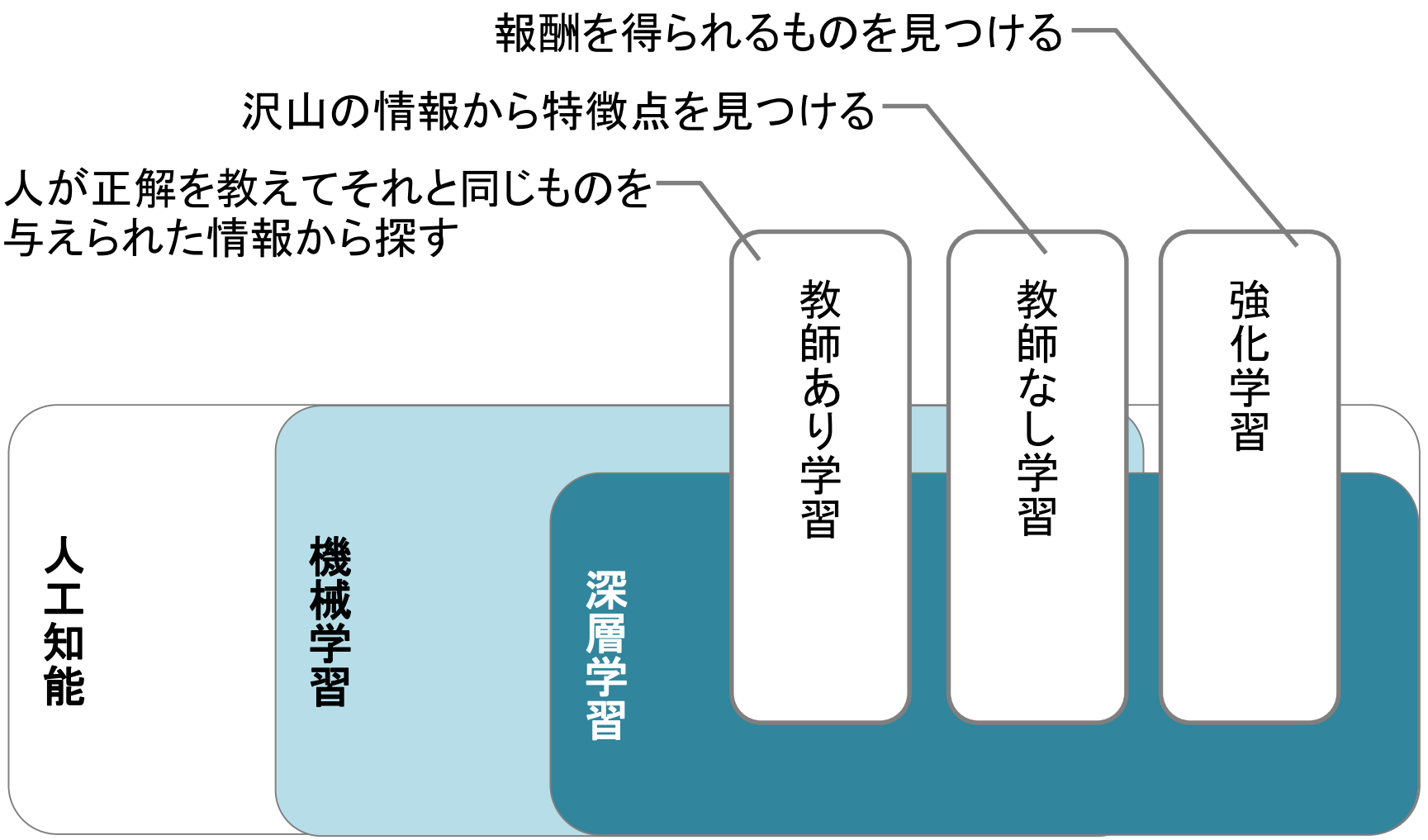
資料2

田辺構成員
提出資料

AIの利活用促進とセキュリティ
（セキュアにAIを利用するために）
～ Road Block ④ 他～

（独）情報処理推進機構　田辺里美

人工知能に関する誤解
何が”脅威”になるのか？

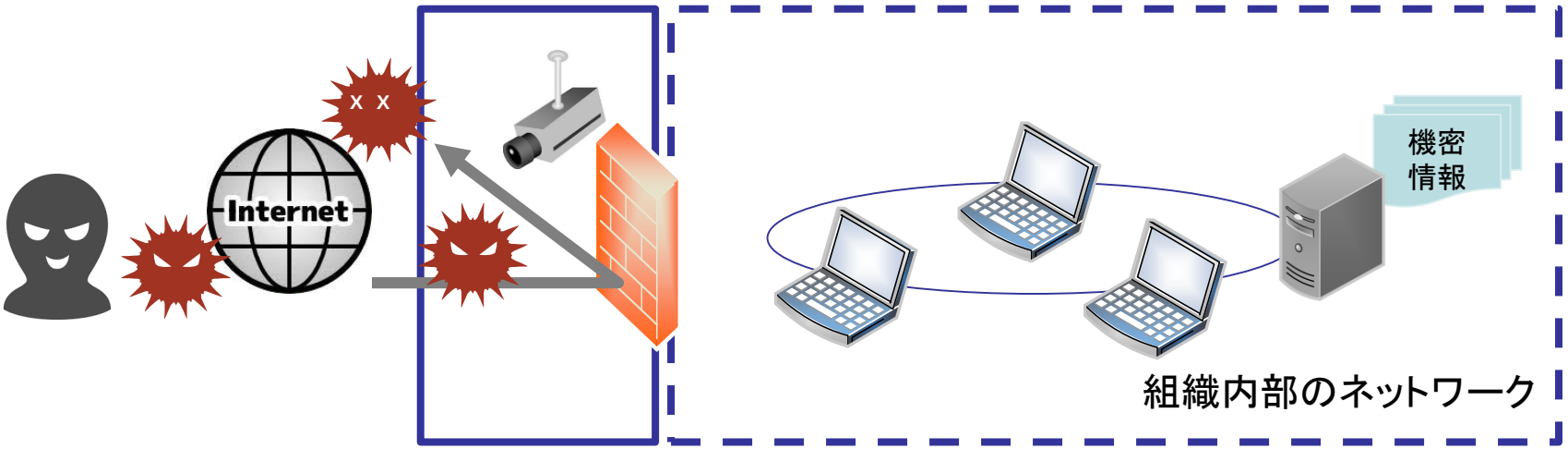


人工知能の学習方法

教師あり学習	正解がわかっている（正解の”ラベル（アノテーション）”が付いている）データを基に学習したうえで、未知のデータに対して同様の予測や識別を行う。	識別、分類（段階に分ける） 等
教師なし学習	正解がわからない（正解がない）データから、共通の特徴を持つものを分類したり、特徴づける情報を抽出したりする。	情報の要約、グループ分け 等
強化学習	ゲームのような試行錯誤を繰り返して結果を得る処理において、実際の処理を繰り返しながら最適な方法を学習する。	囲碁、将棋、ロボットの歩行学習 等

