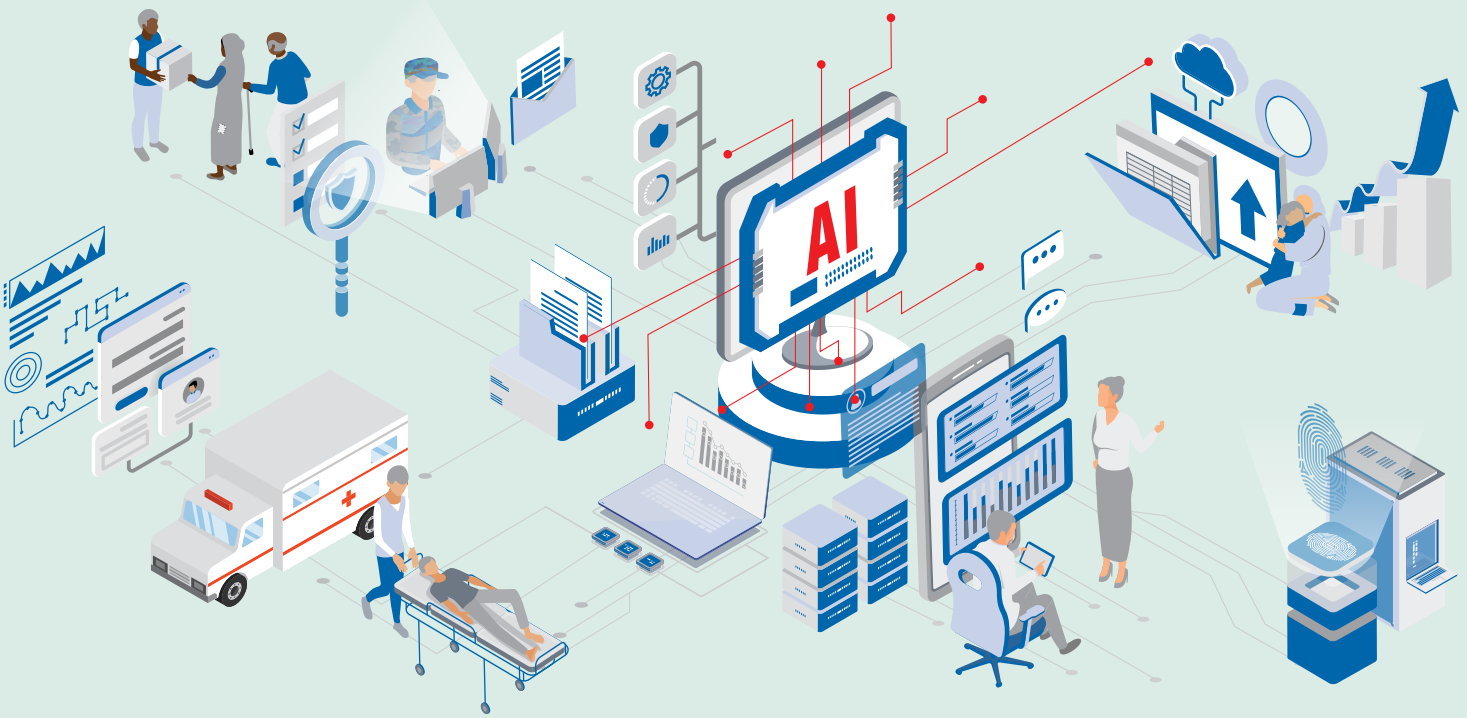


ایجاد رویکردی بشردوستانه و مسئولانه: خط‌مشی کمیته بین‌المللی صلیب سرخ در قبال هوش مصنوعی



توسعه سریع هوش مصنوعی (AI) فرصت‌های مهمی را برای پیشرفت بشر و جامعه، از جمله حوزه بشردوستانه، ایجاد می‌کند. کمیته بین‌المللی صلیب سرخ به بررسی اینکه آیا پیشرفت‌های حاصله در هوش مصنوعی قادرند در تحقق رسالت‌مان مبنی بر حفظ جان و کرامت انسانی کسانی که در میانه مخاصمه یا سایر وضعیت‌های خشونت‌آمیز زندگی می‌کنند، به ما یاری رسانند و اینکه چگونه، چه زمان و کجا اعمال‌پذیرند، خود را متعهد می‌داند.

با این حال، فناوری‌هایی که از هوش مصنوعی بهره می‌برند با ریسک‌های مهم و چندلایه‌ای همراه هستند که شناسایی و کاهش آن‌ها دشوار است.

هیاهو، دبستوبیا، خط‌مشی و رقابت حول هوش مصنوعی درک بی‌طرفانه و عینی آن را دشوار می‌کند. بر همین اساس، خط‌مشی ما در زمینه هوش مصنوعی به رویکردهای در بررسی کاربردهای بالقوه هوش مصنوعی در کار و فعالیت‌هایمان جهت می‌دهد.

خط‌مشی کمیته بین‌المللی صلیب سرخ در زمینه هوش مصنوعی مطابق با رسالت و اصول بنیادینمان، رویکردی صرفاً بشردوستانه دارد. هدف از آن کمک به کارکنان کمیته بین‌المللی صلیب سرخ برای شناخت هوش مصنوعی و شناسایی ظرفیت بشردوستانه آن است.

این خط‌مشی محصول رویکردی همکاری‌محور و چندرشته‌ای است که از تخصص بشردوستانه و عملیاتی کمیته بین‌المللی صلیب سرخ، استانداردهای موجود بین‌المللی هوش مصنوعی و رهنمودها و بازخوردهای کارشناسان برون‌سازمانی بهره می‌گیرد.

نظر به ماهیت هوش مصنوعی که پیوسته در تکامل است، این سند احتمالاً نمی‌تواند به همه پرسش‌ها و چالش‌هایی که در آینده مطرح خواهند شد بپردازد، اما امیدواریم پایه و چهارچوب محکمی فراهم کند تا در استفاده از هوش مصنوعی برای پشتیبانی از رسالت‌مان منطبق با رویکرد سازمانی‌مان برای دوره ۲۰۲۴ تا ۲۰۲۷، از اتخاذ رویکردی مسئولانه و انسان‌محور اطمینان حاصل کنیم.

کمیته بین‌المللی صلیب سرخ با اعلام عمومی این سیاست در صدد است که به تعهد مستمرمان به شفاف‌سازی و پاسخگویی در مقابل اقداماتمان پایبند بماند.

