物流安全分析における生成 AI 出力のオントロジーによる構造化と有効性検証

後藤 大征　鈴木 官優　峰 和仁　アッタウィール アハマド
ユン ジャンソク　リー ボヨン　アルマスリ アハメド　我妻 広明

物流現場では，人・車両・設備が混在する複雑な作業環境において，安全性の向上が重要な課題となっている．生成 AI は物流現場の映像から危険事象を柔軟に抽出できる一方，出力語彙のばらつきや法令との対応関係の曖昧さが課題となる．本研究では，生成 AI による映像分析結果を LHR-SSM に基づくオントロジーへ接続し，危険事象，管理失敗および法令関連候補を構造化する手法を提案する．5 本の動画を対象に生成 AI 単独方式とオントロジー介在方式を比較した結果，オントロジー介在方式は SAPA 対応で 100%，推論経路提示で 91%となり，生成 AI 単独方式の 33%，85%を上回った．一方，語彙の適切性は 73%，ManagementFailure の適切性は 87%であり，生成 AI 単独方式の 92%，92%を下回った．この結果から，オントロジー介在は法令候補の一貫した提示に有効である一方，語彙の被覆性が不足する場合には誤分類を誘発する可能性があることを示した．

1　はじめに

物流現場では，作業者，車両，設備および荷物が同一空間で相互に関係するため，単一の物体や行動の検出だけでは安全上の問題を十分に把握できない．例えば，フォークリフトと作業者の接近を評価するには，両者の位置関係に加え，通路分離，作業手順，教育および管理体制を考慮する必要がある．したがって，物流現場の映像分析には，危険場面の検出だけでなく，その背景にある管理上の問題を整理する機能が求められる．

近年，マルチモーダル大規模言語モデルにより，専用の認識モデルを用意せずに映像から多様な危険場面を自然言語で抽出できるようになった．一方，これらのモデルには，視覚情報に存在しない物体や行動をもっともらしく生成するハルシネーション（hallucination）が報告されている [1]，[2]，[3]．また，同一または類似した状況を異なる語彙で記述する表現揺れも生じるため，複数動画を横断した検索・集計や安全知識との対応付けが困難となる．

特に韓国では，重大災害処罰法（Serious Accidents Punishment Act: SAPA）[4] を背景として，現場の危険事象と事業主・経営責任者による安全保健確保措置との関係を整理する必要がある．ただし，映像のみから実際の法令違反を確定することはできないため，本研究では法的判断ではなく，安全管理者が確認すべき法令関連候補の提示を目的とする．

本研究では，物流現場における危険事象，管理失敗（ManagementFailure）および法令概念を機械可読に表現するオントロジー，物流施設人的資源安全管理標準モデル（Logistics Human Resources Safety Standard Model: LHR-SSM）を構築し，生成 AI による映像分析結果をこれに接続することで，危険事

Ontology-Based Structuring and Evaluation of Generative AI Outputs for Logistics Safety Analysis

Taisei Goto(*1), Kanu Suzuki(*1), Kazuhito Mine(*1), Ahmad Altaweel(*1),
Jang-Sok Yoon(*2), Boyoung Lee(*2), Ahmed Almassri(*3), and Hiroaki Wagatsuma(*1), *1: 九州工業大学 大学院 生命体工学研究科, *2:Logistics Revolution Korea Co. Ltd., *3: 富山県立大学情報工学部知能ロボット工学科, *1: Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, *2: Logistics Revolution Korea Co. Ltd., *3: Department of Intelligent Robotics, Faculty of Information Engineering, Toyama Prefectural University.

