

# 日常生活行動知識グラフと RAG に基づく 家庭内危険行動の理由と根拠提示システム

## A System to Show Reason and Evidence for Dangerous Behaviors in Home Based on Knowledge Graphs of Daily Living Activities and RAG

穴口史将<sup>1</sup> 森田武史<sup>1,2</sup> \*  
Fumikatsu Anaguchi<sup>1</sup> Takeshi Morita<sup>1,2</sup>

<sup>1</sup> 青山学院大学 <sup>2</sup> 産業技術総合研究所  
<sup>1</sup> Aoyama Gakuin University

<sup>2</sup> National Institute of Advanced Industrial Science and Technology

**Abstract:** 本研究では、ナレッジグラフ推論チャレンジ【実社会版 2022】が提供するデータセットを対象として、家庭内危険行動の理由と根拠を提示可能なシステムを提案する。まず、日常生活行動知識グラフから高齢者の家庭内危険行動を抽出する。次に、大規模言語モデル (GPT-3.5) を用いて、その行動が危険である理由を提示する。さらに、高齢者の家庭内事故に関する文献から、RAG を用いてこの理由に類似する文を生成し、検索エンジンの検索結果と共にユーザに根拠として提示する。最後に、根拠オントロジーに基づいて、この根拠の知識グラフを構築する。提案システムが家庭内危険行動の理由を適切に生成できるか、また、この理由に対して提案システムが提示する根拠が適切であるかを評価するために、大学生 15 名に対してアンケートを実施した。アンケートの 5 段階評価の平均値は、それぞれ 3.6 と 2.6 となった。結果として、提案システムは家庭内危険行動の理由に対する一般的な根拠を示すことができた。次に、提案システムが提示する根拠の正確性を評価した。結果として、検索エンジンが家庭内危険行動の理由に類似する箇所を根拠として検索できることを確認した。

## 1 はじめに

近年の大規模言語モデル (LLM: Large Language Model) に基づく文章生成 AI への関心の高まりに伴い、文章生成 AI が普及し様々な社会システムに利用されることが予想される。LLM はプロンプトに対する尤もらしい文章を生成できる。しかし、LLM は深層学習により構築されるブラックボックス型のモデルであるため、根拠となる情報が暗黙的であり、正確性、信頼性、再現性が保証されない欠点がある。文章生成 AI を安全・安心に社会の中で活用していくためには、システムが判断に至った理由とその根拠を、信頼性が高い文献を引用しながら説明する技術が必要になる。

人工知能学会セマンティックウェブとオントロジー研

究会は、説明可能性を有する AI 技術の開発・促進を目的としたコンテスト「ナレッジグラフ推論チャレンジ」を開催している。ナレッジグラフ推論チャレンジ【実社会版 2022】[1] では、高齢者の家庭内の安全を守るという題材のもと開催され、家庭内の危険な行動の検出、理由の説明、解決策の提示をタスクとしていた。

先行研究として、ナレッジグラフ推論チャレンジ【実社会版 2022】で公開された日常生活行動知識グラフ (KG: Knowledge Graph) から検出された家庭内危険行動を文章生成 AI の入力として、その行動が危険な理由を生成する研究 [2] がある。[2] で提案されたシステムは、尤もらしい家庭内危険行動の理由を生成しているが、その理由の正確性は保証されておらず、根拠となる文献などを提示できていないという課題があった。

LLM の学習データに含まれない独自データを利用して、LLM を用いた質問応答や検索を行うための技術として、検索拡張生成 (RAG: Retrieval Augmented Generation) がある。RAG を用いることにより、LLM

\* 連絡先: 青山学院大学理工学部  
〒252-5258 神奈川県相模原市中央区淵野辺 5-10-1  
E-mail: morita@it.aoyama.ac.jp

の欠点を軽減または解消できる可能性がある。

以上より、本研究では、ナレッジグラフ推論チャレンジ【実社会版 2022】が提供する日常生活行動 KG をもとに文章生成 AI が生成した家庭内危険行動の理由に対して、RAG と検索エンジンを用いて信頼性が高い文献を引用しながら、根拠を提示可能なシステムを開発することを目的とする。

## 2 関連研究

VirtualHome2KG[3] は、仮想環境で家庭内を再現した VirtualHome のシミュレーション結果をもとに構築した日常生活行動 KG であり、オープンデータとして公開されている。日常生活行動 KG は、複数のイベントから構成されるアクティビティ（「ソファでくつろぐ」など）とイベントと関連づけられた主体、動作、対象物、場所、時間、状況などをノード、それらの関係をエッジとして定義している。また、対象物や場所を表すノードには、位置座標や状態などが定義されている。VirtualHome2KG は本研究のデータセットとして、提案システムや評価実験に使用する。

「ナレッジグラフ推論チャレンジ【実社会版 2022】」の参考アプローチとして、家庭内危険行動を検出する研究 [4] や、応募作品として、家庭内危険行動の理由を生成する研究 [2] がある。[4] では、VirtualHome2KG を用いて、家庭内事故の報告件数の中で上位二つの高齢者の転倒・転落を対象として、高齢者の家庭内危険行動を推論によって検出する手法を提案している。[2] では、検出された家庭内危険行動を、文章生成 AI が家庭内危険行動の理由を生成する手法を提案している。本研究における家庭内危険行動の検出と理由の生成は、[4] と [2] の研究を参考にした。

