設計初期アーキテクチャ評価のための 大規模言語モデルを活用したモデルベース・ハザード同定

LLM-Integrated Model-Based Hazard Identification for Architecture Evaluation in Early Design Phase

○冨田悠貴（東京大学大学院）*1　　塩苅恵（海上・港湾・航空技術研究所　海上技術安全研究所）*2
青山和浩（東京大学大学院）*3

*1 Yuki Tomita, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, 113-8656, tomita.yuuki@m.sys.t.u-tokyo.ac.jp
*2 Megumi Shiokari, National Maritime Research Institute, MPAT, 6-38-1 Shinkawa, Mitaka, Tokyo, 181-0004, shiokari@m.mpat.go.jp
*3 Kazuhiro Aoyama, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, 113-8656, aoyama@race.t.u-tokyo.ac.jp

キーワード: ハザード同定, 大規模言語モデル, システムアーキテクチャ, MBSE, System of Systems

1. 緒　　言

ICT や制御技術の飛躍的な進展に伴い，複数の人工物システムが連携して機能する System of Systems（SoS）の構築が社会インフラやモビリティ分野で活発に検討されている．特に，宇宙・航空，船舶や自動車などの分野では，ミッション高度化や人員不足への対応のため，自動化・自律化が推進されている．これらのシステムはセーフティクリティカルであり，異常発生時の社会的影響が大きく，開発初期段階のアーキテクチャ選定時には徹底的な安全性評価を行い，品質確保と開発手戻りへのリスクヘッジが重要である．

従来の安全分析では，FMEA，FTA，HAZOP など構成機器単位での故障分析に関する手法が主流であった．しかし，複雑化する SoS 内部の相互作用から創発的に発生するハザードに対しては，このような従来手法では捉えにくく，STAMP/STPA や SysML/UML 等で構築されたシステム記述モデルに基づく SWIFT（Structured What-if Technique）分析[(1)]などを通じ，専門家によるワークショップ形式で進められる人手中心のプロセスが推進されている．これらの手法ではシステム間の制御構造やクラス関係を明確にし，相互作用に起因するハザードを定性的に分析できる一方，安全評価担当者の経験への依存が大きく，またモデリングの作業負荷も高く，専門家の知識と労力に左右される．

このような課題に対し，近年，デジタルエンジニアリングの文脈で適用検討が進められている Model-based Systems Engineering（MBSE）とその記述モデル[(2)]を基盤とし，大規模言語モデル（Large Language Model:LLM）を統合したハザード同定フレームワークを提案する．SysML 等による統合モデルからハザード抽出に必要な情報をクエリし，LLM を用いて分析を半自動化することで，設計初期のアーキテクチャ案比較を効率化し，安全分析を伴う高速なトレードオフ支援を実現することを目的とする．

提案手法の評価のため，実用化に向けて開発が進む自動運航船を対象に，記述されたシステムモデルに対して LLM を用いたハザード同定を実施する．システムのセーフティクリティカルな特性を踏まえた，設計要求としての安全対策/推奨事項の自動抽出も試みる．出力された結果を踏まえ，提案手法の有効性と今後の展望を評価する．

2. モデルベース・ハザード同定への期待と課題

モデルベース・ハザード同定は，システムモデルを利用してハザードを体系的に洗い出す手法である．Shiokari et al.[(1)]は Structure model-based hazard identification（SMB-HAZID）を提案し，UML クラス図を基に開発したシステムとタスクの構成図と，STPA のガイドワードを基に開発したキーワード，並びに自動運航船のハザード同定に必要な観点のチェックリストを活用して自動運航船の危険要因を抽出する手順を示した．Miyake et al.[(3)]はアクティビティ図を用いた Dynamic-task-based HAZID（DTB-HAZID）を開発し，クラス図だけでは捉えにくいタスクフローの失敗モードを抽出できることを報告している．これらの研究は，システム記述モデルの活用が安全分析に有効であることを示しており，設計初期の段階からモデルを活用したハザード同定には期待が高まる．一方で，ここで用いられるモデルは，いずれも安全分析のために新しく専用に作成される必要がある．システムモデルは，MBSE 文脈でも設計者により作成されるため，これらを直接活用することができれば，安全分析への効率的な接続ができると考えられる．

また，LLM を安全分析に活用する試みも始まっている．Qi et al.[(4)]は ChatGPT と協調した STPA によるハザード同定を効果的に実行する研究を報告した．また，塩苅ら[(5)]は LLM を組み込んだ SMB-HAZID によるハザード同定支援ツールを試作し，専門家との協働により効率的かつ網羅的な分析を行える可能性を示した．これらの成果は，人間の負荷を軽減しつつ網羅的なハザード同定を実現する方向性として期待されている．一方で，膨大な知識を事前学習した LLM を用いて特定のコンテクストに限った安全分析を実行するには，対象とするシステムに関する情報の与え方と入力プロンプトには十分な注意が必要である．ハザード同定の対象とするシステムの運用フローや構成要素に関する理解度が低ければ，特定の SoS 内部の相互作用から創発的に発生するハザードを発見することは難しいため，実際の開発対象システムの構成要素間の相互作用を構造化データとして把握させることが重要である．また，対象ドメインの専門家の経験に基づく暗黙知のような情報は，LLM では適切なガイドがなければ，設計開発上で考慮が必要であ