象，管理失敗および SAPA 関連候補を構造化する手法を提案する．具体的には，生成 AI が抽出した時間区間，危険事象，作業区域，重大度および管理失敗候補を OWL 個体として表現し，SWRL ルールを用いて，危険事象と管理失敗との関係，重大事故候補および SAPA 関連候補を導出する．さらに，構築したオントロジーが生成 AI 出力の構造化にどの程度有効であるかを，生成 AI 単独方式との比較によって評価する．

本研究では，以下の研究課題を設定する．

RQ1 オントロジーへの接続により，危険事象を標準語彙で表現できるか．

RQ2 管理失敗および法令関連候補を，機械可読な関係として構造化できるか．

RQ3 導出された結果について，入力事実と適用規則を追跡できるか．

RQ4 同一動画を繰り返し解析した場合，出力の安定性はどのように変化するか．

なお，本研究は危険場面の検出精度ではなく，生成 AI 出力を検索・集計・再利用・追跡可能な安全知識へ変換することを評価対象とする．

2 関連研究

マルチモーダル大規模言語モデルには，視覚情報に基づかない内容を生成するハルシネーションが報告されている [1] [2] [3]．このため，安全管理への利用では，出力を後から確認・修正可能な形式で保持する必要がある．

安全分野では，OWL と SWRL による安全確認支援 [5]，安全知識と BIM の統合 [6]，視覚認識結果とオントロジーまたは知識グラフを接続した危険推論 [7] [8] が提案されている．これらに対し本研究は，生成 AI による自由度の高い映像分析結果を物流安全の標準語彙，管理失敗および法令概念へ接続し，生成 AI 単独方式と比較評価する点に特徴がある．

3 提案手法

3.1 全体構成

提案手法のシステムフローを図 1 に示す．両方式は同一の物流現場映像を Gemini で解析するが，生成 AI 単独方式は結果を自然言語で出力する．一方，オントロジー介在方式は構造化 JSON を OWL 個体へ変換し，LHR-SSM，SWRL および Pellet によって標準語彙，管理失敗，SAPA 候補および推論経路を導出する．

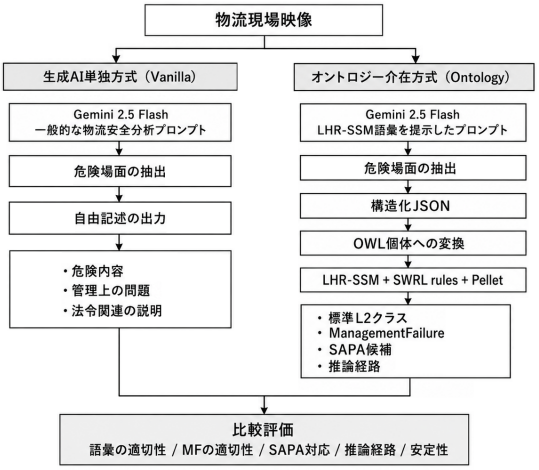

図 1 提案するシステムフロー

3.2 危険場面抽出と OWL 変換

映像解析には Gemini 2.5 Flash を用い，時間区間，状況説明，L2 危険事象クラス，作業区域，重大度，管理失敗候補および映像上の根拠を JSON 形式で出力させた．オントロジー介在方式では，LHR-SSM に定義された 10 個の上位カテゴリ，56 個の L2 クラス，7 区域および 6 個の ManagementFailure クラスを提示し，標準語彙から選択するよう制約した．

各危険事象は OWL の NamedIndividual へ変換し，クラス，区域，重大度，時間区間，場面説明および ManagementFailure との関係を記録する．未定義のクラス名は基底クラス `HazardEvent` へ対応付け，元の名称を `rdfs:comment` として保持する．比較対象の生成 AI 単独方式には標準クラス名，Management-Failure 名および SAPA 条文番号を提示せず，一般的な物流安全の観点のみを与えた．

