人工知能の進化と課題

国立大学法人　長岡技術科学大学
大学院工学研究科　システム安全工学分野
教授　山形　浩史

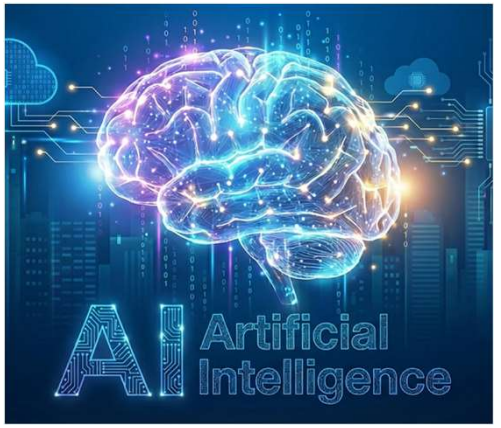
内容

人工知能AIとは

AI開発の歴史
- 機械学習
- 深層学習
- 大規模言語モデル
- 生成AI
- フィジカルAI

生成AIの実力と課題
- 皆さんと勝負
- 高品質な学習データ
- 検索拡張型生成(RAG)
- ニューロシンボリックAI

総合討論に向けて



問１　非常停止機能に関して述べた以下の文中の空欄①から⑩に当てはまる語句を下の選択肢の中から選び、記号で答えなさい。

非常停止機能は、機械の（　①　）において操作可能であり、他の全ての機能及び操作に（　②　）ものでなければならない。非常停止機能によって機械が運転を停止したのち、非常停止機能が（　③　）リセットされるまでいかなる（　④　）も有効となってはならない。

非常停止機能は、（　⑤　）又は他の安全機能の代替手段として採用してはならず、（　⑥　）として設計することが望ましい。また、非常停止機能は、保護機器又は他の安全機能を持つ機器の有効性を（　⑦　）ならない。

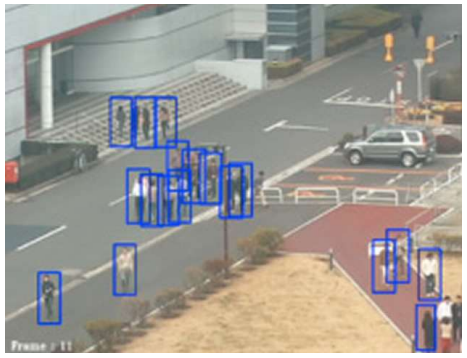
非常停止機能は、非常停止機器の動作後、（　⑧　）が発生することなく、また（　⑨　）なしに、機械の動作を適切な方法で停止するように（　⑩　）に従い設計しなければならない。

選択肢

ア．　新たな危険源	イ．　安全防護方策	
ウ．　一時的に無効化するものでなければ		エ．　規定時間の経過により
オ．　起動信号	カ．手動で	キ．　全ての運転モード
ク．　操作	ケ．　損なっては	コ．　遅延
サ．　電源の再投入	シ．　配慮した	ス．　人の介在
セ．　付加保護方策	ソ．　法令	タ．　本質的安全設計
チ．　優先する	ツ．　ライフサイクル全般	
テ．　リスクアセスメント		

人工知能AIとは？　　定義はまちまち

- 人間の思考プロセスと同じような形で動作するプログラム、あるいは人間が知的と感じる情報処理・技術（日本政府）*1
- ＡＩシステムは、明示的又は暗黙的な目的のために推測する機械ベースのシステム。受け取った入力から、物理環境又は仮想環境に影響を与える可能性のある予測、コンテンツ、推奨、意思決定等の出力を生成（OECD)*2



*3

ChatGPT

*1：日本政府，令和6年版科学技術・イノベーション白書，2024
*2：ＯＥＣＤ，Explanatory memorandum on the updated OECD definition of an AI system,2024
*3：日本信号HP