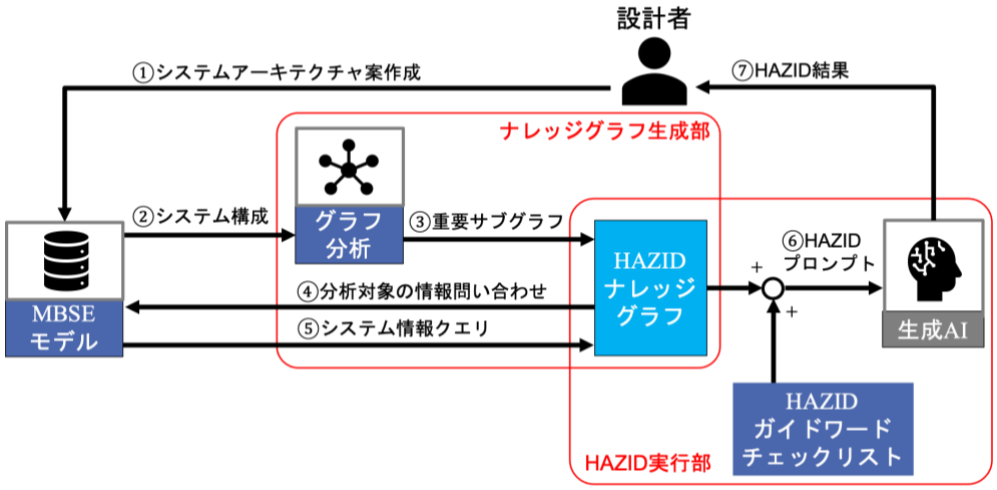

Fig.1　Proposed workflow of LLM-Integrated Model-Based Hazard Identification

る運用設計領域（ODD）や故障モードを考慮したハザード同定出力には至らない．

このような期待と課題に対応するべく，MBSE モデルの主要な利点であるシステムアーキテクチャの情報構造を活用し，LLM と連携させることで，設計初期のハザード同定の効率化・高度化を目指す．本提案における具体的な対処課題は以下のとおりである．

- **MBSE モデルの安全分析への活用：**既存の MBSE システムモデルを活用し，システムアーキテクチャを情報構造化した上で，LLM を活用したハザード同定作業の実施のための情報をクエリする手法を構築する．
- **LLM を用いたハザード分析性能の向上:**LLM を用いて開発対象のシステムに特化したモデルベース・ハザード同定の精度を向上するだけでなく，人間の設計者でも発見に至りにくい，システミックな要因で発生するハザードの同定成功率を高めるためのプロンプト生成ルールセットを構築する．

これらを通じ，MBSE のユースケースを増やすことによる投資対効果を高め，またセーフティクリティカルな SoS における設計初期の安全分析の高密度化による新規性の高い大規模複雑システムの開発効率化を目指す．

3.　LLM と MBSE を統合したハザード同定手法

3.1.　ワークフロー全体像

既存のモデルベース・ハザード同定手法[1][3]を参考とし，MBSE と LLM を統合し，HAZID 実行フローを自動化する手法を提案する．図 1 に検討したワークフロー全体像を示す．提案では，安全分析において可能な限り設計者の負担を低減するため，MBSE 活動で維持更新されることを想定したシステムアーキテクチャを記述したモデルを直接活用するワークフローを目指した．図 1 中にて①システムアーキテクチャ案，と記載されている情報をもとに，MBSE モデルとして整備された，図 1 中の②システム構成は，グラフ分析や情報クエリを通じて HAZID に必要な情報をまとめた HAZID ナレッジグラフとして構造化する．加えて，HAZID ガイドワードやチェックリスト[1]として，HAZID 実行にあたり品質を向上させるために整理されたプロンプトを合わせて LLM に入力する．次節より，本ワークフローの根幹であるナレッジグラフ生成部と HAZID 実行部の詳細を述べる．