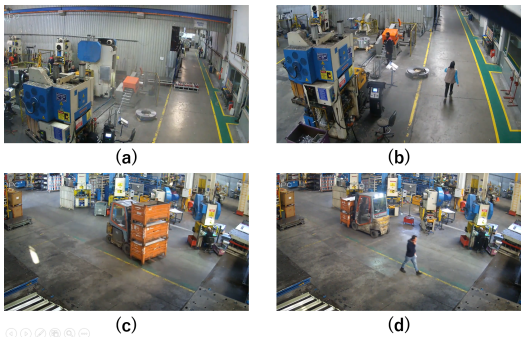

図 2　評価に用いた物流現場映像の例

4　オントロジーおよび推論方法

4.1　LHR-SSM による知識表現

LHR-SSM は，物流現場の安全概念と関係を機械可読に表現するオントロジーである．本研究では，危険事象を `HazardEvent`，その発生・重大化に寄与する管理上の不備を `ManagementFailure` として分離し，作業区域，重大度および法令概念と関連付ける．この分離により，「映像上で何が起きているか」と「どの管理上の問題が関係するか」を異なる論理レベルで記述し，クラス階層とプロパティに基づく検索・集計を可能にする．ただし，入力されたクラスの妥当性をオントロジー自体が保証するものではない．

4.2　SWRL による関係推論

本実験では，LHR-SSM に定義した R1，R2 および R3 の 3 つの SWRL 規則を使用する．

R1 は，重大度が 4 である危険事象を，重大事故候補 `SeriousAccidentCase` として分類する規則である．

$$\begin{aligned}&\mathrm{HazardEvent}(h) \wedge \mathrm{hasSeverity}(h, 4)\\ &\rightarrow \mathrm{SeriousAccidentCase}(h)\end{aligned} \tag{1}$$

R2 は，危険事象から管理失敗への関係が明示されている場合に，その逆方向となる管理失敗から危険事象への関係を導出する規則である．

$$\mathrm{hasContributingFailure}(h, m) \rightarrow \mathrm{triggers}(m, h) \tag{2}$$

ここで，h は危険事象，m は管理失敗を表す．R2 によって，ある危険事象に関連する管理失敗を確認できるだけでなく，特定の管理失敗がどの危険事象に関与したかを逆方向に検索できる．

R3 は，重大度が 3 以上である危険事象を，SAPA 第 4 条に関する確認候補へ対応付ける規則である．

$$\begin{aligned}&\mathrm{HazardEvent}(h) \wedge \mathrm{hasSeverity}(h, s)\\ &\wedge\ \mathrm{swrlb{:}greaterThanOrEqual}(s, 3)\\ &\rightarrow \mathrm{violatesRegulation}(h, \mathrm{SAPAArticle4})\end{aligned} \tag{3}$$

R3 は，SAPA の実際の法的適用条件を完全に表現したものではない．実際の法令適用には，事故結果，事業者の義務，安全管理体制および因果関係等の確認が必要である．したがって，本研究で導出される結果は法令違反の確定ではなく，安全管理者が確認すべき法令関連候補として扱う．

OWL 個体を LHR-SSM および SWRL 規則とともに Pellet へ入力し，明示的な事実と推論によって追加された事実を取得する．本研究では，最終結果だけでなく，結論の導出に用いた入力事実と適用規則も記録し，この性質を推論経路の追跡可能性として評価する．

5　評価実験

5.1　実験目的

本実験では，1 節で設定した RQ1〜RQ4 を検証するため，生成 AI 単独方式とオントロジー介在方式を比較し，LHR-SSM への接続が出力の表現と利用可能性に与える影響を評価する．

5.2　実験条件

実験には，Önal らが公開した Video Dataset for Safe and Unsafe Behaviours [9] から選定した，フォークリフト，荷役作業，荷物の落下および人と車両の接近等を含む 5 本の動画を用いた．同データセットは，生産施設の監視カメラ映像から構成される．映像解析モデルは Gemini 2.5 Flash とし，temperature は両方式とも 0.3 に設定した．各動画について，生成 AI 単独方式とオントロジー介在方式をそれぞれ 5 回実

表 1　標準 L2 クラス形式適合率，構造有無率およびルール発火率

指標	オントロジー介在	生成 AI 単独	ルール発火率	対象件数
標準 L2 クラス一致率	89%	自動判定不可	–	37 / 25
MF 構造有無率	100%	100%	100%	37 / 25
SAPA 対応構造有無率	100%	100%	100%	27 / 18
推論経路構造有無率	100%	100%	100%	37 / 25

