**AI Safety SIG Inaugural Meeting Minutes**

**AI Safety and Red Teaming Special Interest Group (SIG)  
Hong Kong AI Knowledge Management Association**

**Date:** 22nd September 2024  
**Time:** 2:30 PM  
**Venue:** City University (CityU)

## **Attendance**

* Present:
  + Honson Lam
  + John Mok
  + David Wong
  + Anna Han (Google Meet)
* Absent:
  + Gary Wong
  + Kenneth Lam
  + Paul Tse (TBC SIG member)

## **Agenda**

1. Welcome and Opening Remarks
2. Member Introductions
3. Overview of AI Safety and Red Teaming in Hong Kong
4. Mission and Goals of the AI Safety SIG
5. Defining the Scope of Work
6. Establishing AI Safety Standards and Legal Alignment
7. Proposed Projects and Deliverables
8. Methodologies and Tools for AI Safety in Hong Kong
9. Meeting Schedule, Roles, and Communication
10. Any Other Business (AOB)
11. Summary and Closing Remarks

### **1. Welcome and Opening Remarks**

* SIG Lead, David Wong, welcomed all attendees.
* Provided a brief background on the formation of the AI Safety and Red Teaming SIG.
  + Emphasized the importance of collaborative efforts in promoting AI safety and implementing red teaming strategies in Hong Kong.

### **2. Member Introductions**

* Each attendee introduced themselves and their affiliations:
  + Hoson Lam: Chairman & Founder of AIKm, recent focus on BKM Building Knowledge Management and generative 3D models, IoT, Digital Twin and Smart Hotel/facilities management.
  + John Mok: Health sector professional and academic, over 40 years of experience, and
  + Professor Anna Han: Academic: Phd of Law, Phd Candidate for Computing Science with recent focus expertise in AI governance with HKGAI .
  + David Wong: COO for Four Directions Limite, over 30 years of IT Experience on Industries, Great success in Implementation of Technology, SIG Lead and specialist in commercial AI applications.

### **3. Overview of AI Safety and Red Teaming in Hong Kong**

* Presentation Highlights:
  + Current State of AI Safety:
    - Referenced the PCPD's AI Protection Framework:
      * PCPD [AI Protection Framework](https://www.pcpd.org.hk/english/resources_centre/publications/files/ai_protection_framework.pdf) (Jun 2024)
      * HKMA [High-level Principles on Artificial Intelligence](https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2019/20191101e1.pdf) (Nov 2019)
      * HKMA [Use of Artificial Intelligence for Monitoring of Suspicious Activities](https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2024/20240909e1.pdf) (Sep 2024)
      * HKMA [Consumer Protection in respect of Use of Generative AI](https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2024/20240819e1.pdf) (Aug 2024)
    - Professor Anna Han mentioned ongoing work on recommendations or guidelines for government and industries in Hong Kong.
  + Importance of Red Teaming Strategies:
    - Acknowledged that the government has initiated testing of AI safety measures.
    - Discussed aligning with other safety organizations to standardize practices.
    - Differentiated between AI retrieval and generative red teaming approaches.
* Purpose:
  + Set the context for the SIG's work in enhancing AI safety practices and promoting robust red teaming methodologies and implementation tools for Hong Kong. Eventually aligned with International Standards and possible Interoperability.

### 4. **Mission and Goals of the AI Safety SIG**

* Mission Statement Proposals:
  + David Wong: "To provide AIKm a Hong Kong specific safety methodologies and tools for AI knowledge retrieval and generative AI."
  + Professor Anna Han: "To provide classification of risk levels in AI safety to balance implementation and real-life usage."
    - John Mok noted that the medical field uses a 4-level classification system.
  + Honson Lam: Proposed developing a future licensing mechanism for AI safety.
* Group Alignment:
  + Members discussed and reached a consensus to combine the proposals into a cohesive mission statement.
    - Agreed Mission Statement: "To develop and promote safety methodologies, risk classification standards, and tools for AI knowledge retrieval and generative AI, fostering responsible implementation across sectors in Hong Kong with International Alignment."