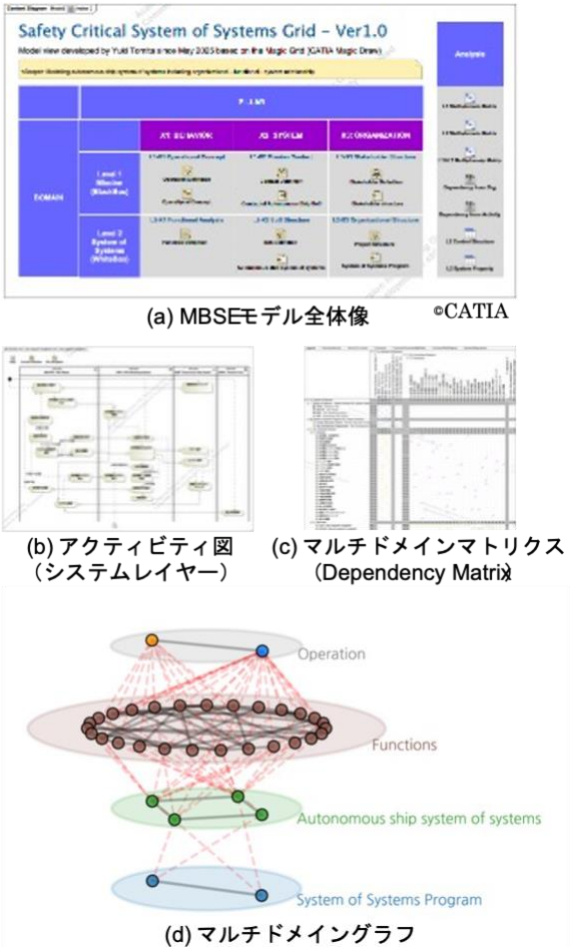

Fig.2　Conversion from MBSE model to multi-domain graph

3.2.　MBSE モデルからの HAZID 知識グラフ抽出

MBSE モデルは SysML 等の記述モデルを束ねた情報構造体を示し，設計者やプロジェクトメンバーは適宜情報を参照し，システム定義やインタフェース定義，機能配分の確認を行うために使用される．設計者は MBSE モデルを用いて，運用，機能，システム，組織等に関する機能配分を確認しながら，モデルベース思考を通じてアーキテクチャを明らかにし，システムの機能配分や運用フローの最適化を目指す．MBSE ではこのようなコンセプトのもと，情報を定義・構造化しながらトレーサビリティを確保の上，システムズエンジニアリングをファシリテーションする．こ

ういった情報は人間中心で実行される設計検討作業には有効であるものの，単純な要求分析やトレーサビリティ管理にとどまらず，様々なライフサイクルにおけるユースケースへの活用に期待が込められているが，いまだ模索が続く状況[2]である．こうした背景の下，MBSE モデルを用いた LLM による分析統合を行うことで，特に人間中心な意思決定判断が求められる機能安全設計を効果的に支援できると考えた．図 2(a)に示すような MBSE モデルの中には SysML を用いて記述された状態遷移図，アクティビティ図，ブロック定義図などがある．本稿では，SoS 内部の相互作用から創発的に生じるハザードを同定する目的から，SoS 内部の機能分担とそのフローを示したアクティビティ図（図 2(b)）を中心的に活用することとした．こういった機能フローや機能配分は，MBSE モデルの中で図 2(c)に示すような Dependency Matrix の形でマルチドメイン“マトリクス”として出力でき，さらにこれを用いて図 2(d)に示すマルチドメイン“グラフ”が構築できる[6]．そこで，グラフにより構造化されたデータから Retrieval-Augmented Generation (RAG)を構築する GraphRAG から着想を得て，アクティビティ図を用いてシステムの動作フローを把握させた状態で LLM によりハザード同定をさせられると考えた．

図 1 中では③重要サブグラフの識別，と記載している，SoS 内部のどの相互作用に着目するか，といった設計者が重点的に評価を行いたいと考えた重要なインタラクションを選定し，それに対してのサブグラフを作成する．（マルチドメイングラフに対するネットワーク中心性解析[6]を用いた複雑性の高いインタラクションに着目する手法もあるが，本稿では詳細は割愛する．）STAMP/STPA や SMB-HAZID[1]，DTB-HAZID[3]では，ある分析対象のインタラクションを選定後，HAZID のためのガイドワードやキーワードを参考として，人間の思考の中で情報を保管しながら安全分析を実行する．一方で，LLM を統合した HAZID でも，適切な分析のためには，LLM にシステム全体像を把握させる必要があるが，システムモデル全体の多くの情報を与えると不用意な HAZID 結果を出力してしまうハルシネーションにつながる恐れがある．そこで，ここでは k-hop 近傍探索を用いて，分析対象のインタラクションのうち，アクティビティ図におけるアクションフローの上流，下流の因果を考慮したサブグラフを作成する手法を採用した．これによって，LLM に情報を与える際に，分析対象のインタラクションがどういった事前アクションに基づきトリガーされており，またどういったアクションをさらに引き起こすかといった情報に焦点を絞ることで，人間の思考が実施しているシステム動作を考慮した発想を AI にも誘発することを目指した．なお，図 1 中にて④分析対象の問い合わせ，⑤システム情報クエリ，と記述している点において，作成したサブグラフに関連する説明文書も MBSE システムモデルから検索できるような仕組みとすることで，システムの意味的な解釈も適切に与えられるように配慮し，サブグラフとまとめて，HAZID ナレッジグラフとして LLM に渡す準備を整える．ナレッジグラフの具体例は第 4 章に示す．

3.3. LLM を用いたハザード同定の自動化

LLM は自然言語を通じて作業指示できるため，基本的には人間中心のワークショップで HAZID を実行するために必要な同様の事項を与えれば良い．ただし，第 3.2 節にて生成する HAZID ナレッジグラフだけでは適切なハザード同定には不十分なため，SMB-HAZID のキーワードも活用する．加えて，SMB-HAZID[1]で提案された，安全解析担当が抽出して欲しい分析観点を HAZID のためのチェックリスト(表 1)として AI に与えることで，人が実施するワークショップと同様に，LLM が実行するハザード分析においても効果的にハザード同定が実行できることを予備検討[5]で確認したため，プロンプト構成を含め流用する．このチェックリストには，特に Man-Machine インタフェース設計を重点的に考慮したハザード同定における観点が整理されている．特に，本稿で取り扱う船舶の自動運航化のようなマシンと人間の機能配分を検討する問題にあたって重要な視点として HAZID 実行で考慮する必要がある．チェックリストでは，人間観点としてタスク失敗や不備のある計画整備，システム観点ではヒューマンマシンインタフェースやソフトウェア，ハードウェアの不適切なインタラクション，また機能安全の安全論証の範囲を示す ODD に対する逸脱状況などがまとめられている．人間が実行するワークショップと同様に，安全分析のヒントを LLM にもプロンプトとして与えることで，品質の高い分析が実行可能となる．

