

セッション E6
「テストと人工知能」

最近の事例から

増田 聡

【所属】

日本アイ・ビー・エム株式会社.東京基礎研究所 シニア・リサーチ・スタッフ・メンバー

【経歴】

日本アイ・ビー・エム株式会社のアプリケーション開発・保守部門において技術支援業務に従事。テストメソドロジー，技法，テストツールなどに専門的に取り組み，プロジェクトにおいて展開をおこなう。2007年からIBMのテストサービス¹の日本市場への展開をおこない，テストコンサルティングなどのテストサービスを提供した。2014年4月から東京基礎研究所においてビジネスアナリティクス，サービス・ソフトウェア・エンジニアリングに従事。

ISO/IEC WG26(ソフトウェアテスト)国内幹事。NPO法人ソフトウェアテスト技術振興協会 (ASTER) 理事。情報処理学会の正会員。

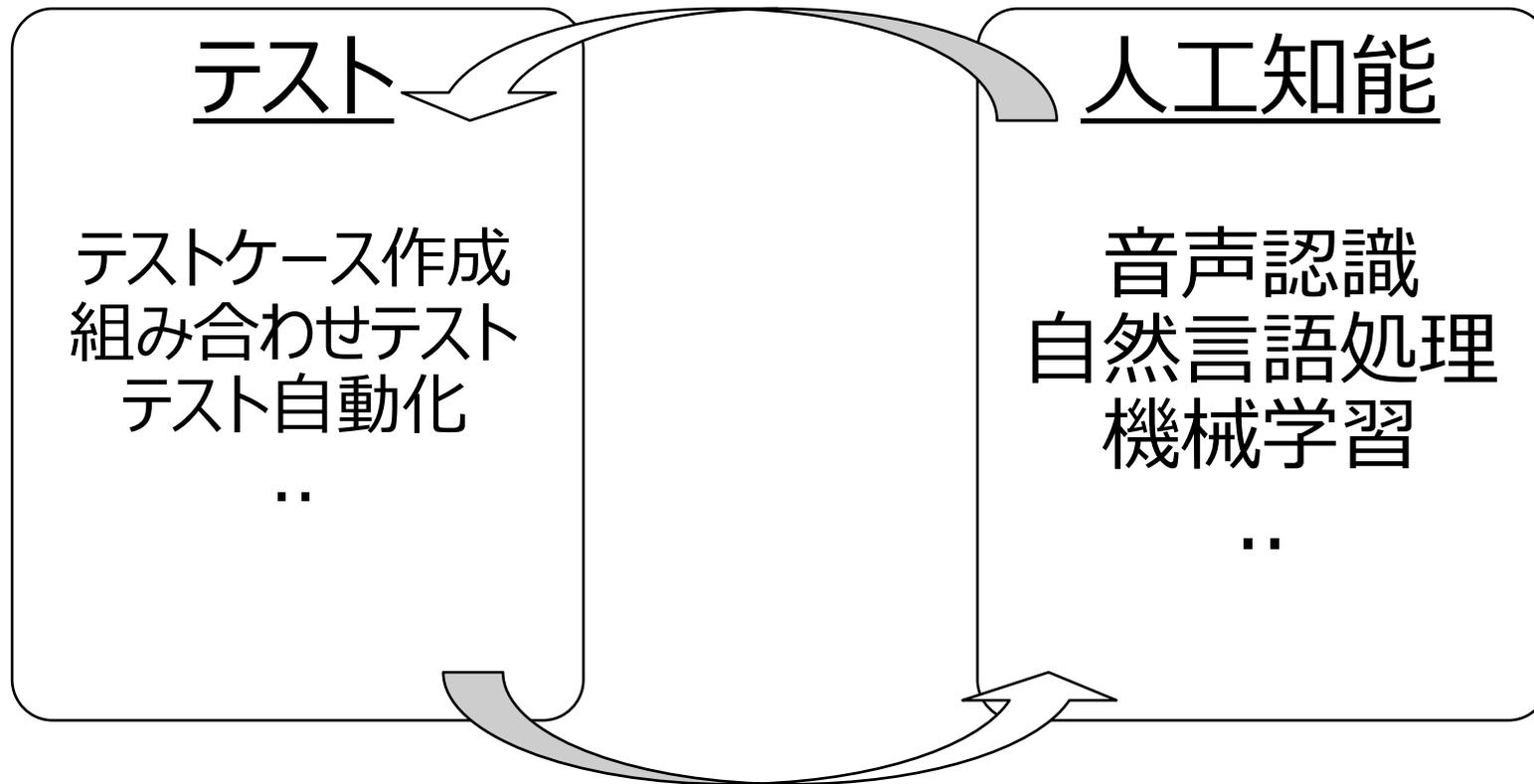
【著書/寄稿】『実践アジャイルテスト』(翻訳、2009、翔泳社)、『テストPRESS Vol.2』に寄稿、『Test ManiaX』に寄稿。



当セッションでは講師個人の考えを述べるもので、IBMの考えを表すものではありません。

- 最近の(私の周辺の)事例から「テストと人工知能」に関連する話をします
 - 音声質問応答システムとテスト
 - 「セルフサービスシステムの実現に向けた回答データ作成手法」
 - 「テストケース作成自動化のための意味役割付与方法」
- より多くの参加者に役立つよう、分かりやすい話を目指します
- 人工知能の技術を利用したアプリケーション開発にはテストのスキルが有用だと思っています