外部コーパスを用いた根拠の提示を行う研究 [5] では、全文検索ソフトウェアを用いて、外部コーパスから GPT-3 の生成物に類似するデータを検索し、質問と回答に結びつけることで、自身のハルシネーションを自己検知する手法を提案している。本提案システムにおける根拠の検索に関するモジュール設計は、[5] の研究を参考にした。

知識グラフ (KG: Knowledge Graph) を活用して、RAG の有用性を示す研究が進められている [6, 7]。[6] では、LLM を用いて自然言語による質問文を SPARQL クエリに変換する手法 (Text-to-SPARQL) について調査している。また、RAG と LLM のコンテキスト内学習による Text-to-SPARQL の結果を比較している。[7] では、KG を情報源とした RAG 手法を試行し、その効

果と限界を明らかにしている。

## 3 提案システム

### 3.1 提案システムの概要

提案システムの概要を図 1 に示す。まず、[4] を参考に、VirtualHome2KG から SPARQL クエリを用いて、高齢者の家庭内危険行動を検出し、家庭内危険行動 KG を構築する。次に、[2] を参考に、家庭内危険行動 KG から文章生成 AI を用いて家庭内危険行動の理由を生成する。次に、家庭内危険行動の理由に対して、検索拡張生成 (RAG: Retrieval Augmented Generation) と Elasticsearch を用いて専門文書から類似文を検索し、根拠としてユーザに提示する。次に、家庭内危険行動の理由に対する根拠の妥当性を検証する。最後に、家庭内危険行動 KG と根拠文から根拠 KG を構築する。

### 3.2 SPARQL クエリによる家庭内危険行動の検出

本モジュールでは、[4] を参考に家庭内危険行動の検出を行う。[4] では、VirtualHome2KG を用いて、家庭内事故の報告件数の中で上位二つの高齢者の転倒・転落を対象として、高齢者の家庭内危険行動を推論によって検出する手法を提案している。まず、安全工学の専門家 1 名により、東京消防庁から提供された救急搬送データから、家庭内の転倒事故・転落事故として多い行動を列举し、分類した。これら家庭内危険行動を VirtualHome2KG の要素と対応付けるための代替表現への変更と、Semantic Web Rule Language (SWRL) による推論ルールの定義を行う。すべての転倒・転落の家庭内危険行動のうち、実行できる動作に限りがあるため、本研究では、「段差を上り下りする」、「高い場所にある物に手を伸ばす」、「低い棚から物を取り出す」の三つを使用する。データ数を増加させた場合に、SWRL を用いて現実的な実行時間で家庭内危険行動を推論することは困難であるため、SPARQL クエリを用いて推論を行う。SWRL で定義した三つの推論ルールを CONSTRUCT クエリに変換する。

この三つの家庭内危険行動は、オブジェクトのサイズ (X 座標, Y 座標, Z 座標) のデータを必要とするため、CONSTRUCT クエリにより補完する。「高い場所にある物に手を伸ばす」と「低い棚から物を取り出す」の SPARQL クエリでは、3D Bounding Box でサイズを補完した、エージェントとオブジェクトの中心座標を取得することで、推論ルールにおける条件の絞り込みを

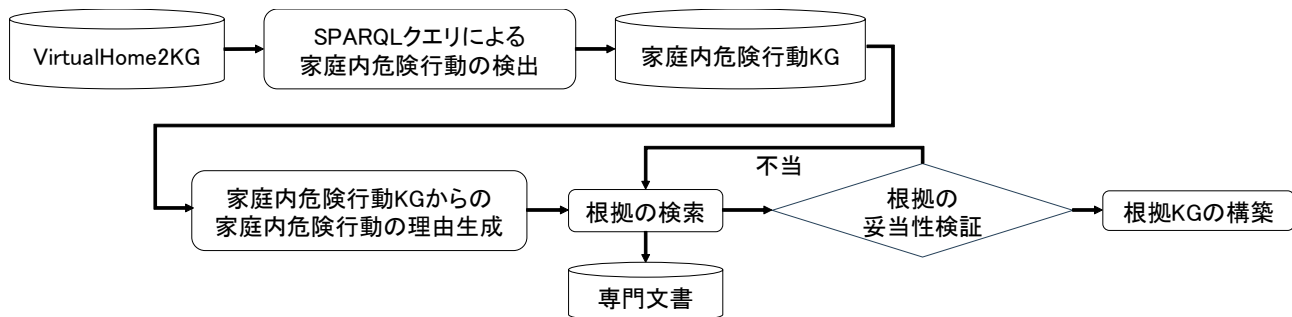


図1 提案システムの概要図

```

CONSTRUCT{
  ?activity hra:riskFactor ?event.
  ?event a hra:ClimbObject.
}
WHERE{
  ?activity :hasEvent ?event.
  ?event :mainObject ?object;
  :action ?action.
  FILTER(?action=ac:climb)
}
  