Table 1　HAZID Checklist [1]

- Human: Task failures
- Human: Faulty diagnosis
- Human: Inadequate plan
- Human: Execution failure
- Interactions among components: Inappropriate human–machine interfaces (HMI)
- Interactions among components: Inappropriate software–software interaction
- Interactions among components: Inappropriate human–human interaction
- Interactions among components: Communication network failure
- ODD deviation: Errors in the detection of deviation from ODD
- ODD deviation: Inadequate notification plan
- ODD deviation: Errors in takeover by humans

4. ケーススタディ

4.1. 対象システム

提案手法のケーススタディとして，近年実用化に向けて開発が進められており，また複雑性の高いセーフティクリティカルの一例である自動運航船を対象に，LLM と MBSE を統合したハザード同定を試みた．自動運航船には様々な方式が提案されているが，遠隔操船の実証実験向けに提案されたシステム[3] [7]に対して提案手法を適用した．また，本稿では，先行研究[3] [7]と同様に，遠隔 Waypoint 運航(以降，WP 運航)フェーズを対象とした．WP 運航におけるシステムと遠隔操船者の役割のベースラインを表 2 に示す．遠隔操船者はシステムが提示する情報をもとに遠隔操船を行い，船上に配置された実験要員は，遠隔操船者へ他船情報等の補足情報の提供および実験中の緊急時対応を行う．このように自動運航船は陸上・船上，またそれぞれに関する人・マシン（Man-Machine）の 4 つのシステム間における

機能配分を伴う検討課題である．設計初期アーキテクチャ検討において，これらの機能配分を実施するにあたり，配分案に対する総合的な評価を行うことになるが，網羅的かつ抜け漏れのないリスク評価がセーフティクリティカルシステム開発における重要な関心事である．また，ハザード同定結果に対して効率的な設計対処方針が立たない場合や，開発ライフサイクル後段において，その機能安全の付与にあたり開発負荷を増大させてしまうようなアーキテクチャを選定することも望ましくない．こういった観点で，設計初期アーキテクチャ評価を様々な機能配分案に対して効率的にハザード同定することを目標とし，そのワークフローの実証を行なった．

Table 2　Man-Machine function allocation baseline for remote waypoint operation [3] [7]

	陸上	船上
人	<遠隔操船者> システムの監視，WP 運航の継続と WP 変針の可否判断，避航の判断と実行	<実験要員> システム監視支援，船上監視状態の収集提供，自動運航の中断判定
マシン	<遠隔操作装置> 船上取得情報の表示，船舶への遠隔操船者の指令送信，異常時の警報発令	<自動運航船> 計画に沿った操船実行，各カメラやセンサ情報の取得と陸上への送信，異常時の警報発令

4.2.　適用結果

4.2.1.　システムモデルからの HAZID ナレッジグラフ生成

対象の自動運航船システムの MBSE モデルとして，システムブロックを定義（図 3(a-1)(a-2)）し，WP 運航の状態遷移(図 3(b))を上位運用フローとして，その中には自動 WP 運航，異常発生時には手動での WP 運航に移行するものとした．この前提のもと，表 2 の機能配分に従い，状態遷移図とアクティビティ図を用いてシステムの動的な振る舞いを図 3(c-1)(c-2)のように記述した．LLM が正しい意図を参照できるよう，アクティビティ図に現れるアクションには，図 3(c-3）のように説明を Property に含めた．提案手法によるハザード同定の有効性を確認するために，本稿ではシステム間のインタラクションが集中する，自動運航船のもつ「目標 WP の設定」→「WP 運航(Autopilot)」におけるアクションの繋がりに対してハザード同定を実行することとした．さらに，このインタラクションに対して 2-hop 近傍探索を施すことで，図 3(d-1)に示すハザード同定対象の周辺アクションを含むサブグラフが抽出できる．また，HAZID ナレッジグラフは直接 Mermaid 形式のコードを用いて，図 3(d-2)に示す形式で LLM へのプロンプトに取り込んだ．

