

`size` int

The size of the array.

JavaArrayList(IReadOnlyList<T>, bool)

Creates a new JavaArrayList<T> object.

```
public JavaArrayList(IReadOnlyList<T> elements, bool handleChildDispose = true)
```

Parameters

`elements` IReadOnlyList<T>

The elements contained in this array.

`handleChildDispose` bool

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

JavaArrayList(JavaArrayList<T>)

Creates a new JavaArrayList<T> from an existing one.

```
public JavaArrayList(JavaArrayList<T> other)
```

Parameters

`other` JavaArrayList<T>

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 the implementation of ElementFactory(T).

Fields

Elements

The elements contained in this array.

```
public readonly IReadOnlyList<T> Elements
```

Field Value

IReadOnlyList<T>

HandleElementsDispose

Is disposal of the elements of Elements handled by this instance?

```
public readonly bool HandleElementsDispose
```

Field Value

bool

Properties

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

Methods

Dispose()

Disposes the handle and its associated resources.

```
public override void Dispose()
```

DisposeArray(IReadOnlyList<T>)

Disposes all elements of `array`, with null-checking.

```
protected void DisposeArray(IReadOnlyList<T> array)
```

Parameters

`array` IReadOnlyList<T>

ElementFactory(AndroidJavaObject)

Creates a new `T` from a native handle.

```
protected abstract T ElementFactory(AndroidJavaObject native)
```

Parameters

`native` AndroidJavaObject

Returns

T

ElementFactory(T)

Creates a new `T` from an existing instance.

```
protected abstract T ElementFactory(T other)
```

Parameters

`other` T

Returns

T

Class JavaObject

Namespace: Uralstech.UAI.LiteRTLM

A C# object backed by a native Java/Kotlin object.

```
public abstract class JavaObject
```

Inheritance

object ← JavaObject

Derived

Content, Conversation, Engine, InputData, JavaArrayList<T>, Message, SamplerConfig, Session

Properties

Handle

The handle to the native object.

```
public abstract AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

IsDisposed

Has the handle been disposed?

```
public bool IsDisposed { get; }
```

Property Value

bool

Methods

Dispose()

Disposes the handle and its associated resources.

```
public virtual void Dispose()
```

ThrowIfDisposed()

```
protected void ThrowIfDisposed()
```

Class Message

Namespace: Uralstech.UAI.LiteRTLM

Represents a message in the conversation. A message can contain multiple Contents.

```
public sealed class Message : JavaObject
```

Inheritance
object ← JavaObject ← Message

Inherited Members
JavaObject.IsDisposed

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(JavaArrayList<Content>).

Fields

Contents

The contents of this message.

```
public readonly ContentArray Contents
```

Field Value

ContentArray

HandleContentsDispose

Is disposal of Contents handled by this instance?

```
public readonly bool HandleContentsDispose
```

Field Value

bool

Properties

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

Methods

Dispose()

Disposes the handle and its associated resources.

```
public override 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 sealed 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

Remarks

This event may be invoked from a background thread.

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. This event may be invoked from a background thread.

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. This event may be invoked from a background thread.

Class ResponseCallbacks

Namespace: Uralstech.UAI.LiteRTLM

Callbacks for receiving streaming responses.

```
public sealed class ResponseCallbacks : AndroidJavaProxy
```

Inheritance

object ← ResponseCallbacks

Constructors

ResponseCallbacks()

```
public ResponseCallbacks()
```

Methods

Invoke(string, nint)

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

Parameters

`methodName` string

`javaArgs` nint

Returns

nint

Events

OnDone

Called when the stream is complete.

```
public event Action? OnDone
```

Event Type

Action

Remarks

This event may be invoked from a background thread.

OnError

Called when an error occurs, with the Kotlin `Throwable` and any error message.

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

Event Type

Action<AndroidJavaObject, string>

Remarks

The error will be a `java.util.concurrent.CancellationException` if the stream was cancelled normally, and a `LiteRtLmJniException` for other errors.

The AndroidJavaObject is disposed of immediately after the event's Invoke is completed. This event may be invoked from a background thread.

OnNext

Called when a new response is available, with the message chunk.

```
public event Action<string>? OnNext
```

Event Type

Action<string>

Remarks

This event may be invoked from a background thread.