```

コード1 「階段を上り下りする」の推論ルールに相当する CONSTRUCT クエリ

行っている。具体的には、エージェントとオブジェクトの中心 Y 座標にそれらの Y 方向のサイズの半分の大きさをそれぞれ足すことで比較を行っている。

本研究では、先行研究のクエリを参考に、「段差を上り下りする」の SPARQL クエリを構築した。「階段を上り下りする」の推論ルールに相当する CONSTRUCT クエリをコード 1 に示す。コード 1 では、エージェント（人間）がオブジェクトに「上る」アクションをしたときの、エージェントの動作情報を推論するクエリである。VirtualHome 上には階段がないため、テーブルなど人間が乗れる家具を段差として、推論ルールにおける条件の絞り込みを行っている。具体的には、「テーブルの上に立つ」や「テーブルにのぼる」といったイベントが推論される。これら「高い場所にある物に手を伸ばす」、「低い棚から物を取り出す」、「段差を上り下りする」の三つの家庭内危険行動を検出するための SPARQL クエリを用いて、これらに関する家庭内危険行動 KG を構築する。

### 3.3 家庭内危険行動 KG からの家庭内危険行動の理由生成

本モジュールでは、まず、コード 2 に示す SELECT クエリを用いて、文章生成 AI による家庭内危険行動の理由生成に必要な VirtualHome の環境情報（二つのオブジェクト間の位置関係や、部屋とオブジェクトの包含関係など）を抽出し、LLM を用いて部屋ごとに分類する。コード 2 において環境情報の抽出に利用するプロパティは、“:inside”, “:on”, “:close”, “:facing” である。“:inside” は、オブジェクト 1 がオブジェクト 2 の中に位置していることを意味する。“:on” は、オブジェクト 1 がオブジェクト 2 の上に位置していることを意味する。“:close” は、オブジェクト間の距離が 1.5m 以内であることを意味する。“:facing” は、オブジェクト 1 からオブジェクト 2 が見えることを意味し、オブジェクトの中心間の距離が 5m 以内であることを意味する。

SELECT クエリを用いて抽出された環境情報を、テンプレートを用いて文章化し、VirtualHome の環境内にあるオブジェクトを部屋ごとに分類する。部屋ごとに分類することで、家庭内危険行動の理由を生成する際の入力トークンを削減することができる。テンプレートは、“The オブジェクト 1 is the inside of the オブジェクト 2.” のように “:inside”, “:on”, “:close”, “:facing” に関するオブジェクトの情報を一文に変換する。これらの文章化された環境情報を元に生成した、VirtualHome の環境内にあるオブジェクトを部屋ごとに分類し、日本語で出力するプロンプトの例をコード 3 に示す。これらの環境情報をプロンプトに含めることで、環境に基づいた家庭内危険行動を LLM が推論できる。

次に、家庭内危険行動 KG から SELECT クエリ（コード 4）を用いて、アクティビティ、オブジェクト、アクションを抽出し、環境情報と共に、家庭内危険行動の理

```

SELECT DISTINCT ?object ?inside ?on ?
close ?facing
WHERE{
  ?activity :hasEvent ?event;
            vh2kg:virtualHome ex:scene1
.
  ?event ho:object ?object.
  ?object a ?objectType.
  ?state :isStateOf ?object;
         :bbox ?shape.
  {?shape vh2kg:inside ?inside.
  }UNION{
    ?shape vh2kg:on ?on.
  }UNION{
    ?shape vh2kg:close ?close.
  }UNION{
    ?shape vh2kg:facing ?facing.
  }
}

```

コード 2 環境情報を抽出する SELECT クエリ

```

The bed is the inside of the bedroom.
The bed is the inside of the closet.
. . .
# ここまで環境情報

ある家庭に関する情報を提供します。
家具などを部屋ごとに分類してください。
部屋はリビングルーム、キッチンルーム、バス
ルーム、ベッドルームの4つです。
出力はすべて日本語で行ってください。

```

コード 3 VirtualHome の環境内にあるオブジェクトを部屋ごとに分類し、日本語で出力するプロンプトの例

由を生成するプロンプトテンプレート（コード 5）に組み込んで、プロンプトを生成する。

コード 5 の変数 “{家庭内危険行動}” に、家庭内危険行動 KG から抽出したアクティビティ、オブジェクト、アクションが代入される。また、Few-Shot 学習を用いて、家庭内危険行動の理由を一つ出力するように指定する。

最後に、このプロンプトをもとに文章生成 AI が転倒・転落に関する家庭内危険行動の理由を生成する。

### 3.4 根拠の検索

本モジュールでは、3.3 節で述べた手法により生成された家庭内危険行動の理由に対して、RAG と Elasticsearch を用いて、その根拠を検索する。

```

SELECT DISTINCT ?activity ?object ?action
WHERE{
  ?activity :hasEvent ?event.
  ?event ho:object ?object;
         :action ?action.
}