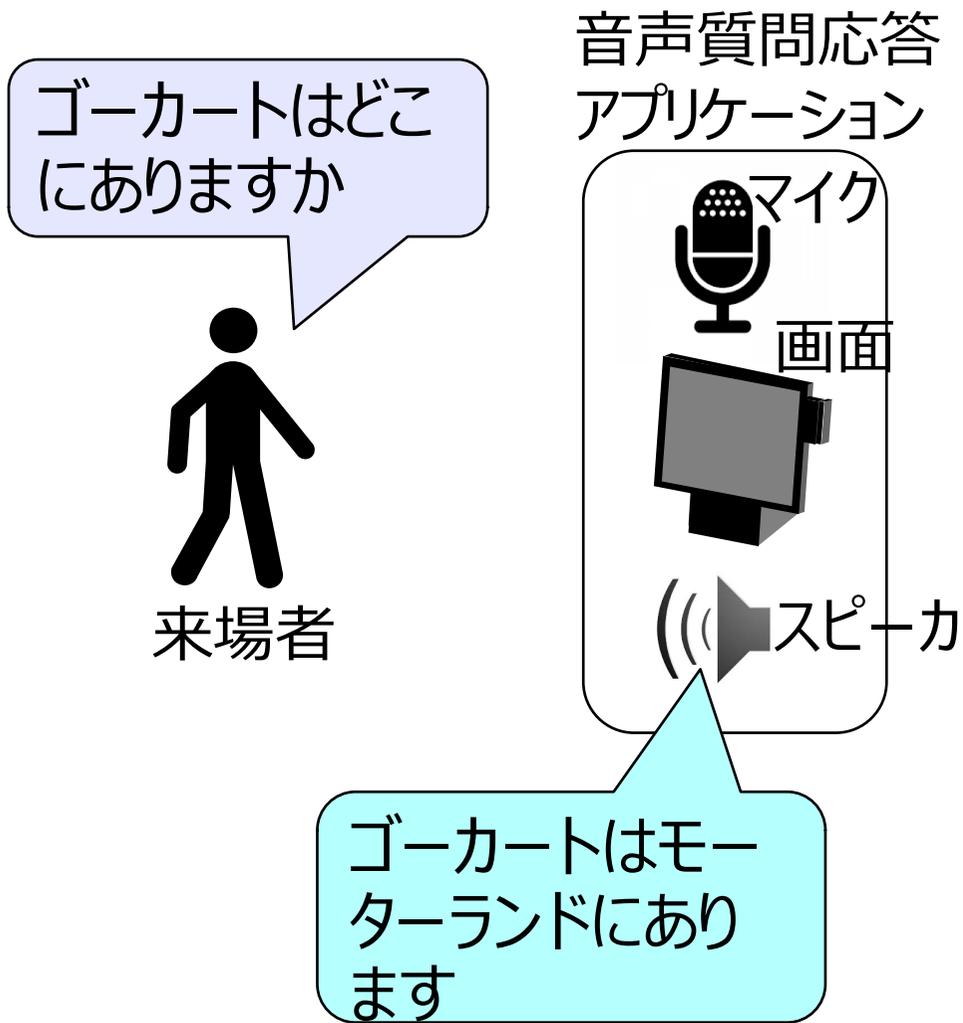
人工知能の技術を利用する



人工知能をテストする

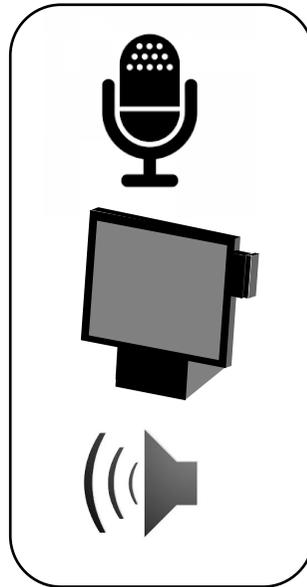
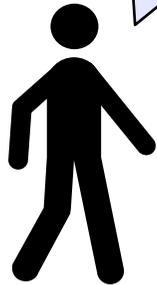
音声質問応答アプリケーション 例

遊園地にある音声質問応答アプリケーション

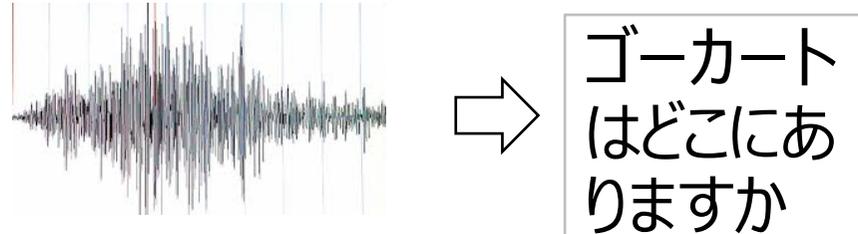


(1) 周辺環境

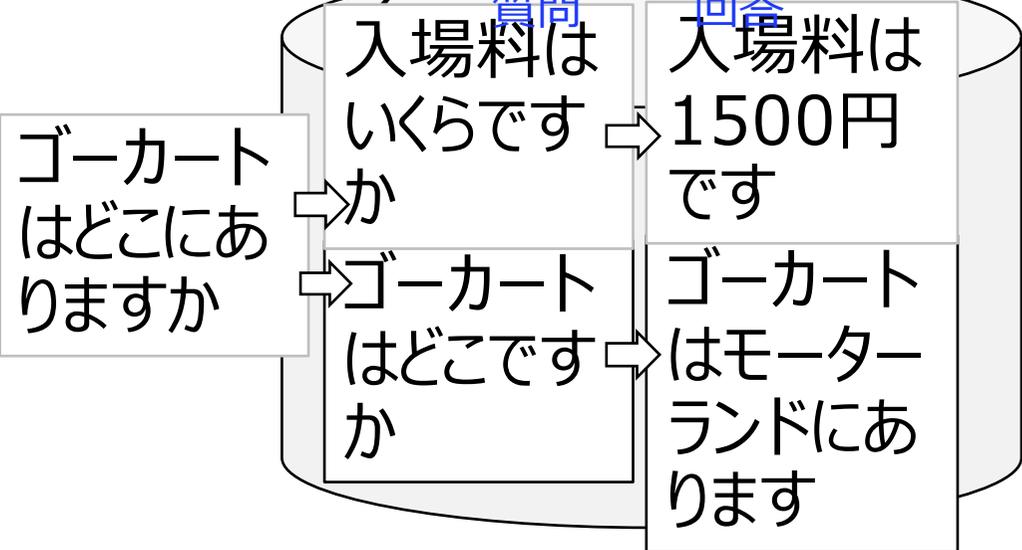
ゴーカートはどこにありますか



(2) 音声テキスト変換



(3) テキスト分類(自然言語処理、機械学習)