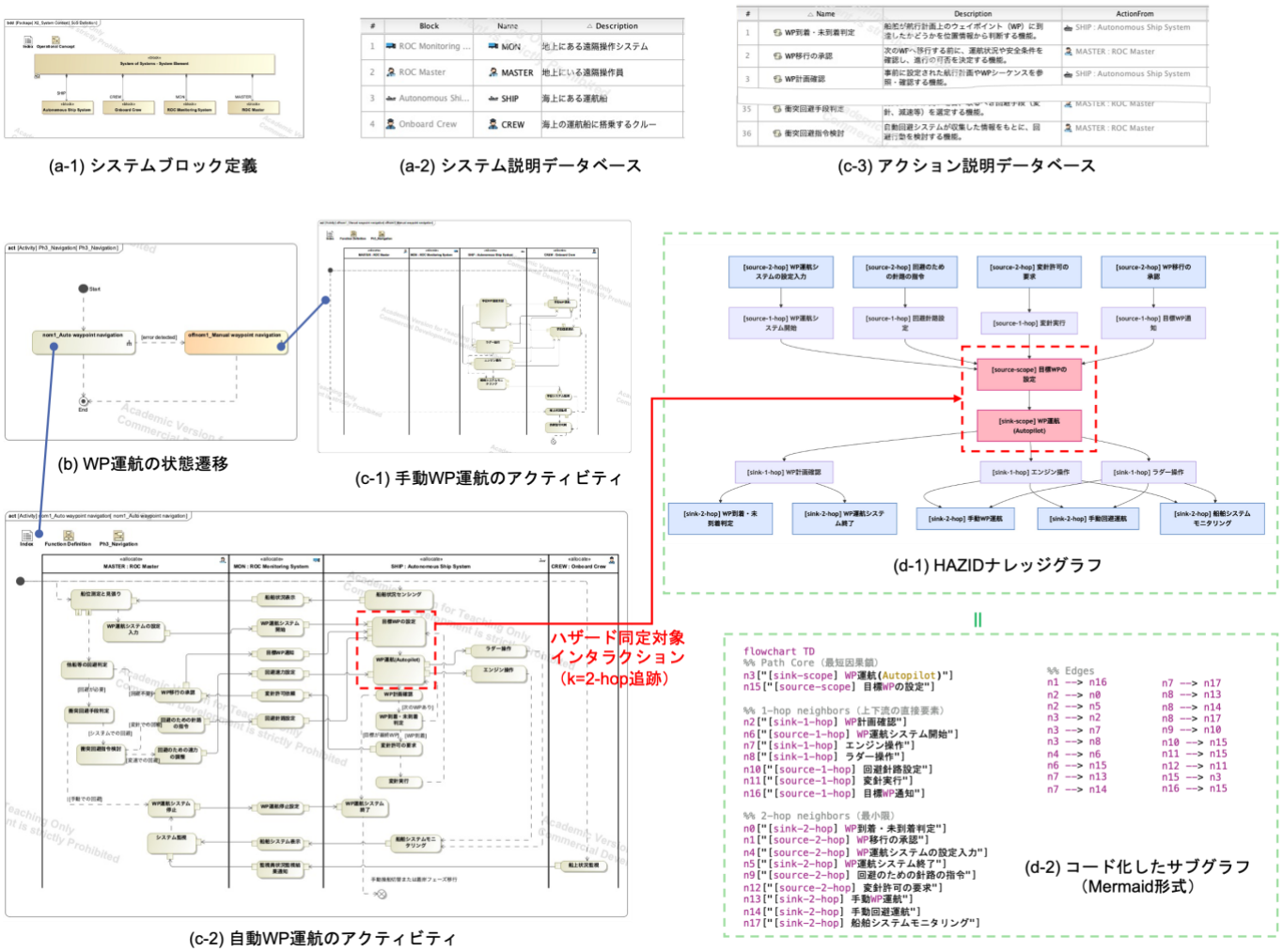

Fig.3　Generating HAZID knowledge graph from SysML diagram

4.2.2. LLM を用いた HAZID 実行

Mermaid 形式のサブグラフに加え，図 1 に示したプロセスの通り，３つのキーワード，そしてハザード同定における発想を支援するための表 1 に示したチェックリストを束ねてプロンプトとして LLM に入力することで，MBSE モデルを入力として，アクションの繋がりを考慮した HAZID が実行可能となった．

実際に図 3 の HAZID ナレッジグラフに対してハザード同定を試みた結果を表 3 に示す．ここでは，３つのキーワードごとに，２つずつのハザード抽出を行った．それぞれのキーワードごとに LLM に HAZID 結果を出力させている．ここで，出力にはハザード，原因（HAZID Checklist に