なぜ今、AIを議論するのか

役割

・従来の「**情報処理**」から「**意思決定支援**」へと**役割が拡大している**

応用

・安全管理、社会インフラ、医療など、**リスク許容度の低い領域**への応用が進んでいる

課題

・**精度の高さだけでは不十分**

・**説明可能性（Explainability）・安全性・責任の所在** が重要な課題となる

AI 活用には「技術の理解」と「適切な設計」が不可欠

AI開発の歴史

1956年　ダートマス会議

第１次ブーム(1960s)
　　ＡＩに関する基礎的な概念、初の対話型自然言語処理プログラム「ELIZA」

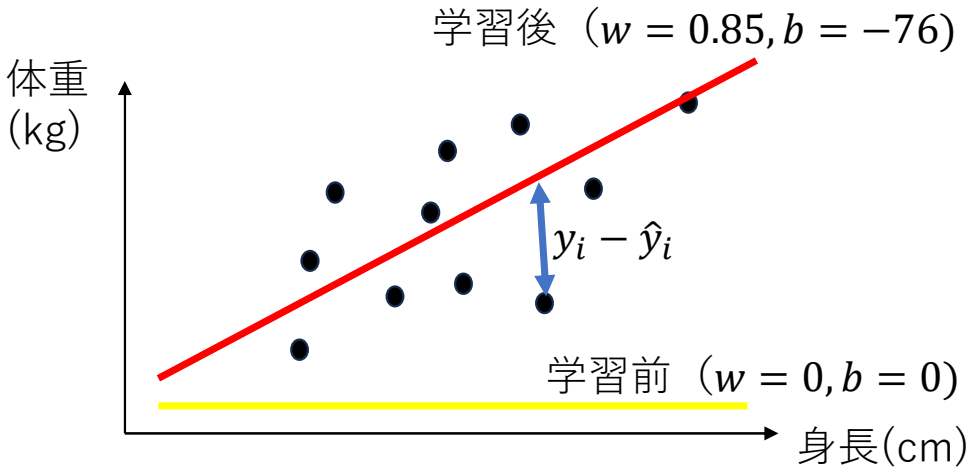
第２次ブーム(1980s-1990s)
　　エキスパートシステム、辞書・ルールベース自然言語処理、第５世代コンピューター
DeepBlue:チェス王者に勝利

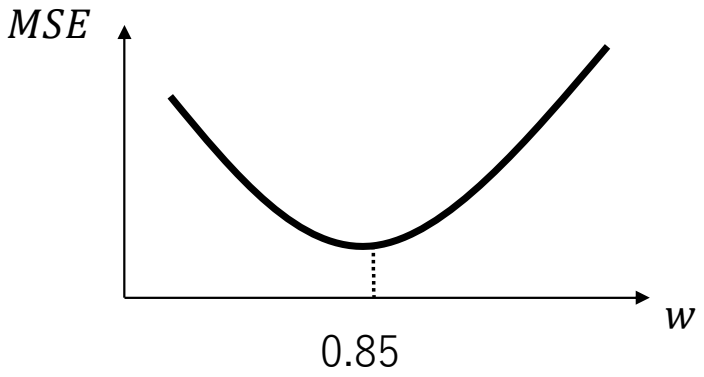
第３次ブーム(2000s-2010s)
　　インターネット、計算能力向上、機械学習、深層学習
Watson:クイズ番組で人間に勝利、　AlphaGo:プロ棋士に勝利

第４次ブーム(2020s)
　　Transformer、生成AI（ChatGPT、画像生成AIなど）
Gemini2.5Pro,　ChatGPTo3:司法試験短答式合格

機械学習の原理：線形回帰モデル（教師あり学習）

身長から体重を推定する





1. 教師データの収集
身長と体重のデータ収集

2. モデルの設定
予測値を $y = wx + b$で表す

3. 損失関数の定義
予測値と実際の値の誤差を評価するため、平均二乗誤差（MSE）を用いる

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2$$

4.最適化（学習）
収集したデータから損失関数が最小になるように wと bを調整

5. 推定
得られたwとbの回帰式に身長を入れ、体重を推定

機械学習の原理：線形重回帰モデル（教師あり学習）

身長、胸囲、腹囲、体脂肪率から体重を推定する

1. 教師データ

氏名	身長 (cm)	胸囲 (cm)	腹囲 (cm)	体脂肪率 (%)	体重 (kg)
Aさん	172	95	82	18.5	68
Bさん	160	88	76	24.0	60
Cさん	180	102	90	21.0	78
Dさん	165	85	70	17.0	58
Eさん	175	98	85	26.5	75