二つのベクトル $\mathbf{a}=(a_1, a_2, \dots, a_n)$ と $\mathbf{b}=(b_1, b_2, \dots, b_n)$ に対して

$$\frac{a_1b_1 + \dots + a_nb_n}{\sqrt{a_1a_1 + \dots + a_na_n}\sqrt{b_1b_1 + \dots + b_nb_n}}$$

をcos類似度という。

cos類似度が1に近いほど似ている。

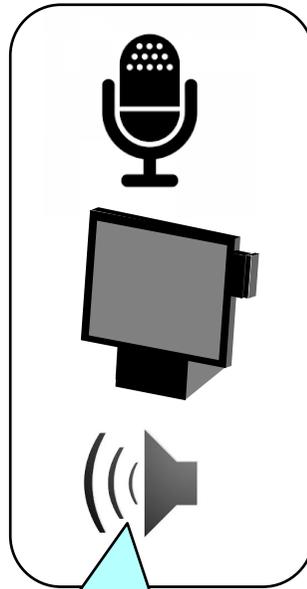
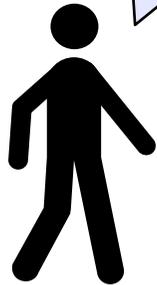
	ゴーカート	は	どこ	に	あり	ます	か	です	入場料	いくら
ゴーカートはどこにありますか	1	1	1	1	1	1	1	0	0	0
ゴーカートはどこですか	1	1	1	0	0	0	1	1	0	0
入場料はいくらですか	0	1	0	0	0	0	1	1	1	1

「ゴーカートはどこにありますか」と「ゴーカートはどこですか」の
cos類似度は **0.676**

「ゴーカートはどこにありますか」と「入場料はいくらですか」の
cos類似度は **0.338**

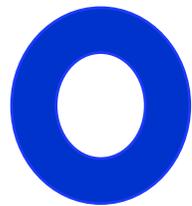
(1) 周辺環境

ゴーカートはどこにありますか



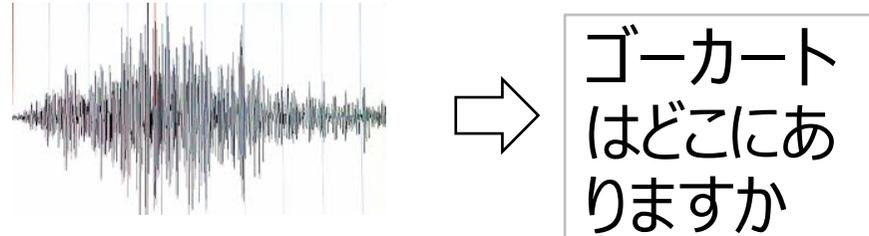
ゴーカートはモーターランドにあります

(4) テキスト音声変換

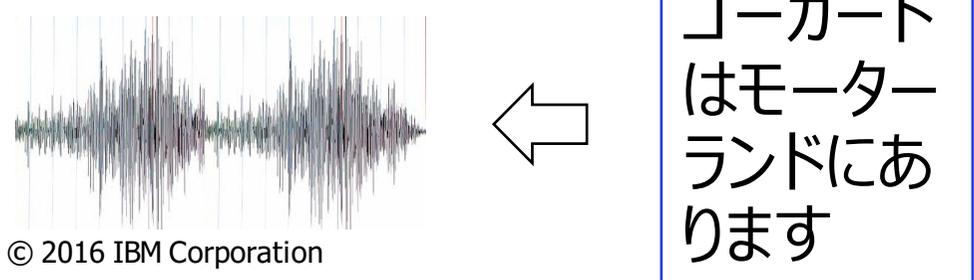
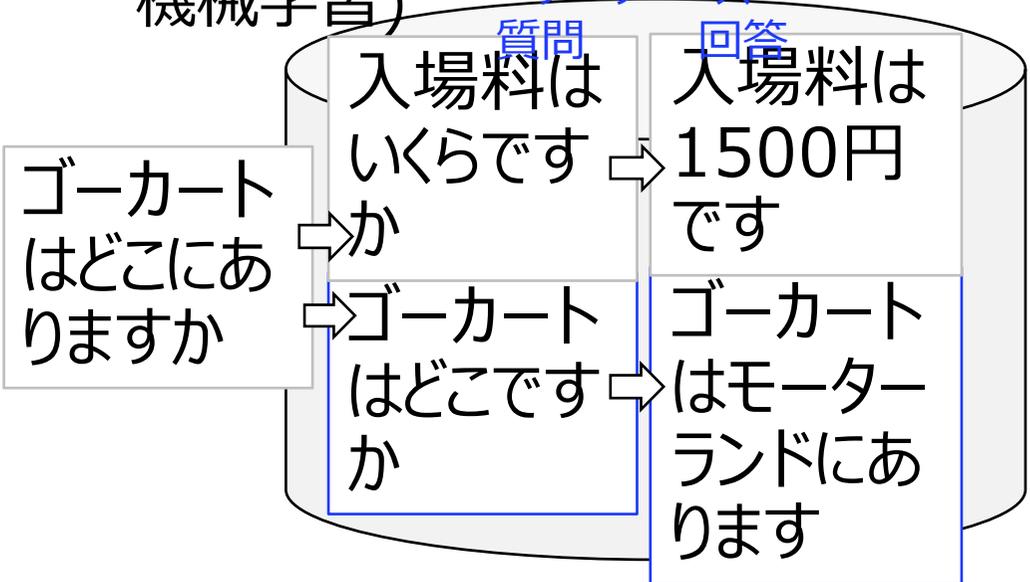