表 2　質的評価による内容品質

指標	オントロジー介在	生成 AI 単独
語彙の適切性	73%	92%
MF の適切性	87%	92%
SAPA 対応	100%	33%
推論経路提示	91%	85%

表 3　同一動画を 5 回解析した場合の安定率

項目	オントロジー介在	生成 AI 単独
Severity	78%	80%
L2/危険カテゴリ	76%	未計測
MF	92%	未計測
SAPA 候補の有無	100%	100%

行したため，全実行回数は 5 動画 ×2 方式 ×5 回の計 50 回である．

両方式から計 62 件の危険イベントが抽出された（生成 AI 単独方式 25 件，オントロジー介在方式 37 件）．プロンプトの違いにより抽出場面数が一致しないため，イベント検出数は性能指標としていない．

図 2 に，評価に用いた映像の例を示す．(a)，(b) は通路上に金属類の静止障害物が存在する場面，(c)，(d) はフォークリフトを含む作業場面である．

5.3　評価指標

標準 L2 クラス形式適合率，ManagementFailure 構造化，SAPA 対応，推論経路提示および出力安定性を評価した．定量評価では，L2 クラス名が 56 個の標準クラスに含まれる割合，各項目の構造有無率，および前提条件を満たす個体に対する R2・R3 の発火率を算出した．これらは処理系の正常性を示す参考値であり，内容の正しさを直接示さない．

質的評価では，評価者 1 名が 62 件のイベントについて語彙，ManagementFailure，SAPA 関連候補および推論経路を 0〜3 点で採点し，最大得点に対する割合を求めた．SAPA 対応は重大度 3 以上のイベントのみを対象とした．また，同一動画を 5 回解析した結果を scene group ごとに対応付け，各項目の最頻出値の出現割合を安定率とした．

6　実験結果

6.1　定量評価結果

標準 L2 クラス形式適合率，構造有無率およびルール発火率を表 1 に示す．オントロジー介在方式では 37 件中 33 件が標準 L2 クラスに対応し，残り 4 件は L2 クラス名が空欄であった．また，R2 および R3 は対象となる全個体で発火した．一方，構造有無率は両方式とも 100%であり，項目の存在のみでは内容品質を評価できないことが分かった．

6.2　質的評価結果

質的評価による内容品質を表 2 に示す．

語彙と ManagementFailure の適切性は，生成 AI 単独方式がそれぞれ 92%，92%で，オントロジー介在方式の 73%，87%を上回った．一方，SAPA 対応と推論経路提示はオントロジー介在方式が 100%，91%で，生成 AI 単独方式の 33%，85%を上回った．したがって，提案方式の効果は特に法令関連候補の明示に現れた．

6.3　出力安定性

同一動画を 5 回解析した場合の出力安定率を表 3 に示す．Severity 安定率はオントロジー介在方式 78%，生成 AI 単独方式 80%であり，提案方式による改善は確認されなかった．提案方式の L2 クラス，ManagementFailure，SAPA 候補の有無はそれぞれ 76%，92%，100%であったが，単独方式の自由記述項目は未計測である．

表 4　図 2(b) に対する両方式の出力例

生成 AI 単独方式（Vanilla）	オントロジー介在方式（Ontology）
hazard_text: 作業現場の整理整頓不良，不適切な動線，保護具の不着用により，転倒や機械との接触，飛来物による負傷などの潜在的なリスクが存在する．	hazard_text: 工場内の通路付近に AGV が停止しており，男性作業員がその近くを歩いて通過している．その後，女性作業員も通路を歩いている．
severity_score: 2	l2_class: AutomatedVehicleHazard
management_issue_text: 5S 活動の不徹底；安全衛生教育の不足または不徹底；作業手順書や安全ルールの策定・周知・遵守の監督不足；...	severity_score: 3
sapa_candidates: 事業者の安全配慮義務；労働者の危険防止義務；保護具着用の義務；...	mf_class: OperationalProcedureFailure
	r2_inferred: true
	sapa_candidates: SAPA_Article4
	r3_inferred: true