2. モデルの設定

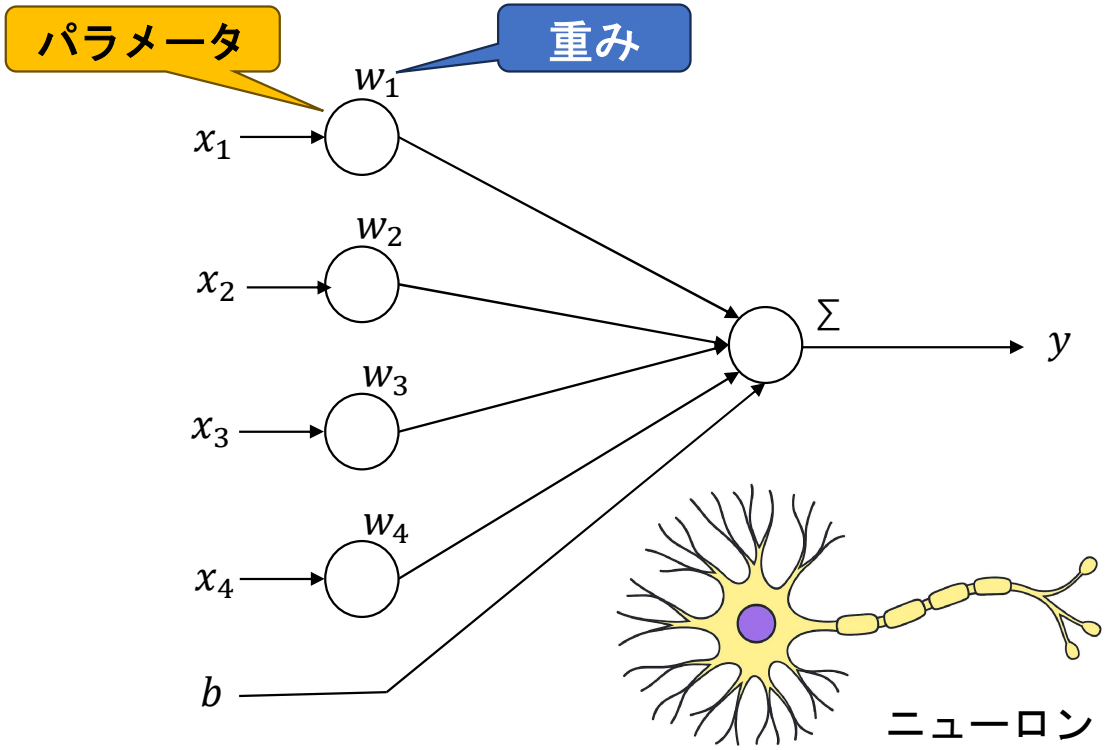
$$y = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b$$