正解

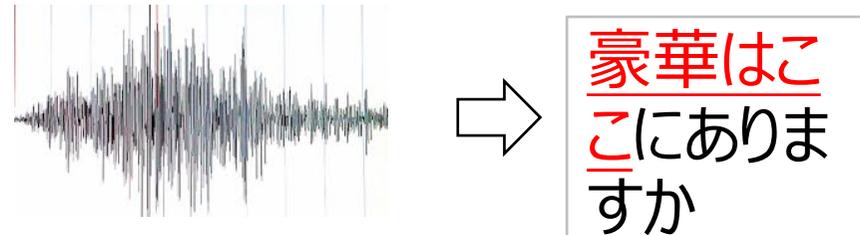
(2) 音声テキスト変換



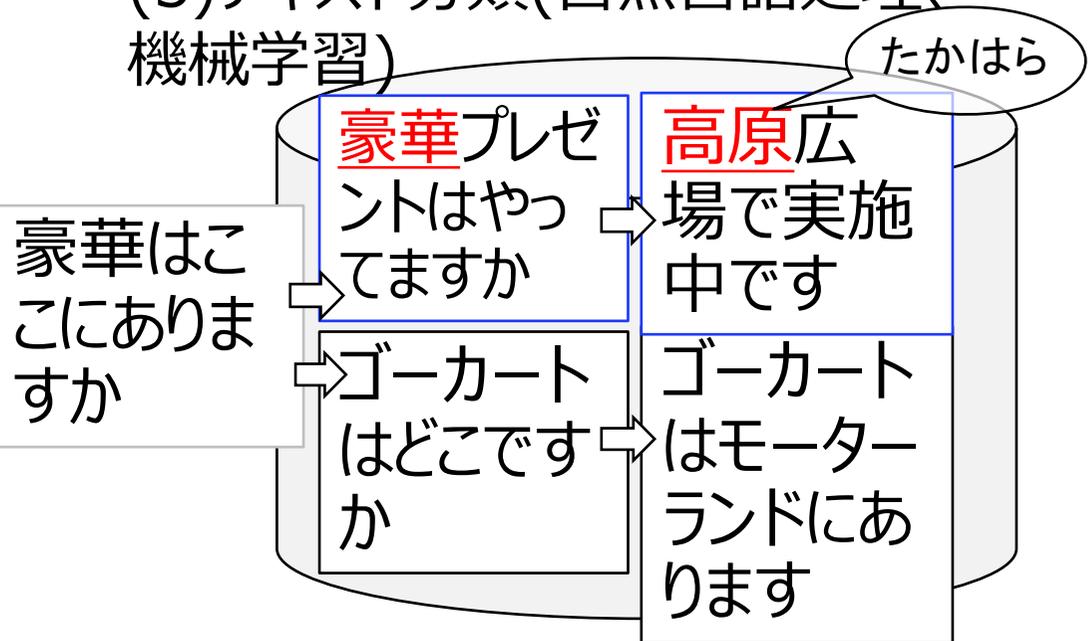
(3) テキスト分類(自然言語処理、機械学習)



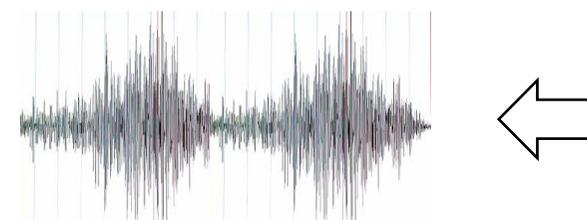
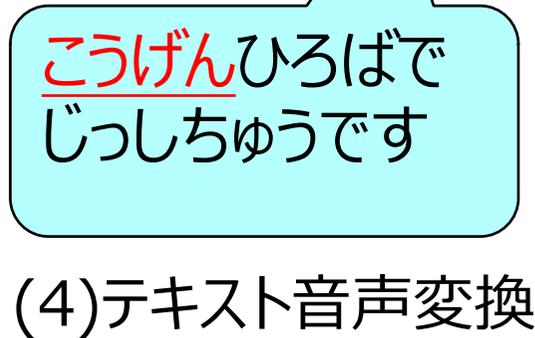
(2) 音声テキスト変換



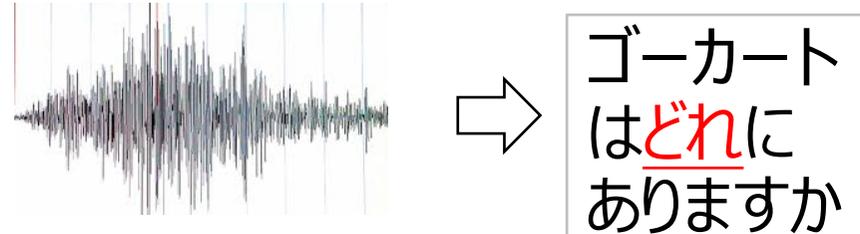
(3) テキスト分類(自然言語処理、機械学習)