Table 3　Output result of LLM-integrated model-based auto hazard identification

ID	GR-1-A-1	GR-1-A-2	GR-1-B-1	GR-1-B-2	GR-1-C-1	GR-1-C-2
キーワード	Key1 タスクの実行がされない		Key2 タスクが遅れて実行される		Key3 タスクが誤って実行される	
ハザード	目標 WP の設定が行われないことで WP 運航が所要の目標を受け取れない	目標 WP の設定が行われないことで WP 移行や変針の承認・実行が停止する	目標 WP の設定の遅延により Autopilot が新しい目標 WP に切り替わらず回避動作が遅れる	目標 WP の設定遅延により WP 移行承認や WP 運航開始が遅れ、到着が大幅に遅延する	目標 WP の誤設定により自動操船が誤った航路へ進行する	目標 WP が ODD 外や到達不可な位置に設定され自動運航が継続不能または迂回を余儀なくされる
原因	①WP 運航システム開始の失敗により目標 WP 設定のトリガが発生しない ②目標 WP 通知の通信障害により目標 WP が船側に届かない	①WP 運航システムの設定入力ミス（誤った航行パラメータや制約）により目標 WP の自動設定が阻害される ②WP 計画確認での不一致により WP 移行の承認が行われない	①通信ネットワークの遅延や切断により目標 WP 通知が遅延する ②目標 WP 設定処理を行うソフトウェアの負荷やバグで目標 WP 反映が遅延する	①WP 移行の承認（人または上位システム）処理が遅延する ②目標 WP 通知がシステム間でキュー遅延や処理待ちで遅延する	①誤入力や誤操作 ②入力表示や確認不足により誤 WP が確定される	①ODD 逸脱判定の誤り・遅延 ②運用者が設定の不適合に気付かない
結果	目標 WP が設定されず WP 運航が適切に更新されないため、自動操船が誤った航路を維持して他船や障害物と衝突する可能性がある	目標 WP が設定されないため WP 移行や変針が実行されず、自動運航が継続不能となり手動操船へ切替えが必要になる	目標 WP への切替が遅れて Autopilot が回避針路を採れず、他船や障害物との衝突リスクが高まる	自動による WP 移行が遅延して予定通りの航行ができず到着予定が大幅に遅れるか、手動運転への切替が必要となる	誤った WP に向けて WP 運航(Autopilot)が舵・エンジンを操作し、他船や障害物と衝突する可能性がある	自動運航が WP 遷移や到着判定で問題を生じ、手動介入や大幅な到着遅延を招く可能性がある
対策・推奨事項	①目標 WP 通知の冗長通信経路と到達確認応答(ACK)を実装する ②WP 未設定時に安全停止または自動的に回避針路を生成するフォールバック機能を導入する	①WP 運航システムの設定入力時に整合性チェックと入力ミス検出のバリデーションを追加する ②WP 計画確認と WP 移行承認の自動照合と異常時アラートを実装する	①目標 WP 通知に対する ACK/再送と冗長通信経路を実装し遅延検出を可能にする ②目標 WP 設定にタイムアウトとフェイルオーバー（手動介入または安全停止）を導入する	①WP 移行承認プロセスにタイムアウトと自動エスカレーション（自動承認の条件設定等）を導入する ②目標 WP 通知の優先度制御と遅延検出用の監視アラートを追加し冗長経路で配送する	①目標 WP 入力時に航路・周辺障害物照合と衝突リスク評価を自動で実行し入力をブロックする ②重要な WP 変更には二重確認（表示と承認）や多要素承認を要求する	①目標 WP を ODD 条件（航路、海域、視程等）と照合し不適合時は設定を拒否または警告を強制する ②WP 移行の承認フローを必須化し、WP 移行前に WP 到着判定やシステム状態確認を要求する
確信度	2	3	3	2	3	2
関連ノード	[source-1-hop] 目標 WP 通知 [source-1-hop] WP 運航システム開始 [sink-1-hop] ラダー操作	[sink-1-hop] WP 計画確認 [source-2-hop] WP 運航システムの設定入力 [sink-2-hop] 手動 WP 運航	[source-1-hop] 目標 WP 通知 [sink-1-hop] ラダー操作 [sink-1-hop] エンジン操作	[source-1-hop] 目標 WP 通知 [source-2-hop] WP 移行の承認 [sink-2-hop] 手動 WP 運航	[sink-1-hop] WP 計画確認 [sink-1-hop] ラダー操作 [source-1-hop] 目標 WP 通知	[source-2-hop] WP 移行の承認 [source-1-hop] WP 運航システム開始 [sink-2-hop] WP 到着・未到着判定

基づき最大 2 つ識別することを指示)，結果，どのような情報を使用したかのトレーサビリティ確保のための対象キーワードと関連するアクションノード，そしてハザードへの設計での対策・推奨事項（最大２つ識別することを指示）を記載した json 形式で構造化出力させた．

試行結果を見ると，ハザード同定対象のコントロールアクション（目標 WP の設定（自動運航船））やコントロールアクションの接続先（WP 運航(Autopilot)）だけでなく，そのアクションに至るまでの他システムのアクションや，分析対象のアクションの結果後段で実行されるタスクに渡り，アクションの前後の因果関係を網羅的に考慮した状態での具体的なハザード同定（例 : GR−1-C-2 のハザードにおける「目標 WP が ODD 外や到達不可な位置に設定され自動運航が継続不能または迂回を余儀なくされる」では，2-hop 下流の WP 到着・未到着判定との連携が考慮されている）が実施できた．また，設計対応方針例を見ると，同様に具体的な対応策（例 : 2-hop 上流の WP 移行の承認フローでの対策強化）が生成されていることがわかり，これらもチェックリストから連想されたものであることも推測できる．このように，MBSE モデルに保管したアクティビティ図を考慮したハザード同定を LLM により行うことで，シームレスなワークフローを達成できるだけでなく，システム構成やタスクの流れを把握した人間の解析者が実行する場合と同等に，分析対象アクションの前後の因果関係を考慮した結果を出力可能なことを確認した．一方で，検討されたハザードの中には，関連ノードとの順序因果を一部誤って出力しているものも見られた（例 : GR−1-B- ２のハザードにおける「目標 WP の設定遅延により WP 移行承認や WP 運航開始が遅れ、到着が大幅に遅延する」では，2-hop 上流の WP 移行の承認のタスクと，目標 WP の設定の順序が逆転している）．プロンプトでは，このようなタスクフローの順序因果を考慮するように指示しているものの，指定数のハザード同定を実施するにあたり，強制的に指示に従わない連想が出力されている場合も見られた．このようなケースに備え，LLM にはそのハザード同定結果に対する確信度（ここでは，０～３ : ３が最大，とした.）を申告させるようにした．実際に，確信度が低い値となる場合には一部意図とは異なる出力を見せることがあった．しかしながら，このようなハザード同定結果や対策・推奨事項もただちに無効な評価ではなく，システムの連続的動作を考慮した結果として設計者に重要な洞察を与えるレベルには到達しているため，人間によるチェックが未だ必要ではあるが，設計者のハザード発想の支援には資する出力ができた．