y：体重

x_1：身長
x_2：胸囲
x_3：腹囲
x_4：体脂肪率

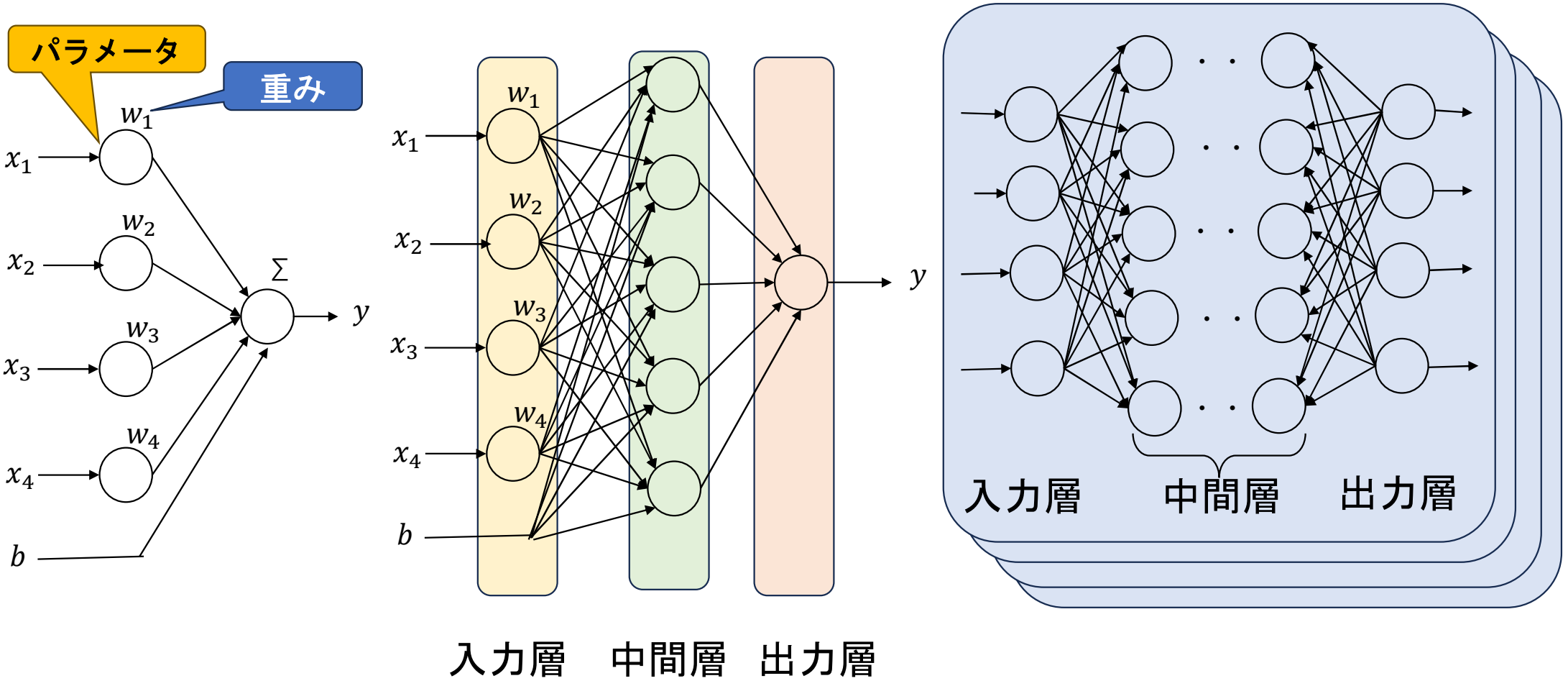
b：バイアス

3. 損失関数が最小になるようw_iとbを調整



教師データが多いほど
パラメータが多いほど
体重の推定値の精度が上がる

機械学習⇒深層学習：ニューラルネットワーク



深層学習の応用分野：識別系

1. 画像認識・分類
•顔認識（本人確認、監視カメラ）
•医療画像診断（腫瘍の有無、骨折検出）
•品質検査（製造ラインでの不良品検出）

2. 音声認識・音声分類
•音声コマンドの認識（「電気をつけて」など）
•話者識別（誰が話しているか）
•感情認識（怒り・喜びなどの音声トーン分析）

3. 自然言語処理（NLP）の識別系
•感情分析（レビューがポジティブかネガティブか）
•文書分類（ニュース記事のジャンル分け）
•スパム検出（迷惑メールの判定）

4. 異常検知・予測モデル
•機械の故障予測（センサーデータから異常を検出）
•サイバーセキュリティ（不正アクセスの検知）
•金融リスク評価（信用スコア、融資判断）

5. 自動運転の認識系
•車線検出、標識認識
•歩行者や障害物の認識
•周囲環境のセグメンテーション（画像の領域分割）

大量のデータを事前に学習することにより、
人間でも間違えることに対しては、より正確に判断する

正答の *確率をアップ* させる

大規模言語モデル（LLM）

基盤モデル名	開発組織	発表年	パラメータ数
BERT	Google	2018	3.4億
GPT-1	OpenAI	2018	1.2億
GPT-2	OpenAI	2019	15億
GPT-3	OpenAI	2020	1,750億
PaLM	Google	2022	5,400億
GPT-4	OpenAI	2023	非公表

日本政府，令和6年版科学技術・イノベーション白書

私はAIです。

1. 入力：テキストをトークンに分割
文章を単語やサブワードに分解
（例：「私」「は」「AI」「です」）

2. 埋め込み（Embedding）
各トークンを数値ベクトルに変換

事前にトークンの
種類数（語彙数、例：50,000）
ベクトル次元数（例：1024次元）
を決め、公開文献を教師に学習

位置情報を数値ベクトルに変換

3. 文脈理解
単語間の関係性を計算
複数の視点で文脈を捉える
単語の順序情報を補完

トランスフォーマー

4. 出力予測（次の単語を生成）
文脈に基づき、最も自然な次の単語を確率的に予測

I am ?

AI の確率が高いよ

5. 応答生成
自然な文章として出力

I am AI.