خط‌مشی هوش مصنوعی

۱. هدف

این خط‌مشی رهنمودهایی را درباره رویه جاری و تصمیم‌گیری در کمیته بین‌المللی صلیب سرخ درباره شناسایی، استقرار، کاربری و مدیریت هوش مصنوعی و فناوری‌ها، راهکارها و ابزار مربوط به یادگیری ماشین که از کار و فعالیت‌های کمیته بین‌المللی صلیب سرخ پشتیبانی می‌کنند، ارائه می‌دهد. در این خط‌مشی، رویکردی ارزش‌محور تعیین شده و مجموعه‌ای از اصول راهنمای کارمان در ارتباط با هوش مصنوعی را در اختیارمان قرار می‌دهد. همچنین به‌عنوان چهارچوبی سازمانی در راستای کاربری انسان‌محور، مسئولانه و اخلاقی هوش مصنوعی در اقدامات بشردوستانه عمل می‌کند. رهنمودهای ارائه‌شده در این خط‌مشی آرمان‌گرایانه بوده و با هدف پشتیبانی از تلاش‌های مستمر کمیته بین‌المللی صلیب سرخ برای استفاده مسئولانه از فناوری‌های دیجیتال و در راستای رسالت، اصول، ارزش‌ها و رویه‌های کاری بشردوستانه‌مان طرح شده‌اند. اجرای آن مستلزم سرمایه‌گذاری در قابلیت‌های عملیاتی و حکمرانی خاص، در محدوده ظرفیت‌های عملیاتی و مالی کمیته بین‌المللی صلیب سرخ است.

۲. تعریف

هیچ تعریف جهانی واحدی از هوش مصنوعی وجود ندارد. تعاریف بسیاری در زمینه‌های مختلف و برای اهداف مختلف موجود بوده و این واژه همگام با تحولات فناورانه، پیوسته در تکامل است. از نگاه کمیته بین‌المللی صلیب سرخ و منطبق با خط‌مشی حاضر، هوش مصنوعی این‌گونه تعریف می‌شود: «هر سامانه، ابزار یا فناوری‌ای که به‌منظور انجام وظایفی که در حالت عادی مستلزم شناخت، برنامه‌ریزی و استدلال بشر باشد و از سامانه‌های رایانه‌ای بهره‌بردار^۱، یادگیری ماشین عبارت است از «سامانه‌هایی که برای توسعه کارکرد خود از داده‌های زیادی استفاده کرده و تجربی یاد می‌گیرند». در این خط‌مشی، مقصود از «هوش مصنوعی»، هم سامانه‌های هوش مصنوعی و هم یادگیری ماشین است.

۳. حدود اختیارات

این خط‌مشی در مورد استفاده کلیه اعضا از هوش مصنوعی و همچنین فعالیت‌ها، خدمات داخلی و پاسخ‌های عملیاتی از جمله شراکت‌ها اعمال می‌گردد. خط‌مشی حاضر در مورد همه فناوری‌ها، راهکارها و ابزارهای مرتبط با هوش مصنوعی (مانند هوش مصنوعی مولد، سامانه‌ها و ربات‌های توانمندسازی‌شده به یادگیری ماشین، ابزار دیجیتال مجهز به هوش مصنوعی) و حوزه‌های کاربردی (مانند علوم داده، بینایی رایانه‌ای، پردازش زبان طبیعی، پیش‌بینی و مدل پیش‌بینی‌کننده، مدیریت اطلاعات و داده، تصمیم‌گیری، پاسخ‌های برنامه‌ای، برنامه‌ریزی و بهینه‌سازی، ارتباطات درون‌سازمانی و برون‌سازمانی) موضوعیت دارد.

محدودیت‌ها - رهنمودهای ارائه‌شده در این سند ماهیتی آرمان‌گرا و عمومی داشته و قادر به پرداختن به تمامی مسائل راهبردی، فنی یا عملیاتی ممکن حاصل از استفاده از هوش مصنوعی نیست. خط‌مشی و رهنمودهای عملیاتی مفصل‌تر برای ابزارها و موارد استفاده ویژه همگام با پیدایش نیاز و تکامل رویه کمیته بین‌المللی صلیب سرخ، همسو با رویکرد و اصول مندرج در این خط‌مشی از سوی بخش‌ها طرح خواهد شد. با عنایت به سرعت بالای نوآوری‌های مرتبط با هوش مصنوعی و پیدایش مستمر و تکامل چارچوب‌های نظارتی و هنجارهای قابل‌اعمال، خط‌مشی حاضر سند زنده‌ای است که بنا به اقتضا و طبق ارزیابی نهادهای سازمانی مربوطه، مرتب به‌روزرسانی خواهد شد تا تحولات در چشم‌انداز هوش مصنوعی و رویه کمیته بین‌المللی صلیب سرخ را منعکس کند.

منابع - رهنمودهای ارائه‌شده در این خط‌مشی بر استانداردها، اصول و توصیه‌های اتخاذشده از سوی سازمان‌های بین‌المللی پیشرو و متخصصان دانشگاهی و علمی در حوزه هوش مصنوعی استوار است (به پیوست ۱، «منابع و اسناد مربوطه» رجوع کنید). این رهنمودها، استانداردهای حداقلی بین‌المللی را ترکیب و آن‌ها را با نیازها، اهداف، رسالت، فرهنگ یکپارچگی، ارزش‌ها، اصول و رویه‌های کاری کمیته بین‌المللی صلیب سرخ سازگار می‌کند. این خط‌مشی از سایر متون مرجع داخلی مربوط به اهداف راهبردی کمیته بین‌المللی صلیب سرخ در زمینه رویه‌های کاری و فناوری‌های اطلاعات و ارتباطات، الهام گرفته و بر این اساس استوار است.

از آنجا که توسعه و کاربرد هوش مصنوعی در بیشتر موارد با کاربست داده‌های شخصی همراه است، خط‌مشی حاضر بدون ایراد هیچ خدشه‌ای، مکمل اعمال **قواعد** کمیته بین‌المللی صلیب سرخ **راجع به حفاظت از داده‌های شخصی** (که من بعد «قواعد» گفته می‌شود) بوده و برای سازمان الزام‌آور است. رهنمودهای ویژه درباره اجرای قواعد در زمینه هوش مصنوعی در **کتابچه حفاظت از داده در اقدامات بشردوستانه** ارائه گردیده و شکایت‌های مربوط به عدم پای‌بندی به قواعد را می‌توان از طریق «اداره حفاظت از داده» به عرض «کمیسیون حفاظت از داده‌های کمیته بین‌المللی صلیب سرخ» رساند.