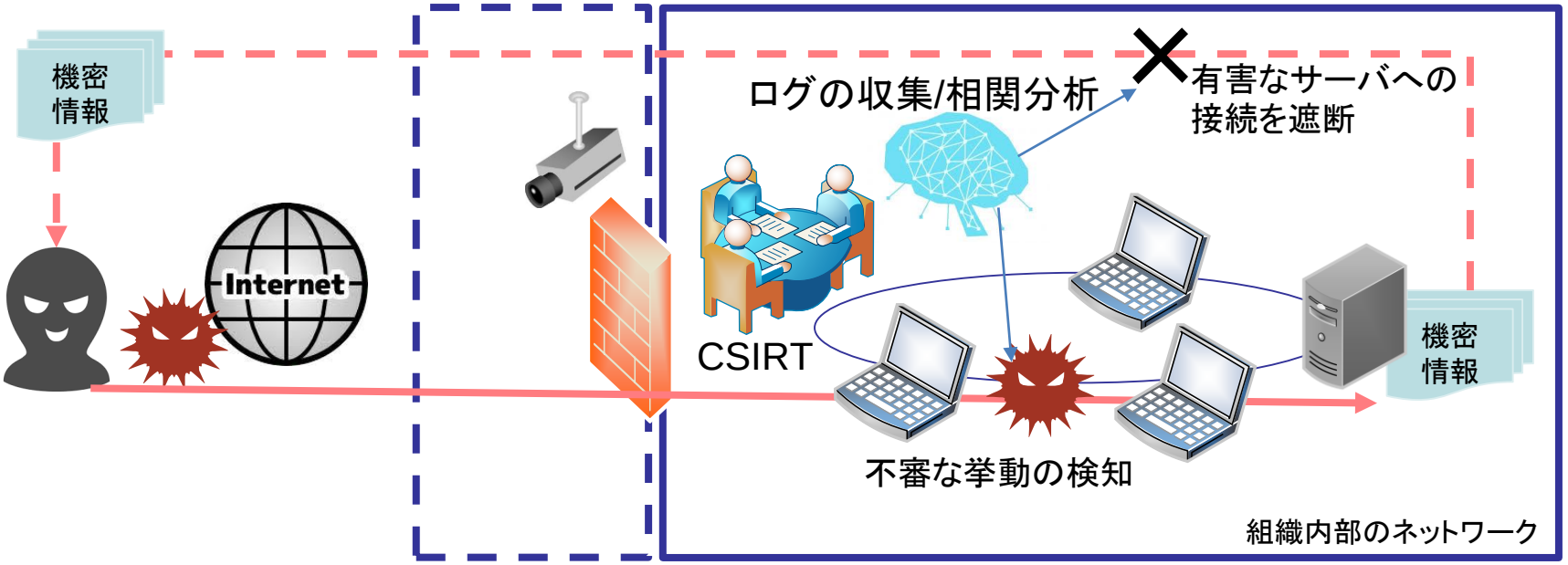
情報セキュリティ対策の方向性

＜これまでの防御の主流＞
　組織内部のネットワークの手前での防御
　⇒侵入されないようにする



情報セキュリティ対策の方向性

＜現在の防御の主流＞
侵入されることを前提に被害を局所化・最少化する
●問題が発生した場合即時に対応できる体制整備⇒CSIRT（シーサート）
●的確な対応のための情報収集と分析⇒ログの収集と相関分析
※即時対応や的確な情報分析を支援するためAIが活用されています。



セキュリティ対策に係るガイドラインへの期待

- セキュリティインシデント発生を前提とした対策
  - 情報侵害発生時の緊急対応手順や体制（CSIRT)の整備―GDPR、HIPAA、HITECH
  - 緊急対応が可能なログ収集と分析

- クラウド環境の利用を前提とした場合の対策
  - 認定制度の活用
  - クラウド提供事業者との契約モデル（サービスレベル（ペナルティを含む）、非機能要件）

- 費用と効果のバランスのとり方
  - 人の負荷を軽減するための方策（人工知能による監視等の自動化）
  - リスクアセスメントによる対応優先度検討

プライバシーの保護とデータの加工

プライバシー保護の方法

個人を特定できる情報の削除または一般化、多様化、情報分布の差の調整

- データの利用目的・研究テーマによって、手法やパラメータは異なる
  ⇒例示やある程度の標準化（ガイドライン化）は可能

- 加工後のデータに対し”脆弱性” の実検証を行う公的な組織等が必要
  ⇒継続的な、実検証の実施及び標準（ガイドライン）の維持