生成AIの応用分野

1. 文章生成（自然言語生成）
- チャットボット
- 自動要約
- 記事・ブログ生成
- プログラムコード生成
- ストーリー・詩の創作
- メール・レポート作成

Transformer
LLM
Chain-of-Thought

2. 画像生成
- イラスト・アート生成
- プロダクトデザイン
- ファッション・インテリア提案
- 医療画像の補完
- ゲーム・映画のコンセプトアート

Diffusion Models
Generative Adversarial Networks
Transformer

3. 音声・音響生成
- 音声合成（TTS）
- ボイスクローン
- 音楽生成
- 効果音生成
- ナレーション

Variational Autoencoder
GAN

文章・画像・音声を統合的に扱う「マルチモーダルAI」

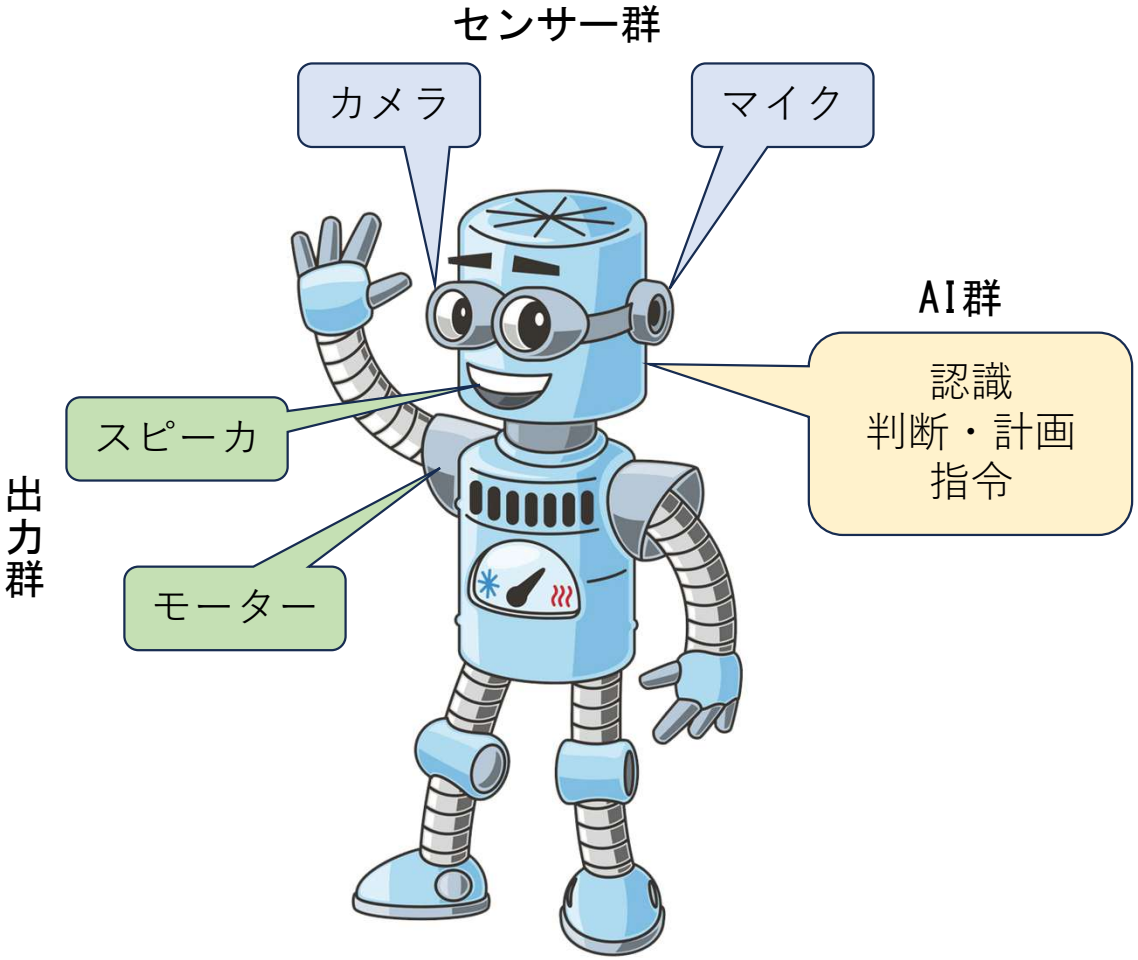
自律的に「動作」できる「フィジカルAI」

フィジカルAI

生成AIとフィジカルAIの比較

項目	生成AI	フィジカルAI
対象	デジタル空間	物理空間
出力	文章・画像・音声	動作・制御・物理的反応
主な技術	Transformer・Diffusion	センサー・ロボティクス・強化学習
例	ChatGPT、DALL·E	Tesla自動運転、ヒューマノイドロボット

配膳ロボット
（双日）



SSEアソシエイト試験（高専レベル）の解答を

問１　非常停止機能に関して述べた以下の文中の空欄①から⑩に当てはまる語句を下の選択肢の中から選び、記号で答えなさい。

非常停止機能は、機械の（　①　）において操作可能であり、他の全ての機能及び操作に（　②　）ものでなければならない。非常停止機能によって機械が運転を停止したのち、非常停止機能が（　③　）リセットされるまでいかなる（　④　）も有効となってはならない。