۱ ابزار یا بسترهای مورد استفاده کمیته بین‌المللی صلیب سرخ پس از پذیرش، کارکرد هوش مصنوعی را اضافه می‌کنند، اما این کارکرد مضاعف می‌تواند صرفاً جنبه زیبایی داشته و مثلاً به‌عنوان چت‌بات پشتیبانی یا توصیه‌های طراحی مورد استفاده قرار گیرد. در مواردی، افزودن هوش مصنوعی ممکن است به‌حدی بنیادی باشد که ماهیت ابزار یا بستر را تغییر دهد. یک نمونه را می‌توان در توصیه‌های جستجو که هوش مصنوعی به‌جای تطبیق متن تعیین می‌کند، یافت. در چنین مواردی، استفاده از بستر فوق ممکن است مستلزم ارزیابی مجدد در چهارچوب این خط‌مشی باشد.

خط‌مشی حاضر به **موضع‌گیری** کمیته بین‌المللی صلیب سرخ در قبال استفاده طرفین مخاصمه مسلحانه از هوش مصنوعی در زمینه حقوق بین‌المللی بشردوستانه نمی‌پردازد.

۴. منطق

هوش مصنوعی فناوری جدیدی نیست. کمیته بین‌المللی صلیب سرخ از ابزارها و فناوری‌های هوش مصنوعی برای پشتیبانی از فعالیت‌هایش (مانند بهبود **لجستیک** یا ارائه **خدمات درمان**) بهره برده و تعدادی سند و رهنمود درون‌سازمانی اساس رویه‌های آن را تشکیل می‌دهند. با این حال، سرعت و مقیاس نوآوری‌های مربوط به هوش مصنوعی طی چند سال اخیر، پیدایش ابزارهای فناورانه جدید، دسترسی تسهیل شده به آن‌ها، عوامل قبض و بسط درون‌سازمانی و برون‌سازمانی کمیته بین‌المللی صلیب سرخ، و بحث‌های جاری درباره فرصت‌ها و ریسک‌هایی که از جمله برای اقدامات بشردوستانه، ایجاد می‌کنند، نیاز به حصول اطمینان از برخورداری سازمان از امکانات کافی را برای شناسایی و مدیریت این تحولات در زمان مناسب و به شکل منسجم و مسئولانه برانگیخته است.

از منظر درون‌سازمانی، جریان‌های کاری و ابتکارات متعدد مرتبط با هوش مصنوعی در بخش‌ها، طرح‌ها و مشاغل مختلف جریان داشته و ظهور می‌کنند. با این حال، کمیته بین‌المللی صلیب سرخ در حال حاضر فاقد یک چهارچوب خط‌مشی مشترک برای حصول اطمینان از این است که این فعالیت‌های مختلف طبق یک روش مشخص، متناسب، رسمی و حرفه‌ای سازماندهی و مدیریت می‌شوند. هدف از خط‌مشی حاضر کمک به رفع این شکاف و بنیان‌نهادن تلاش‌های کمیته بین‌المللی صلیب سرخ به‌منظور بهره‌برداری کارآمد از نوآوری فناورانه از طریق سازوکارهای حکمرانی مؤثر و رویکردی مسئولانه است که بر ایجاد تأثیر مرتبط و پایدار بشردوستانه و تأیید اصول و ارزش‌های سازمان تمرکز دارد.

از منظر برون‌سازمانی، مباحث مربوط به هوش مصنوعی - راجع به فرصت‌ها، ریسک‌ها، نیازها و گزینه‌های نظارتی - در زمینه رقابت راهبردی جهانی بین دولت‌ها که نظارت‌ها و محدودیت‌ها و نیروهای بازار را که مشوق خطرپذیری کنشگران بخش خصوصی و ارزیابی‌های متنوع از ارتباط و تأثیر هوش مصنوعی بر جوامع و مردم بوده، تضعیف می‌کنند، ماهیت سیاسی یافته‌اند. موضع‌گیری‌ها اغلب تحت تأثیر یک الگوی دوگانه و بیش‌ازحد ساده‌انگارانه قرار می‌گیرند که مشتاقان هوش مصنوعی را در برابر بدبینان، دیدگاه‌های آرمان‌شهری را در برابر دیدگاه‌های ویران‌شهری قرار داده که غالباً با هیاهوی ساختگی یا هراس‌پراکنی نسبت به هوش مصنوعی تشدید می‌شوند. این شامل بحث‌های مربوط به ریسک‌های ادعایی «آسیب موجودیتی» برای بشریت و نیاز به لحاظ‌کردن آن‌ها در تعادل با آسیب‌های قابل‌پیش‌بینی‌تر و رایجی‌ست که سامانه‌های هوش مصنوعی ایجاد یا تقویت می‌کنند.

خط‌مشی حاضر یک رویکرد بشردوستانه، اخلاقی و آموزشی مروج استفاده مسئولانه، ایمن، منسجم و انسان‌محور از هوش مصنوعی را مشخص می‌کند. چنین رویکردی حکمرانی کمیته بین‌المللی صلیب سرخ و اعتبار و توانایی آن را در مشارکت در بحث‌های مربوط به هوش مصنوعی با پیامدهای بشردوستانه و تأثیرگذاری بر آن‌ها در راستای مصالح مردم متأثر از مخاصمه و اقدام بشردوستانه اصولی تقویت می‌کند.

۵. رویکرد ارزش‌محور

۵.۱. رویکرد بشردوستانه و اصولی

بشردوستی - رویکرد کمیته بین‌المللی صلیب سرخ در قبال هوش مصنوعی با اهداف صرفاً بشردوستانه و اصل بنیادی بشردوستی هدایت می‌شود. بر این اساس، احترام برای اصول بشردوستانه، حقوق، ایمنی، امنیت، کرامت، رفاه و پویایی بشر در مرکز سیاست‌ها و رویه‌های ما قرار می‌گیرد. همسو با این رویکرد، خود را به کاربست و شناسایی ظرفیت راهکارهای هوش مصنوعی به‌منظور پشتیبانی از کیفیت و کارآمدی کارمان در چهارچوب رسالت بشردوستانه، ظرفیت‌های عملیاتی و مالی و قواعد داخلی‌مان متعهد می‌دانیم.