```

コード 4 家庭内危険行動 KG からアクティビティ、オブジェクト、アクションを抽出する SELECT クエリ

まず、OpenAI 社の OpenAI Embeddings<sup>\*1</sup>を用いて、PDF ファイル形式の専門文書をベクトル化し、専門文書ベクトル DB に保存する。次に、LangChain の RetrievalQA を用いて、専門文書ベクトル DB から家庭内危険行動の理由に最も類似する記述を根拠として専門文書中から出力するように、文章生成 AI にプロンプトを与える。類似する文が存在しない場合を想定し、類似する文が無い場合は “No Data” と出力するようにプロンプトに指示を与える。

根拠文を検索するプロンプトテンプレートをコード 6 に示す。コード 6 の変数 “{家庭内危険行動の理由}” に、3.3 節で述べた手法により生成された家庭内危険行動の理由を代入し、プロンプトを生成する。

本研究で用いる専門文書のデータ量は比較的軽量かつ、複雑な処理を行う必要はないため、LangChain の RetrievalQA におけるパラメータのうち、chain\_type には、ベクトル DB を単純に検索する “stuff” を使用する。

最後に、Elasticsearch を用いて根拠となる記述を専門文書から全文検索し、検索結果とそのメタデータを、家庭内危険行動の理由とその根拠と共に出力する。提案システムにおける Elasticsearch のインデックスは、根拠文、専門文書のタイトル、発行日、著者、ページである。データ型はそれぞれ、“text”、“text”、“text”、“text”、“integer” である。発行日には、年月日などの文字もデータに含めるため、テキスト型として作成した。インデックスは専門文書ごとに作成する。インデックスの構成要素である、ドキュメント（レコード）は文ごとに作成する。これによって一文ごとにメタデータを付与することができる。

家庭内危険行動の理由に対する根拠の検索結果例をコード 7 に示す。コード 7 の「危険な理由」は、文章生成 AI が生成した家庭内危険行動の理由である。「根拠文」は、LangChain の RetrievalQA を用いて専門文書

<sup>\*1</sup> <https://platform.openai.com/docs/guides/embeddings>

家庭内の環境情報は以下になります。

この家には、キッチン・リビング・寝室・バスルームの4つの部屋があります。それぞれの部屋ごとに、家具などの情報を提供します。

キッチン：  
お酒、アルコール、・・・、ジュース、ワインがあります。

リビング：  
・・・

# ここまで環境情報

この家庭内で起きる「状況、対象物、行動」を与えます。この条件下で高齢者が家庭内で起こりうる、転倒・転落・衝突に関する危険を一つ教えてください。

「{家庭内危険行動}」

出力テンプレートは以下のようになります。

状況：  
対象物：  
行動：  
高齢者の家庭内危険行動：

出力方法の例を示します。

例1)

「キッチンの掃除、冷蔵庫、開ける」

状況：キッチンの掃除

対象物：冷蔵庫

行動：開ける

高齢者の家庭内危険行動：冷蔵庫のドアを開けた際に、中の物が出てきて足元に落ち、転倒するから

例2)

「きれいなソファでくつろぐ、枕、つかむ」

状況：きれいなソファでくつろぐ

対象物：枕

行動：つかむ

高齢者の家庭内危険行動：枕をつかもうとしてソファから落ちる可能性があるから

コード5 家庭内危険行動の理由を生成するプロンプトテンプレート

「{家庭内危険行動の理由}」

に最も関連する記述を1文で文書中から抜き出してください

最も関連する記述のみ出力してください。文書中にない記述を出力するのはやめてください

記述が無い場合は、'No Data'と出力してください

コード6 根拠文を検索するプロンプトテンプレート

ベクトルDBから検索した類似文である。「原文」、「タイトル」、「発行日」、「著者」、「ページ」は、Elasticsearchの検索結果である。

危険な理由：テーブルの上に乗っている際にバランスを崩し、転倒してしまう可能性があるから

根拠文：階段などの段差でつまずく、足がもつれて家具にぶつかる、ベッドから降りるときに転落する、靴下が引っかかって転落する、バスマットやじゅうたん、毛布などに足をとられて転倒するなど、ちょっとしたことが転落・転倒の原因になっている。

原文：階段などの段差でつまずく、足がもつれて家具にぶつかる、ベッドから降りるときに転落する、靴下が引っかかって転落する、バスマットやじゅうたん、毛布などに足をとられて転倒するなど、ちょっとしたことが転落・転倒の原因になっている

タイトル：医療機関ネットワークか事業からみた家庭内事故—高齢者編—

発行日：平成25年3月28日

著者：独立行政法人国民生活センター

ページ：3

コード7 家庭内危険行動の理由に対する根拠の検索結果例

### 3.5 根拠の妥当性検証

3.4節で述べた手法により検索された根拠文が、家庭内危険行動の理由に対する根拠として妥当であるかを検証する。検証後に根拠が不当であると判断された場合、再検索を行い、妥当な根拠を検索する。根拠の妥当性の検証を自動的に行うことは難しいため、本研究では手動で検証する。

### 3.6 根拠 KG の構築

3.5節で述べた手法より妥当性が認められた根拠をKG化する。根拠オントロジーにおけるクラスとプロパティの関係を図2に示す。また、接頭辞と名前空間を表1に示す。根拠オントロジーに基づいて、家庭内危険行動KGを拡張し、根拠KGを構築する。根拠KGは、家庭内危険行動の“Activity”の下位クラスである、“Event”を中心に構築している。“Event”は、“Activity”の一場面を定義している。“Event”は、家庭内危険行動の理由を示す“Cause”を定義している。RAGの検索結果である根拠文は、“RiskEvidence”として“Cause”に結びついている。“RiskEvidence”には、文書の各文、文書タイトル、発行日、著者、ページといったメタデータがクラスとして定義されている。