非常停止機能は、（　⑤　）又は他の安全機能の代替手段として採用してはならず、（　⑥　）として設計することが望ましい。また、非常停止機能は、保護機器又は他の安全機能を持つ機器の有効性を（　⑦　）ならない。

非常停止機能は、非常停止機器の動作後、（　⑧　）が発生することなく、また（　⑨　）なしに、機械の動作を適切な方法で停止するように（　⑩　）に従い設計しなければならない。

選択肢

ア．　新たな危険源	イ．　安全防護方策	ウ．　一時的に無効化するものでなければ	エ．　規定時間の経過により	
オ．　起動信号				
カ．　手動で	キ．　全ての運転モード	ク．　操作	ケ．　損なっては	コ．　遅延
サ．　電源の再投入	シ．　配慮した	ス．　人の介在	セ．　付加保護方策	ソ．　法令
タ．　本質的安全設計	チ．　優先する	ツ．　ライフサイクル全般	テ．　リスクアセスメント	

生成AIの実力:SSEアソシエイト試験（高専レベル）

問１　非常停止機能に関して述べた以下の文中の空欄①から⑩に当てはまる語句を下の選択肢の中から選び、記号で答えなさい。

非常停止機能は、機械の（　①　）において操作可能であり、他の全ての機能及び操作に（　②　）ものでなければならない。非常停止機能によって機械が運転を停止したのち、非常停止機能が（　③　）リセットされるまでいかなる（　④　）も有効となってはならない。

非常停止機能は、（　⑤　）又は他の安全機能の代替手段として採用してはならず、（　⑥　）として設計することが望ましい。また、非常停止機能は、保護機器又は他の安全機能を持つ機器の有効性を（　⑦　）ならない。

非常停止機能は、非常停止機器の動作後、（　⑧　）が発生することなく、また（　⑨　）なしに、機械の動作を適切な方法で停止するように（　⑩　）に従い設計しなければならない。

選択肢

ア. 新たな危険源	イ. 安全防護方策		
ウ. 一時的に無効化するものでなければ		エ. 規定時間の経過により	
オ. 起動信号	カ. 手動で	キ. 全ての運転モード	
ク. 操作	ケ. 損なっては	コ. 遅延	
サ. 電源の再投入	シ. 配慮した	ス. 人の介在	
セ. 付加保護方策	ソ. 法令	タ. 本質的安全設計	
チ. 優先する	ツ. ライフサイクル全般		
テ. リスクアセスメント			

	Copilot	Gemini	正解
①	ツ	キ	キ
②	ウ	チ	チ
③	カ	カ	カ
④	オ	オ	オ
⑤	イ	タ	イ
⑥	タ	セ	セ
⑦	ケ	ケ	ケ
⑧	ア	コ	ア
⑨	ス	ス	ス
⑩	テ	シ	テ
正答率	70%	70%	

生成AIの実力:SSEアソシエイト試験（高専レベル）

CopilotもGeminiも合格（６０点以上）
暗記問題は完璧、少し考える問題は苦手
上位バージョンで正答率アップ

問	問題の内容	Copilot	Gemini 2.5Frash	Gemini 2.5Pro	満点
1	非常停止機能	14	14	18	20
2	インターロック式ガード	16	20	20	20
3	リスクアセスメントの基礎用語	10	10	10	10
4	リスクアセスメントの正誤問題	10	10	10	10
5	危険源の分類	10	10	10	10
6	製造物責任法	8	10	10	10
7	労働安全衛生法（各役割の名称）	10	10	10	10
8	技術者倫理	6	9	9	10
	合計点	84	93	97	100

生成AIの原理と課題

原理	課題	問題	対応
大量の学習 （ネット上の公開情報を学習）	全ての情報を食べ尽くした	著作権侵害	
	間違った情報も学習	間違った回答 バイアス ハルシネーション	高品質な学習データの使用（信頼性の高い情報源からデータを収集し、誤情報を排除） ファクトチェック（出力された情報を人間や外部ツールで検証）
	非公開情報を学習できない	秘密情報をAIに提供すると漏洩の可能性 再学習に多大なコスト	RAG
確率で計算	間違える確率が残る	ルールを破る可能性 ハルシネーション	ニューロシンボリックAI RAG