- کمیته بین‌المللی صلیب سرخ بر راهکارهای هوش مصنوعی که قادر به کاستن از رنج مردم متأثر از مخاصمه مسلحانه و سایر وضعیت‌های خشونت‌آمیز بوده و بتواند کارایی و اثربخشی سامانه‌ها و فرآیندهای عملیاتی و سازمانی پشتیبانی اقدامات حفاظتی و یاری‌رسانی ما را بهبود بخشد، تمرکز دارد.
- کمیته بین‌المللی صلیب سرخ همسو با اصل «عدم آسیب‌رسانی» تأثیرات بالقوه منفی ثانویه کوتاه‌مدت و بلندمدت راهکارهای هوش مصنوعی بر ایمنی، کرامت، پویایی و احترام به حقوق بنیادی افراد آسیب‌دیده را به‌شکلی نظام‌مند مورد ارزیابی قرار داده و تمامی اقدامات لازم را به‌منظور پیشگیری و کاهش آن‌ها اتخاذ می‌کند.
- کمیته بین‌المللی صلیب سرخ توجه ویژه‌ای مبذول داشته تا اطمینان حاصل کند که کاربری و توسعه راهکارهای هوش مصنوعی توانایی ما در ابراز بشردوستی و همدلی از طریق مشارکت مستقیم و حضور بشردوستانه در ارائه فعالیت‌ها و برنامه‌های خارجی و داخلی را به خطر نمی‌اندازد. ما بر راهکارهای هوش مصنوعی که قادر باشند خودآیینی و پویایی بشر را تقویت و تعاملات انسانی متقابل و همچنین حضور فیزیکی و نزدیکی با افراد آسیب‌دیده را تکمیل کنند، متمرکز هستیم.

- نظر به موارد فوق، کمیته بین‌المللی صلیب سرخ از کاربری، استقرار یا ارائه سامانه‌های هوش مصنوعی در موارد زیر خودداری خواهد کرد: سامانه‌هایی که از تکنیک‌های زیرآستانه‌ای، دستکاری یا فریبده برای تحریف رفتار و آسیب‌زدن به تصمیم‌گیری آگاهانه استفاده کنند؛ سامانه‌هایی که با بهره‌برداری از آسیب‌پذیری‌های مربوط به سن، ناتوانی یا شرایط اجتماعی-اقتصادی، تحریف رفتار کنند؛ سامانه‌هایی که از طبقه‌بندی بیومتریک که ویژگی‌های حساس شخصی را تشخیص می‌دهند استفاده کنند، مگر در مواردی که برچسب‌زنی یا پالایش مجموعه داده‌های بیومتریکی که طبق قانون حاصل شده‌اند؛ سامانه‌هایی که بر استنباط امتیازدهی اجتماعی تکیه کنند (یعنی ارزیابی یا طبقه‌بندی افراد یا گروه‌ها بر اساس رفتار اجتماعی یا ویژگی‌های شخصی). سامانه‌هایی که به‌جز در موارد استفاده برای بهبود ارزیابی‌های بشردوستانه، خطر ارتکاب جرم از سوی فرد را صرفاً بر اساس مشخصات فردی یا ویژگی‌های شخصیتی ارزیابی کنند؛ سامانه‌هایی که با استخراج غیرهدفمند تصاویر از اینترنت یا دوربین‌های مداربسته، پایگاه‌های داده‌ای تشخیص چهره گردآوری کنند؛ یا سامانه‌هایی که جز به دلایل پزشکی یا ایمنی، احساسات را تشخیص دهند.

بی‌غرضی - ما از راهکارهای هوش مصنوعی به‌منظور تقویت توانایی‌هایمان برای شناسایی، تحلیل، اولویت‌بخشی و پاسخگویی به نیازهای افراد آسیب‌دیده طبق مقیاس و سطح فوریت بهره می‌بریم بدون اینکه به دلیل نژاد، ملیت، جنسیت، دین، طبقه اجتماعی، دیدگاه سیاسی یا دلایل دیگر، تبعیض قائل شویم. کمیته بین‌المللی صلیب سرخ همواره به جمعیت‌ها و کاربران آسیب‌دیده دارای نیازها و آسیب‌پذیری‌های خاص و همچنین الزامات خاص زمینه‌هایی که هوش مصنوعی در آن‌ها لحاظ شده و مورد استفاده قرار می‌گیرد، توجه خاصی معطوف می‌دارد.

- کمیته بین‌المللی صلیب سرخ به ریسک‌های مهم سامانه‌های هوش مصنوعی که ممکن است داده‌های نادرست یا غیردقیق تولید کرده و تبعیضات، غرض‌ورزی‌ها و کلیشه‌های نهادی و اجتماعی را به‌ویژه حول مسئله جنسیت و نژاد، تکرار یا تقویت می‌کنند، اذعان می‌کند. ما به منظور درک، پیشگیری و کاهش اثر این ریسک‌ها با گزینش راهکارهای هوش مصنوعی و استفاده از آن‌ها و همچنین ممانعت از کاربری آن‌ها در مواردی که از غرض‌ورزی‌ها نتوان به‌شکل مؤثری پیشگیری کرد یا از اثر آن‌ها کاست، تلاش‌های فعالانه و پیوسته‌ای را در پیش می‌گیریم.

- کمیته بین‌المللی صلیب سرخ به منظور حصول اطمینان از اینکه راهکارها و ابزارهای هوش مصنوعی مورد استفاده ما از نظر طرح و همچنین در اثرگذاری، جامع بوده و با لحاظ کردن درجات مختلف نیاز و ظرفیت، یاری‌رسانی و خدمات‌رسانی منصفانه غیرتبعیض‌آمیز، برابر و عادلانه را به تمامی افراد و کاربران آسیب‌دیده ممکن می‌سازد، کلیه تلاش‌های ممکن را صورت خواهد داد.