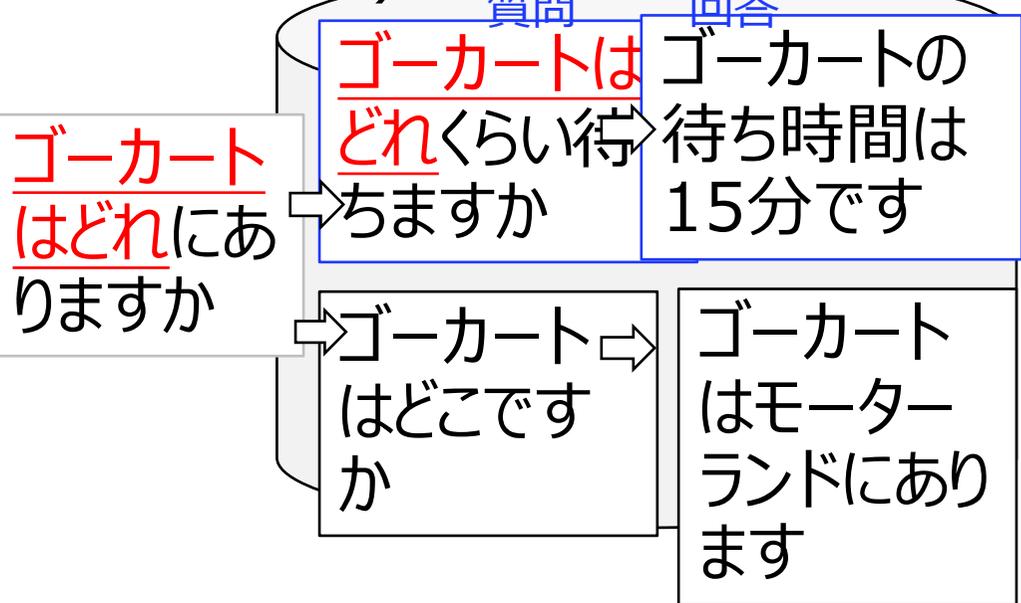
不正解



(2) 音声テキスト変換

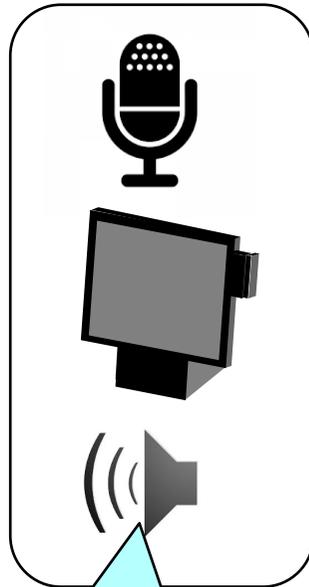
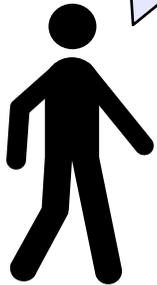


(3) テキスト分類(自然言語処理、機械学習)



(1) 周辺環境

ゴーカートはどこにありますか



×
不正解
ゴーカートの待ち時間は15分です

(4) テキスト音声変換



「セルフサービスシステムの実現に向けた回答データ作成手法」

■ 回答データはどのような構造が良いか？

人工知能学会
第18回知識流通ネットワーク研究会
SIG-KSN-018-06

表 1: 購買行動の段階と求められるサービス知識の関係性 (H:高 M:中 L:低)

	認知段階	感情段階	行動段階
事実・定義	H	H	M
方法・手段	M	H	H
理由・結果	L	L	H
助言・推薦	H	M	L

表 2: サービスの構成要素と必要となる知識タイプの関係 (Y:必要あり N:必要なし)

	事実・定義	方法・手段	理由・結果	助言・推薦
サービス機能	Y	Y	Y	Y
可変な属性	Y	Y	Y	N
不変な属性	Y	N	N	N

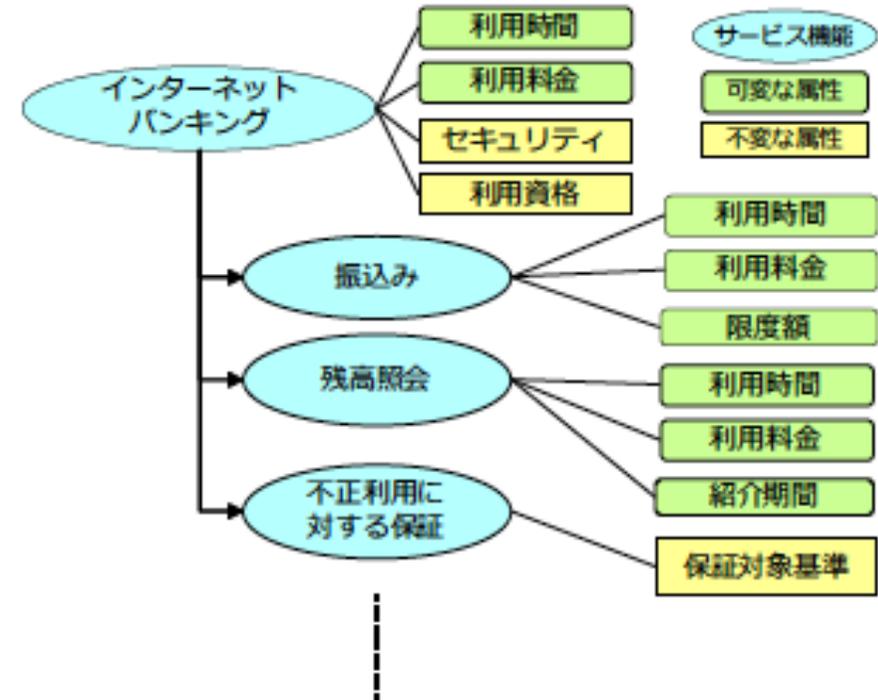
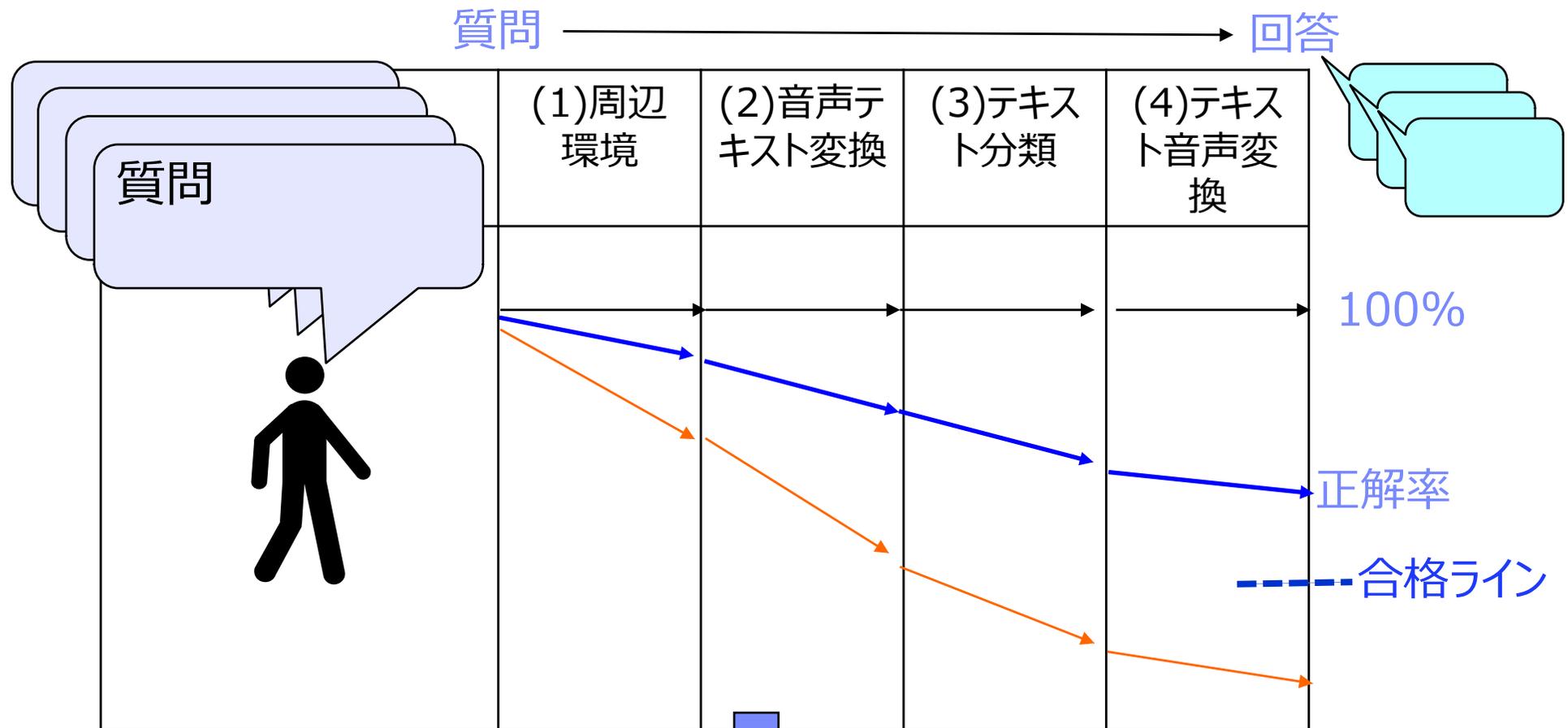