高品質な学習データ

データをそのまま学習させると

・**間違った情報を学習する**

間違った答えをする

質の悪いデータを学習すると
答えも質が悪い
例：質の低いRAシートを学習
しても質の低い答え

・**バイアスがかかる**

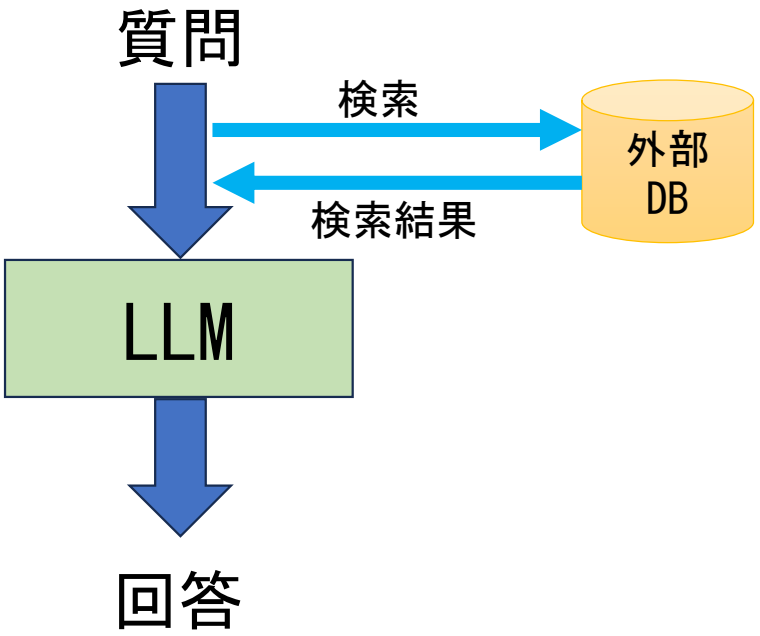
過去の傾向を引きずる
例：男性社員が多いと男性を
多く採用する
大きな母集団の傾向を引きずる

対策

1. 目的と要件の明確化
2. データ収集
•信頼性の高いソースから収集（自社データ、公開データ、スクレイピングなど）
•偏りやノイズの少ないデータを選定する
3. データ前処理（クレンジング）
•欠損値の補完、異常値の除去、フォーマットの統一
4. ラベル付け（教師あり学習の場合）
•専門知識を持つ人によるラベル付けが理想
5. データの多様性とバランス
•特定のパターンに偏らないよう、多様なシナリオや条件を含める
6. 検証・テストデータの分離
•学習用と検証・テスト用データを明確に分ける
•過学習を防ぐための評価指標を設定
7. 品質管理と継続的改善
•定期的にデータの品質をレビューし、古い・不正確なデータを更新

検索拡張型生成(RAG: Retrieval-Augmented Generation)

RAGとは、生成AIが外部知識を検索して活用することで、より正確で信頼性の高い回答を生成する技術
幻覚（ハルシネーション）を抑え、最新情報や専門知識を取り込む



	RAG	Fine-tuning
概要	外部データベースから関連情報を検索し、生成時に活用	モデルのパラメータ自体を再学習させて知識を埋め込む
目的	最新・動的な情報を活用して回答の正確性を向上	特定のドメインやタスクにモデルを最適化
データ更新	外部データを更新すれば即時反映可能	再学習が必要（コスト・時間がかかる）
ハルシネーション対策	検索結果を根拠にするため抑制しやすい	埋め込まれた知識に依存するため幻覚のリスクあり
導入コスト	比較的低い（ベクトルDBと検索エンジン構築）	高い（GPU環境、学習コスト、MLOps体制）
柔軟性	ドメインや用途の切り替えが容易	用途ごとに再学習が必要
応答速度	検索処理が入るためやや遅くなることも	高速（モデル内に知識があるため）

RAG+生成AIの実力：SSEアソシエイト試験（高専レベル）

問１　非常停止機能に関して述べた以下の文中の空欄①から⑩に当てはまる語句を下の選択肢の中から選び、記号で答えなさい。

非常停止機能は、機械の（　①　）において操作可能であり、他の全ての機能及び操作に（　②　）ものでなければならない。非常停止機能によって機械が運転を停止したのち、非常停止機能が（　③　）リセットされるまでいかなる（　④　）も有効となってはならない。

非常停止機能は、（　⑤　）又は他の安全機能の代替手段として採用してはならず、（　⑥　）として設計することが望ましい。また、非常停止機能は、保護機器又は他の安全機能を持つ機器の有効性を（　⑦　）ならない。