بی‌طرفی و استقلال - هوش مصنوعی مانند سایر فناوری‌ها فی‌نفسه خوب، بد یا بی‌طرف نیست. می‌تواند ارزش‌های شخصی یا سیاسی افراد و یا شرکت‌های خلق‌کننده، طراح و مروج آن‌ها را نمایانده یا ترویج کند. هوش مصنوعی موضوع سیاسی‌کاری و رقابت فزاینده بین دولت‌ها در سطح جهانی و همچنین کنش‌گران بخش خصوصی و غیردولتی است. به‌علاوه، بسیاری از راهکارهای هوش مصنوعی از سوی شرکت‌های فناوری‌ای ارائه می‌شوند که فعالیت فزاینده‌ای در حوزه سیاسی داشته یا در درگیری‌های مسلحانه جاری، جانب‌داری می‌کنند. از این‌رو، ما به‌عنوان یک سازمان، توجه خاصی مبذول داشته و تمامی اقدامات لازم را به‌منظور حصول اطمینان از اینکه ابزارها و ارائه‌دهندگان منتخبمان برای هوش مصنوعی برای به‌چالش کشیدن بی‌طرفی و استقلالمان مورد استفاده ابزاری قرار نگیرند، به کار می‌بندیم.

- کمیته بین‌المللی صلیب سرخ جهت حصول اطمینان از چهارچوب‌های اخلاقی (چهارچوب مشارکت با بخش خصوصی در راستای جذب پشتیبان، اصول اخلاقی راهنمای شراکت‌های کمیته بین‌المللی صلیب سرخ با بخش خصوصی) و فرآیندهای تأمین (نظام‌نامه رفتار تأمین‌کننده) بهره می‌برد تا اطمینان حاصل شود که ما تأمین‌کنندگان هوش مصنوعی را که (الف) مستقیماً در فعالیت‌های سهیم در مخاصمه مسلحانه یا بسیار مرتبط با فعالیت‌های نظامی طرف‌های مخاصمه مشارکت داشته، یا (ب) مروج منافع سیاسی یا ارزش‌های خاصی باشند که مغایر با رسالت بشردوستانه و ارزش‌های اخلاقی کمیته بین‌المللی صلیب سرخ است، به کار نخواهیم گرفت.

ما بنا بر اقتضا، تلاش می‌کنیم راهکارهای هوش مصنوعی رایگان و یا منبع‌باز قابل‌دسترسی، ایمن و کافی برای کاهش ریسک‌های ناشی از وابستگی زیاد به ابزار و سامانه‌های سیاسی‌شده یا اختصاصی را که ممکن است (الف) درک ما از بی‌طرفی و استقلال از نهادهای درگیر جنجال‌های سیاسی یا مرتبط با مخاصمه را به خطر انداخته؛ یا (ب) استقلال عملیاتی و خودآیینی‌مان را به خطر بیندازد؛ یا (ج) توانمان در انتخاب گزینه بدون واردکردن هزینه‌های نامتناسب را محدود کند، بررسی کنیم.

۵.۲. رویکرد توانمندساز و آموزشی

تقویت نوآوری اخلاقی و اثرگذار - کمیته بین‌المللی صلیب سرخ متعهد به ایجاد یک فضای سازمانی است که شناسایی و کاربری ابتکارات و رویه‌های اثرگذار، پایدار و نوآورانه هوش مصنوعی را به شکلی ایمن و مسئولانه تسهیل و پشتیبانی کند.

- کمیته بین‌المللی صلیب سرخ به شکلی نظام‌مند با تکیه بر تجزیه و تحلیل به هم پیوستگی محیط‌ها، داده‌ها و سامانه‌های پشتیبان هوش مصنوعی و عوامل انسانی و اجتماعی، رویکردی زمینه‌محور و زیست‌بومی اتخاذ می‌کند. ما به استفاده مسئولانه از داده؛ سازوکارهای قوی حکمرانی، طبقه‌بندی، مقررات‌گذاری و نظارتی و همچنین ارزیابی و انطباق مستمر از طریق رویکردی هماهنگ، همکارانه، چندرشته‌ای و چندذی‌نفعی مبادرت می‌ورزیم.

- کمیته بین‌المللی صلیب سرخ به منظور حصول اطمینان از اینکه سرمایه‌گذاری‌ها در حوزه تحقیق و توسعه هوش مصنوعی از نظر مالی مسئولانه و پایدار بوده، چشم‌انداز بلندمدتی را اتخاذ کرده که بر مرتبط‌بودن، کارایی و تأثیرگذاری متمرکز است، و همچنین ارزش‌افزوده، هم‌کنش‌پذیری، مقیاس‌پذیری و کارایی هزینه (شامل هزینه‌های مربوط به نگهداری، آموزش، راستی‌آزمایی و انطباق) را در بر می‌گیرد.

- کمیته بین‌المللی صلیب سرخ به همسویی بین تأثیر راهکارهای لحاظ‌شده هوش مصنوعی و اهداف کوتاه‌مدت و بلندمدت سازمان، محدودیت‌ها و الزامات عملیاتی و نهادی و ریسک‌های برنامه‌ای و سازمانی مربوطه در سطوح امنیتی و خوش‌نامی، توجه ویژه‌ای مبذول می‌دارد.

شایستگی و ظرفیت‌سازی - پیچیدگی فنی و ریسک‌های بالقوه مرتبط با سامانه‌های هوش مصنوعی و کارکردشان مستلزم سطح قابل توجهی از دانش و تخصص برای گزینش، عملیاتی‌سازی و مدیریت اثربخش و مسئولانه آن‌هاست. کمیته بین‌المللی صلیب سرخ متعهد است تضمین نماید که اعضای کمیته در سطوح مختلف سازمان برای شناسایی، توسعه و کاربری هوش مصنوعی به شکل مسئولانه و منطبق با رسالت بشردوستانه، اصول بنیادی، نظام‌نامه اخلاقی و فرهنگ یکپارچگی، از اطلاعات، تجهیزات و پشتیبانی مناسب برخوردار باشند.

- کمیته بین‌المللی صلیب سرخ در سطح نهادی، میزان کافی از سواد هوش مصنوعی و دیجیتال و فراگیری و تقویت مهارت‌ها برای اعضا در همه سطوح را تشویق و تسهیل می‌کند. ما از خدمات ارزیابی هم‌تایان، پشتیبانی و همکاری با کارشناسان ذی‌صلاح و معتمد برون‌سازمانی و شرکایمان برای ترویج و تداوم رویکرد فراگیری پیوسته به استفاده از هوش مصنوعی در اقدامات بشردوستانه بهره می‌بریم.