Class SamplerConfig

Namespace: Uralstech.UAI.LiteRTLM

Configuration for the sampling process.

```
public sealed class SamplerConfig : JavaObject
```

Inheritance

object ← JavaObject ← SamplerConfig

Inherited Members

JavaObject.IsDisposed , JavaObject.Dispose()

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

Properties

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

Class Session

Namespace: Uralstech.UAI.LiteRTLM

Manages the lifecycle of a LiteRT-LM session, providing an interface for interacting with the native library.

```
public sealed class Session : JavaObject
```

Inheritance
object ← JavaObject ← Session

Inherited Members
JavaObject.IsDisposed

Remarks

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

Properties

Handle

The handle to the native object.

```
public override AndroidJavaObject Handle { get; }
```

Property Value

AndroidJavaObject

IsAlive

Returns true if the session is alive and ready to be used; false otherwise.

```
public bool IsAlive { get; }
```

Property Value

bool

Methods

CancelProcess()

Cancels any ongoing inference process.

```
public bool CancelProcess()
```

Returns

bool

Dispose()

Disposes the handle and its associated resources.

```
public override void Dispose()
```

GenerateContent(InputDataArray)

Generates content from the provided `input` and any previous input data added by RunPrefill(InputDataArray) or RunPrefillAsync(InputDataArray, CancellationToken).

```
public string? GenerateContent(InputDataArray input)
```

Parameters

`input` InputDataArray

An array of InputData to be processed by the model. If the user wants to run the decode loop only, they can pass an empty array.

Returns

string

The generated content as a string if successful; null otherwise.

Remarks

This handles both the prefilling and decoding steps.

GenerateContentStream(InputDataArray, ResponseCallbacks)

Generates content from the provided `input` and any previous input data added by RunPrefill(InputDataArray) or RunPrefillAsync(InputDataArray, CancellationToken).

```
public bool GenerateContentStream(InputDataArray input, ResponseCallbacks callbacks)
```

Parameters

`input` InputDataArray

An array of InputData to be processed by the model. If the user wants to run the decode loop only, they can pass an empty array.

`callbacks` ResponseCallbacks

The callback to receive the streaming responses.

Returns

bool

true if the request was executed successfuly; false otherwise.

Remarks

This handles both the prefilling and decoding steps.

GenerateContentStreamAsync(InputDataArray, ResponseCallbacks?, CancellationToken)

Generates content from the provided `input` and any previous input data added by RunPrefill(InputDataArray) or RunPrefillAsync(InputDataArray, CancellationToken).

```
public IAsyncEnumerable<string> GenerateContentStreamAsync(InputDataArray input,
ResponseCallbacks? callbacks = null, CancellationToken token = default)
```

Parameters

`input` InputDataArray

An array of InputData to be processed by the model. If the user wants to run the decode loop only, they can pass an empty array.

`callbacks` ResponseCallbacks

Optional cached callback to receive the streaming responses.

`token` CancellationToken

Returns

IAsyncEnumerable<string>

The streamed response chunks.

Remarks

This handles both the prefilling and decoding steps. CancelProcess() is automatically called if this method is cancelled using `token`.

RunDecode()

Runs the decode step for the model to predict the response based on the input data added by RunPrefill(InputDataArray) or RunPrefillAsync(InputDataArray, CancellationToken).

```
public string? RunDecode()
```

Returns

string

The generated content as a string if successful, null otherwise.

Remarks

This is a blocking call and the function will return when the decoding process is done.

RunDecodeAsync(AsyncDecodeCallback?, CancellationToken)

Runs the decode step for the model to predict the response based on the input data added by Run Prefill(InputDataArray) or RunPrefillAsync(InputDataArray, CancellationToken).

```
public Awaitable<string?> RunDecodeAsync(AsyncDecodeCallback? callback = null,
CancellationToken token = default)
```

Parameters

`callback` AsyncDecodeCallback

Optional cached callback proxy.

`token` CancellationToken

Returns

Awaitable<string>

The generated content as a string if successful, null otherwise.

Remarks

Cancellation of this operation **does not** cancel the decode task, but cancels the awaiting of it.

RunPrefill(InputDataArray)

Adds the `input` and starts the prefilling process.

```
public bool RunPrefill(InputDataArray input)
```

Parameters

`input` InputDataArray

An array of InputData to be processed by the model.

Returns

bool

true if successful; false otherwise.

Remarks

User can break down their InputData into multiple chunks and call this function multiple times. This is a blocking call and the function will return when the prefill process is done.

RunPrefillAsync(InputDataArray, CancellationToken)

Adds the `input` and starts the prefilling process.

```
public Awaitable<bool> RunPrefillAsync(InputDataArray input, CancellationToken token
= default)
```

Parameters

`input` InputDataArray

An array of InputData to be processed by the model.

`token` CancellationToken

Returns

Awaitable<bool>

true if successful; false otherwise.

Remarks

User can break down their InputData into multiple chunks and call this function multiple times. Cancellation of this operation **does not** cancel the prefill task, but cancels the awaiting of it.