### **5. Defining the Scope of Work**

* Focus Areas Identified:
  + Government Sector
  + Industry Sector
  + Academic Sector
  + Pilot Citizen Engagement
* Priority Ranking Exercise:
  + Members rated each sector (1 = highest priority, 4 = lowest priority):
    - Government Sector: Total Score 13 (Average 3.25)
    - Academic Sector: Total Score 10 (Average 2.5)
    - Industry Sector: Total Score 9 (Average 2.25)
    - Public Citizen Engagement: Total Score 8 (Average 2)
* Results & Interpretation:
  + Highest Priority: Pilot Public Citizen Engagement
    - Indicates a strong interest in engaging and educating the general public on AI safety.
  + Second Priority: Industry Sector
    - Highlights the need to collaborate with businesses to enhance AI safety practices.
  + Third Priority: Academic Sector
    - Emphasizes involving educational institutions in research and methodology development.
  + Fourth Priority: Government Sector
    - Suggests focusing resources on other sectors first while maintaining future government engagement plans.

### **6. Establishing AI Safety Standards and Legal Alignment**

* Existing Standards in Hong Kong:
  + Professor Anna Han mentioned:
    - Research-based databases being created.
    - Reports on social data.
    - Existing IT company rules.
    - Security standards from OGCIO.
* Alignment with International Standards:
  + Agreed on the importance of aligning with international standards.
  + [AI standards by Singapore Government](https://aisingapore.org/innovation/ai-standards/)
  + Noted the relevance of ISO/IEC TR 5469:2024:
    - [ISO Standard](https://www.iso.org/standard/81283.html)
  + [IEEE Standards](https://www.linkedin.com/pulse/ieee-ai-standards-roadmap-ethical-responsible-digital-moore/)
* Next Steps:
  + Working Group Formation:
    - To delve deeper into standards development and alignment.
  + Considerations:
    - Creating or aligning with a standard committee (none currently exist in Hong Kong for AI safety).
    - Learn from Singapore's AI Verification Application Standards Setting.
    - US AI Safety
      * Executive\_Order\_14110
      * https://en.wikipedia.org/wiki/Executive\_Order\_14110
    - Exploring the provision of professional certifications or training programs.

### **7. Proposed Projects and Deliverables**

* Potential Initiatives:
  + Newsletters: Regular updates on AI safety developments.
  + Workshops and Seminars: Educational events for various stakeholders.
  + Whitepapers: In-depth research documents on AI safety topics.
  + Policy Reviews: Analyzing and providing feedback on regulatory policies.
  + Tool Development: Creating tools tailored for AI safety in Hong Kong.
  + Fundraising: Securing funds for SIG projects if necessary.
* Volunteer Sign-Up:
  + Members expressed interest in leading or participating in specific projects.

### **8. Methodologies and Tools for AI Safety in Hong Kong**

* Discussion Points:
  + Reviewed existing methodologies and tools.
  + Emphasized the development of red teaming processes specific to Hong Kong's context.
  + Honson Lam mentioned sharing ML (Machine Learning) common references with the group.
* Action Items:
  + Professor Anna Han to inquire if HKGAI (Hong Kong Government AI initiatives) needs assistance with red teaming industry pilots.
  + Plan to develop custom tools for Hong Kong's needs.
  + Begin collecting data pertinent to red teaming purposes.

### **9. Meeting Schedule, Roles, and Communication**

* Meeting Frequency:
  + Agreed on holding meetings monthly.
* Member Roles and Responsibilities:
  + Government Sector Lead: Professor Anna Han
  + Industry Sector Leads:
    - Construction: Honson Lam
    - Health: John Mok
    - Commercials: David Wong
  + Academic Sector Lead: Kenneth Lam (to be confirmed)
  + Public Citizen Engagement Lead: David Wong
  + Newsletter:
    - David Wong to draft.
    - All SIG members to review and contribute.
  + Workshops and Seminars:
    - David Wong and Professor Anna Han to collaborate on agenda and organization.
* Communication Channels:
  + Decided to use WhatsApp for quick communication.
  + Email to be used for formal communications and document sharing.

### **10. Any Other Business (AOB)**

* Academic Sector Initiatives:
  + John Mok proposed:
    - Developing research, teaching materials, and certification programs related to AI safety.
* User Participation and Partnerships:
  + John Mok suggested:
    - Engaging with user groups and pilot co-development partners (e.g., HKCS, HKMU).
* Government Engagement:
  + John Mok to explore opportunities for:
    - Government sponsorship (Part 1).
    - Engagement with regulators (Part 2).