- کمیته بین‌المللی صلیب سرخ در برنامه‌های فراگیری و توسعه سازگار سرمایه‌گذاری می‌کند تا فهم اعضايمان را از چیرستی سامانه‌های هوش مصنوعی، چگونگی کارکرد آن‌ها و فرصت‌ها و ریسک‌هایی را که برای فعالیت‌های حرفه‌ای اعضايمان دارند پشتیبانی کنیم. این شامل آگاهی و فهم ملاحظات اخلاقی و قواعد و مقررات درون‌سازمانی متفاوتی‌ست که به ابزارها و فناوری‌های درون‌سازمانی و برون‌سازمانی هوش مصنوعی اعمال می‌گردد. کمیته بین‌المللی صلیب سرخ در صورت لزوم، دسترسی به ابزارهای هوش مصنوعی را به اعضای مجاز که آموزش خاص در این زمینه دیده باشند محدود می‌کند.

۶. اصول راهنما

۶.۱. تناسب، احتیاط و «عدم آسیب‌رسانی»

کمیته بین‌المللی صلیب سرخ، هوش مصنوعی را ابزاری بالقوه در کمک به پاسخ‌دهی بهتر به نیازهای بشردوستانه و سازمانی، مبتنی بر شناسایی و تحلیل مسئله که در آن سطح مشخصی از ریسک‌ها، اثربخشی نتایج مورد انتظار و راهکارهای جایگزین مربوطه لحاظ شده، به شمار می‌آورد.

- راهکارهای بالقوه هوش مصنوعی صرفاً زمانی لحاظ شده و به کار گرفته می‌شوند که جهت‌دستیابی به اثر مثبت و پاسخ به اهداف روشن و ازپیش‌تعریف‌شده، لازم و کافی باشند.

انتخاب و کاربست سامانه‌های هوش مصنوعی توسط کمیته بین‌المللی صلیب سرخ تحت هدایت یک رویکرد محتاطانه، عمل‌گرا و علم‌محور انجام می‌شود که ریسک‌های چندبعدی محتمل برای جمعیت‌های آسیب‌دیده، کاربران و سازمان، ازجمله از جنبه حقوقی را با احتیاط لحاظ می‌کند. راهکارهای بالقوه هوش مصنوعی به کمک رویکردی تفکیکی که نوع، هدف، حدود اختیارات، محل استقرار و اثر و میزان ریسک‌های مربوطه (زیاد، متوسط یا کم) را تعیین می‌کند، بررسی و ارزیابی می‌شوند.

- تحلیل ویژه ریسک‌های مضاعف و اقدامات کاهش‌دهنده زمانی به کار گرفته می‌شوند که راهکارهای هوش مصنوعی که تحت بررسی هستند، سطوح بالاتری از ریسک‌ها را بروز داده یا تأثیر بالقوه‌ای بر ایمنی، کرامت و حقوق جمعیت‌ها و کاربران آسیب‌دیده داشته باشند، یا مستلزم کاربست داده‌های داخلی شخصی محرمانه، کاملاً محرمانه یا حساس باشند و یا با ریسک‌های حقوقی یا حسن شهرت برای سازمان مواجه باشند.

- زمانی که منافع بالقوه هوش مصنوعی بر ریسک‌های شناسایی‌شده سنگینی کند، راهکارهای تحت بررسی پیش از ارائه‌شدن کاملاً آزمایش می‌شوند. شواهد لازم از طریق آزمایش‌های علمی و آزمایش «جعبه شنی» برای حصول اطمینان از اثربخشی و ایمنی راهکارها تا حد ممکن، در این زمینه گردآوری می‌شود.
- سامانه‌های نظام‌مند بازبینی و ممیزی که از پشتیبانی سازوکارهای منظم بازخورد درون‌سازمانی و برون‌سازمانی برخوردارند بدین منظور اجرایی می‌شوند که از تحلیل مستمر و چندبعدی ریسک‌ها بر اساس احتمال و شدت اثر بالقوه‌شان بر جمعیت‌ها و کاربران آسیب‌دیده، یا سازمان از جنبه حقوقی، اطمینان حاصل شود. زمانی که اقدامات کاهشی کافی و مؤثر برای ریسک‌های شناسایی‌شده موجود یا کافی نباشد، کمیته بین‌المللی صلیب سرخ، استفاده از سامانه‌ها و راهکارهای مربوطه هوش مصنوعی را مردود دانسته یا به آن خاتمه می‌دهد.

۶.۲. ایمنی و امنیت

کمیته بین‌المللی صلیب سرخ نظر به ریسک‌های چندگانه و قابل توجه راهکارهای هوش مصنوعی برای محرمانگی و حفاظت از داده‌هایی که در راستای فعالیت‌های کمیته گردآوری، پردازش و مدیریت شده و همچنین با در نظر گرفتن عواقب منفی یا مضر احتمالی ناشی از این موارد، از طریق یک رویکرد انسان‌محور با بالاترین سطح ممکن ایمنی، امنیت و حریم خصوصی در طراحی و تضمین نظارت و راستی‌آزمایی انسانی بر خروجی‌های هوش مصنوعی در طول چرخه عمرشان، راهکارهای هوش مصنوعی را شناسایی کرده و به کار می‌بندد.

- ریسک‌های سوءاستفاده، استفاده نادرست، نفوذ و حمله به سامانه‌های هوش مصنوعی و آسیب‌ها و صدمات بالقوه مربوطه به‌شکل نظام‌مند لحاظ می‌شوند و بالاترین استانداردهای ممکن برای حفاظت از داده و اقدامات امنیت سایبری به‌منظور پیشگیری و کاهش آن‌ها منطبق با سطح ریسک ارزیابی‌شده اجرایی می‌شود.

- کنترل کیفیت نظام‌مند و مضاعف داده‌ها و سازوکارهای نظارتی اثربخش انسانی با هدف کنترل و بررسی الگوهای آموزشی مربوطه، سامانه‌های الگوریتمی، ورودی‌ها و خروجی‌ها در طول چرخه عمر راهکارهای هوش مصنوعی که برای پشتیبانی از فعالیت‌های عملیاتی و فرآیندهای تصمیم‌گیری تأثیرگذار که با ریسک آسیب‌رسانی به افراد و کاربران آسیب‌دیده استفاده می‌شوند، به اجرا در می‌آیند.

کمیته بین‌المللی صلیب سرخ اطمینان حاصل می‌کند که استفاده از راهکارهای درون‌سازمانی و برون‌سازمانی هوش مصنوعی با قواعد و اصول قانون‌مداری، انصاف، شفافیت، محدودیت هدف، حداقل‌سازی داده‌ها، صحت، محدودیت ذخیره‌سازی، امنیت و پاسخگویی همسو باشد.