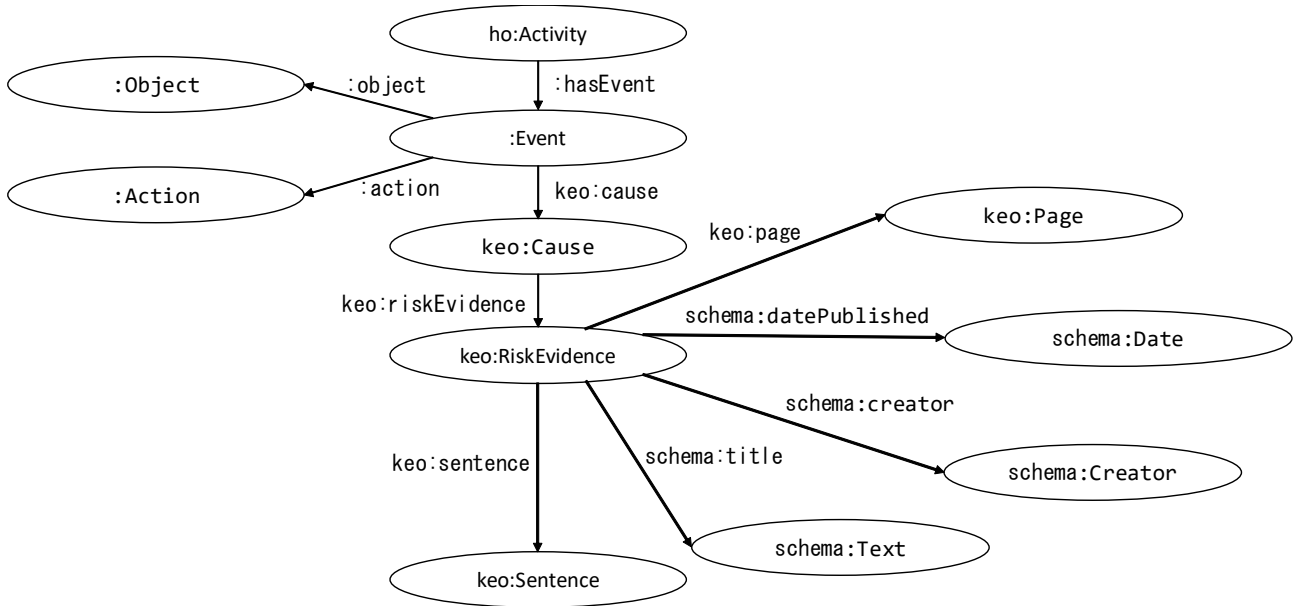


図2 根拠オントロジーにおけるクラスとプロパティの関係

表1 接頭辞と名前空間

接頭辞	名前空間
:	http://kgrc4si.home.kg/virtualhome2kg/ontology/
ho	http://www.owl-ontologies.com/VirtualHome.owl#
keo	http://ke.it.aoyama.ac.jp/ontology/
schema	http://schema.org/

## 4 評価実験

### 4.1 評価実験概要

本評価実験では、文章生成 AI が家庭内危険行動の理由および理由に対する根拠文を適切に生成できるかを被験者アンケートにより評価する。また、RAG と Elasticsearch が、それぞれ専門文書の文とメタデータを正確に出力できているかを評価する。

VirtualHome には、七つの家庭環境がある。環境によって存在する部屋や物、また、それらの位置などが異なる。本研究では、Scene1 のみを対象として評価実験を行う。Scene1 には、キッチン、リビング、寝室、バスルームの四部屋があり、テーブルや椅子などの生活に必要な基本的な家具が配置されている。

表2 評価用データセットから抽出した家庭内危険行動の三つ組

識別子	アクティビティ	オブジェクト	アクション
R1	テーブルの上に立つ	テーブル	のぼる
R2	歯磨きの準備	歯ブラシ	掴む
R3	冷蔵庫に食料品をしまう	冷蔵庫	開ける
R4	クローゼットの整理整頓	シャツ	置く
R5	ソファでくつろぐ	ソファ	座る
R6	リビングルームの掃除	テレビ	拭く
R7	テレビを見ながらジュースを飲む	テレビ	スイッチオン
R8	寝る	ライトスイッチ	スイッチオフ
R9	コンロを使ってジャガイモを調理する	フライパン	油を注ぐ

### 4.2 評価用データセット

評価用データセットには、VirtualHome2KG から 6 SPARQL クエリにより検出した高齢者の家庭内危険行

表3 評価用データセットから抽出した家庭内危険行動の説明文

識別子	家庭内危険行動の説明文
R1	テーブルの上に立つためにテーブルにのぼる
R2	歯磨きの準備をするために歯ブラシを掴む
R3	冷蔵庫に食料品をしまうために冷蔵庫を開ける
R4	クローゼットの整理整頓のためにシャツを置く
R5	ソファでくつろぐためにソファに座る
R6	リビングルームの掃除のためにテレビを拭く
R7	テレビを見ながらジュースを飲むためにテレビのスイッチを押す
R8	寝るためにライトのスイッチを押す
R9	コンロを使ってジャガイモを調理するためにフライパンに油を注ぐ