7　考察

7.1　語彙制約の効果と限界

定量評価による標準クラス一致率と，質的評価による語彙の適切性には差がみられた．これは，標準語彙として表現できていても，映像内容に対して適切なクラスが選択されるとは限らないことを示している．

図 2(b) は，通路上に金属類の静止障害物が存在する場面である．この場面での両方式の出力例を表 4 に示す．実際の出力から主要項目のみを抜粋している．

オントロジー介在方式では，ManagementFailure，SAPA 候補および推論結果が構造化された一方，5 回すべてで危険事象が AutomatedVehicleHazard へ誤分類され，生成 AI も AGV が存在するものとして記述した．実際には自律走行機器は映っておらず，「通路上の静止障害物」に対応する適切な L2 クラスが語彙内に存在しなかったため，最も近いクラスへの誤分類が誘発されたと考えられる．これに対して生成 AI 単独方式では，整理整頓不良や動線管理の問題として記述する例が確認された．

したがって，閉じた語彙は表現を統一できる一方，語彙の被覆性が不足している場合には，誤った具体化を強制する可能性がある．本結果は，生成 AI との接続以前に，LHR-SSM の危険事象語彙自体を継続的に拡張・検証する必要性を示している．今後は，本実験で不足が確認された静止障害物等の語彙を追加し，追加前後で語彙の適切性および出力安定性を比較することで，語彙被覆性の改善効果を検証する．また，適切な候補が存在しない場合に分類を保留し，元の自然言語記述を保持する仕組みについても検討する．

7.2　構造化と内容品質

構造有無率は両方式とも 100%であり，生成 AI 単独方式でも，管理上の問題や法令関連候補を自然言語として出力できた．したがって，オントロジーの利点を，単に「生成 AI には出せない情報を追加できること」と説明するのは適切ではない．

提案手法の差は，危険事象，管理失敗および法令関連候補を，標準クラスと明示的なプロパティで接続できる点にある．自然文の表現揺れに対して追加の自然言語処理を行わず，WorkforceCompetencyFailure 等の共通クラスとして検索・集計できる．

ただし，構造化されていることと，割り当てられたクラスが正しいことは別である．実験では，同一実行内の同一場面に対して，OperationalProcedureFailure と WorkforceCompetencyFailure という異なる管理失敗が割り当てられた例が確認された．後段の推論器は入力されたクラスを前提として動作するため，生成 AI による意味付けが誤っていれば，推論結果もその誤りを引き継ぐ．質的評価でも，オントロジー介在方式の ManagementFailure の適切性は 87%となり，生成 AI 単独方式の 92%を下回った．この結果は，構造化によって検索可能性は得られる一方，分類内容の妥当性が自動的に保証されるわけではないことを示している．

7.3　SAPA 対応における効果

SAPA 対応は本実験で最も大きな差が得られた指標である．オントロジー介在方式では，重大度 3 以上の危険事象に対して，R3 により SAPA 第 4 条の確認候補が一貫して付与され，法令候補とその導出根拠を

明示的に提示できた．一方，生成 AI 単独方式では，一般的な安全配慮義務や安全管理の必要性に関する説明にとどまり，具体的な条文候補との対応が示されない例が多かった．この結果から，オントロジーと規則の介在は，生成 AI 出力を法令概念へ一貫して接続する上で有効であると考えられる．ただし，R3 は重大度のしきい値に基づく簡略規則であり，実際の SAPA 適用条件を完全に表現していない．したがって，本結果は法的判断の正確性ではなく，安全管理者が確認すべき法令関連候補の提示能力を示すものである．