非常停止機能は、非常停止機器の動作後、（　⑧　）が発生することなく、また（　⑨　）なしに、機械の動作を適切な方法で停止するように（　⑩　）に従い設計しなければならない。

選択肢

ア.	新たな危険源	イ.	安全防護方策
ウ.	一時的に無効化するものでなければ	エ.	規定時間の経過により
オ.	起動信号	カ.	手動で
キ.	全ての運転モード	ク.	操作
ケ.	損なっては	コ.	遅延
サ.	電源の再投入	シ.	配慮した
ス.	人の介在	セ.	付加保護方策
ソ.	法令	タ.	本質的安全設計
チ.	優先する	ツ.	ライフサイクル全般
テ.	リスクアセスメント		

Copilotに外部DBとしてJIS B 9703を連携

日本工業規格　JIS

B 9703：2019
(ISO 13850：2015)

機械類の安全性−非常停止機能−設計原則

Safety of machinery-Emergency stop function-Principles for design

序文

この規格は，2015年に第3版として発行されたISO 13850を基に，技術的内容及び構成を変更することなく作成した日本工業規格である。

この規格は，機械類の安全性規格群の一つであり，その構成は次による。

タイプA規格（基本安全規格）−全ての機械類に適用できる基本概念，設計原則及び一般的側面を規定する規格

タイプB規格（グループ安全規格）−広範な機械類に適用できる安全面又は安全防護物を規定する規格

タイプB1規格−特定の安全面（例えば，安全距離，表面温度，騒音）に関する規格

タイプB2規格−安全防護物（例えば，両手操作制御装置，インターロック装置，圧力検知装置，ガード）に関する規格

タイプC規格（個別機械安全規格）−個々の機械又は機械群の詳細な安全要求事項を規定する規格

この規格はタイプB2規格である。

正答率は７０％から１００％にアップ！

ニューロシンボリックAI（Neuro-symbolic AI）

安全分野ではルールを確実に守る必要がある
「確率的に９０%程度守ってます」ではいけない

ニューロシンボリックAIは、
- 生成AI（大規模言語モデルなど）に代表されるニューラルネットワーク（データからパターンを直感的に学習するのが得意）と、
- ルールベースAIに代表されるシンボリックAI（明確なルールや論理に基づいて推論するのが得意）

を融合させるアプローチ

期待される効果　生成AIの「ハルシネーション」を抑制する
AIの「説明可能性」を高める
より少ないデータで効率的に学習

今後の研究テーマ

他分野でのニューロシンボリックAI研究

1. 法令解釈（Statutory Interpretation）
法令の条文を解釈し、特定の事案に適用する研究です。

`Neuro-Symbolic AI For Legal Reasoning: A Hybrid Approach To Statutory Interpretation`（法的推論のためのニューロシンボリックAI：法令解釈へのハイブリッドアプローチ）
この種の研究では、Transformerベースのニューラルネットワーク（生成AIの基盤技術）が法的文書の曖昧な表現や文脈を理解し、シンボリックロジック（ルール）が法律の厳密な論理構造（「AかつBの場合、Cである」といったルール）を適用します。米国の税法などを対象に、ルールを組み込むことで、純粋なニューラルモデル（生成AIのみ）よりも高い解釈の正確性を達成したと報告されています。

2. 規制コンプライアンス（Regulatory Compliance）
金融機関のマネーロンダリング対策（AML）など、厳格なルール遵守が求められる分野での研究です。

`Neuro-Symbolic Reasoning for Automated Regulatory Compliance Systems`（自動化された規制コンプライアンスシステムのためのニューロシンボリック推論）
規制文書からナレッジグラフ（知識の構造化）を構築し、論理ルール（シンボリック）として組み込みます。ニューラルネットワーク（生成AI）が膨大な取引データや顧客情報を処理し、パターンを認識します。シンボリックAIが、そのパターンが規制ルールに抵触するかどうかを厳密に判定します。これにより、説明可能性の向上や誤検知の削減が期待されています。

総合討論に向けて

人工知能AIを使って、やってみたいことは何ですか？

Zoom参加の人はチャットで

ご清聴ありがとうございました

国立大学法人
長岡技術科学大学
Nagaoka University of Technology

山形　浩史　　博士（工学）

国立大学法人　長岡技術科学大学
大学院工学研究科　システム安全工学分野
教授

〒940-2188 新潟県長岡市上富岡町1603-1
E-mail:　yamagata ＠マーク vos.nagaokaut.ac.ジェーピー
HP:　http://safety-management.na.coocan.jp/