表4 評価実験1の結果

家庭内危険行動	平均	標準偏差
R1	4.1	1.1
R2	3.6	1.4
R3	3.0	1.6
R4	3.1	1.7
R5	3.5	1.5
R6	3.5	1.4
R7	3.1	1.5
R8	4.3	1.1
R9	4.1	1.1
全体	3.6	1.4

表5 評価実験2の結果

家庭内危険行動	平均	標準偏差
R1	2.9	1.7
R2	2.8	1.7
R3	1.6	1.0
R4	2.5	1.6
R5	2.5	1.5
R6	2.6	1.5
R7	3.2	1.6
R8	3.5	1.4
R9	2.1	1.3
全体	2.6	1.6

動を使用する。部屋数とアクション数が網羅的になるように、九つの家庭内危険行動を選択した。

表2に評価用データセットから抽出した家庭内危険行動の三つ組(アクティビティ, オブジェクト, アクション)を示す。また, アンケート用に, 表2に示す三つ組を「アクティビティのために, オブジェクトにアクションする」といった説明文に手作業で変換した結果を表3に示す。予備実験において, 家庭内危険行動の三つ組を

被験者に提示したところ, 被験者に家庭内危険行動の意図が伝わらなかったため, 家庭内危険行動の三つ組の代わりに説明文を被験者に提示することにした。表2と表3の1列目には, 家庭内危険行動の識別子として, R1からR9を付与する。評価実験結果では, 1列目の識別子を用いて説明する。

### 4.3 評価方法

提案システムを用いて, 評価用データセットから抽出した家庭内危険行動に対する理由を生成し, 家庭内危険行動の理由に対する根拠文を専門文書から検索する。使用するLLMは, 「gpt-3.5-turbo-1106」である。また, 本評価実験では, 専門文書中に根拠文が存在することを前提とする。これらを踏まえ, 以下の三つの評価実験を行う。

**評価実験1** 「家庭内危険行動」に対して「危険な理由」は適切であるか

**評価実験2** 「危険な理由」に対して「根拠文」は適切であるか

**評価実験3** 専門文書から検索された根拠文は正確であるか

評価実験1と評価実験2に関しては, 被験者アンケートを行う。被験者は大学生15人である。被験者の認知バイアスによる影響を抑えるために, 被験者には使用したLLMを伝えず, 質疑応答システムの返答に関して評価するように伝える。また, 被験者の主観により評価が揺れないように, 評価の基準を被験者に提示した。評価実験1の基準は, 「“指定した家庭内危険行動について, 日常生活で想定される危険な理由”を適切とする」とする。評価実験2の基準は, 「“文の種類が異なっても, 根拠文から行動が危険である理由を推測できる文”を適切とする」とする。

評価指標には, リッカート尺度に基づく5段階評価を使用する。具体的には, 適切, やや適切, どちらとも言えない, やや不適切, 不適切のうち, 最も当てはまる評価値を被験者に回答してもらう。すべての設問に自由記述欄を設けたため, 回答に至った経緯を収集できる。

評価実験1では九つの家庭内危険行動に対して危険な理由が適切であるかの評価値を回答してもらう。評価実験2では四つの専門文書それぞれについて, 危険な理由に対する根拠文が適切であるかの評価値を回答してもらう。評価実験1と評価実験2では, 適切から不適切の評価値を5から1の数値に置き換え, それぞれの家庭内危険行動ごとに被験者による評価の平均値と標準偏差を小数第一位まで求める。

表6 評価実験2の専門文書別の結果

家庭内 危険行動	専門文書1		専門文書2		専門文書3		専門文書4	
	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
R1	3.3	1.6	2.3	1.6	2.1	1.6	3.8	1.3
R2	3.5	1.6	1.6	1.1	3.0	1.9	2.9	1.7
R3	1.7	1.1	1.8	1.2	1.2	0.6	1.9	1.1
R4	3.5	1.6	1.3	0.8	1.9	1.3	3.4	1.5
R5	3.5	1.4	2.2	1.3	2.2	1.6	2.2	1.6
R6	2.5	1.4	1.7	1.2	3.9	1.1	2.3	1.5
R7	3.5	1.6	1.5	0.8	4.1	1.3	3.5	1.6
R8	3.9	1.2	3.1	1.6	3.7	1.5	3.5	1.5
R9	1.6	1.1	1.8	1.0	2.9	1.4	2.2	1.5
全体	3.0	1.6	1.9	1.3	2.8	1.7	2.9	1.6