### **11. Summary and Closing Remarks**

* Recap of Action Items:
  + David Wong to prepare and circulate meeting minutes.
  + David Wong to draft the initial newsletter.
  + David Wong and Professor Anna Han to develop the workshop agenda.
  + Professor Anna Han to check with HKGAI about red teaming pilots.
  + Honson Lam to share ML common references.
  + John Mok to investigate user participation and government sponsorship opportunities.
* Closing:
  + David Wong thanked all attendees for their valuable contributions.
  + Encouraged ongoing participation and collaboration.
  + Reminded members of the importance of their roles in advancing AI safety in Hong Kong.

## **Action Items and Target Dates**

1. David Wong
   * Prepare Meeting Minutes
     + Deadline: 29th September 2024
     + Deliverable: Circulate minutes to all members for review.
   * Draft Initial Newsletter
     + Deadline: 6th October 2024
     + Deliverable: Share draft newsletter for member feedback.
2. David Wong & Professor Anna Han
   * Develop Workshop Agenda
     + Deadline: 13th October 2024
     + Deliverable: Provide detailed workshop agenda and proposed dates.
3. Professor Anna Han
   * Inquire with HKGAI
     + Deadline: 29th September 2024
     + Deliverable: Report on whether HKGAI requires assistance with red teaming industry pilots.
4. Honson Lam
   * Share ML Common References
     + Deadline: 29th September 2024
     + Deliverable: Distribute list of machine learning references to SIG members.
5. John Mok
   * Explore User Participation and Partnerships
     + Deadline: 13th October 2024
     + Deliverable: Report on potential pilot co-development partners (e.g., HKCS, HKMU).
   * Investigate Government Sponsorship and Regulator Engagement
     + Deadline: 13th October 2024
     + Deliverable: Provide proposal for approaching government sponsorship and regulatory bodies.
6. All Members
   * Review Mission Statement Draft
     + Deadline: Before next meeting (20th October 2024)
     + Action: Provide feedback and suggestions.
   * Consider Participation in Projects
     + Deadline: Ongoing
     + Action: Express interest in initiatives and volunteer for roles.
7. Formation of Working Groups
   * Interested Members
     + Deadline: Before next meeting
     + Action: Sign up for subcommittees focused on specific sectors or projects.
8. Kenneth Lam (Pending Confirmation)
   * Develop Academic Sector Initiatives
     + Deadline: 20th October 2024
     + Deliverable: Outline potential academic collaborations and programs.

## **Next Meeting**

* Date: 20th October 2024
* Time: 2:30 PM
* Venue: To be confirmed
* Agenda Items:
  + Review progress on action items.
  + Finalize mission statement.
  + Discuss detailed plans for prioritized projects.
  + Establish subcommittees and working groups.

## **Deliverables Summary**

* Meeting Minutes: Prepared by David Wong (Due: 29th September 2024)
* Draft Newsletter: Prepared by David Wong (Due: 6th October 2024)
* Workshop Agenda: Prepared by David Wong & Professor Anna Han (Due: 13th October 2024, Workshop to be held on Jan 2025)
* ML Common References: Provided by Honson Lam (Due: 29th September 2024)
* HKGAI Assistance Report: Provided by Professor Anna Han (Due: 29th September 2024)
* Partnerships Report: Provided by John Mok (Due: 13th October 2024)
* Government Engagement Proposal: Provided by John Mok (Due: 13th October 2024)
* Academic Initiatives Outline: Provided by Kenneth Lam (Due: 20th October 2024)

## **Closing Notes**

* Members are encouraged to stay engaged via the WhatsApp group for updates and discussions.
* Formal communications and documents will be shared via email.
* David Wong will send out calendar invites for the next meeting once the venue is confirmed.
* Members who were absent will be briefed and brought up to date on discussions and action items.

We look forward to the collaborative efforts of all members in promoting AI safety and implementing effective red teaming strategies across Hong Kong's sectors.

Minutes Prepared By: David Wong  
Date: 22 September 2024

**- End -**