図 2: サービス機能の分解に基づく分析結果

音声質問応答アプリケーションの開発

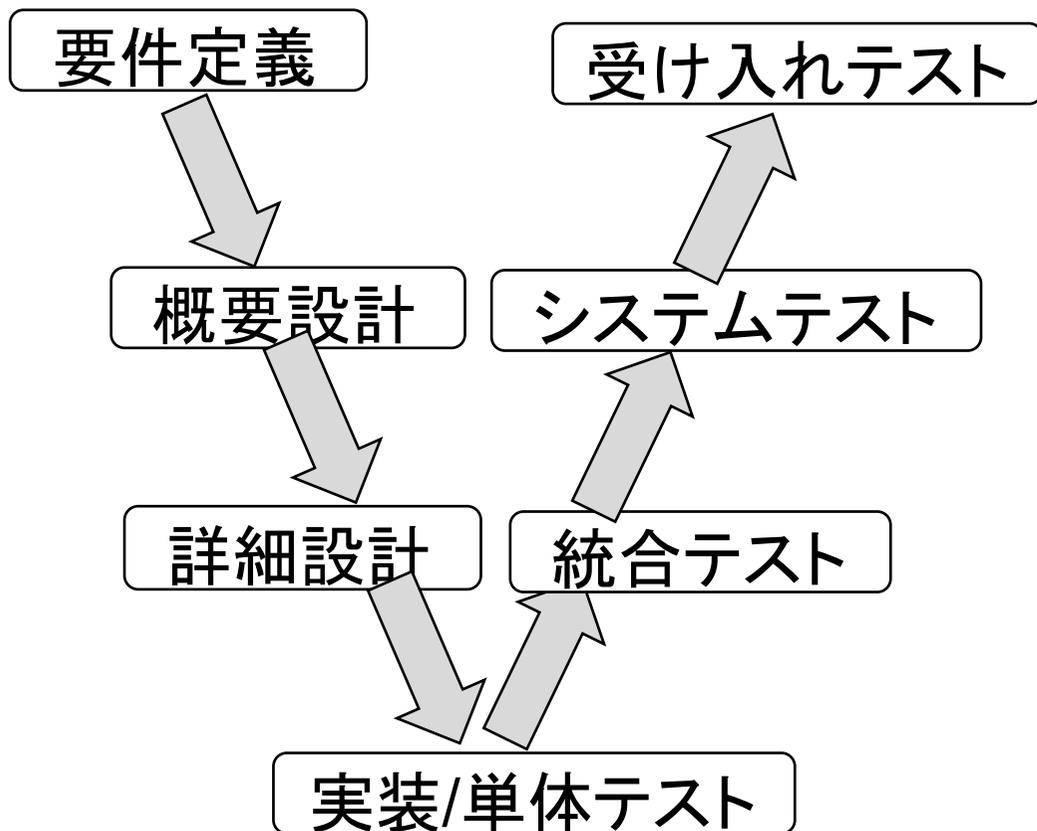


この例のようなアプリケーションの開発作業は質問に対する正解率が合格ラインに達するまで、データやロジックを調整していくこと

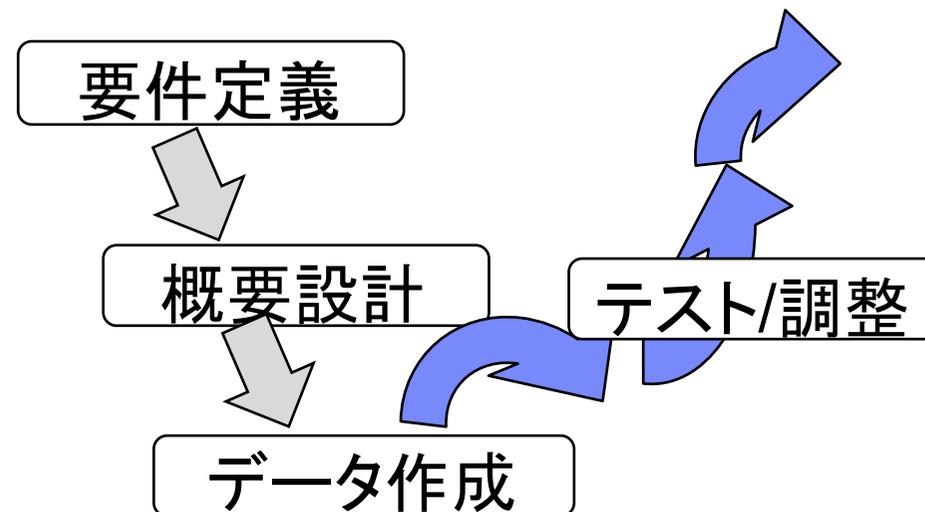
テスト、結果確認、調整、再テストの繰り返し
=> テストの技術が有益である

V字モデルとの比較

V字モデル



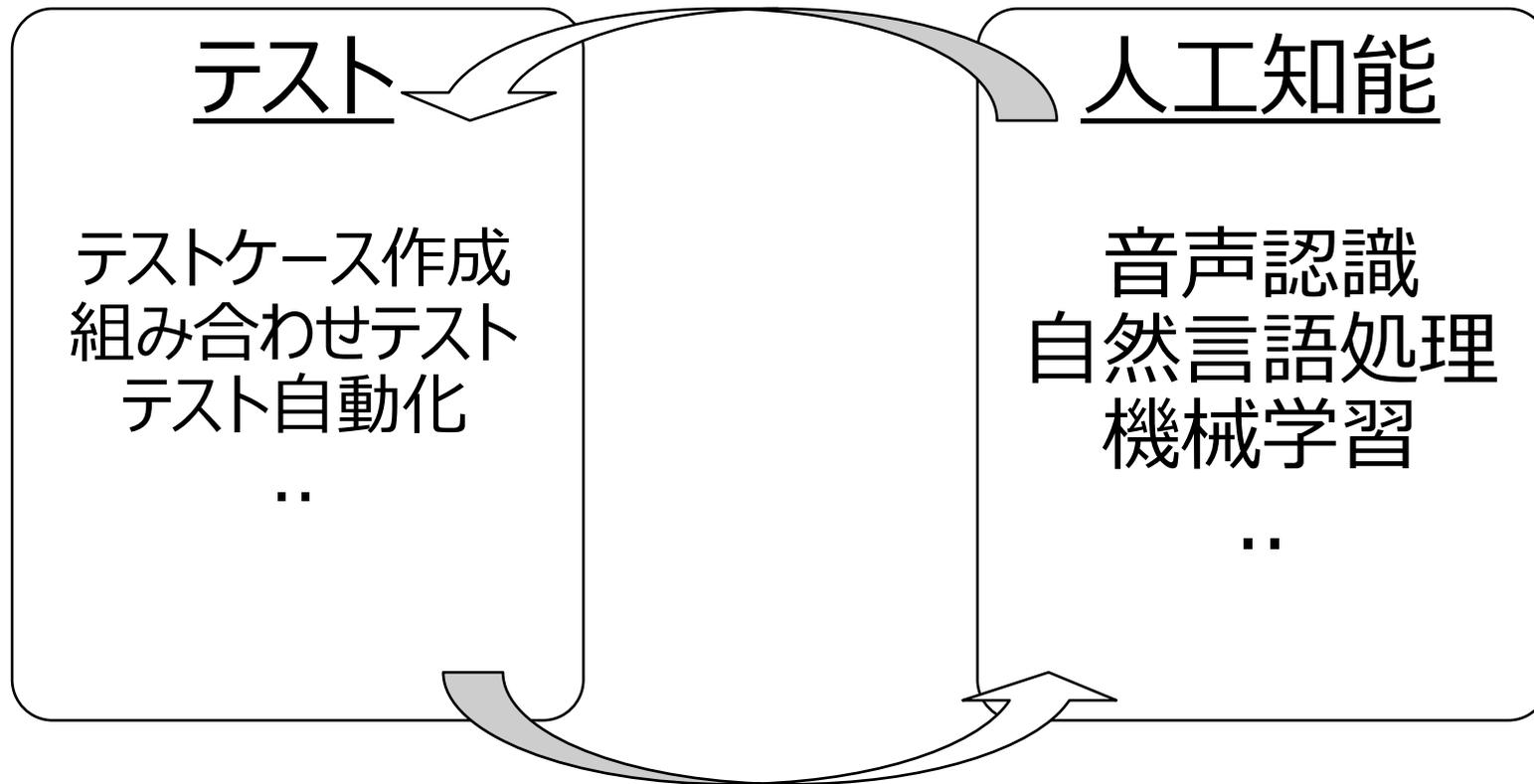
事例の開発モデル



次はこちらの方の話



人工知能の技術をテストに利用する



人工知能をテストする

「テストケース作成自動化のための意味役割付与方法」から

■ ISO/IEC/IEEE 29119-4のDecision Table TestingのStep

1. 仕様を確認する

- 「年齢が20歳以上の場合、入場料1000円を表示する。」

2. Step 2: テスト条件(条件と動作)を抽出する

条件(Conditions)

- TCOND1(C1): 年齢が20歳以上の場合

動作(Actions)

- TCOND4(A1): 入場料1000円を表示する

「テストケース作成自動化のための意味役割付与方法」から

Step 3. 条件と動作の組み合わせを挙げる

		Test Coverage Item	
	Decision Rules	1	2
条件	(C1):年齢が20歳以上の場合	T	F
動作	(A1):入場料1000円を表示する	T	F

Step 4. テストケースを作成する

テスト ケース 番号	入力	出力	Test Coverage Item
	年齢	表示	
1	25歳	入場料1000円を表示する	1
2	10歳	入場料1000円を表示しない	2