表7 危険な理由と適切な根拠文

家庭内危険行動	危険な理由	根拠文	平均
R1	テーブルの上に立っている際にバランスを崩し、転倒してしまう可能性があるから	転倒は、つまずき、滑り、踏み外しといった足元のトラブルで発生しているケースが7割を超えています。	3.8
R2	歯ブラシを掴もうとしてバスルームのシンクに手をついて滑り、転倒する可能性があるから	風呂場での転倒、やけど風呂場は転倒のほか、浴槽で熱い湯に漬かってしまったことによるやけど、溺でき水すい等も起こりやすい場所である	3.5
R3	冷蔵庫のドアを開けた際に、中の物が出てきて足元に落ち、転倒する可能性があるから	床面に置きっぱなしになっている衣類や紙類で滑るというケースも少なくありません。	1.9
R4	シャツをクローゼットに置こうとしてバランスを崩し、転倒する可能性があるから	居間での電球の交換時や高い位置にある押し入れの物を取ろうとしたときのような生活空間での事故が多いのです。	3.5
R5	ソファに座る際にバランスを崩して転倒する可能性があるから	階段等でつまづく、足がもつれて家具にぶつかる、ベッドから降りるときに転落する、靴下が引っかけた転落する、バスマットやじゅうたん、毛布などに足をとられて転倒するなど、ちょっとしたことが転落・転倒の原因になっている。	3.5
R6	テレビを拭こうとしてバランスを崩し、転倒する可能性があるから	■滑る ■つまずく ■ぐらつく ■ベッド等から移動時に ■引っ掛かる加齢に伴って、日常生活の中にも転倒事故のきっかけとなる危険性が高まってきます。	3.9
R7	テレビのスイッチをオンにする際に、コードや家具につまづいて転倒する可能性があるから	電源コードに足を取られたり、カーペットやこたつ、座布団に引っ掛かったりして、転倒に至る事故が多く見られます。	4.1
R8	ライトスイッチを探そうとする際、家具につまづいて転倒する可能性があるから	階段などの段差でつまづく、足がもつれて家具にぶつかる、ベッドから降りるときに転落する、靴下が引っかけた転落する、バスマットやじゅうたん、毛布などに足をとられて転倒するなど、ちょっとしたことが転落・転倒の原因になっている。	3.9
R9	油を注ぐ際に手が滑り、油がこぼれて床が滑りやすくなり、転倒する可能性があるから	事故の状況が分かっている218件を分類すると、「滑る」43件、「つまずく」43件が共に多く	2.9



表8 評価実験3の結果

家庭内危険行動	専門文書 1				専門文書 2				専門文書 3				専門文書 4			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
R1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R2		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
R4		✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R5		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R6		✓		✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
R7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R8	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R9		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓

評価実験3に関しては、客観的に判断できる基準を決めて筆者が評価する。以下に基準となる条件を示す。

- 条件1 専門文書中の記述を一言一句変わらず、一文で出力できている
- 条件2 高齢者の転倒・転落に関する家庭内危険行動の理由の根拠を出力できている
- 条件3 性別の指定や病気・怪我を患っている人など、一部の高齢者に関する情報ではない
- 条件4 Elasticsearchの結果がLangChainの結果のメタデータとして正しい

条件1に関して、箇条書きの行頭文字や段落番号は、欠損していても条件を満たしていると定義する。条件4に関して、メタデータであるページ数は、文書中に記載されたページ数ではなく、PDFファイルが示すページ数と定義する。そのため、メタデータのページ数と文書中の実際のページ数が異なる場合がある。

#### 4.4 評価実験結果

評価実験1の結果を表4に示す。評価実験1の全体の平均値は「3.6」となり、「どちらとも言えない」と「やや適切」の中間に位置する結果となった。次に、自由記述欄についてまとめる。適切・やや適切な理由として、「高齢者の身体的・精神的能力の低下を踏まえると危険である」や「行動が危険であることが容易に想定される、事故が起こる可能性がある」などが挙げられていた。不適切・やや不適切の理由として、「こじつけや非現実的なことであると感じる」や「理由から行動が発生する状況が想定しにくい」などが挙げられていた。

評価実験2の結果を表5に示す。評価実験2の全体の平均値は「2.6」となり、「やや不適切」と「どちらとも言えない」の中間に位置する結果となった。次に、自由記述欄についてまとめる。適切・やや適切な理由として、

「行動などキーワードが共通している」や「根拠文の内容は対策であるが、根拠文から行動が危険な理由を推測できる」や「多少の違いはあるが、的外れな内容ではない」などが挙げられていた。不適切・やや不適切の理由として、「全く関連がない」や「抽象的な内容であるため、ほとんどの家庭内危険行動に当てはまる」や「着目点の違いすぎて、的外れであると感じる」などが挙げられていた。

また、評価実験2の専門文書別の結果を表6に示す。表6より、四つの専門文書のうち最もスコアが高い根拠文を、危険な理由の適切な根拠文として選定する。その結果を表7に示す。また、そのスコアはスコアは2.9～4.1の間に収まる。

評価実験3の結果を表8に示す。表8では、条件1から条件4をC1からC4で表記する。条件を満たす項目にそれぞれチェックマークを付けた。四つすべての条件を満たす根拠文を正確と定義する。正確な根拠文は、24/36(66.7%)となった。四つの条件のうち、最も条件を満たさなかったのは条件1である。この原因として、平仮名を漢字に変換してしまうこと、文書中の振り仮名を出力してしまうこと、接続詞を欠損してしまうこと、半角空白を数字の間に表示してしまうことが挙げられる。次に多かった条件3では、文の一部を出力したことで、一部の高齢者のみに偏った内容を、すべての高齢者に当てはまる内容であるように出力していたことが原因として挙げられる。ここでは、性別を指定した転落事故の事例を出力していた。条件2と条件4に関しては、すべての根拠文において条件を満たすという結果になった。

