AI-WG 設立趣意書

2023 年 9 月 9 日
諸角　昌宏

この度、以下の通りＡＩワーキンググループの設立を行いたく、ご確認いただき、この趣旨にご賛同いただきますよう、皆様のご協力を是非ともお願い致します。

背景

CSA 本部は、生成 AI について、今日の AI の使い方を規定するベストプラクティスの開発に取り組み、その経験を次世代の改善に役立てることを目標に取り組むことを宣言しました。　この取り組みを継承し CSA ジャパンとしても、日本~~市場~~におけるＡＩ規制やインフラ確立に対して AI の使い方のベストプラクティスを提供できるように活動することが求められていると考えている。

CSA ジャパンにおける AI ワーキンググループでは、CSA 本部の取り組み、および日本やグローバルの動きを調査・研究することを進めていきたい。

1. WG 活動方針
   ① 第 1 期インプット：AI および生成 AI の各種ガイドライン等のスタディ
      - CSA 本部資料に基づくスタディ
        - Security Implications of ChatGPT　翻訳版のスタディ（勉強会形式）
        - CSA Generative AI Usage Policy 現在 draft。Draft の内容を含めてスタディ（勉強会形式）
      - その他の規制/規格のスタディ
        - EU-AI act の理解と生成 AI に向けての拡張の理解。
        - OWASP Top10 for LLM のスタディ
        - IEEE Position Statement Artificial Intelligence のスタディ。
      - 国内状況のスタディ
        - 国内法と規制の動向調査（※年内は立法化の動きはない）
        - 業界法と規制の調査（医師法、弁護士法、著作権法、個人情報保護法）
        - 全省庁からガイドラインが出ている。これのスタディ。
      - プライバシーWG との連携

   ② 第 1 期アウトプット；ブログで中間成果物、web で最終成果物

- CSA 本部資料の翻訳版の提供
- 上記①でスタディした成果物の提供
  ブログ優先で素早く情報発信し、その後資料にまとめていく。

③ 第 2 期インプット；来年度以降の課題
- 政府 AI 戦略会議の提示課題を検討（サイバー攻撃、偽情報、知財リスク）
- 職業資格と試験の検討（マイナビ DX、デジタルスキル標準、IT パスポート試験）
- AI の敵対的テストと標的サンプルの検討
- AI の法的能力の検討（会話能力、評価能力、判断能力、意思決定）
- AI の安全性評価と信頼性評価の検討
- データの偏りと AI 学習による破壊的忘却
- AI 脅威の検討
  - 脅威 1（ＡＩモデルと学習データの非透明性）
  - 脅威 2（偽情報とデジタルコンテンツの e 公証性）
  - 脅威 3（ハルネーションの事実らしさの不正確性）
  - 脅威 4（脅威インテリジェンスの共有性）

④ 第 2 期アウトプット；ブログで中間成果物、web で最終成果物
- 上記③で検討した成果物の提供
  ブログ優先で素早く情報発信し、その後資料にまとめていく。

2. 今後のステップ
① 運営委員会での承認願い（9/5）
② WG メンバー募集（会員向け ML /Slack）
③ キックオフの実施（10 月想定）
④ 活動開始（活動報告は、ＷＧリーダ会を経由して運営委員会に報告）

3. 人員、メンバー
① WGリーダー
- リーダー：　諸角昌宏

② WGコアメンバー
- 笠松隆幸、他

以上