5. 考　　察

5.1. MBSE モデルの活用

本稿では MBSE モデルとして，産業にて共通的に使用されているフレームワークとして知られる MagicGrid 方式を用いてモデル化されたシステムを想定し，それを直接的に HAZID ナレッジグラフに活用するワークフローを提案した．設計初期におけるシステムアーキテクチャ検討では，運用コンセプトから識別された機能を各システムにどう配分するかが論点となる．また，セーフティクリティカルシステムにおいてはその機能配分により生じるシステム間の不具合や不整合から生じるハザードイベントに対しての処置もトレードオフ検討対象となる．そのため，機能配分を検討する上で利便性の高いアクティビティ図を用いて，要すれば適宜アクティビティ図を修正し，容易に HAZID ナレッジグラフを再生成できるため，LLM を用いた HAZID は MBSE との親和性が高いと考える．加えて，デジタルエンジニアリングの文脈を通じて設計開発における MBSE モデルの活用が進められるものの，そのユースケースは未だ限定的である中，このような HAZID に直接活用でき，設計者の負荷低減と具体性のあるハザード同定結果を得られることは，設計効率の向上に繋がるといえるだろう．

5.2. HAZID ナレッジグラフ

HAZID ナレッジグラフの生成には，対象コントロールアクションを起点として k-hop 追跡を活用することで，グラフ解析により関連アクションを MBSE モデルから効率的に抽出することが可能となった．今回は k=2 としたが，陸上と船上，またそれぞれに人間とマシンを含む４つのシステムに横断する因果接続を満遍なく拾い上げることができたと考えられる. hop 数は少なすぎると HAZID 結果の具体性が欠け，一方で多すぎると，ハルシネーションのリスクも高まることに留意が必要である．また，システムモデルのアクション粒度はモデラーによって大きくばらつきが出るが，k=2 ほどの追跡を行うことで，モデリング粒度に対してロバストになる効果が得られたと考える．今回は対象アクションの k-hop 追跡のみをプロンプトに盛り込んだが，MBSE モデルから作成された知識グラフに対して Graph Neural Networks (GNNs)による分析[8]を施し，より広い知識グラフを評価するような手法も研究されており, HAZID 実行のために必要な情報を効果的に検索し，プロンプトに盛り込む手法は広く検討したい．

また，最終的に Mermaid 形式を用いてナレッジグラフを出力し，プロンプトに直接埋め込むことで，MBSE モデルから容易に LLM 向け入力を作成することができた．今回はツールの都合上，SysML での図を Mermaid に変換したが，SysML Ver2 として開発が進められている SysML モデルとコードの直接変換を可能とするオントロジーの確立に伴い，このようなワークフローは様々なツールにおいても汎用的に組み上げることができることが想定される．

5.3. 生成 AI モデルの活用

今回の分析では OpenAI 社の GPT５-nano[9]を API 経由で呼び出して使用した. GPT-5 は推論モデルであり, GPT-4 などと比べると複雑なタスクへより長く検討を行うため，大量のあいまいな情報に基づいて意思決定を行うのに効果的である．LLM へのプロンプトの中には前述の通り Mermaid 形式を用いたが，予備検討[5]ではうまくハザード同定ができていた GPT-4 を用いて分析をした際には十分に因果が考慮されない結果となってしまった．Mermaid 形式はアクション間の接続エッジ表現が特殊であり，人間が直接文字ベースで読解するにもやや難解なことから，GPT-4 が正確にフローを理解するにはイタレーション的な推論が必要であったことが要因と推定している. GPT-5 は推論を伴うため，GPT-4 に比べ１０倍近くのトークンを使用するため，網羅的なハザードを高速に回すためには，Mermaid 形式のアクティビティをより解釈しやすいように Chain of Thought を明示して書き下すことで，低負荷なモデルでも動作が可能になると考えられる．いずれにせよ，複数のアーキテクチ

ャ比較を要する設計初期フェーズにおいて，人間が実施するには飽きやすい安全分析の繰り返しを安定した品質で出力を継続できるため，LLM の活用はセーフティクリティカルなシステム検討には非常に有力である．