#### 4.5 考察

評価実験1の結果より、被験者が家庭内危険行動に対して危険な理由は適切であると判断するには限界があ

るという考察が得られた。一部の家庭内危険行動では、こじつけや非現実的であるといったコメントも見られた。また、危険な理由が適切であると判断するには、現状の説明方法では難しいため、改善する必要があると考える。具体的には、詳細に危険な理由を生成するように文章生成 AI に指示することや、環境情報に対してオブジェクト間の位置関係や状況などを詳細にプロンプトに含めることで、LLM がより具体的に危険な理由を生成できるようにすることが考えられる。

評価実験 2 の結果より、提案システムが提示した根拠文は、家庭内危険行動が危険であることの信頼性を示す、一般的な根拠文であり、根拠として扱うには具体性に欠ける内容も多いという考察が得られた。一般的な根拠文には、足元のトラブルやバランス能力の低下などによる転倒の危険性が記述されていた。また、専門文書によって精度が変わるため、複数の専門文書で根拠付けを行い、適切さを人間が行うのであれば、文章生成 AI の出力に根拠文を付与することも可能であると考えられる。

評価実験 3 の結果より、専門文書中の記述を一言一句正確に出力できないものの、欠損箇所は接続詞などが多く、意味的には問題ない文が多かった。意味的な欠損では、性別を指定した転落事故の事例をすべての高齢者に当てはまる内容であるように出力していた。専門文書中の記述を一言一句変わらず出力することは難しいが、メタデータの出力に関しては間違いがなく、付与されたメタデータは信頼できると考えられる。

## 5 おわりに

本研究では、ナレッジグラフ推論チャレンジ【実社会版 2022】が提供する日常生活行動知識グラフを元に GPT-3.5 が生成した家庭内危険行動の理由に対して、RAG と Elasticsearch を用いて信頼性が高い文献を引用しながら、根拠を提示可能なシステムを提案した。

評価実験では、提案システムが家庭内危険行動の理由を適切に生成できるか、また、この理由に対して提案システムが提示する根拠が適切であるかを評価するために、大学生 15 名に対してアンケートを実施した。アンケートの 5 段階評価の平均値は、それぞれ 3.6 と 2.6 となった。結果として、提案システムは家庭内危険行動の理由に対する一般的な根拠を示すことができた。次に、提案システムが提示する根拠の正確性を評価した。結果として、検索エンジンが家庭内危険行動の理由に類似する箇所を根拠として検索できることを確認した。

本研究の今後の課題として、様々な形式で記述された専門文書や知識グラフを根拠の知識源とすることや、他

の LLM との比較が挙げられる。また、根拠 KG の応用についても検討する。

## 謝辞

本研究は JSPS 科研費 23K11221 の助成を受けたものです。本研究成果の一部は、国立研究開発法人新エネルギー・産業 技術総合開発機構 (NEDO) の委託業務 (JPNP20006) の結果得られたものです。

## 参考文献

- [1] 鶴飼孝典, 江上周作, 大野美喜子, 窪田文也, 福田賢一郎, 川村隆浩, 古崎晃司, 松下京群. 高齢者の家庭内事故予防に役立つ AI システムの開発 一産業版ナレッジグラフ推論チャレンジに向けて一. 人工知能学会第二種研究会資料, SIG-SWO-056-15, 2022.
- [2] 浅野歴, 濱道光希, 牟耕, 森俊人, 矢野翔平, 森田武史. 知識グラフと GPT を用いた家庭内の危険行動の検知と説明. 2023 年度人工知能学会全国大会 (第 37 回), 3G1-OS-24a-03, 2023.
- [3] Shusaku Egami, Takanori Ugai, Mikiko Oono, Koji Kitamura, and Ken Fukuda. Synthesizing Event-Centric Knowledge Graphs of Daily Activities Using Virtual Space. *IEEE Access*, Vol. 11, pp. 23857–23873, 2023.
- [4] 江上周作, 鶴飼孝典, 窪田文也, 大野美喜子, 北村光司, 福田賢一郎. 家庭内の事故予防に向けた合成ナレッジグラフの構築と推論. 人工知能学会第二種研究会資料, SIG-SWO-056-14, 2022.
- [5] Siqing Huo, Negar Arabzadeh, and Charles Clarke. Retrieving Supporting Evidence for Generative Question Answering. In *SIGIR-AP*, 2023.
- [6] 江上周作, 福田賢一郎. 大規模言語モデルを用いた SPARQL クエリ生成の予備の実験. 人工知能学会第二種研究会資料, SIG-SWO-060-04, 2023.
- [7] 松澤有三, 高橋陽一. ナレッジグラフを用いた RAG 手法の検証. 人工知能学会第二種研究会資料, SIG-SWO-061-03, 2023.