「テストケース作成自動化のための意味役割付与方法」から - 実施内容:雛形文体

表 1 有識者判断と構文的類似度の比較

例文
年齢が20歳以上の場合、入場料1000円を表示する。
個数が10個以上の場合、配送料500円を徴収する。
ログオン・ユーザーが対象会社ではない場合、システムは、対象ユーザーではない旨を表示する。
年齢が20歳以上で女性の場合、500円を表示する。
ユーザーは、ログインし、イントラネットIDおよびパスワードを入力する。
納期が3日未満の場合、追加料800円を徴収し、係員が手配をする。
その際に担当が代わる場合は、「担当メンテ」を使用し、発行の担当者名を変更する必要あり。
当日受付確認処理された患者情報一覧を表示すること(主に医師が診察室で使用する)。
時間管理が出来ること(オーダー時間、受付時間、採血時間等)
何もせずに終わる場合→終了。

「テストケース作成自動化のための意味役割付与方法」から

- 前処理：雛形文体知識を利用した構文的類似度比較により文の振り分け
 - 条件と動作が明確に記述された文を雛形文体とする
 - 雛形文体を「年齢が20歳以上の場合、入場料1000円を表示する。」とした
 - 仕様書の文の文体を、雛形文体と比較し条件と動作の抽出に適した文を振り分ける
 - 構文木の類似度比較は木カーネルを用いる

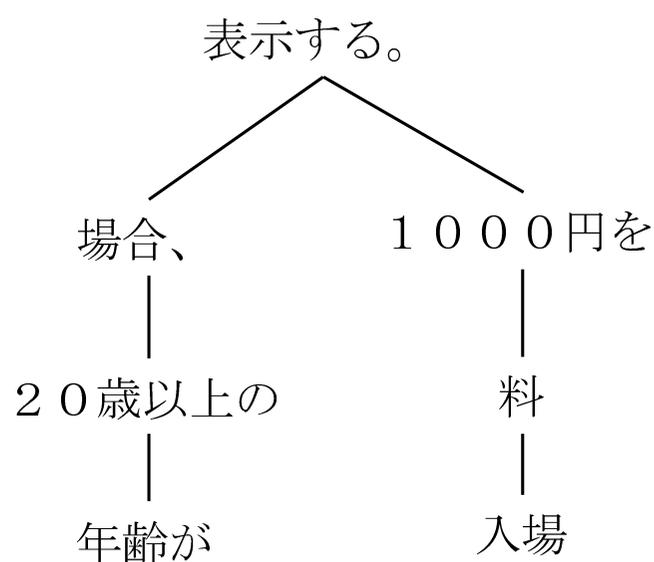


図 2 雛形文体の係り受け構造

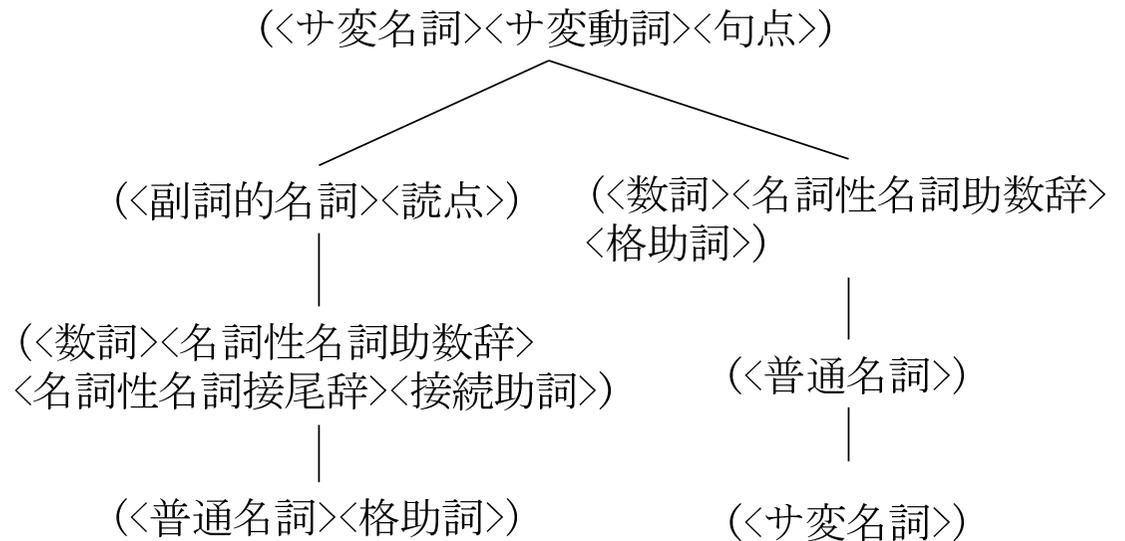


図 3 雛形文体の品詞の構造

「テストケース作成自動化のための意味役割付与方法」から - 実施内容:雛形文体

適と不適の構文的類似度の閾値は一般的[高瀬2014,グエン2011]な0.3とした

表 1 有識者判断と構文的類似度の比較

例文	有識者判断	構文的類似度
年齢が20歳以上の場合、入場料1000円を表示する。	適	1.000
個数が10個以上の場合、配送料500円を徴収する。	適	1.000
ログオン・ユーザーが対象会社ではない場合、システムは、対象ユーザーではない旨を表示する。	適	0.594
年齢が20歳以上で女性の場合、500円を表示する。	適	0.563
ユーザーは、ログインし、イントラネットIDおよびパスワードを入力する。	適	0.500
納期が3日未満の場合、追加料800円を徴収し、係員が手配をする。	適	0.313
その際に担当が代わる場合は、「担当メンテ」を使用し、発行の担当者名を変更する必要あり。	不適	0.292
当日受付確認処理された患者情報一覧を表示すること(主に医師が診察室で使用する)。	不適	0.271
時間管理が出来ること(オーダー時間、受付時間、採血時間等)	不適	0.200
何もせずに終わる場合→終了。	不適	0.125

- 最近の(私の周辺の)事例から「テストと人工知能」に関連する話をします
 - 音声質問応答システムとテスト
 - 「セルフサービスシステムの実現に向けた回答データ作成手法」
 - 「テストケース作成自動化のための意味役割付与方法」
- より多くの参加者に役立つよう、分かりやすい話を目指します
- 人工知能の技術を利用したアプリケーション開発にはテストのスキルが有用だと思っています

The end of presentation