- کمیته بین‌المللی صلیب سرخ همواره اطمینان حاصل می‌کند که اطلاعات داخلی محرمانه یا حساس کمیته یا داده‌های شخصی با بیشترین احتیاط مورد استفاده قرار گرفته و طبق قاعده رازداری مندرج در نظام‌نامه اخلاقی کمیته بین‌المللی صلیب سرخ در سامانه‌های برخط برون‌سازمانی غیرمجاز هوش مصنوعی استفاده نشده و به این سامانه‌ها منتقل نمی‌گردند.

- به‌منظور حصول اطمینان از اینکه داده‌های مورد نیاز برای سامانه‌های هوش مصنوعی، به شیوه‌ای متناسب با حق حریم خصوصی افراد و کاربران آسیب‌دیده، گردآوری، استفاده، هم‌رسانی، بایگانی و حذف می‌شوند، سازوکارهایی به کار بسته می‌شود.

- ارزیابی‌های نظام‌مند اثر حفاظت از داده پیش از دست‌یابی به راهکارهای مورد نظر هوش مصنوعی و استقرار یا کاربست آن‌ها صورت می‌گیرد و راهکارها در صورت مربوط و ممکن بودن با هدف محدود کردن پردازش و توانمندسازی اصلاح و یا زدودن داده‌های استفاده‌شده، اجرایی می‌گردند.

۶.۳. شفافیت و «تبیین‌پذیری»

شفافیت و «تبیین‌پذیری»، عناصر ضروری رویکرد ایمن، اخلاقی و مسئولانه به هوش مصنوعی بوده و برای رعایت اصول، قواعد و ارزش‌های کمیته بین‌المللی صلیب سرخ ضرورت دارند. از این‌رو، کمیته بین‌المللی صلیب سرخ همواره نسبت به بالاترین سطح ممکن از شفافیت در استفاده از راهکارهای هوش مصنوعی و مدیریت آن، منطبق با زمینه، ریسک‌ها و اثر سامانه‌ها و ابزارهای لحاظ‌شده و رعایت اصول ایمنی، امنیت و حریم خصوصی اطمینان حاصل می‌کند.

- تا حد ممکن و مطابق با سطح ریسک، زمینه و مورد استفاده، اقدامات معنادار در زمینه قابلیت ردیابی، افشا و اطلاع‌رسانی مربوط به کاربست سامانه‌های هوش مصنوعی که از فعالیت‌ها و فرآیندهای تصمیم‌گیری کمیته بین‌المللی صلیب سرخ پشتیبانی می‌کنند، از جمله بنا به اقتضا، انتشار سامانه‌های هوش مصنوعی به‌شکل منبع باز، صورت می‌گیرد.

- اطلاعات مربوط و قابل فهم درباره کاربست سامانه‌های هوش مصنوعی به‌شکل نظام‌مند به افراد و کاربران آسیب‌دیده، در صورتی که این سامانه‌ها، آن‌ها یا حقوق‌شان را تحت تأثیر قرار دهد یا اینکه تصمیماتی با عواقب قابل توجه برای سازمان مدنظر باشد، ارائه می‌گردد.

- این اطلاعات، اطلاعات کافی و معنادار درباره چرایی، زمان و چگونگی کاربست راهکارهای هوش مصنوعی و توضیحات درباره پیوند علی بین ورودی‌ها و خروجی‌های هوش مصنوعی و کارکرد فرآیندهای الگوریتمی را در بر می‌گیرد.

۴.۶. مسئولیت‌پذیری و پاسخگویی

کمیته بین‌المللی صلیب سرخ با پایش، ارزیابی، گزارش‌دهی مستمر درباره تهیه، کاربست، مدیریت و بررسی راهکارهای هوش مصنوعی، از جمله از طریق گزارش‌دهی با دسترسی عمومی در صورت اقتضا، بالاترین سطح ممکن از مسئولیت‌پذیری و پاسخگویی را در کاربست و مدیریت راهکارهای هوش مصنوعی تضمین می‌کند.

- در صورت ارتباط و امکان، ارزیابی‌های نظام‌مند تأثیرگذاری، تأییدپذیری و تکرارپذیری اجرایی شده و با پشتیبانی فرآیندهای اثربخش ارزیابی و ممیزی، از سازگاری سامانه‌ها و ابزارهای هوش مصنوعی کمیته بین‌المللی صلیب سرخ با این خط‌مشی اطمینان حاصل می‌کند.

- در صورت ارتباط و امکان، سازوکارهای مشارکتی معنادار و در دسترس، با هدف ارائه مشاوره به افراد و کاربران آسیب‌دیده در زمینه طراحی، کاربست و تأثیر راهکارهای مدنظر هوش مصنوعی، اجرایی می‌شود. مشروط بر اینکه بدین ترتیب، افراد آسیب‌دیده از فرصت‌های مشابه محروم نشوند، امکان انصراف از سامانه‌های هوش مصنوعی و دسترسی به راهکارهای جایگزین و سازوکارهای شکایت و جبران خسارت در اختیار آنان قرار می‌گیرد.

- مسئولیت‌پذیری و پاسخگویی در قبال تصمیم‌گیری‌های تحت پشتیبانی هوش مصنوعی همواره بر عهده کمیته بین‌المللی صلیب سرخ و یا کارکنانی‌ست که طبق نظام‌نامه اخلاقی کمیته بین‌المللی صلیب سرخ و سایر قواعد داخلی ناظر، از ابزارها یا سامانه موردنظر استفاده می‌کنند.

کمیته بین‌المللی صلیب سرخ به‌منظور حصول اطمینان از اینکه استفاده ما از راهکارهای هوش مصنوعی همسو با الزامات و تعهدات فراگیر سازمانی مان بوده و با قواعد و مقررات ناظر منطبق باشد، سازوکارهای مؤثر حکمرانی به کار می‌بندد.

- کارکنان نهادهای سازمانی ذی‌صلاح و شایسته جهت حصول اطمینان از حکمرانی خوب و اجرای اثربخش و انطباق با خط‌مشی حاضر و سایر قواعد و مقررات داخلی ناظر بر سامانه‌های هوش مصنوعی، منصوب و تجهیز می‌شوند.

- تصمیمات مربوط به کاربست سامانه‌ها، ابزار یا فناوری‌های هوش مصنوعی یا ارائه به دیگران باید منطبق با ملاحظات حقوقی بوده و در تمامی فرآیندهای تصمیم‌گیری مرتبط با هوش مصنوعی، نظرات کارشناسی حقوقی کسب شود.