一方で，今回の検討では表 3 の通り各キーワードに対して２つずつのハザード出力を試みたが，実際には 2 つ以外にも見落とすべきではないハザードがある可能性に留意が必要である．しかしながら，LLM は指定した数だけいくらでもハザード同定結果を出力できるが，数が増えるほど類似のハザードを出力するほか，ハルシネーションリスクが増加することも試行の中で確認した．こういった観点でも，LLM に申告させたハザード同定結果に対する自身の度合いを示す確信度の評価は，出力結果のレビューにおいて有効であった．どのような出力結果が有用かつ着目に値するかを自動で判定できれば，ハザード同定もより効果的に実施できるため，検討を進めたい．

5.4. HAZID キーワードとチェックリスト

HAZID キーワードとチェックリスト，またそれらを埋め込み，HAZID を実行するためのプロンプトは予備検討[5]にて設計したものを流用した．出力された HAZID 結果は概ね良好な結果を見せた．今回使用したチェックリストのうち，特に ODD 条件は自動運航船に特化したものを活用したが，異なる分野のセーフティクリティカルシステムでも本提案手法を試行することで，チェックリストの更なる評価が必要である．ハザードへの設計対策/推奨事項については特段明確な要求をプロンプトで指定しなかったが，十分な品質で洞察の得られる出力がなされた．チェックリストの明確な定義が役立ったことに加え，推論モデルによる思考イタレーションの中で，LLM 自身の持つ知識が十分に盛り込まれたことが推定される．出力の安定化や設計要求設定への直接的な活用ができるよう，設計対策/推奨事項に関するプロンプト検討も今後進めたい．

6. 結　　言

本稿では，MBSE モデルの更なる活用価値向上とデジタルエンジニアリングを通じた開発支援のため，システムモデルから直接，抽出したナレッジグラフを活用し，LLM との連携により半自動的にハザード同定に活用するワークフローを提案し，自動運航船を対象とするケーススタディで有効性を示した．試行を通じて，提案手法では以下に示す有意性を確認した．

- **LLM によるハザード同定の性能向上**：アクティビティ図を基盤に k-hop 近傍でサブグラフを抽出し，対象インタラクションの前後因果を明示したうえで LLM に入力することで，過剰・過少な文脈供給を抑制しつつ，専門家の思考プロセスに近い発想と出力を実現した．
- **設計初期の比較検討支援**：MBSE モデルを直接活用することを可能とし，Mermaid 記法によるグラフ表現，キーワード／チェックリストによるプロンプトのルール化，および構造化出力により，反復が容易で比較可能な HAZID 実行を実現した．これにより LLM を用いて，アーキテクチャ代替案の迅速な比較検討に活用が可能となる．

今後は，本提案手法をさらに発展し，LLM からの HAZID 出力結果の定量化や安全分析を考慮したアーキテクチャ最適化への活用を検討する予定である．引き続き，設計初期の反復的な安全分析を可能にし，セーフティクリティカルな SoS のアーキテクチャ決定を実務レベルで支援する基盤への発展を目指す．

文　　献

(1) Shiokari, M., Itoh, H., Yuzui, T., Ishimura, E., Miyake, R., Kudo, J. and Kawashima, S.: Structure model-based hazard identification method for autonomous ships. Reliability Engineering & System Safety, 247, 110046, 2024. https://doi.org/10.1016/j.ress.2024.110046

(2) Wolny, S., Mazak, A., Carpella, C., Geist, V. and Wimmer, M. : Thirteen years of SysML: a systematic mapping study. Software & Systems Modeling, 19, 111–169, 2020.

(3) Miyake, R., Kudo, J., Ishimura, E., Itoh, H., Yuzui, T., Shiokari, M., Kawashima, S., Hirata, K., Niki, Y., Kobayashi, M., Sawada, R. and Inaba, S.: Application of dynamic-task-based hazard identification method to remote operation of experimental ship Shinpo. Journal of Physics: Conference Series. Vol. 2311. No. 1. IOP Publishing, 2022. https://doi.org/10.1088/1742-6596/2311/1/012013

(4) Qi, Y., Zhao, Y., Khastgir, S. and Huang, X.: Safety analysis in the era of large language models: A case study of STPA using ChatGPT, Machine Learning with Applications, Vo.19, 100622, 2025.

(5) 塩苅恵，冨田悠貴，青山和浩：大規模言語モデルを活用したハザード同定支援ツールの開発及び人間と人工知能の協働作業の展望－仮想の自動運航船への適用を通して－，日本船舶海洋工学会講演会論文集第 41 号.

(6) 冨田悠貴，青山和浩：セーフティクリティカルシステムの組織横断設計におけるアーキテクチャ記述と連携リスク評価，日本機械学会 第 35 回設計工学・システム部門講演会講演論文集，2025.

(7) 三宅里奈，稲葉祥梧，塩苅恵，石村惠以子，伊藤博子，柚井智洋，工藤潤一，平田宏一，仁木洋一：小型実験船「神峰」の遠隔操船実験に基づくタスク分析，第 91 回マリンエンジニアリング学術講演会講演論文集，2021.

(8) Karagoz, E., Fischer, O.J., Mavris, D.: Identification of Missing Knowledge in MBSE System Models Using Graph‐Based Machine Learning, Systems Engineering, 2024 Dec 16:e70013.

(9) OpenAI: GPT-5 nano. https://platform.openai.com/docs/models/gpt-5-nano