一方，Severity 安定率には改善がみられず，生成 AI による入力判断の揺らぎは後段の推論にも引き継がれた．今後は，複数回解析結果の統合等による前段での品質向上を検討する．実務上，本手法は法令違反を自動判定するものではないが，危険事象と確認すべき法令候補，およびその導出根拠を一貫した形式で記録できるため，安全管理者による確認対象の整理や分析結果の追跡を支援する基盤として利用できる可能性がある．

7.4 本実験条件での限界

本実験は 5 本の動画から抽出した 62 件を対象とし，質的評価も評価者 1 名によるものであるため，結果の一般化には限界がある．今後は対象動画と評価者を増やし，複数評価者による評価者間一致度を確認する必要がある．また，本実験は Gemini 2.5 Flash のみを対象としており，他の生成 AI モデルに対する再現性は未検証である．さらに，R3 は重大度に基づく簡略規則であるため，SAPA の実際の適用条件を表すには，安全管理体制や因果関係等を含む規則の精緻化が必要である．

8 おわりに

生成 AI による物流現場映像の分析結果を LHR-SSM へ接続し，危険事象，管理失敗および SAPA 関連候補を構造化する手法を提案した．評価の結果，SAPA 対応と推論経路提示では有効性が確認された一方，語彙被覆性の不足による誤分類や入力判断の揺らぎが課題として残った．今後は，不足語彙追加の効果検証，分類保留・複数候補の許容，複数評価者による質的評価，推論規則の精緻化を行う．

謝辞　本研究の一部は，九州工業大学ーLogistics Revolution Korea Co. Ltd. 間の国際共同研究の支援を受けた．

参 考 文 献

[1] T. Guan, F. Liu, X. Wu, R. Xian, Z. Li, X. Liu, X. Wang, L. Chen, F. Huang, Y. Yacoob, D. Manocha, and T. Zhou, “HallusionBench: An Advanced Diagnostic Suite for Entangled Language Hallucination and Visual Illusion in Large Vision-Language Models,” *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14375–14385, 2024.

[2] P. Kaul, Z. Li, H. Yang, Y. Dukler, A. Swaminathan, C. J. Taylor, and S. Soatto, “THRONE: An Object-based Hallucination Benchmark for the Free-Form Generations of Large Vision-Language Models,” *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27228–27238, 2024.

[3] A. Seth, U. Tyagi, R. Selvakumar, N. Anand, S. Kumar, S. Ghosh, R. Duraiswami, C. Agarwal, and D. Manocha, “EGOILLUSION: Benchmarking Hallucinations in Egocentric Video Understanding,” *Proc. of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 28461–28480, 2025.

[4] Korea Ministry of Government Legislation, “Serious Accidents Punishment Act,” Korean Law Information Center, Act No. 17907, 2021.

[5] Y. Lu, Q. Li, Z. Zhou, and Y. Deng, “Ontology-Based Knowledge Modeling for Automated Construction Safety Checking,” *Safety Science*, Vol. 79, pp. 11–18, 2015.

[6] S. Zhang, F. Boukamp, and J. Teizer, “Ontology-Based Semantic Modeling of Construction Safety Knowledge: Towards Automated Safety Planning for Job Hazard Analysis (JHA),” *Automation in Construction*, Vol. 52, pp. 29–41, 2015.

[7] W. Fang, L. Ma, P. E. D. Love, H. Luo, L. Ding, and A. Zhou, “Knowledge Graph for Identifying Hazards on Construction Sites: Integrating Computer Vision with Ontology,” *Automation in Construction*, Vol. 119, Article 103310, 2020.

[8] X. Mei, F. Xu, Z. Zhang, and T. Yu, “Unsafe Behavior Identification on Construction Sites by Combining Computer Vision and Knowledge Graph-Based Reasoning,” *Engineering, Construction and Architectural Management*, Vol. 32, No. 12, pp. 8360–8389, 2025.

[9] O. Önal and E. Dandıl, “Video Dataset for the Detection of Safe and Unsafe Behaviours in Workplaces,” *Data in Brief*, Vol. 56, Article 110791, 2024.