- تلاش‌های مربوط به آگاهی‌بخشی، آموزش و انطباق با قوانین به‌منظور حصول اطمینان از موارد زیر صورت می‌گیرد: (الف) کاربست راهکارهای هوش مصنوعی توسط کمیته بین‌المللی صلیب سرخ، محدودیت‌های قانونی اجرایی کشوری، منطقه‌ای و بین‌المللی را نقض نکند؛ (ب) کارکنان و کاربران کمیته بین‌المللی صلیب سرخ هنگام استفاده از خدمات برخط برون‌سازمانی یا تجاری هوش مصنوعی، هیچ‌گونه اطلاعات حیاتی بشردوستانه، سایر اطلاعات محرمانه و یا بسیار محرمانه، یا هرگونه اطلاعات اختصاصی تحت حفاظت حقوق مالکیت معنوی را بارگذاری نکنند.

- تمامی تلاش‌های ممکن، به‌منظور یکپارچگی و کاهش اثرات زیست‌محیطی و ردپای کربن حاصل از راهکارهای مدنظر هوش مصنوعی با رویکرد چرخه عمر صورت می‌گیرد که مشتمل بر ترجیح راهکارها و ابزارهای دارای ردپای کربن پایین‌تر منطبق با برنامه اقدام کمیته بین‌المللی صلیب سرخ تحت «[منشور اقلیم و محیط‌زیست برای سازمان‌های بشردوست](#)» و همچنین «[چهارچوب توسعه پایدار](#)» کمیته بین‌المللی صلیب سرخ است.

منابع منتخب

تعدادی از منابع متعدد مربوط به جنبه‌های اخلاقی هوش مصنوعی که کمیته بین‌المللی صلیب سرخ در تدوین این خط‌مشی از آن‌ها بهره برده در زیر می‌آید:

<https://unesdoc.unesco.org/ark:/48223/pf0000381137> توصیه‌هایی در باب اخلاق در هوش مصنوعی، یونسکو:

<https://digitalprinciples.org/principles/> «اصول توسعه دیجیتال»:

درک اخلاق و ایمنی هوش مصنوعی: راهنمایی برای طراحی و اجرای مسئولانه سامانه‌های هوش مصنوعی در بخش دولتی، مؤسسه آلن تورینگ: <https://www.turing.ac.uk/news/publications/understanding-artificial-intelligence-ethics-and-safety>

چهارچوبی برای استفاده اخلاقی از روش‌های علوم داده‌های پیشرفته در بخش بشردوستانه، گروه علوم داده‌ها و اخلاق: <https://www.hum-dseg.org/sites/g/files/tmzbd1476/files/2020-10/Framework%20for%20the%20ethical%20use.pdf>

«رهنمودهای اخلاقی برای هوش مصنوعی قابل اعتماد»، کمیسیون اتحادیه اروپا:

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

<https://artificialintelligenceact.eu/the-act/> قانون هوش مصنوعی اتحادیه اروپا:

کاغذ سفید درباره هوش مصنوعی قابل اعتماد، آکادمی فناوری اطلاعات و ارتباطات چین: <https://cset.georgetown.edu/publication/white-paper-on-trustworthy-artificial-intelligence/>

«بیانیه اخلاقی آمریکای لاتین برای طراحی»، توسعه و استفاده از هوش مصنوعی: <https://aichallenge.utec.edu.uy/community/foro-bai-biases-in-ai-consultas-generales/declaracion-de-etica-ia-latam-para-el-diseno-desarrollo-y-uso-de-la-inteligencia-artificial/>

راهبرد قاره‌ای هوش مصنوعی: مهار هوش مصنوعی برای توسعه و رونق آفرینا: https://au.int/sites/default/files/documents/44004-doc-EN-Continental_AI_Strategy_July_2024.pdf

راهنمای اتحادیه کشورهای جنوب شرق آسیا راجع به حکمرانی و اخلاق هوش مصنوعی، اتحادیه کشورهای جنوب شرق آسیا: https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf

اخلاق دیجیتال: راهنما برای متخصصان عصر دیجیتال: <https://www.cigref.fr/wp/wp-content/uploads/2019/02/Cigref-Syntec-Digital-Ethics-Guide-for-Professionals-of-Digital-Age-2018-October-EN.pdf>

هوش مصنوعی و حفاظت از داده، شورای اروپا: <https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-des-donnees/168098e1b7>

حکمرانی بر هوش مصنوعی: تقویت حقوق بشر و کرامت انسان، داده‌ها و جامعه: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

«بیانیه تورنتو: حفاظت از حق برابر و عدم تبعیض در سامانه‌های یادگیری ماشین»، عفو بین‌الملل و سازمان دسترسی اکنون: <https://www.torontodeclaration.org/declaration-text/english/>

«چگونگی طراحی هوش مصنوعی برای خیر جامعه: هفت عامل اساسی»، اخلاق علوم و مهندسی: <https://link.springer.com/article/10.1007/s11948-020-00213-5>

«بازنگری در هوش مصنوعی بشردوستانه: بهره‌گیری از ظرفیت و گذر از چاله‌ها»: <https://odihpn.org/publication/humanitarian-ai-revisited-seizing-the-potential-and-sidestepping-the-pitfalls/>

ما به مردمی که در سراسر جهان تحت تأثیر مخاصمه مسلحانه یا سایر خشونت‌ها واقع شده باشند کمک کرده و در این راستا اغلب با شرکایمان در صلیب سرخ و هلال احمر، هر آنچه را که در توان داریم در جهت حفاظت از جان و کرامت آن‌ها و کاهش رنج‌هایشان ترتیب می‌دهیم. ما همچنین با ترویج و تقویت حقوق بشردوستانه و پشتیبانی از اصول جهان‌شمول بشردوستانه، به‌دنبال پیشگیری از سختی هستیم.

کمیته بین‌المللی صلیب سرخ

تهران، الهیه، خیابان شهید شریفی‌منش، جنب بیمارستان اختر،

بن‌بست آذر، پلاک ۴

تلفن: ۲۲۶۰۰۵۳۴ / فکس: ۲۲۶۴۵۸۲۱-۴

پست الکترونیکی: teheran@icrc.org

کمیته بین‌المللی صلیب سرخ، اردیبهشت ۱۴۰۴

 icrc.org

 blogs.icrc.org/ir

 aparat.com/icrc_iran

 t.me/icrc_iran